FINANCIAL TIME SERIES FORECASTING USING CONDITIONAL RESTRICTED BOLTZMANN MACHINE

LAI KWOK CHUNG

Ph.D

The Hong Kong Polytechnic University

2017

The Hong Kong Polytechnic University

Department of Computing

Financial Time Series Forecasting using Conditional Restricted Boltzmann Machine

Lai Kwok Chung

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

December, 2015

CERTIFICATE OF ORIGINALITY

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

____________________ (signed)

Lai Kwok Chung ______ (Name of student)

Abstract

Inspired by the success of deep learning in big data image recognition in Restricted Boltzmann Machine, Conditional Restricted Boltzmann Machine which was the original design to forecast human motion movement had been modified to forecast financial time series. As far as the author is aware of, this is the first attempt to apply deep learning in financial time series forecasting. Conventionally, deep learning is applied in image classification and several layers of deep learning in the huge dataset could increase its accuracy. The traditional forecasting method is using Euclidean distance to map the dataset into a higher dimension which facilitates to draw a hyperplane to separate the data. The more the cluster of the data in the hyperplane, the closer the distance of those neighbour data. As a result, those cluster data are the foundation to forecast. A new approach in Restricted Boltzmann Machine is to assign low energy based on probability concept to those connections that are relevant to each other while high energy is assigned to those that are irrelevant. The advantage of this method over Euclidean distance is that the probability energy assignment can be done one layer at a time and extend to many layers. Each layer information is retained and passed on to another layer to be trained again. As a consequence, all the information in the dataset is carefully scrutinized to obtain the best result.

In this research, it has been demonstrated in the following Chapters that deep learning using modified Conditional Restricted Boltzmann Machine is able to handle high dimensionality data which is over 100 with the dataset array as big as 600000x100. This setup enables it to capture the information of the high dimensions in each layer. Eventually, it will improve the forecasting accuracy. This was not possible before as our previous research has experienced. Historical records are not as important as the dimension of the financial time series problem domain. 30 or 20 years of stock history may not have that much impact on the current stock price in one stock. As the financial market is closely related to other markets, the stock price of a particular security is heavily dependent on other stocks in the same market as well as the performance of other markets. Hence, it is more important to increase the dimensionality or features of the data. In other words, including more factors such as the price of others stocks, economic factors such as interest rates and GDP can enhance the performance.

The algorithm based on Conditional Restricted Boltzmann Machine has demonstrated remarkable forecasting accuracy as reported in Chapter 4 and 5.

Publications resulted from this research

1. Lai, K.C. L. & Liu, N.K. J. (2008). WIPA: A Neural Network and CBR-based Model for Allocating Work in Progress, *5th International Conference on Information Technology and Applications*, 533-538.
2. Lai, K.C. L. & Liu, N.K. J. (2009). A Neural Network and CBR-based Model for Sewing Minute Value, *International Joint Conference on Neural Networks, IJCNN 2009,* 1696-1701. DOI:10.1109/IJCNN.2009.5178803.
3. Lai, K.C. L. & Liu, N.K. J. (2009). ALBO: An Assembly Line Balance Optimization Model Using Ant Colony Optimization, *Fifth International Conference on Natural Computation, ICNC2009*, 8-12. DOI:10.1009/ICNC.2009.693.
4. Lai, K.C. L. & Liu, N.K. J. (2009). Stock Forecasting Using Support Vector Machine, *International Conference on Machine Learning and Cybernetics (ICMLC),* 4, 1607-1614. DOI:10.1109/ICMLC.2010.5580999.
5. Lai, K.C. L. & Liu, N.K. J. (2012). WIPA: Neural Network and Case Base Reasoning Models for Allocating Work In Progress, *Journal of Intelligent Manufacturing,* 23(3), 409-421. DOI:10.1007/s 10845-010-0379-2.
6. Lai, K.C. L. & Liu, N.K. J. (2013). Support Vector Regression with Lévy Distribution Kernel for Stock Forecasting, *International Conference on Advanced Computer Science and Electronics Information (ICACSEI 2013),* 654-657. DOI:102.8331/icacsei.2013.16.
7. Lai, K.C. L., Hu, Y.X. R. & Liu, N.K. J. (2012). A weighted Support Vector Data Description Based on Rough Neighbourhood Approximation, *2012 IEEE 12th International Conference on Data Mining Workshops,* 635-642.
8. Lai, K.C. L., & Liu, N.K. J. (2013). Chart Pattern and Least Square Support Vector Machine in Stock Forecasting, *International Journal of Advanced Engineering Applications,* 45-55.
9. Lai, K.C. L. & Liu, N.K. J. (2014). Support Vector Machine and Least Square Support Vector Machine Stock Forecasting Models, *Computer Science and Information Technology, Horizon Research Publishing Corporation,* 30-39. DOI:10.13189/csit.2014.020102.

Acknowledgements

Mind and intelligence nurturing is a very long and tedious process. The age to understand the decree of Heaven when a man reaches the age of 50 is a famous Confucius remark . To be able to pick up academic research at that age after leaving school for many years is not only intriguing but also mind-boggling in my life. All these marvellous works will not be possible without the guidance from my patient and knowledgeable supervisors Dr. James Liu and Professor Jane You. Needless to say, the insurmountable support from my family especially my wife Anna and my two little daughters Agnes and Anna cannot be neglected. I have deprived them of their valuable family time at the expenses of my own selfish utopian dream.

I also wish to dedicate this work to my mother who passed away 20 years ago and her teachings are still lingering in my mind. As a teacher for more than 45 years, she had nourished many students which include my sister who got her Doctor of Philosophy 30 years ago. I am proud to be the second child in our family to pursue this academic path.

Without statistical analysis, this thesis would not be completed. Hence, without the assistance of Professor Wong Heung and Dr. Stan Yip who allowed me to take their courses as auditing student on statistical inference and forecasting, I would not be able to pick up the long forgotten statistical theories and complete this

thesis. The guidance from the mathematical department on my work is crucial.

Last but not least, I wish to thank Hong Kong Polytechnic University to accept a mature student like me and provide the necessary facilities both the hardware and the academic environments so that this research work can be accomplished.

Table of Contents

Major Contributions

The candidate believes the following are the major contributions in this thesis.

The first is the technique of windowize in Chapter 3.2.2.1. It has doubled or trebled the size of the original input. This technique outperforms the Relative Difference in Percentage of price (RDP) or 15-day Exponential Moving Average (EMA) methods which are commonly applied in financial time series study. The uni-dimensional data is normally transformed into different dimensions but still in a single dimension using RDP or EMA methods. This is very similar to the normalization technique which transforms the data into certain value range, for example, o to 1 or -1 to 1. In windowize method, one dimension of data has been multiplied into 2 or 3 dimensions and the information embedded in the data can be fully exploited by an appropriate algorithm.

The second is the introduction of Lévy Distribution Kernel in Chapter 3.4 for the Support Vector Regression. It is a challenge of the Normal distribution kernel which is widely used in Support Vector Regression. This doctrine has never been challenged even the defect in the Noble-price-winning Black-Sholes formula has been discovered after the financial tsunami, and it is still irreplaceable. The candidate has demonstrated that there are advantages using Lévy Distribution Kernel and the potential and the application of it is wide and extensive. Financial time series movement is not always in normal distribution which has been proved by Lévy but there are still very few applications to use Lévy distribution.

The third is from Table XXV showing which model is a better one under what situations or criteria. There are totally 20 different models and 3 data sets in this thesis. It is a very tedious process to test every model under different criteria and situations in order to improve the accuracy of the forecasting results. As most of the models have a long history in the literature review, the candidate wishes to point out some models that can be improved by using different approaches such as Lévy distribution kernel in Support Vector Regression. The table gives a summary of all the works the candidate has been doing throughout the period to pursue a research career.

The last one is the application of Conditional Restricted Boltzmann Machine in Financial Time Series Forecasting with a remarkable result. It allows multiple dimensions dataset to be used as input for the algorithm and simultaneously output the forecast. As financial markets are closely related to each other, it seems necessary to look into the inter-relationship of each market and use as many as possible different data from different markets to forecast the future events. The ability to predict many events at the same time using multiple dimension has many applications but not limited to financial forecasting domain.

List of Figures

List of Tables

Abbreviation

ANN	Artificial Neural Network
APT	Arbitrage Pricing Theory
ARIMA	Autoregressive Integrated Moving Average
CAPM	Capital Asset Pricing Model
CBR	Case Based Reasoning
CRBM	Conditional Restricted Boltzmann Machine
DBM	Deep Boltzmann Machine
DJ	Dow Jones Industrial Average Index
DWT	Discrete Wavelet Transform
EGARCH	GARCH to capture asymmetry in volatility clustering
ELM	Extreme Learning Machine
EMA	Exponential Moving Average
EMH	Efficient Market Hypothesis
FLOPS	Floating Point Operations per Second
FT	Fourier Transform
GA	Genetic Algorithms
GARCH	Generalized Autoregressive Conditional Heteroskedasticity

GJR	Glosten, Jagannathan, and Runkle (GJR) mode
GRNN	Generalized Regression Neural Networks
HIS	Hang Sang Index
IPO	Initial Public Offering
LSSVM	Least Square Support Vector Machine
MA	Moving Average
MAE	Mean Absolute Error
MAPE	Mean of the absolute percentage error
MNIST	Mixed National Institute of Standards and Technology database
MSE	Mean Square Error
NaN	Not a number
NN	Neural Network
PNN	Probabilistic Neural Networks
RB	Radial Basis Neural Networks
RBM	Restricted Boltzmann Machine
RDP	Relative Difference in Percentage of Price
RMSE	Root Mean Square Error
SARIMA	Seasonal Autoregressive Integrated Moving Average
SLFNs	Single Hidden Layer Feedforward Networks

SMAPE	Symmetric Mean of the absolute percentage error
SSE50	Shanghai Stock Exchange 50 composite Index
SVM	Support Vector Machine
SVR	Support Vector Regression
WT	Wavelet Transform

Notations in Equations

Equation 1 to 5

y_t, $\hat{y}_t$, e_t, n	y_t actual value, $\hat{y}_t$ forecast value, e_t difference between actual and forecast, n is the number of input historical records in forecasting

Equation 6 to 8 & 58 to 60

G, g	G, g output function of the i^{th} hidden node,
β_i	Output weight vector connecting the i^{th} hidden node and the output nodes
R, C	R real number C integer constant

Equation 9 to 17

ω^T, $\varphi(x)$, b	ω^T, transpose of a matrix or vector $\varphi(x)$ is a mapping function, b is an integer constant
ξ, ξ^*	Slack variables
α, α^*, η, η^*	Lagrange multipliers
$K(x_k, x_l)$	Kernel function mapping into higher dimension

J_D	$J_D(.)$ is the objective function for the dual problem obtained after elimination of the primal variables
$L(\varepsilon)$	Vapnik ε-insensitive loss function
e_k	an error variable for least square SVM
$\phi(x_k)$	the mapping to the high dimensional feature space

Equation 18 to 23

$\Psi_{m,n}$	Orthonormal wavelet basis
$T_{m,n}$	Inverse discrete wavelet transform
$\Phi_{m,n}$	Scaling functions
$S_{m,n}$	Approximation coefficients

Equation 24 to 28

X_t	Discrete time stochastic process
ε_t	White noise
ϕ	Parameter
∇	$1 - B$
B^2X_t	$B(BX_t)$
σ	Time dependent standard deviation
α	GARCH equation parameter

β	GARCH equation parameter

Equation 29 to 38

W_t	Standard Brownian motion
C(S,T)	Stock price, price of the call option and time
P_t	Trading value
B_t	Certain amount of money put in a bank
S_t	Certain amount of money to buy stock

Equation 39 to 42

B_i	Output weight vector connecting the i^{th} hidden node and the output nodes
a_i, b_i	Learning parameters in extreme learning machine
L	Hidden Nodes
n	Input Nodes

Equation 43 to 55

v,h	Visible variables v and hidden variable h
Z	Normalization constant
W	A matrix of pairwise weights between elements of v and h

$l(\theta)$	log-likelihood
b^v, b^h	Biases for the visible and hidden variables

Chapter 1 Introduction and Objectives of the Research

1.1 Introduction

Forecasting is sometimes regarded as an art rather than science. Statistical theory has provided a solid foundation on modern forecasting theory. Estimation of parameters and testing of the hypothesis are the two cornerstones of modern statistics (He, Liu, Wang, & Hu, Optimal bandwidth selection for re-substitution entropy estimation, 2012). The parameter could be dependent or independent which is defined by the nature of the study domain. In financial time series forecasting, stock price can be taken as one of the parameters. From the statistical point of view, all data are classified into different types of distribution using statistical analysis. Not all data are available such as the income of each individual in society, statistical samples are taken and a parameter to represent the population such as the mean (average income of the society) is estimated from the sample using statistical inference. Here, the estimation of the parameter is for present phenomena. In financial time series forecasting, historical samples are known and there is no need to estimate the parameter. The goal is to forecast the future movement of stock price.

With the help of the modern computer technology, many algorithms and theories have been established in the last 20 to 30 years.

They are either from statistical background such as Autoregressive Integrated Moving Average (ARIMA), General Autoregressive Heteroskedasticity(GARCH) or computational background such as Neural Network (NN) or Support Vector Machine (SVM). Many statistical forecasting methods such as ARIMA or GARCH are still playing a key role in the forecasting society. In fact, statistical significance is an important measurement on the outcome of prediction. However, the forecasting society is still debating which approach, statistical or computational intelligent, is better to do the job. The Forecasting Competition for Artificial Neural Network (ANN) and Computational Intelligence recruit talents and nurture new brains to get better forecasting model and result. It is essentially a competition between the performance of statistical and computational intelligent forecasting method. We depend very much on forecasting events ranging from the weather forecast to economic trend. The world would be very different if we are not presented with the possible future events that we can prepare in advance. However, the methodology that provides forecasting events for us is still debatable and controversial. Whether the statistics can forecast better than Artificial Intelligence or vice versa is not the scope of this research for discussion but we would like to explore the best from our research.

The argument over the practical use of Artificial Intelligence to forecast financial time series is a very sensitive and controversial issue. In the book of Forecasting economic time series (Clements & Hendry, Forecasting Economic Time Series, 1998), there is clear definition of what is predictable and unpredictable. The definition of unpredictability is equivalent to the statistical independence of an m-dimensional stochastic variable V_t (with non-degenerate density $Dv_t(.)$) from an information set denoted T_{t-1}. Then, V_t is unpredictable with respect to T_{t-1} if the conditional and unconditional distributions coincide: $Dv_t(V_t|T_{t-1})=Dv_t(V_t)$. The concept does not connote to erratic or wild that nothing useful can be said. Rather, it entails that knowledge of T_{t-1} does not improve prediction or reduce any aspect of the uncertainty about V_t. The main message is that we cannot forecast the unpredictable. In the book of Investment (Bodie, Kane, & Marcus, Investments, 2005), it says the prices of securities fully reflect available information in the Efficient Market Hypothesis (EMH), a cornerstone of finance theory for over 40 years. Investors buying securities in an efficient market should expect to obtain an equilibrium rate of return. The strong form hypothesis asserts that stock prices reflect all relevant information including insider information. The semi-strong form hypothesis asserts that stock prices already reflect all publicly available information. It is the boldest testable version of EMH (Aronson, Evidence-Based Technical Analysis, 2007) which asserts that no information in the public domain, either fundamental or technical, can be used to generate risk-adjusted returns in excess of the market index. Both strong and semi-strong forms of EMH defining the market is unpredictable. However, it cannot pass the statistical definition of unpredictability as described in (Clements & Hendry, Forecasting Economic Time Series, 1998). From (Aronson, Evidence-Based Technical Analysis, 2007) cross-sectional time series studies, price movements are predictable to some degree with stale public information, and excess risk-adjusted returns are possible. Here, stock prices and other data used in technical indicators are useful. This is very good news for technical analysis and very bad news for EMH.

1.2 Problem Statement

In financial time series market, the Holy Grail is to formulate an algorithm that can predict the market movement which could generate huge profit from it. This notion is not that far-fetched as companies with huge resource and deep pocket have been doing it for many years. We are not aware of it because the market somehow has been manipulated by so-called algorithm transaction. The first man-made financial crisis in 1988 was believed to be the result of the computer-generated transaction to dump the stock at the predetermined level set by the designer, which led to a catastrophic consequence or the so-called butterfly effects on the world financial market. The necessity of a proper forecasting tool to assist the financial manager to make decisions in trading is beyond any doubt. However, it is a constant struggle with the latest trading technology and information available to the market. Big data has pushed the limit of forecasting to another level where not only new hardware is required but also new software must be available to cope with the change.

1.3 General Definition

The general definition of the terminology forecasting in time series Y_t is best depicted in Figure 1 Forecasting Time Line

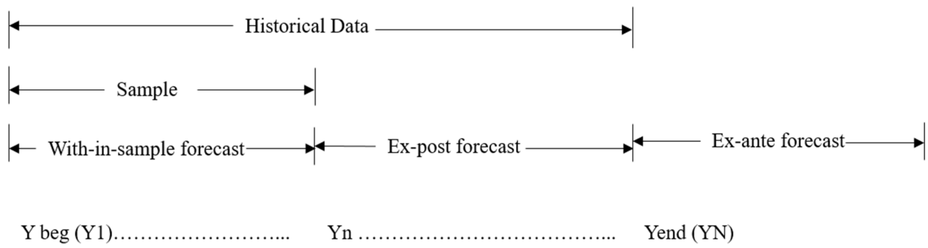

Figure 1 Forecasting Time Line

The Ex-ante forecast accuracy is the benchmark to judge the performance of the algorithm. The forecast error is the difference between the actual and the forecast values, expressed as $e_t = Y_t - \hat{Y}_t$. The notation $\hat{Y}_t$ is the predicted value which falls in Ex-ante forecast period in which no observations on the time series variable exists.

The following definitions are to measure the forecasting error

Mean absolute error

$$MAE = \left(\sum_{t=1}^{n} |e_t| \right) / n$$

Equation 1

Mean of the absolute percentage error

$$MAPE = \left(\sum_{t=1}^{n} \frac{|e_t|}{Y_t} \right) / n$$

Equation 2

Symmetric Mean of the absolute percentage error

$$SMAPE = \left(\sum_{t=1}^{n} \frac{|e_t|}{(Y_t + \hat{Y}_t)/2} \right) / n$$

Equation 3

Mean square error

$$MSE = \left(\sum_{t=1}^{n} e_t^2 \right) / n$$

Equation 4

Root mean square error

$$RMSE = \sqrt{MSE = \sqrt{\left(\sum_{t=1}^{n} e_t^2 \right)} / n}$$

Equation 5

Throughout this research, MAPE and SMAPE will be the major benchmark to compare the performance of the models. These two are selected because, in forecasting society such as Artificial Neural Network & Computational Intelligence Forecasting Competition NN3 (Artificial Neural Network & Computational Intelligence Forecasting Competition NN3, 2007), these two benchmarks are used to judge the accuracy of all the participants.

1.4 Objective of The Research

The objective of this research is to seek out the most appropriate model that can encapsulate the big data in financial time series analysis and produce a fruitful result. Like all models which depend on the quality and quantity of the input, big data has a higher chance to provide the high quality and huge quantity of data as more resources and the focus are in this domain. The notion that knowledge is power has begun to spread as whoever left behind will not be able to compete and eventually have to step out. However, it is not easy for any organization to let individuals master the technique in big data analysis. To commence with, the hardware requirement is the first obstacle. In order to cope with the computational requirement, a new computer with 2 hard disks each with 4T memory running on an i7-4770 CPU@3.40 GHz RAM 32GB 64-bit operating system x64-based processor together with a NVIDIA GeForce GTX670 2GB graphic card is acquired for this research. The computer speed of this machine is 102.3 Giga FLOPS for CPU and 3,120,223 GFLOPS for GPU. It is impossible for this research to conduct an experiment on actual big data due to the limitation of resources. Accordingly, only a fraction of the big data has been selected. Chinese stock market with high transaction historical records from 2000 to 2013 is employed in this research and the new computer is able to handle this. Conditional Restricted Boltzmann Machine as one of the deep learning techniques is used to analyse the data. The future stock price is predicted and MAPE and SMAPE of the result will be calculated to compare with other models which the author has been investigated in previous research. The results in the competition of forecasting in ENUNITE (Lin, Chen, & Chang, Load Forecasting using Support Vector Machines: A study on ENUNITE Competition 2001, 2001) and Neural Network Competition NN3 (Artificial Neural Network & Computational Intelligence Forecasting Competition NN3, 2007), have inspired our research to set the forecasting accuracy to below 2 of the MAPE value.

1.5 Motivation of The Research

The candidate's original research domain was in manufacturing optimization. The objective was to reduce the idle time in a manufacturing operation in order to cut cost. Optimization and forecasting are closely related. Classification is the first step in optimization analysis. Once the problem domain is properly classified, the forecasting technique is required to do the optimization. Manufacturing optimization is not an easy task and the research is confined to the data available from the manufacturing environment. Needless to say, each manufacturing environment has its own problem domain and it may not be applicable to other sectors. As the research hit to an obstacle that the problem domain data was not enough to carry on, the candidate decided to refocus on the financial time series forecasting domain.

As the world is relying more and more on forecasting news ranging from natural phenomena like an earthquake, eclipse or storm, to social behaviours like political vote decision and to financial market direction, it has intrigued the candidate to investigate on what the theoretical background and techniques in forecasting are. Financial time series is selected as the new problem domain simply because there are many electronic data available on the Internet. The two domains are very similar in nature and it would be effortless to transfer to another domain. Unlike the previous manufacturing domain which limited the data resources to a single industry and in fact restricted to a single enterprise, financial time series domain has a wider source of data accessible through the Internet. In the beginning of this new problem domain research, the candidate is using only the available financial data from the website with a limited size of data and dimension in order to experiment with the prevailing forecasting techniques such as NN, SVM, GARCH, ARIMA and wavelet transform. The research and experiment are conducted in MATLAB environment which equips with the proper tool on all the forecasting techniques mentioned above. As high dimensional data is employed at a later stage of this research, there is a constant struggle to reduce the dimensions of the data by using feature reduction method such as IsoMAP in order to fit the requirement of the forecasting model.

After a few years of research on this path, the candidate has discovered that financial time series is a very complicated problem domain. Although it is easy to get a pretty good forecasting result from one stock in one market using the prevailing techniques, it would be ignorant not to consider the other attributes such as the other stocks in the same market and the effect of other markets towards a particular market. For example, to forecast the close value of the Hang Seng Index in Hong Kong market using historical records of the Open, High, Low and Close values, this simple setup with 4 attributes can accurately forecast the future close value around 0.5 MAPE value on average according to our experiments in Chapter 3.2. However, it would be difficult for the tools which we have experienced to handle more attributes and high volume data. Big data is the current hot topic and almost every sector of the industry that involved in digital management must deal with it. The message hidden in big data collect either in text format from online chat like Twitter or Facebook to high volume transaction in exchange centre is a treasure waiting to be discovered. In this new era of a digital generation, raw data is abundant but useful information is getting harder and harder to find and most people get lost in this digital fortress. The demand to get new tools to discover information from big data is strong. Inspired by this request, this research attempts to find meaningful information from big data.

1.6 Outline of The Thesis

The problem domain in this research has been clearly defined in Chapter one and the scope of the experiments has been laid out. The complexity to work on big data and the potential benefit of getting useful information in financial time series from the gigantic digital fortress has been briefly depicted too. The various methodologies either from statistical or from computational intelligence background have been stated based on the current literature review in Chapter Two. The purpose of utilizing these methods is focused in financial time series forecasting. In Chapter Three, various models which have been experienced by the previous researchers relating to financial time series analysis have been reported. It contains all the experimental results using limited dataset and narrow data dimension. Analysis on the limitation of each approach has been specified. It is the path where the candidate has been guided from limited datasets to big data. Deep learning method using Conditional Restricted Boltzmann Machine (CRBM) method in the miniature of big dataset is scrutinized in Chapter Four. The advantage and disadvantage of this method have been pointed out. The ultimate performance and comparison of the new method with the previous methods have been analysed in Chapter Five. Finally, the conclusion and future development are summarized in Chapter Six.

Chapter 2 Literature Reviews on Forecasting Methodologies

2.1 Neural Network

NN has been used with success in many areas, in pattern classification (Lai & Liu, A Neural Network and CBR-based Model for Allocating Work in Progress, 2008) (Lai & Liu, WIPA:Neural network and case base reasoning models for allocating work in progress, 2012), pattern recognition (Majumdar, Majumdar, & Sarkar, An investigation on yarn engineering using Artificial Neural Networks, 2006), weather forecasting (Lee & You, iJade WeatherMan - A Multiagent Fuzzy-Neuro Network Based Weather Prediction System, 2001), data mining and knowledge discovery (Lai & Liu, A Neural Network and CBR-based Model Sewing Minute Value, 2009) (Hui & Ng, A new approach for prediction of sewing performance of fabrics in apparel manufacturing using Artificial Neural Networks, 2006) (Wong, Prediction of clothing sensory comfort using Neural Networks and fuzzy logic, 2002), stock prices (Lee, Kim, Jang, & Lim, Forecasting Short-Term KOSPI Time Series Based on NEWFM, 2008) (Wang W. , Zhao, Li, & Liu, A Novel Hybrid Intelligent Model for Financial Time Series Forecasting and its Application, 2009), foreign exchange forecasting (Lee & Liu, iJade Stock Predictor - An Intelligent Multi-Agent Based Time Sereis Stock

Prediction System, 2001) and have shown themselves to be more accurate than other AI tools, such as Genetic Algorithms (GA) and Fuzzy Logic.

In the following diagram, a simple Neural Network is a 3-layer structure which consists of input, hidden and output layers. Basically, it is a mesh wire of nodes interconnected.

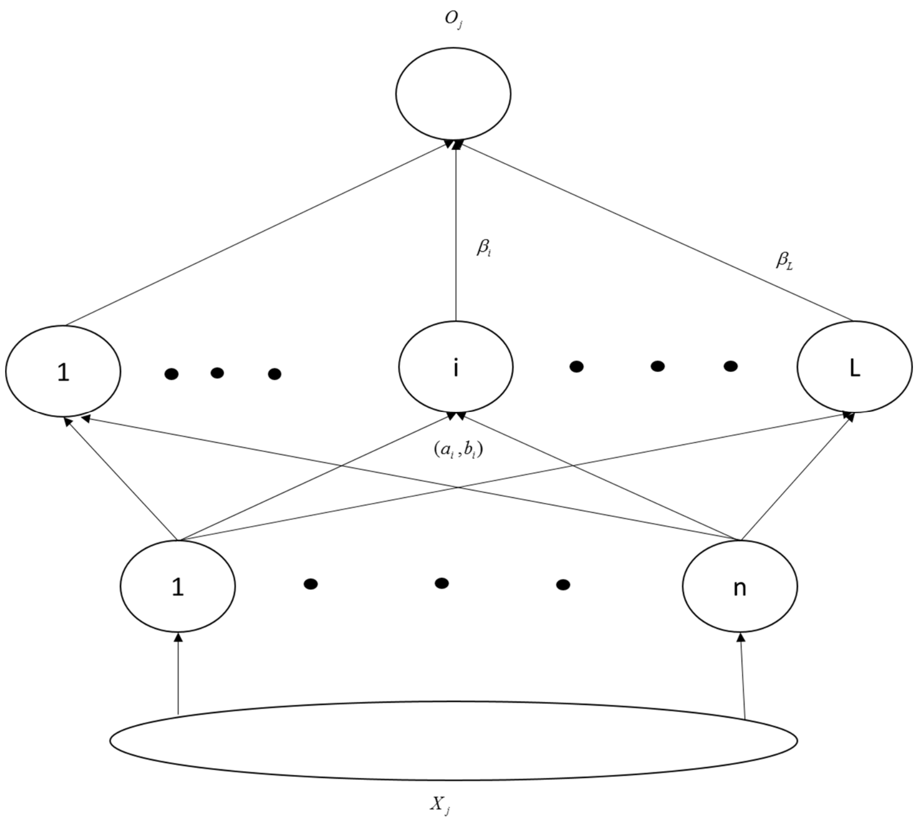

Figure 2 Schematic Diagram of Neural Network

The above feedforward Neural Network is presented in order to compare with the Extreme Learning Machine concept later on. The precise description is single hidden layer feedforward networks (SLFNs) (Liang, Huang, Saratchandran, & Sundararajan, A Fast and Accurate Online Sequential Learning Algorithm for Feedforward Networks, 2006)and the following equations define the relationship of the nodes and the corresponding weight vector.

Output of additive hidden nodes:

$$G(a_i, b_i, x) = g(a_i * x + b_i)$$

Equation 6

Output of Radial Basis Function hidden nodes:

$$G(a_i, b_i, x) = g(b_i \| x + a_i \|)$$

Equation 7

The output function of SLFNs is:

$$f_L(x) = \sum_{i=1}^{L} \beta_i G(a_i, b_i, x)$$

β_i: Output weight vector connecting the i^{th} hidden node and the output nodes

Equation 8

2.2 Support Vector Machine and Least Square Support Vector Machine

Support Vector Machine (SVM) has been used in many machine learning tasks such as pattern recognition, object classification, and with regression analysis in time series prediction in Support Vector Regression, or SVR, a methodology in which a function is estimated using observed data which in turn is used to train the SVM (Vapnik, Golowich, & Smola, Support Vector method for function approximation, regression estimation, and signal processing, 1997). It differs from traditional time series prediction methodologies in that there is no model in the strict sense – the data drives the prediction (Burges, A Tutorial on Support Vector Machines for Pattern Recognition, 2005). (Malyscheff, Trafalis, & Raman, From Support Vector Machine learning to the determination of the minimum enclosing zone, 2002) used SVR to determine the minimum enclosing zone and (Fernandex, Irma, Zanakis, Stelios, & Walczak, Knowledge discovery techniques for predicting country investment risk, 2002) used SVR to predict investment risk in a country. In (Lv & Zhang, Application of least squares support vector machine in futures price forecasting, 2011), the future price is forecasted using least squares support vector machines. The contract prices of stock index futures were forecasted in (Yang, Su, Zhou, & He, Support Vector Mchine Based Forecasting of the Contract Prices of Stock Index Futures, 2011) using support vector machine-based forecasting model. The freight volume is forecasting using a hybrid model of support vector machine and least squares vector machines (Wang, Zhao, & Zhgn, 2011).

SVR has been used in long-term stock market forecasting. (Pasila, Ronni, & Wijaya, Long-term Forecasting in Financial Stock Market using accelerated LMA on Neuro-Fuzzy structure and additional Fuzzy C-Means Clustering for optimizing the GMFs, 2008) used an accelerated Levenberg-Marquardt algorithm to predict the stock market series of the Jakarta Stock Indices over 10 months, achieving an RMSE of 1.96%. (Bao, Lu, & Zhang, Forecasting Stock Price by SVMs Regression, 2004) applied SVR to forecast the price trend for a single Chinese stock.

(Mitsdorffler & Diederich, Prediction of First-Day Returns of Initial Public Offering in the US Stock Market using rule extration from Support Vector Machines, 2008) used SVR to predict the first day returns of US stock market IPOs, but found to be accurate in only 18% of cases. (Zhai, Hsu, & Halgamuge, Combining News and Technical Indicators in Daily Stock Price Trends Prediction, 2007) claimed a profit over two months using a methodology that combined news and technical indicators. Huang et al (Huang, Nakamori, & Wang, Forecasting Stock market movement direction with Support Vector Machine, 2004) used SVR to forecast the direction of stock movements which was correct 73% of the time. (Sivakumar & Mohandas, Modeling and Predicting Stock Returns using the ARFIMA-FIGARCH a case study on Indian stock data, 2009) reported the use of SVR in financial time series prediction over a 5-day forecasting horizon.

The following is a brief description of the book (Ingo & Christmann, Support Vector Machine, 2008) SVR for nonlinear function estimation such as the financial time series. In the primal weight space, the model takes the form

$$f(x) = \omega^T \varphi(x) + b$$

Equation 9

With the given training data $\{x_k, y_k\}_{k=1}^{N}$ and $\varphi(.) : R^n \rightarrow R^{nh}$ a mapping to a high dimensional feature space which can be infinite dimensional and is only implicitly defined. Note that in this nonlinear case the vector ω can also become infinite dimensional. The optimization problem in the primal weight space becomes

$$\min_{\omega,b,\xi,\xi^*} J_P(\omega,\xi,\xi^*) = \frac{1}{2}\omega^T\omega + C\sum_{k=1}^{N}(\xi_k + \xi_k^*)$$

subject to:

$$y_k - \omega^T\varphi(x_k) - b \le \varepsilon + \xi_k, \quad k = 1, \ldots, N$$
$$\omega^T\varphi(x_k) + b - y_k \le \varepsilon + \xi_k^*, \quad k = 1, \ldots, N$$
$$\xi_k, \xi_k^* \ge 0, \text{ k=1, \ldots, N,}$$

Equation 10

Applying the Lagrangian and conditions for optimality, the following is the dual problem

$$\max_{\alpha,\alpha^*} J_D(\alpha,\alpha^*) =$$
$$-\frac{1}{2}\sum_{k,l=1}^{N}(\alpha_k - \alpha_k^*)(\alpha_l - \alpha_l^*)K(x_k, x_l) - \varepsilon\sum_{k=1}^{N}(\alpha_k + \alpha_k^*) + \sum_{k=1}^{N} y_k(\alpha_k - \alpha_k^*)$$

Such that:

$$\sum_{k=1}^{N}(\alpha_k - \alpha_k^*) = 0$$
$$\alpha_k, \alpha_k^* \in [0, c]$$

Equation 11

Here the kernel trick has been applied with $K(x_k, x_l) = \varphi(x_k)^T\varphi(x_l)$ for k, l = 1,...,N. The dual representation of the model becomes

$$f(x) = \sum_{k=1}^{N}(\alpha_k - \alpha_k^*)K(x, x_k) + b$$

Equation 12

Consider the following Vapnik's ε-insensitive loss function

$$L_{\varepsilon}(y - f(x)) = \begin{cases} 0, & \text{if } |y - f(x)| \le \varepsilon \\ L(y - f(x)) - \varepsilon & \text{otherwise} \end{cases}$$

Equation 13

Equation 13 is a convex cost function where L(.) is convex.

Primal problem

$$\min_{w,b,\varepsilon,\varepsilon^*} \frac{1}{2} w^T w + C \sum_{k=1}^{N} (L(\varepsilon_k) + L(\varepsilon_k^*))$$

subject to
$$y_k - w^T \varphi(x_k) - b \le \varepsilon + \varepsilon_k$$
$$w^T \varphi(x_k) + b - y_k \le \varepsilon + \varepsilon_k^*$$
$$\varepsilon_k, \varepsilon_k^* \ge 0$$

Equation 14

where $\varepsilon_k, \varepsilon_k^*$ are slack variables. Here, x_k is mapped to a higher dimensional space by the function φ and ξ_k is the upper training error (ξ_k^* is the lower) subject to the ε–insensitive tube $|y_k - w^T \varphi(x_k) - b| \le \varepsilon$. The parameters which control the regression quality are the cost of error C, the width of the tube ε, and the mapping function φ.

The constraints imply that we should put most data x_k in the tube $|y_k - w^T \varphi(x_k) - b| \le \varepsilon$. If x_k is not in the tube, there is an error ξ_k or ξ_k^* which we must minimize the objective function SVR to avoid under-fitting or over-fitting of the training data by minimizing the training error $C\sum_{k=1}^{N}(L(\varepsilon_k) + L(\varepsilon_k^*))$ as well as the regularization term $\frac{1}{2}w^T w$.

The Lagrangian for this problem is

$$L(\omega, b, \varepsilon, \varepsilon^*; \alpha, \alpha^*, \eta, \eta^*) = \frac{1}{2}\omega^T\omega +$$
$$c\sum_{k=1}^{N}(L(\xi_k) + L(\xi_k^*)) - \sum_{k=1}^{N}\alpha_k(\varepsilon + \varepsilon_k - y_k + \omega^T\phi(x_k) + b) -$$
$$\sum_{k=1}^{N}\alpha_k^*(\varepsilon + \varepsilon_k^* + y_k - \omega^T\phi(x_k) - b) - \sum_{k=1}^{N}(\eta_k\varepsilon_k + \eta_k^*\varepsilon_k^*)$$

Equation 15

With Lagrange multipliers $\alpha_k, \alpha_k^*, \eta_k, \eta_k^* \geq 0$ for k=1,…,N.

Dual problem

$$\max_{\alpha, \alpha^*, \eta, \eta^*} J_D(\alpha, \alpha^*, \eta, \eta^*)$$

subject to $\sum_{k=1}^{N}(\alpha_k - \alpha_k^*) = 0$

$$cL'(\varepsilon_k) - \alpha_k - \eta_k = 0, \quad k=1,\ldots,N$$
$$cL'(\varepsilon_k^*) - \alpha_k^* - \eta_k^* = 0, \quad k=1,\ldots,N$$
$$\alpha_k, \alpha_k^*, \eta_k, \eta_k^* \geq 0, \quad k=1,\ldots,N.$$

Equation 16

Equation 10 to Equation 12 are the SVM for linear function estimation mathematical expression. Equation 13 to Equation 16 are the SVM for nonlinear function estimation mathematical expression.

LSSVM regression (Suyken, et al., 2002) is closely related to regularization networks, Gaussian processes and reproducing kernel Hilbert spaces but with emphasis on primal-dual interpretations in the context of constrained optimization problems. It is relatively a new tool, there is very little research in financial forecasting using LSSVM such as (Gen & Ma, Least Squares Support Vector Regression Based CARRX

Model for Stock Index Volatility Forecasting, 2008) (Gestel, et al., Financial Time Series Prediction using Least Squares Support Vector Machines within the Evidence Framework, 2001) (Shen, Zhang, & Ma, Stock Return Forecast with LS-SVM and Particle Swarm Optimization, 2009) (Zhang & Shen, Stock Yield Forecast based on LS-SVM in Bayesian inference, 2009).

The following is a brief description of LSSVM mechanism on regression problems by the book (Ingo & Christmann, Support Vector Machine, 2008). Given a training data $\{x_k, y_k\}_{k=1}^{N}$, we can formulate the following optimization problem in the primal weight space

$$\min_{w,b,e} J_p(w,e) = \tfrac{1}{2}\omega^T\omega + C\,\tfrac{1}{2}\sum_{k=1}^{N} e_k^2$$

such that $y_k = \omega^T \phi(x_k) + b + e_k, \quad k = 1,\ldots,N$

Equation 17

Equation 17 is modified here at two points comparing with Equation (10). First, instead of inequality constraints, one takes equality constraints where the value y_k at the left hand side is rather considered as a target value than a threshold value. Upon this target value, an error variable e_k is allowed such that misclassifications can be tolerated in the case of overlapping distributions. These error variables play a similar role as the slack variables ξi in SVR. Secondly, a squared loss function e_k^2 is taken for this error variable. These modifications will greatly simplify the problem which has been demonstrated in the book (Ingo & Christmann, Support Vector Machine, 2008).

2.3 Wavelet Transform

The wavelet transform (WT) has been found to be particularly useful for analysing signals which can best be described as aperiodic, noisy, intermittent and transient (Addison, The Illustrated Wavelet Transform Handbook, 2002). It really began in the mid-1980s where they were developed to interrogate seismic signals. The application of wavelet transform analysis in science and engineering really began to take off at the beginning of the 1990s. WT and Fourier transform (FT) are very similar in nature especially FT has been around since the 1800s (Crowley, A guide to wavelets for economists, 2007). FT is built from sine and cosine functions which are periodic waves that continue forever. This approach is only good for signals that have time-independent wave-like features, signals which have more localized features for which sines and cosines do not model very well. WT is a different set of building blocks to model these types of signals (Boggess & Narcowich, A first course in wavelets with fourier analysis, 2009). WT will be tested if it can improve the forecasting accuracy of financial time series which by definition is not with time-independent wave-like features. Based on (Zhou & Tian, Predicting Corporate Financial Distress based on Rough Sets and Wavelet Support Vector Machine, 2007) work, we have developed an algorithm that combines SVM and WT to perform the test. Wavelet is a mathematical function used to divide a given function or continuous-time signal into different scale components. A wavelet transform is the representation of a function by wavelets. The wavelets are scaled and translated copies (known as “daughter wavelets”) of a finite-length or fast-decaying oscillating waveform (known as “mother wavelet”). It is widely applicable to time series analysis. In (Dai & Lu, Financial Time Series Forecasting using a Compound Model Based on Wavelet Frame and Support Vector Regression, 2008), multi-resolution discrete wavelet transforms combining with SVR technique was applied to forecast the opening cash index of Nikkei 225 with MAPE value at 0.31 which is a very good result. (Rua, A wavelet approach for factor-augmented forecasting, 2010) predicted GDP growth one- and two-quarter-ahead of Germany, France, Italy and Spain using multi-resolution discrete wavelet transforms. The best mean squared error was 65% better relative to the autoregressive benchmark in Spain but it was 10% worst in Italy. However, GDP growth cannot be compared with the financial index as the latter is more volatile.

DWT is any wavelet transform for which the wavelets are discretely sampled. It was invented by the Hungarian mathematician Alfred Haar. The most commonly used set of DWT was formulated by the Belgian mathematician Ingrid Daubechies in 1988 which is one of the methods considered in this research. This formulation is based on the use of recurrence relations to generate progressively finer discrete samplings of an implicit mother wavelet function; each resolution is twice that of the previous scale. There are a number of families in Daubechies and Haar is the first one. Daubechies wavelets are quite asymmetric, in order to improve symmetry while retaining simplicity, Daubechies proposed Symmlets as a modification to her original wavelets (also symmlets). The Daubechies and Symmlets wavelets are employed here in this research.

(Rua, A wavelet approach for factor-augmented forecasting, 2010) described the conventional factor model, and the data-generating process of each variable is the sum of two components: a component associated with factors common to all series and an idiosyncratic term. The underlying idea is that one can summarize the large information set into a small number of variables, the common factors, which retain the main features. Wavelet multi-resolution analysis allows one to decompose a time series into a low-frequency base scale and higher-frequency scales. A more detailed description on wavelet multi-resolution can be found on (Bjorn & Wim, An Overview of Wavelet Based Multiresolution Analyses , 2010). Those frequency components can be analysed individually or compared across variables. Firstly, time series are decomposed to orthogonal components of different frequencies. Then, each time scale uses a model to fit in. Finally, the overall forecast is obtained by recombining the components. (Rua, A wavelet approach for factor-augmented forecasting, 2010) only used Symmlet wavelet at level 4. Here, we used Symmlet wavelet functions with coefficients from 2 to 8 and Daubechies wavelet function coefficients from 1 to 20 for comparison. The selections of such coefficients are based on the work (Kong, Wong, Lee, & Liu, Fuzz- IEEE, 2009). The application of wavelet methodology in Financial Time Series Forecasting is also rare such as (Zhou & Tian, Predicting Corporate Financial Distress based on Rough

Sets and Wavelet Support Vector Machine, 2007) and it is always a combination with the other forecasting techniques such as rough set, support vector machine or neural network.

The discrete wavelet transform (DWT) can be written as:

$$T_{m,n} = \int_{-\infty}^{\infty} x(t)\psi_{m,n}(t)dt$$

Equation 18

where the integers *m* and *n* control the wavelet dilation and translation respectively. By choosing an orthonormal wavelet basis, $\psi_{m,n}(t)$, we can reconstruct the original signal in terms of the wavelet coefficients, $T_{m,n}$, using the inverse discrete wavelet transform as follows:

$$x(t) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} T_{m,n}\psi_{m,n}(t)$$

Equation 19

The orthonormal discrete wavelets are associated with scaling functions and their dilation equations as follows:

$$\phi_{m,n} = 2^{-m/2}\phi(2^{-m}t - n)$$

Equation 20

They have the property

$$\int_{-\infty}^{\infty} \phi_{0,0}(t)dt = 1$$

Equation 21

The scaling function can be convolved with the signal to produce approximation coefficients as follows:

$$S_{m,n} = \int_{-\infty}^{+\infty} x(t)\phi_{m,n}(t)dt$$

Equation 22

We can represent a signal x(t) with a combined series expansion using both the approximation coefficients and the wavelet coefficients as follows:

$$x(t) = \sum_{n=-\infty}^{\infty} S_{m_0,n}\phi_{m_0,n}(t) + \sum_{m=-\infty}^{m_0} \sum_{n=-\infty}^{\infty} T_{m,n}\psi_{m,n}(t)$$

Equation 23

2.4 ARIMA and GARCH

The only useful function of a statistician has been defined by William Edward Deming to make predictions (1900-1993). Autoregressive Integrated Moving Average (ARIMA) model or Box-Jenkins model has been the golden standard in prediction with strong statistical background. Despite the glory of the ARIMA model is not as shine as it used to be, this is still a very important model which this research must compare to.

The following description of ARIMA model is based on the work of (Asteriou & Hall, ARIMA Models and the Box–Jenkins Methodology, 2011). ARIMA(p,d,q) is the general form but most time series in practice do not exceed 2. Consider the following ARIMA(1,1,1)

$$(1-\phi_1 B)\nabla x_t = (1-\theta_1 B)\varepsilon_t$$

Equation 24

Which simplifies to

$$x_t = (1+\phi_1)x_{t-1} - \phi_1 x_{t-2} + \varepsilon_t - \theta_{t-1}$$

Equation 25

And ARIMA(2,1,0) process

$$(1+\phi_1 B^1 - \phi_2 B^2)\nabla x_t = \varepsilon_t$$

Equation 26

Which can be written as

$$x_t = x_{t-1} + \phi_1(x_{t-1} - x_{t-2}) + \phi_2(x_{t-2} - x_{t-3}) + \varepsilon_t$$

Equation 27

If an autoregressive moving average model (ARMA model) is assumed for the error variance, the model is a General Autoregressive Conditional Heteroskedasticity (GARCH, Bollerslev 1986) (Bollerslev, Glossary to ARCH (GARCH), 2007) model. ARIMA model or Box-Jenkins model (Box & Jenkins, Time Series Analysis: Forecasting and Control, 1970) is a standard textbook material in econometrics and finance for many years. There are many families of GARCH as described in (Hentschel, Nesting symmetric and asymmetric GARCH models, 1997) and its application is throughout the financial institutes. GARCH models are designed to capture certain characteristics that are commonly associated with financial time series such as fat tails, volatility clustering leverage effects. One branch of GARCH called Ngarch as described in (Posedel, Analysis of the exchange rate and pricing foreign currency options on the Coration market : The NGARCH models as an alternative to the Black-Scholes model, 2006) is an alternative approach to the famous Black Scholes Model. ARFIMA-FIGARCH from (Sivakumar & Mohandas, Modeling and Predicting Stock Returns using the ARFIMA-FIGARCH a case study on Indian stock data, 2009) that could predict the Indian Stock Data during the period 3 July, 1990 to 18 September 2009 accurately. In the paper by (Huang & Wu, Wavelet-Based Relevance Vector Machines for Stock Index Forecasting, 2008), GARCH prediction on NK225 has the RMSE value of 0.2013 while that of the pure SVM is 0.1820 and the best RMSE value from Wavelet-based RVM is 0.0202 while the pure SVM value is 0.182.

The following description of GARCH model is based on the work by (Engle, GARCH 101: The Use of ARCH/GARCH Models in Applied Econometric, 2011).The GARCH(p,q) model (where p is the order of the GARCH terms and q is the order of the ARCH terms) is given by

$$\sigma^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \ldots + \alpha_q \varepsilon_{t-q}^2 + \beta_1 \sigma_{t-1}^2 + \ldots + \beta_p \sigma_{t-p}^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2$$

Equation 28

2.5 Chart Pattern Matching

Chart patterns are the earliest and still a very popular tool for technical analysis. Yet, how to identify and discover the chart pattern is the most subjective part of this body of knowledge. The skill to identify chart pattern is the part that most students and even professionals have problems mastering. Sometimes, that it is referred as an art rather than science. There are 53 patterns according to the book encyclopaedia of stock patterns (Bulkowski, Encyclopedia of Chart Patterns, 2005) but only 7 patterns have been selected for study due to the limitation of the resources. They are No. 2 Broadening Formations, Right-Angled and Ascending, No. 3 Broadening Formations, Right-Angled and Descending, No. 24 Head-and-Shoulders Bottoms, No. 26 Head-and-Shoulders Tops, No. 47 Triangles, Ascending, No. 48 Triangles, Descending, No. 49 Triangles, Symmetrical, please refer to Figure 3 for their shape. These 7 patterns can be classified into 2 categories which are triangles and head and shoulder. It is rather difficult to classify how many categories in the 53 patterns from the literature but these 2 are the most common. They are chosen because the frequency of appearance in time series is very high. The identification guidelines are provided in Appendix. Chart reading takes a bit of intuition and observation. It is a skill to be developed and honed with experience. Algorithms have been written in an attempt to spot the above 7 patterns from 90 different Hong Kong equities. The results are quite promising as it will be explained in Chapter 3 Section 3.4. In the following Figure 3, 7 patterns have been selected from the book (Bulkowski, Encyclopedia of Chart Patterns, 2005).

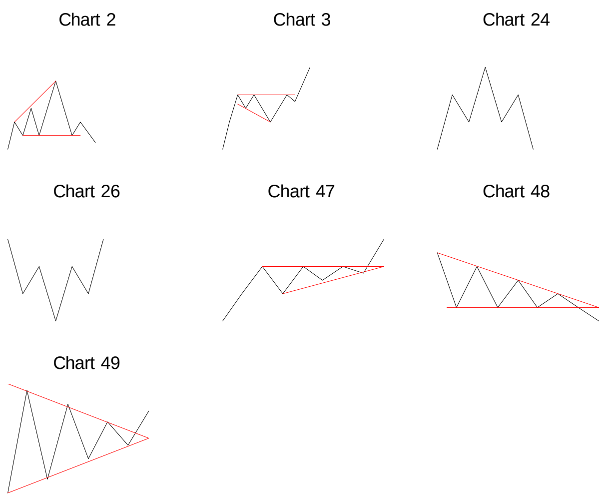

Figure 3 Seven Chart Patterns Easily Identified

2.6 Black-Scholes Formula

The following is the Black-Scholes pricing model which is based on (Chance, Derivation and interpretation of the Black-Scholes-Merton Model, 2011)

$$dS_t = \mu S_t dt + \sigma S_t dW_t$$

Equation 29

where W_t is a standard Brownian motion. It is assumed that interest rates are constant. C(S,T) denotes (stock price, price of the call option and time) the value of a call option at time t.

By Ito's lemma

$$dC(S,t) = (\mu S_t \frac{\partial C}{\partial S} + \frac{\partial C}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 C}{\partial S^2})dt + \sigma S_t \frac{\partial C}{\partial S} dW_t$$

Equation 30

Now consider a self-financing trading strategy where at each time t we hold x_t units of the cash account and y_t units of the stock. Then P_t the time 0 value of the trading strategy satisfies

$$P_t = x_t B_t + y_t S_t$$

Equation 31

Equation 31, B_t represents a certain amount of money put in a bank, S_t represents a certain amount of money to buy stock which is a representation of an option. We can choose x_t and y_t in such a way that the strategy replicates the value of the option. The self-financing assumption implies that

$$dP_t = x_t dB_t + y_t dS_t$$

Equation 32

$$dP_t = r x_t B_t dt + y_t (\mu S_t dt + \sigma S_t dW_t)$$

$$dP_t = (r x_t B_t + y_t \mu S_t) dt + y_t \sigma S_t dW_t)$$

Equation 33

Rewriting terms. We can equate terms in Equation 30 with the corresponding terms in Equation 32 to obtain

$$y_t = \frac{\partial C}{\partial S}$$

Equation 34

$$rx_t B_t = \frac{\partial C}{\partial t} + \frac{1}{2}\sigma^2 S_t^{\ 2} \frac{\partial^2 C}{\partial S^2}$$

Equation 35

P is the amount of money before replication should be the same as the call option. If we set P0 = C0 then it must be the case that Pt = Ct for all t since C and P have the same dynamics. This is true by construction after we equate terms in Equation 30 with the corresponding term in Equation 33.

We get the Black-Scholes PDE by substituting Equation 30 into Equation 35 to get:

$$C = \frac{1}{r}\left(\frac{\partial C}{\partial t} + \frac{1}{2}\sigma^2 S_t^{\ 2} \frac{\partial^2 C}{\partial S^2}\right) + S_t \frac{\partial C}{\partial t}$$

Equation 36

$$rS_t \frac{\partial C}{\partial S} + \frac{\partial C}{\partial t} + \frac{1}{2}\sigma^2 S_t^{\ 2} \frac{\partial^2 C}{\partial S^2} - rC = 0$$

Equation 37

In order to solve the partial differential Equation 36, the following original formula is copied from the book of (Bodie, Kane, & Marcus, Investments, 2005).

$$C_0 = S_0 N(d_1) - Xe^{-rT} N(d_2)$$

$$d_1 = \frac{Ln(S_0 / X) + (r + \sigma^2 / 2)T}{\sigma\sqrt{T}}$$

$$d_2 = d_1 - \sigma\sqrt{T}$$

C_0 = Current call option value

S_0 = Current stock price

N(d) = The probability that a random draw from a standard normal distribution will be less than d.

X = Exercise price

r = Risk-free interest rate (the annualized continuously compounded rate on a safe asset the same maturity as the expiration of the option, which is to be distinguished from rf, the discrete period interest rate.)

T = Time to maturity of option, in years

σ = Standard deviation of the annualized continuously compounded rate of return of the stock.

Equation 38

It is quite complicated to develop Black-Sholes formula but the fundamental concept is still normal distribution. The impact of such assumption has a serious impact on stock forecasting. It has been discovered by Lévy that stock price movement is not always normal but Lévy distribution.

2.7 Extreme Learning Machine

Sequential learning algorithm for single hidden layer feedforward network (SLFN) as per Figure 2 is the backbone of Extreme Learning Machine (ELM) (Huang, Chen, & Siew, Universal Approximation Using Incremental Constructive Feedforward Networks with Random Hidden Nodes, 2006). The following equations outline the major characteristic of the algorithm. The following description on ELM is based on (Huang, Zhu, & Siew, Extreme learning machine: theory and applications, 2006).

Output function of "generalized" SLFNs

$$f_L(x) = \sum_{i=1}^{L} \beta_i G(a_i, b_i, x)$$

Equation 39

The hidden layer output function (hidden layer mapping):

$$h(x) = [G(a_1, b_1, x), \ldots, G(a_L, b_L, x)]$$

Equation 40

The output functions of hidden nodes can be but are not limited to:

Sigmoid: $G(a_i, b_i, x) = g(a_i * x + b_i)$

Equation 41

RBF: $G(a_i, b_i, x) = g(b_i \| x - a_i \|)$

Equation 42

Here is the new learning theory – learning without iterative tuning: Given any non-constant piecewise continuous function g, if continuous target function F(x) can be approximated by SLFNs with adjustable hidden nodes g, then the hidden node parameters of such SLFNs needn't be tuned (Huang, Chen, & Siew, Universal Approximation Using Incremental Constructive Feedforward Networks with Random Hidden Nodes, 2006). There are two advantages. The first proves the existence of the networks and the learning solutions. The second is these hidden node parameters can be randomly generated without the knowledge of the training data. As a result, ELM is a very efficient algorithm as it does not rely on the iterative tuning of the parameters in the hidden nodes, compared with the NN or SVR algorithm it requires very less computational power. This latest forecasting technique has many applications such as (Cao, Lin, & Huang, Self-Adaptive Evolutionary Extreme Learning Machine, 2012) and it has attracted many researchers to conduct new experiments. NN has lost its moment for quite sometimes and many conferences and journals have lost its interest using NN. But this ELM has revived NN and provides a new perspective for the NN researchers to redefine the role and structure of the classical NN.

2.8 Conditional Restricted Boltzmann Machine

Entia non sunt multiplicanda praeter necessitatem (Entities should not be multiplied beyond necessity) by William of Ockham (1288-1348) is a famous statistical remark on the dimension of data that is necessary to do forecasting. Is the dimension a curse or blessing? When computational ability is weak, feature extraction algorithm like ISOmap is employed to reduce the dimension in order to handle the complexity. As computational power is growing stronger, the more dimensions are input to search for better results. Today, the Big data era has come and demands for new algorithm to work with the gigantic volume of data which is unprecedented in the short history of computer science. One of the prevailing algorithms is Boltzmann machine (Minh, Larochelle, & Hinton, Conditional Restricted Boltzmann Machines for Structured

Output Prediction, 2010)which is basically an energy based algorithm. It uses energies concept to define probabilities of the hidden nodes similar to the diagram as depicted in the neural network. The node with more information or strong connection has low energy while others will be assigned with high energy. The probability of the final configuration over both visible and hidden units depends on the energy of that joint configuration compared with the energy of all other joint configurations. The Restricted Boltzmann Machine (RBM) has one hidden and one visible layer with only interlayer connections. Each layer is trained independently without any reference to the second layer and the weight vector is updated with another layer after it has been trained. Deep learning is the result of training multiple layers of networks and it involves normally 6 to 15 layers by definition. Deep Boltzmann Machine (DBM) is a type of binary pairwise Markov Random Field with multiple layers of hidden random variables. In the paper – A better Way to Pre-train Deep Boltzmann Machines by (Salakhutdinov & Hinton, A Better Way to Pretrain Deep Boltzmann Machines, 2010), the authors applied the DBM to train the MNIST and NORB image datasets and demonstrated that the new pre-training algorithm can learn much better than generative models. It is the frequently used algorithm in deep learning research. RBMs have primarily been used for learning new representations or classification of data while CRBM is generally used in prediction area.

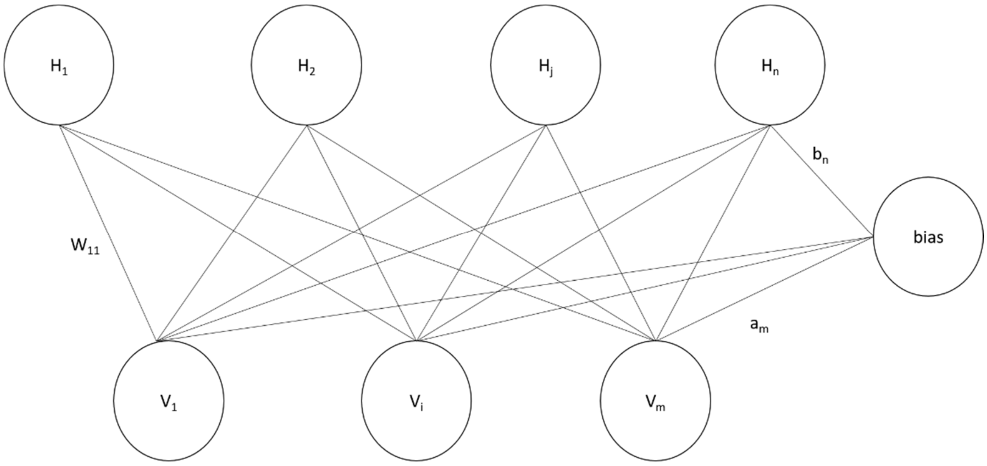

Figure 4 Restricted Boltzmann Machine

In Figure 4 m and n are the number of visible and hidden units while weight w_{ij} is assigned to each connect unit pair (v_i, h_j) and every unit has an associated bias term a_i and b_i for v_i and h_j respectively. RBM is an undirected graphical model that defines a probability distribution over a vector v and h as depicted in the above diagram.

The following description on the RBM is based on the definition of a joint probability over v and h in RBM as follows

$$p(v,h) = \exp(-E(v,h))/Z,$$

Equation 43

The energy function of a given joint configuration (v, h) is defined as

$$E(v,h) = -v^T Wh - v^T b^v - h^T b^h$$

Equation 44

To obtain p(v) one simply marginalizes out h from the joint distribution:

$$p(v) = \sum_h \exp(-E(v,h))/Z = \exp(-F(v))/Z$$

Equation 45

F(v) is called the free energy and can be computed in time linear in the number of elements in v and h:

$$F(v) = -\log \sum_h \exp(-E(v,h))$$

$$F(v) = -v^T b^v - \sum_j \log(1+\exp(b_j^h + v^T W_j))$$

Equation 46

RBMs have generally been trained using gradient descent in negative log-likelihood $-l(\Theta)$ for some set of training vectors V.

$$\log p(v) = \log \exp(-F(v)) - \log \sum_{v'} \exp(-F(v',))$$

Differentiating $-l(\theta)$

$$\frac{\partial - l(\theta)}{\partial \theta} = \frac{\partial F(v)}{\partial \theta} - \sum_{v'} \frac{\partial F(v')}{\partial \theta} p(v')$$

Equation 47

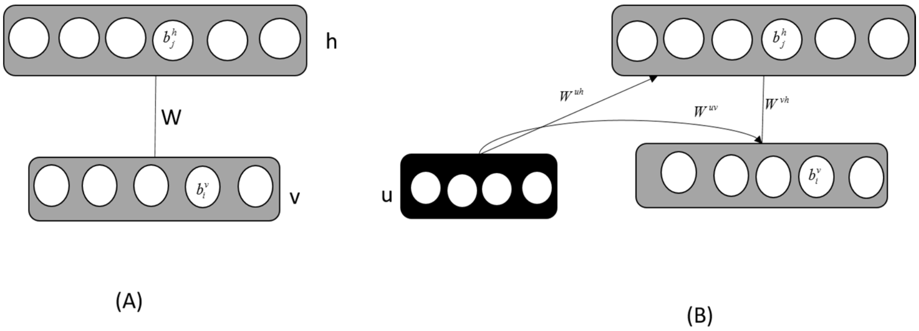

Figure 5 Schematic Diagram for Restricted Boltzmann Machine

Figure 5 is extracted from the paper, page 2 of (Mnih, Larochelle, & Hinton, 2011). RBMs have generally been trained using gradient descent in negative log-likelihood $-l(\theta)$ for some set of training vectors V with the following definition

$$\log P(v) = \log \exp(-F(v)) - \log \sum_{v'} \exp(-F(v',))$$

Equation 48

And differentiating $-l(\theta)$ with respect to some parameter θ, we get the gradient

$$\frac{\partial - l(\theta)}{\partial \theta} = \frac{\partial F(v)}{\partial \theta} - \sum_{v'} \frac{\partial F(v')}{\partial \theta} p(v')$$

Equation 49

RBM and CRBM can be compared with the following equations

$$E(v,h,u) = -v^T W^{vh} - v^T b^v - u^T W^{uv} v - u^T W^{uh} h - h^T b^h$$

Equation 50

With the associated free energy

$$F(v,u) = -\log \sum_h \exp(-E(v,h,u))$$

$$= -\sum_j \log(1 + \exp(b_j^h + v^T W_j^{vh} + u^T W_j^{wh})) - v^T b^v - u^T W^{uv} v$$

Equation 51

The CRBM model defines the following probability distribution:

$$p(v \mid u) = \frac{\exp(-F(v,u))}{\sum_{v'} \exp(-F(v',u))}$$

Equation 52

The gradient of the negative log conditional likelihood for a CRBM is given by

$$\frac{\partial - l(\theta)}{\partial \theta} = \frac{\partial F(v \mid u)}{\partial \theta} - \sum_{v'} \frac{\partial F(v',u)}{\partial \theta} p(v' \mid u)$$

Equation 53

Here, CRBM models the distribution p(v|u) by using an RBM to model v and using u to dynamically determine the biases or weights of the RBM.

Chapter 3 Analysis on Existing Methods and Limitations

3.1 The Accuracy of Three Different Stock Forecasting Models

3.1.1 Introduction

NNs and SVMs are both standard, mature machine learning approaches with applications in prediction based on times series data. They are well known to handle non-linear prediction data such as stock forecasting. NNs have been used with success in pattern classification and recognition, weather forecasting, data mining and knowledge discovery, and in time series prediction tasks such as financial market prediction of stock prices and foreign exchange forecasting. Lee iJade Stock Predictor (Lee & Liu, iJade Stock Predictor - An Intelligent Multi-Agent Based Time Sereis Stock Prediction System, 2001)and iJade WeatherMan (Lee & You, iJade WeatherMan - A Multiagent Fuzzy-Neuro Network Based Weather Prediction System, 2001) are good examples exploring the use of agent technology and artificial intelligence techniques. They have shown themselves to be more accurate than other

AI tools, such as Genetic Algorithms and Fuzzy Logic. (Wang & Li, A Novel Nonlinear RBF Neural Network Ensemble Model for Financial Time Series Forecasting, 2009) have demonstrated that a single RBF-NN model prediction on S&P500 NMSE value is 0.2389 while the ν–SVMR model NMSE value is 0.0670. In our experiment, NN model has advantages in long-term forecast during the volatile period.

SVR has been used in long-term stock market forecasting. (Pasila, Ronni, & Wijaya, Long-term Forecasting in Financial Stock Market using accelerated LMA on Neuro-Fuzzy structure and additional Fuzzy C-Means Clustering for optimizing the GMFs, 2008) used an accelerated Levenberg-Marquardt algorithm to predict the stock market series of the Jakarta Stock Indices over 10 months, achieving an RMSE of 1.96%. (Bao, Lu, & Zhang, Forecasting Stock Price by SVMs Regression, 2004) applied SVR to forecast the price trend for a single Chinese stock. (Mitsdorffler & Diederich, Prediction of First-Day Returns of Initial Public Offering in the US Stock Market using rule extration from Support Vector Machines, 2008) used SVR to predict the first day returns of US stock market IPOs, but found to be accurate in only 18% of cases. (Zhai, Hsu, & Halgamuge, Combining News and Technical Indicators in Daily Stock Price Trends Prediction, 2007) claimed a profit over two months using a methodology that combined news and technical indicators. Huang et al (Huang, Nakamori, & Wang, Forecasting Stock market movement direction with Support Vector Machine, 2004) used SVR to forecast the direction of stock movements which was correct 73% of the time. (Sivakumar & Mohandas, Modeling and Predicting Stock Returns using the ARFIMA-FIGARCH a case study on Indian stock data, 2009) reported the use of SVR in financial time series prediction over a 5-day forecasting horizon.

The performances of GARCH, NN and SVM in stock price prediction on the Dow Jones Industrial Average Index (DJ) based on 30 stocks, Hang Seng Index (HSI) based on 50 stocks and Shanghai Composite based on 1038 stocks over a 5-day and 22-day horizon respectively are thoroughly examined in this section. It is inspired by the work of (Bao, Lu, & Zhang, Forecasting Stock Price by SVMs Regression, 2004) and Lin et al (Lin, Chen, & Chang, Load Forecasting using Support Vector Machines: A study on ENUNITE Competition 2001, 2001) who both applied SVM technique in stock and power consumption prediction respectively. We carried out experiments in SVR using the software system from (Lin & Chang, LIBSVM, 2001), and that NN & GARCH using standard MATLAB command.

3.1.2 Empirical Modeling

The objective is to investigate different means of predicting the 5-day and 22-day horizons of the 3 indices market value given their historical values. The historical data of the 3 markets from years 2002 to 2007 were downloaded from the Yahoo financial website. They are organized in four datasets. The first two sets, corresponding to year 2006 and years 2002 to 2006 are used to predict the 3 market values for January of year 2007. The third and fourth sets, corresponding to year 2007 and years 2003 to 2007 are used to predict the 3 market value for January of year 2006. It has been mentioned in the early section that these 3 models are mature ones in literature but it has not been examined under an extreme situation such as the financial tsunami. The candidate has explored different parameters in the 3 models to seek out the best output.

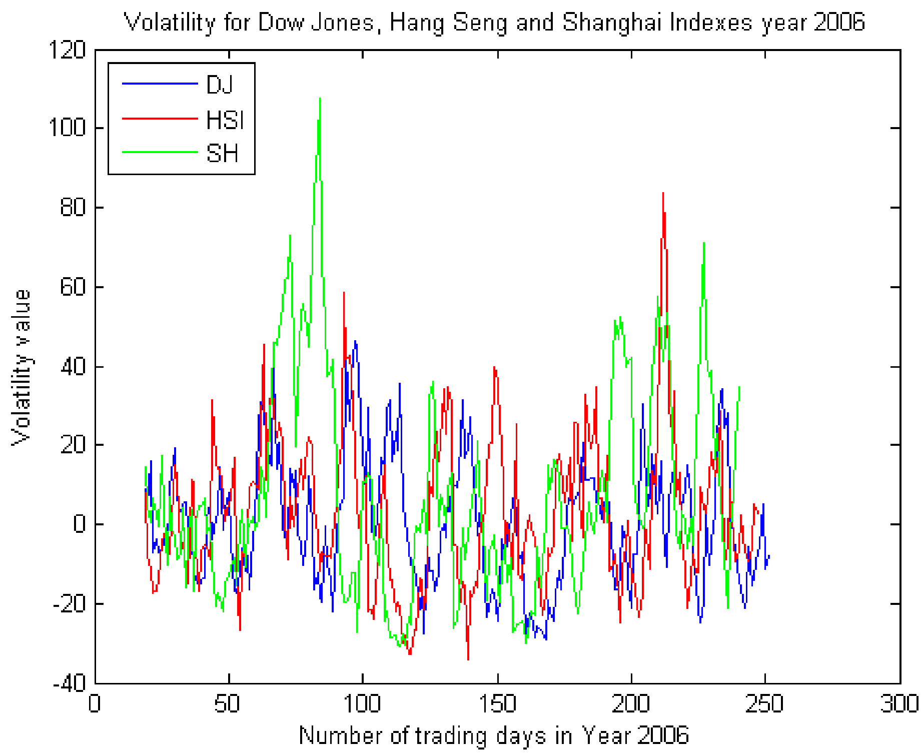

Figure 6 Data Set of Year 2006

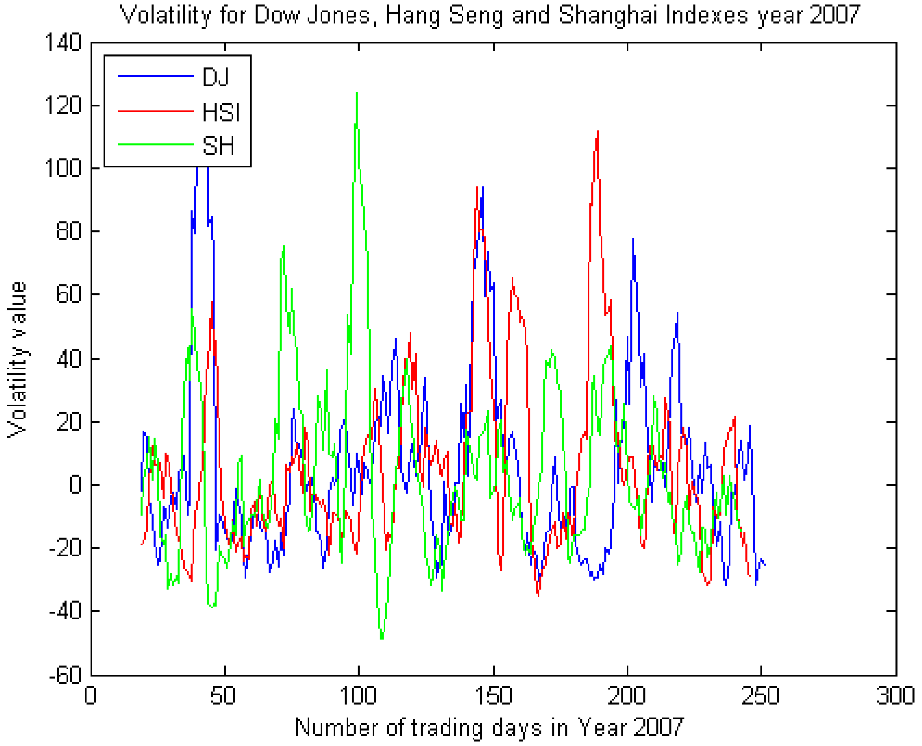

Figure 7 Data Set of Year 2007

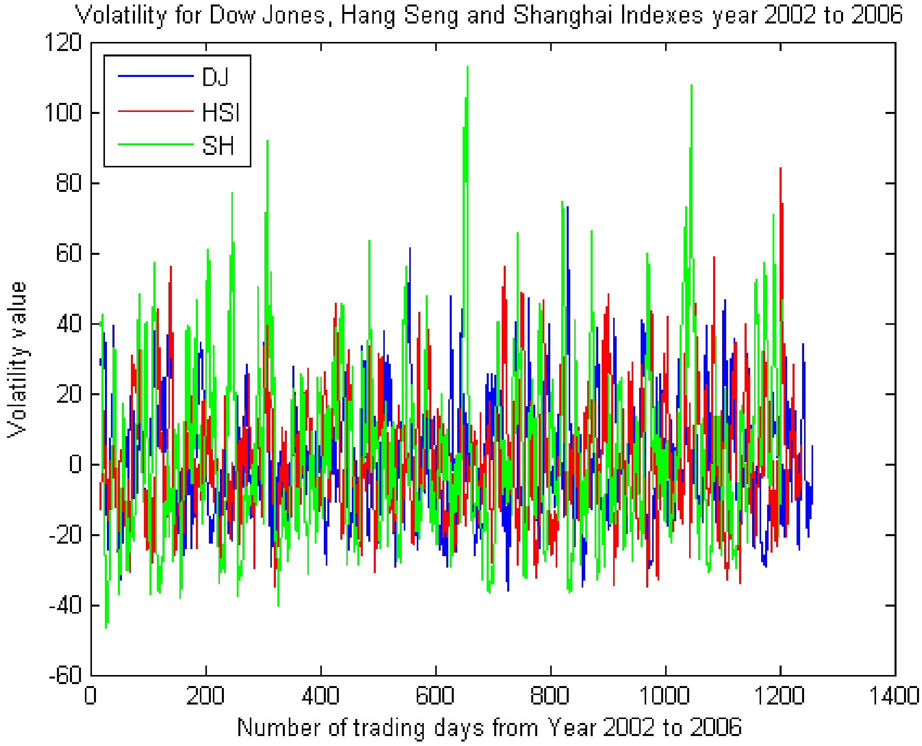

Figure 8 Data Set for Year 2002 to 2006

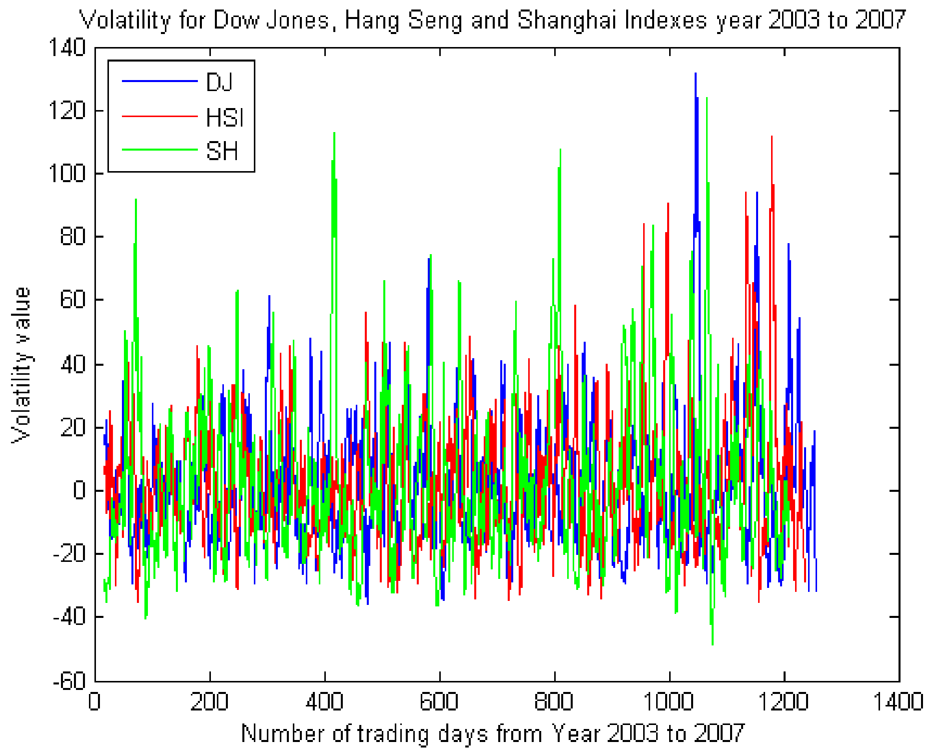

Figure 9 Data Set for Year 2003 to 2007

Table I Three Stock Indices Volatility Range and Average

		2006	2002 to 2006	2007	2003 to 2007
Dow Jones	Max	46.522	72.977	131.386	131.729
	Min	-29.196	-36.009	-31.669	-36.009
	Average	0.661	1.000	6.167	1.829
Hang Seng	Max	83.889	83.889	111.539	111.539
	Min	-34.028	-34.855	-35.116	-35.116
	Average	3.057	1.475	5.891	3.170
Shanghai	Max	107.728	113.068	124.091	124.091
	Min	-30.911	-46.634	-48.695	-48.695
	Average	7.897	2.920	3.848	3.814

Table II Three Stock Indices Kurtosis and Skewness

		2006	2002 to 2006	2007	2003 to 2007
Dow Jones	Kurtosis	2.821	2.360	2.710	2.080
Hang Seng	Kurtosis	2.333	2.254	4.042	2.198
Shanghai	Kurtosis	5.273	3.428	5.092	1.742
Dow Jones	Skewness	-0.387	0.763	0.289	-0.308
Hang Seng	Skewness	0.326	0.612	1.092	0.774
Shanghai	Skewness	0.766	0.741	1.830	-0.016

The above figures are from the high and low values of the relative datasets and MATLAB command “chaikvolat” was used to calculate on 10-period exponential moving average and 10-period difference volatility value. Despite the fact that the command “chaikvolat” is a standard tool, the value of volatility obtained from such command is still important to analyse the relationship between the volatility and the predicted result. As the 4 datasets have different records, it is impossible to put all of them in the same scale along the horizontal and vertical axes. The volatility value as depicted in Table I is a fair comparison for all these datasets. The average volatility value in Figure 6 for 2006 is 3.87 and the individual index is DJ 0.66, HSI 3.06 and SH 7.9. The average volatility value in

Figure 7 is 5.3 for 2007 and the individual index is DJ 6.17, HSI 5.9 and SH 3.85. The maximum and minimum volatility range in Figure 6 is 141.76 and in Figure 7 is 180.08. It can be concluded that the 2007 dataset is more volatile than 2006. The average volatility value in Figure 8 for 2002 to 2006 is 1.8 and the individual index is DJ 1.0, HSI 1.47 and SH 2.29. The average volatility value in Figure 9 is 2.94 for 2003 to 2007 8 and the individual index is DJ 1.83, HSI 3.17 and SH 3.81. The maximum and minimum volatility range in Figure 8 is 159.7 and in Figure 9 is 180.42. It can be concluded that the 2003 to 2007 datasets are more volatile than 2002 to 2006 datasets.

The short-term forecasting goal in this work is to predict a window of 5-day closing values for the first 5 trading days of January 2007 and January 2008 respectively. The long-term forecasting goal is to make prediction for a window of 22-day closing values for the first 22 trading days of January 2007 and January 2008.

3.1.3 Experiments and Results

The following, describes three experiments using MATLAB programming with scripts written by the candidate. In GARCH model, 3 different Variance Models are provided in the MATLAB software which is GARCH, EGARCH and GJR. The best results are selected from each model to compare with NN and SVR model. In NN model, there are 4 different types of networks provided in the MATLAB software which are

RBE, RB, GRNN and PNN. Please refer to the MATLAB manual for the details of the Variance Models in GARCH and different types of networks in NN. In SVR model, standard RBF kernel is selected.

3.1.3.1 Experiments on GARCH

The stock prices are transformed from prices to returns so that it can produce a stationary time series. This process is similar to normalization which can limit the data volatility into a narrow range. It is a technique to fit the datasets into the GARCH model. We performed the GARCH experiment in MATLAB software and selected the EGARCH model which has additional leverage terms to capture asymmetry in volatility clustering. As the dataset is from financial tsunami period, thus asymmetry in volatility clustering is assumed. The other families of the GARCH model performance are not as good as EGARCH.

Table III Prediction Result of GARCH Model

Market	Horizon	Model	Kurtosis	Skewness	Volatility	MAPE5	MAPE22
Dow Jones	2002-2006	EGARCH	2.360	0.763	1.00	1.447	4.458
	2006	EGARCH	2.821	-0.387	0.66	**1.280**	**4.063**
	2003-2007	EGARCH	2.080	-0.308	1.83	6.369	19.082
	2007	EGARCH	2.710	0.289	6.17	6.790	21.993
Hang Sang	2002-2006	EGARCH	2.254	0.612	1.47	2.054	9.914
	2006	EGARCH	2.333	0.326	3.06	1.825	9.180
	2003-2007	EGARCH	2.198	0.774	3.17	6.605	28.620
	2007	EGARCH	4.042	1.092	5.90	7.317	34.864
Shanghai	2002-2006	EGARCH	3.428	0.741	2.92	2.404	13.640
	2006	EGARCH	5.273	0.766	7.90	1.330	10.556
	2003-2007	EGARCH	1.742	-0.016	3.81	2.977	25.241
	2007	EGARCH	5.092	1.830	3.83	4.943	36.428
Average value of 3 markets on MAPE5 and MAPE23 respectively						**3.779**	**18.170**
Average value of 3 markets on MAPE5 and MAPE23							**10.974**

The data set of 2006 is used for short term forecast with 2007 5-day forecasting horizon (MAPE5 the forecasting result) while the data set of 2002 to 2006 is used for 2007 5-day forecast horizon. Both target the same 5-day forecasting horizon while the former only inputs 1 year of historical values but the latter involves 5 years of inputs. It is the same principle for data set 2007 and 2003 to 2007. The reason to pick 5-day forecasting horizon is that it is the number of trading days in 1 week. On the other hand, the long-term forecast is 22-day horizon and MAPE22 is the forecasting result. Again, the reason to pick 22-day forecasting horizon is because the number of trading days in a month is roughly 22 days. In financial time series forecasting, weekly and monthly forecasts give a general forecasting pattern.

The lowest MAPE5 value is 1.280 from Dow Jones market using 2006 data set. It is the same for the lowest MAPE22 value which is 4.063

also from Dow Jones market using 2006 data set. This is consistent with the lowest volatility value in Dow Jones 2006 dataset. In general, the higher the volatility value, the higher the MAPE value. Also, the group volatility value average in years 2006 of the 3 markets is 3.87 while the average MAPE value is 4.71. The group volatility value average in years 2007 of the 3 markets is 5.3 while the average MAPE value is 18.72. The group volatility value average in years 2002 to 2006 of the 3 markets is 1.8 while the average MAPE value is 5.65. The group volatility value average in years 2003 to 2007 of the 3 markets is 2.94 while the average MAPE value is 14.82. The co-relationship of volatility value and the MAPE value is quite obvious. However, the overall result of the GARCH model which takes the average of all the result in Table III is 10.974. Compared to the best 1.280 from the lowest MAPE5, it is not ideal at all, The GARCH model is a famous mean reverting forecasting model and it is designed to handle linear time series. But the 3 markets are definitely not linear which could explain why its prediction result is not up to the expectation. In terms of volatility ranking, the lowest volatility value is in the 2006 Dow Jones Index which is 0.66 and the model predicted MAPE5 and MAPE22 value is 12.8 and 4.063 which is also the best. The highest volatility values all happen in 2007 in all three markets and all their predicted MAPE5 and MAPE22 are almost the highest. There is an anomaly data which is 2006 Shanghai Composite Index that has the highest volatility value 7.9 but the second best MAPE5 value of 1.33 and even the MAPE5 value of 10.536 ranked the fifth. From Table II, the kurtosis value of 2006 Shanghai Composite Index is 5.273 which is highest in the group and since the value is greater than 3, it is not a normal distribution. The skewness value is 0.766, it is a positive skew which means the right tail is longer and the mass of distribution is concentrated on the right of the dataset. A possible explanation of this could be Shanghai Composite Index has its unique market characteristic and is less affected by the market of Hang Seng and Dow Jones in 2006. In 2007, the year before the financial tsunami crisis, the influence of both Hang Seng and Dow Jones to Shanghai market is higher. However, the relationship between volatility value and MAPE value is not always proportional to the ranking scale. Finally, the average MAPE value in Dow Jones is 8.185, Hang Seng is 12.547 and Shanghai 12.19.

3.1.3.2 Experiments on Neural Network

We performed the NN also in MATLAB environment. Three types of neural networks are selected for comparisons. The first one is Probabilistic Neural networks (PNN) which are used for classification problem and matching the input to a training input in order to produce a probabilistic output vector. The second is Radial Basis Neural Networks (RB). A radial basis network is a network with two layers. A hidden layer of radial basis neurons and an output layer of linear neurons. With the correct weight and bias values for each layer, and enough hidden neurons, a radial basis network can fit any function with any desired accuracy. The third is the Generalized Regression Neural Networks (GRNN) which is used for function approximation. It has a radial basis layer and a special linear layer. Open, High, Low and Close values of each trading day are the input parameters. There are more than 20,000 iterations at each run. As the results for each kernel function vary between runs, we provide the results for only the best of ten runs. The following table selected the best result from each kernel of the corresponding datasets.

Table IV Prediction Result of NN models

Market	Horizon	Kurtosis	Skewness	Volatility	Model	MAPE5	Model	MAPE22
Dow Jones	2002-2006	2.360	0.763	**1.00**	grnn	**0.273**	grnn	**1.209**
	2006	2.821	-0.387	0.66	grnn	0.351	rb	1.222
	2003-2007	2.080	-0.308	1.83	pnn	2.094	pnn	4.893
	2007	2.710	0.289	6.17	pnn	2.094	pnn	4.893
Hang Sang	2002-2006	2.254	0.612	1.47	rb	2.635	grnn	5.088
	2006	2.333	0.326	3.06	rb	2.635	grnn	5.183
	2003-2007	2.198	0.774	3.17	pnn	1.070	pnn	6.450
	2007	4.042	1.092	5.90	grnn	1.549	grnn	8.485
Shanghai	2002-2006	3.428	0.741	2.92	rb	7.321	rb	18.248
	2006	5.273	0.766	7.90	rb	7.321	rb	18.248
	2003-2007	1.742	-0.016	3.81	rb	7.593	grnn	7.891
	2007	5.092	1.830	3.83	grnn	7.436	pnn	6.929
Average value of 3 markets on MAPE5 and MAPE22 respectively						**3.827**		**7.957**
Average value of 3 markets on MAPE5 and MAPE23								**5.892**

The forecasting result in NN is much better than GARCH. The overall average in NN is 5.463 while GARCH is 10.974. Like GARCH model, Dow Jones has the lowest MAPE5 value which is 0.273 using GRNN and lowest MAPE22 value which is 1.209 also using GRNN but it has the second-lowest volatility value 1 in the 2002 to 2006 Dow Jones Market. The anomaly in section 3.1.3.1 2006 Shanghai market which has the highest volatility value 7.9 but the second best MAPE5 value of 1.33

has not been found using NN model. The average MAPE5 and MAPE22 value from GRNN is 4.163, for PNN is 5.791 and RB is 8.153. It is concluded not only GRNN has the lowest MAPE5 value of 0.273, it is also the best performer. As GRNN has both linear and non-linear function approximation abilities embedded in its structure, it seems that it is quite suitable to handle the above datasets. In terms of the best performance, EGARCH has 0.66 in MAPE5 and 1.28 in MAPE22 in Dow Jones 2006 while GRNN has 0.273 in MAPE5 and 1.209 in MAPE22. In terms of percentage there is 31% improvement and as a whole, there is 50% better than EGARCH. In each market, Dow Jones average MAPE value is 3.054 which is 168% better than EGARCH, Hang Seng average MAPE value is 6.301 which is 99% better than EGARCH and Shanghai average MAPE value is 12.829 which is 4% worse than EGARCH. Again, the relationship between volatility value and the MAPE value is not always proportional to the ranking scale.

It would seem not fair to compare GARCH model as there is only 1 input (close value) while NN has 4 inputs (open, high, low and close). However, these 2 models have different designs as the first one is mean reverting while the latter uses hidden layers concept which in theory must use more inputs to build up hidden layers. The famous Black-Scholes formula in Section 2.8 only uses the stock price as 1 input into the model. From the pragmatic point of view, there is no harm to compare as our research objective is to seek out the most useful one not necessary the best one.

3.1.3.3 Experiments on Support Vector Regression

Table V Prediction Result of SVR Model

Market	Horizon	Model	Kurtosis	Skewness	Volatility	MAPE5	MAPE22
Dow Jones	2002-2006	SVR	2.360	0.763	1.00	0.700	1.300
	2006	SVR	2.821	-0.387	**0.66**	**0.380**	**1.300**
	2003-2007	SVR	2.080	-0.308	1.83	1.140	2.740
	2007	SVR	2.710	0.289	6.17	1.940	4.550
Hang Sang	2002-2006	SVR	2.254	0.612	1.47	2.240	4.800
	2006	SVR	2.333	0.326	3.06	2.240	4.830
	2003-2007	SVR	2.198	0.774	3.17	0.800	6.120
	2007	SVR	4.042	1.092	5.90	0.820	4.330
Shanghai	2002-2006	SVR	3.428	0.741	2.92	7.320	18.250
	2006	SVR	5.273	0.766	7.90	9.740	19.240
	2003-2007	SVR	1.742	-0.016	3.81	1.850	8.890
	2007	SVR	5.092	1.830	3.83	1.850	8.630
Average value of 3 markets on MAPE5 and MAPE22 respectively						**2.585**	**7.082**
Average value of 3 markets on MAPE5 and MAPE23							**4.833**

From Table V, SVR model is the best among the three as it has the lowest average value of 4.833. Again, Dow Jones has the lowest MAPE5 value 0.38 and MAPE22 value 1.30. It also has the lowest volatility value 0.66. Despite the best forecasting result in SVR, it is the most difficult model to implement. GARCH model only requires converting the data set from daily value to daily return while NN can directly feed the daily value into the model. GARCH model has 3 different Variance Models while NN has 4 different network types for selection. It is still very easy to implement and it does not take too long to practice it. However, SVR is a completely different story. As explained in Chapter 2 Section 2.2, C is the value in Equation 10 and g is the parameter of the mapping function ϕ. These 2 parameters are very important in selection in order to gain a fruitful result. Based on our previous work (Lai & Liu, Stock forecasting using Support Vector Machine, 2010) (Lai, Hu, & Liu, A weighted Support Vector Data Description based on Rough Neighborhood approximation, 2012) there are 50 combinations of C & g parameters to test in the algorithm.

In each market, Dow Jones average MAPE value is 1.756 which is 366% better than EGARCH, Hang Seng average MAPE value is 3.273 which is 283% better than EGARCH and Shanghai average MAPE value is 9.471 which is 88% better than EGARCH. In fact, SVR has all the best average MAPE in the 3 markets. It has the lowest 4.833 MAPE value among the 3 forecasting methods. Once again, the relationship between volatility value and MAPE value is not always proportional to the ranking scale. SVR is easily compared with NN as both use 4 inputs while it is controversial to compare with GARCH model. Using the same reason in section 3.1.2.3, there is no harm to make the comparison as our research objective is to seek out the most useful one not necessary the best one.

Table VI Average MAPE Value of Each Model in Each Market

Model	Market	Average MAPE5	Average MAPE22
SVR	Dow Jones	**1.445**	**2.267**
NN	Dow Jones	3.105	3.065
GARCH	Dow Jones	3.972	10.713
SVR	Hang Seng	**2.431**	4.502
NN	Hang Seng	5.059	**6.053**
GARCH	Hang Seng	4.450	17.405
SVR	Shanghai	6.584	12.319
NN	Shanghai	9.485	**12.160**
GARCH	Shanghai	**2.914**	17.756

Table VI is a summary of the performance of 3 models in the 3 markets. Each MAPE value is the average of all the values from the data sets 2006, 2002 to 2006, 2007 and 2003 to 2007. SVR model is the best in Dow Jones MAPE5 and MAPE22 and in Hang Seng markets MAPE5. NN is the best in MAPE22 Hang Seng market and MAPE22 in Shanghai composite index. GARCH is the best in Shanghai market MAPE5. SVR is a controversial concept despite its strong statistical background which many neural network computer scientists believe it is just an extension of a new type of neural network. Nevertheless, it is not the focus here to discuss the root of the problem but to show from our experiments that both models perform very similar. From the above, SVR has 3 scores while NN has 2 scores and GARCH 1 score. Hence, it is difficult to judge whether SVR is better than NN or not. But from our experiments, NN is much easier to handle while SVR is very complicated.

Table VII Performance of Each Model in The 3 Markets

Market	Period	GARCH MAPE5	GARCH MAPE22	NN MAPE5	NN MAPE22	SVR MAPE5	SVR MAPE22
Dow Jones	2002-2006	1.447	4.458	**0.273**	**1.209**	0.700	1.300
Dow Jones	2006	1.280	4.063	**0.351**	**1.222**	0.380	1.300
Dow Jones	2003-2007	6.369	19.082	2.094	4.893	**1.140**	**2.740**
Dow Jones	2007	6.790	21.993	2.094	4.893	**1.940**	4.550
Hang Sang	2002-2006	**2.054**	9.914	2.635	5.088	2.240	**4.800**
Hang Sang	2006	**1.825**	9.180	2.635	5.183	2.240	**4.830**
Hang Sang	2003-2007	6.605	28.620	1.070	6.450	**0.800**	**6.120**
Hang Sang	2007	7.317	34.864	1.549	8.485	**0.820**	**4.330**
Shanghai	2002-2006	**2.404**	13.640	7.321	18.248	7.320	18.250
Shanghai	2006	**1.330**	10.556	7.321	18.248	9.740	19.240
Shanghai	2003-2007	2.977	25.241	7.593	**7.891**	**1.850**	8.890
Shanghai	2007	4.943	36.428	7.436	**6.929**	**1.850**	8.630

3.1.4 Conclusions

As a conclusion, the best result in the above experiments is SVR with the lowest average MAPE value of 4.833. This conclusion is consistent with the literature review that SVR outperforms many other methodologies. The average volatility value in the Dow Jones index in Figure 6 to Figure 8 is 2.414, 3.398 in Hang Seng Index and 4.62 in Shanghai Composite Index. Dow Jones Index MAPE value is the best in all 3 algorithms and except EGARCH has better performance in Shanghai composite Index than Hang Seng, the others basically follow the trend that Dow Jones, Hang Seng and the last Shanghai Composite Index. Despite the relationship of volatility value to MAPE value is not strong, it is a good indicator. In GARCH model, when the volatility value is greater than 3, there are 3 MAPE5 less than 3 MAPE value. In NN, there are 4 MAPE5 less than 3 MAPE value. In SVR, there are 6 MAPE5 less than 3 MAPE value. Based on this fact, it seems SVR has better ability to deal with volatile datasets. From Table VII, SVR has 12 best results, NN has 6 and GARCH has 6. In this respect, SVR is also a winner. As a conclusion, SVR is a good candidate that we would like to carry on with our research.

However, the overall performance of SVR in the above data sets is not consistent. GARCH outperforms SVR both in short term and long term forecast in Shanghai 2002 to 2006 datasets. Both GARCH and SVR have strong statistical backgrounds. It is inconclusive to judge which model supersedes the other as the result is not supportive enough. On the surface, SVR is better with volatile data because of its robustness in transforming data into hyperplane without knowing the mechanism of the transformation. GARCH stipulates stationary data which could be the reason why it performs better in non-volatile data especially in 2002 to 2006 with average volatility value 2.92. From the literature review, it is difficult to improve the accuracy of GARCH. As far as the NN model is concerned, it performs no good at all in Hang Seng Index. SVR and NN are very similar in structure but NN is an evolutionary algorithm which can handle a huge amount of data. Even though we only use 1 hidden layer in the NN model, it seems the complex structure has its disadvantage in short-term forecast.

It is necessary to point out that despite these three mature models, we have provided a very critical and sensitive conclusion in this section. As expected, each model is sensitive to certain data range, Table III and Table IV have demonstrated which dataset is the best for which model in both GARCH and NN models. SVR is very sensitive towards C & g parameters and in our previous work (Lai & Liu, Stock forecasting using Support Vector Machine, 2010) the C & g parameters corresponding to the result in Table V has been published. Without the critical method and parameters set in the above models, the outcome will be very different. The limitations of the above models are their heavy dependency on parameters and tuning of the models.

3.2 The wavelet-based Stock Forecasting Models Reviews

3.2.1 Introduction

(Bjorn, Questioning the Inefficient Market Hypotheses, 2003) explained that US equity returns have been predictable for many years especially in the long run. Earnings yield has had clear empirical advantages over dividend yield. Earnings yield is the benchmark on how well the company performs while dividend yield is the ability of the company to distribute its profit. It is not always a good indicator as banking and utilities sectors have steady dividend yield while new Initial Public Offering (IPO) will not be so generous. The use of dividend yield as a predictive variable leads to a basis in forecasting regression. (Leroy, Risk Aversion and the Martingale Property of Stock Prices, 1973) proved that random walk is not a sufficient and necessary condition for EMH. (Zhang Z. , Is China a weak-form Efficient Market Hypothesis market, 2001) found out that Chinese stock market cannot be classified as weak-form EMH. (Famma, Efficient Capital Markets II, 1991) proved that the β parameter of a company (which measures the extent to which returns on the stock and the market moves together) and the stock return lacks significant relationship. Capital Asset Pricing Model – CAPM is based on market portfolio but in reality, it is difficult to find. (Li Y. F., Research on Stock Value Investment Based on Artificial Intelligence, 2008) stated that CAPM is not applicable to recent Chinese stock market. (Li Y. F., Research on Stock Value Investment Based on Artificial Intelligence, 2008) also mentioned CAPM is robust but Arbitrage Pricing Theory (APT) can easily analyse all factors affecting the stock price. The proof of CAPM is rigid but not APT. In 1992, using NYSE, AMEX and NASDAQ, (Li Y. F., Research on Stock Value Investment Based on Artificial Intelligence, 2008) found out β has nothing to do with the company size. All these findings using modern investment theories could be confusing as it is difficult to draw a conclusion on how to use it. This is probably because the market is not easy to be defined and there is no single market that would not be affected by others. Today's economic model is quite different from that 10 or 20 years ago and it would make the financial forecast even more challenging. As the above result from the 3 models cannot reach lower than MAPE 2 even in 5-day forecasting horizon, it has not reached the objective of this research. The MAPE5 in GARCH model is 3.779, NN is 3.531 and SVR is 2.585. Hence, it is necessary to develop new tools and methodologies in the financial forecast as the markets are becoming more robust and complicated.

The above section 3.1 is our first attempt in our research on financial time series forecasting. The employment of NN, GARCH and SVR methodologies is pretty standard. Many of the tools have a relatively long history in the artificial intelligence society. Even the latest SVR has over 30 years of history. It is obvious from the result of section 3.1 that we need to improve the prediction accuracy of SVR by finetuning the parameters c and g. The latest SVR algorithm has provided an easy approach to optimize those parameters. However, the application is not that successful in the real world. There is not much we can improve the accuracy of the forecast by simply tuning the parameters and another approach must be deployed to seek out a better model. From the literature review, there are many research papers which combine NN, GARCH and SVR with other tools such as Wavelet transform, Genetic algorithm, Particle swarm optimization, Ant colony optimization and Relevance Vector Machine to form a hybrid algorithm. In view of this trend, GARCH, SVR and LS-SVM will be combined with wavelet-transform in this section to form hybrid algorithm. Wavelet transform technique is applied as the key hybrid algorithm in order to study the difference between hybrid algorithm and standalone algorithm. The expectation is to get a better result after the data has been transformed by wavelet technique. This approach is encouraged by the success of wavelet transform in financial forecasting in the literature review (Huang, Huang, & Hang, Financial Time Series Forecasting based on Wavelet kernel Support Vector Machine, 2012). Other financial theory such as Copula (Cherubini, Luciano, & Vecchiato, Copula Methods in Finance, 2004) or efficient market theory (Wunder & Mayo, Study Supports Efficient Maket Hypothesis, 1995) (Malkiel, The Efficient Market Hypothesis & Its Critics, 2003) will be incorporated into the new models if it is appropriate.

The objective of this section is to review the wavelet-based forecasting models through which we would like to test the predictability of the models and compare those without the wavelet-based models. The models are based on GARCH, SVR and LSSVM. They are set to forecast the actual daily close value of Hong Kong Hang Seng Index (HSI) given the past 5-year records. HSI has been selected because it reflects the semi-strong-form EMH (Bodie, Kane, & Marcus, Investments, 2005). Hong Kong being the third largest financial trading centre cannot be compared with the US market which has a very long history, enormous trading volume, pioneer of financial reform and impeccable securities law. Before Hong Kong was a follower of the US market until recently that Chinese market has a significant impact on it. Hong Kong investment advisor (Wong, Winner of Bull and Bears market, 2010) has pointed out that the Hong Kong Stock market is not efficient and in lack of volume like the US stock market to support the development of other approaches like artificial intelligence method. Wong's theory will be challenged and it will be demonstrated that the proposed models can accurately predict Hong Kong Stock market using the latest forecasting techniques. (Chen, Hardle, & Jeong, Forecasting Volatility with SVM-Based GARCH Model, 2010) predicted the volatility of stock index and (Olson & Mossman, Cross-correlations and Predictability of Stock Returns, 2001) predicted the stock returns, which is an indirect approach for the actual index value. The actual index value from these approaches may not be useful. It is well known throughout the literature that financial time series particularly stock index is non-linear. The three main factors of such time series are trend, seasonal and stochastic. These 3 factors affect the prediction result in the stock index as it is impossible to develop a model to integrate all these factors. (Kong, Wong, Lee, & Liu, Fuzz- IEEE, 2009) used Chaotic Oscillatory-based Neural Networks and Lee Oscillator to successfully catch the variability period of HSI between 2007 and 2008. But it was a pattern prediction rather than actual value forecast. The application of the stochastic factor in the stock forecast is limited, hence we focus on the trend and season and our challenge is to find out the best model for the prediction task. Despite the fact that stock index forecast has been conducted for many decades, the latest artificial intelligence techniques such as GARCH, SVR and LSSVM have improved the degree of prediction accuracy. Our objective is to seek the best algorithm from the current techniques and apply it to recent financial time series.

The following section explores the prediction performances of wavelet-based models such as WL_GARCH, WL_SVR and WL_LSSVM in stock price prediction on the Hang Seng Index (HSI) over a 4-day and 20-day forecasting horizons respectively. There are 5 trading days in a week but wavelet-based models can only deal with even number of days and hence a 4-day cycle is chosen to represent a week. In order to compare the 4-day short-term forecast, a 20-day long-term forecast is selected which is 4 weeks to represent a month. The model will give a 4-day and a 20-day ahead forecast respectively. In addition, the same datasets were employed in GARCH, SVR and LSSVM without the wavelet-based kernel as comparison. As an extension of our previous work from (Lai & Liu, Stock forecasting using Support Vector Machine, 2010) and (Lai & Liu, The Accuracy of Stock Forecasting Models -

GARCH, Neural Network and Support Vector Machine, 2012) which employed SVR in stock forecasting, here wavelet-based kernel is introduced. SVR was conducted with the software system from (Lin & Chang, LIBSVM, 2001), LSSVM was conducted using the LS-SVMLAB toolbox which was provided by Katholieke Universiteit Leuven (Katholieke Universiteit Leuven, n.d.) while the experiment of GARCH was conducted with MATLAB GARCH toolbox. The three wavelet-based algorithms, WL_GARCH, WL_SVR and WL_ LSSVM, are developed by the candidate under MATLAB environment using GARCH, SVR and LSSVM as the basic kernel.

3.2.2 Empirical Modeling

3.2.2.1 Data

The objective of this model is to predict the 4-day and 20-day horizon of HSI closing value given the historical data of HSI. The historical data of HSI during July 2005 till June 2011 is downloaded from the financial website Yahoo. They are organized per time period, a total of four datasets. The first and second datasets, during 3 July 2009 till 30 June 2010 with 248 records and during 5 July 2005 till 30 June 2010 with 1232 records are used to predict the July HSI value of the year 2010. The third and fourth datasets, during 5 July 2010 till 30 June 2011 with 248 records and during 30 June 2006 till 30 June 2011 with 1236 records are used to predict the July HSI value of 2011. It is obvious the datasets covered the notorious financial tsunami period during October 2008 till March 2009. The selection of such data range is an extension of our previous work (Lai & Liu, Stock forecasting using Support Vector Machine, 2010) which we chose the dataset during January 2002 till December 2007. Basically, the 4 datasets are separated into 5-year and 1-year records which are trained to predict the next 4-day ahead and 20-day ahead value.

A 4-day instead of 5-day forecast horizon is applied here. It is because that the discrete wavelet transform (DWT) function only accepts even numbers. In our previous work (Lai & Liu, Stock forecasting using Support Vector Machine, 2010), we adopted 5- and 22-day forecast horizon for one week and one month forecast but odd numbers cannot apply DWT. Hence, we changed it from 5 to 4. In addition, we used 5 parameters: the 15-day Exponential Moving Average (EMA15), 5-day Relative Difference in Percentage of price (RDP5), RDP10, RDP15 and RDP20 of the Hang Seng Index Close value as input into our previous models. As mentioned by (Thomason, The practitioner methods and tools, 1999), the advantage of using RDP is that the distribution of the transformed data will become more symmetrical and will follow more closely to a normal distribution. In (Cao, Zhan, & Wu, Application of SVM in Financial Research, 2009), EMA15 is used to maintain as much information contained in the original closing price as possible since the application of the RDP transform to the original closing price may remove some useful information.

The historical data of HSI during August 2003 till June 2009 is downloaded from Yahoo financial website and it is separated into two datasets. The first during 5 July 2007 till 30 June 2009 with 488 records is used to predict the 4-day with a sliding window of 248 days which is roughly one-year records. The first shift-window during 5 July 2007 till 8 July 2008 is used to predict the next 4-day from 9 July 2008 onward. The next shift-window during 11 July 2007 till 14 July 2008 is used to predict the next 4-day from 15 July 2008 onward. Totally, there are 60 results. Another set, during 18 August 2003 till 30 June 2009 with 1448 records is used to predict the 20-day with a sliding window of 248 days. The first shift-window during 18 August 2003 till 16 August 2004 is used to predict the next 20-day from 17 August 2004 onward. The next shift-window during 16 September 2003 till 13 September 2004 is used to predict the next 20-day from 14 September 2004 onward. Totally, there are 60 results. The above data range is a test of the model robustness to highly volatile market as it ended near the financial tsunami. As a summary, one-year sliding window of 248 days is applied to the 488 records (5.7.2007-30.6.2009) to predict the stock price in the next 4 days, and to the 1448 records (18.8.2003-30.6.2009) in order to predict the stock price in the next 20 days. The purpose is to test the general forecasting ability of each model.

Using the same methodologies, two sets of index values of Shanghai composite Index and Dow Jones Index with the same record length and roughly the same period (Shanghai composite index 17.7.2003-30.6.2009 & 3.7.2007-30.6.2009 and Dow Jones 30.9.2003-

30.6.2009 & 25.7.2007-30.6.2009) were analysed by these models. As mentioned in the introduction, Shanghai composite index – China stock market is a weak-form EMH, HSI – Hong Kong stock market is semi-strong-form EMH and Dow Jones Index – US stock market is a strong-form EMH. Our purpose is to put these 3 markets to test the above models and hypothesis that strong form EMH should perform better than weak form. It also provides a foundation that our models can handle all kinds of markets and their robustness in handling extreme data values during financial tsunami. The unprecedented financial tsunami is once-in-a-lifetime experience for all financial institutions to handle. Compared with the financial crisis back in 1997 due to the collapse of Long Term Capital Management, the magnitude is far greater. The following figures are the characteristics of these data ranges.

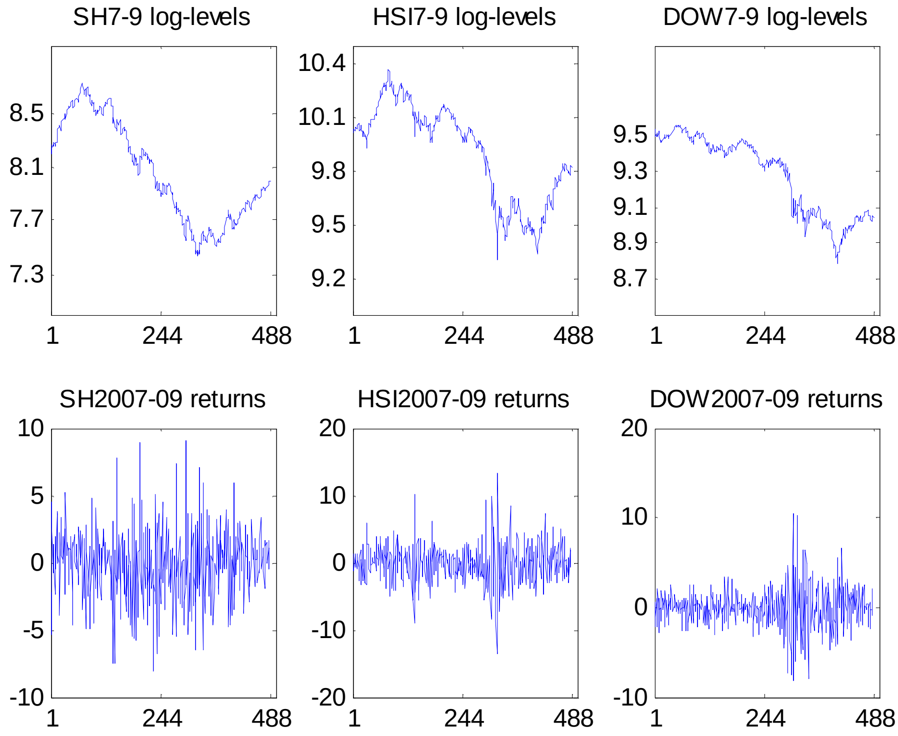

Figure 10 Shanghai, HSI and Dow Jones Indices from 2007 to 2009

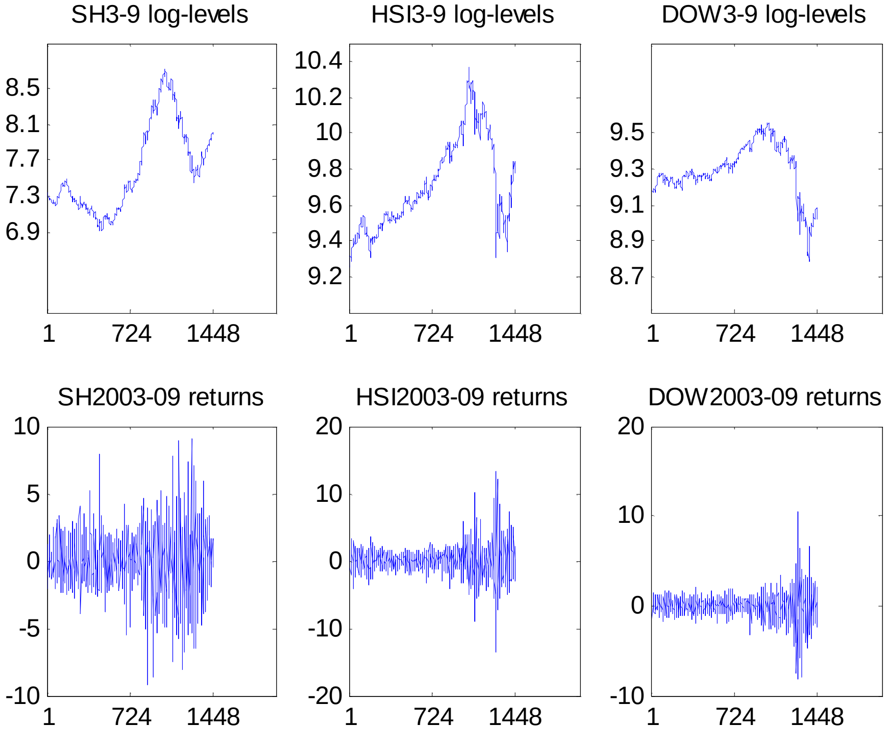

Figure 11 Shanghai, HSI and Dow Jones Jones Indices from 2003 to 2009

Only one parameter, the daily close value is used and a new data pre-processing technique – windowize is considered. It makes a nonlinear Auto Regressive predictor with a nonlinear regressor. The last elements of the resulting matrix will contain the future values of the time series, the others will contain the past inputs. The following is a simple example.

$$A = \begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \\ d_1 & d_2 & d_3 \\ e_1 & e_2 & e_3 \\ f_1 & f_2 & f_3 \\ g_1 & g_2 & g_3 \end{bmatrix}$$

Figure 12 Matrix A

W=windowize(A,[1 2 3])

$$W = \begin{bmatrix} a_1 & a_2 & a_3 & b_1 & b_2 & b_3 & c_1 & c_2 & c_3 \\ b_1 & b_2 & b_3 & c_1 & c_2 & c_3 & d_1 & d_2 & d_3 \\ c_1 & c_2 & c_3 & d_1 & d_2 & d_3 & e_1 & e_2 & e_3 \\ d_1 & d_2 & d_3 & e_1 & e_2 & e_3 & f_1 & f_2 & f_3 \\ e_1 & e_2 & e_3 & f_1 & f_2 & f_3 & g_1 & g_2 & g_3 \end{bmatrix}$$

Figure 13 Matrix A Transform to W after Windowize

Windowize is the relative index of data points in matrix A, data points are selected to make a window. Each window is put in a row of matrix W. The matrix W contains as many rows as there are different windows selected in A. The matrix A with dimension 3x7 is transformed into matrix W with dimension 9x5. Obviously, it is not dimensionality reduction. The novelty of this technique is that the last elements of the resulting matrix W will contain the future values of the time series, others will contain the past. In this way, the data distribution is widening more than the RDP and EMA method. It has been discovered this method outperforms the RDP as it is easier to apply. (Cao, Tay, & Francis, Support Vector Machine With Adaptive Parameters in Financial Time Series Forecasting, 2010) employed RDP5, RDP10, RDP15 and RDP20 to perform the same function as the windowize.

3.2.2.2 Forecasting Models and Parameters

Appendixes A to F has illustrated the 6 algorithms of the models which are developed by the candidate based on the guideline provided by the literature. There are parameters in each model that require the algorithm to search in order to get the best result. Based on our previous work (Lai & Liu, Stock forecasting using Support Vector Machine, 2010), the C and g parameters are set to 500, 1,000, 5,000, 10,000, 20,000, 40,000 and g 1, 2 for the SVR and WL_SVR model. For the wavelet-based kernel, discrete wavelet transform is used and two types of methods are employed. The first is Daubechies with coefficients from 1 to 20 and the other is Symmlet with coefficients from 2 to 8.

3.2.2.3 Empirical Results

The following tables described the result from the simulation of the models.

Table VIII Three Indices Markets Performance

Index Average	SH	SH	Hang Sang	Hang Sang	Dow Jones	Dow Jones	Average 3 markets	
Data sets range	2007_9	2003_9	2007_9	2003_9	2007_9	2003_9	2007_9	2003_9
Volatility	3.521	5.365	4.355	2.937	0.391	1.252		
Kurotsis	3.212	3.646	1.859	2.994	1.534	2.816		
Skewness	0.676	1.299	-0.106	0.928	-0.311	-0.170		
Forecast Horizon	MAPE4	MAPE20	MAPE4	MAPE20	MAPE4	MAPE20	MAPE4	MAPE20
SVR	**1.376**	6.490	2.879	4.139	2.090	**2.524**	2.115	4.384
WL_db_svm	1.510	21.487	3.537	9.218	2.578	6.003	2.542	12.236
WL_sym_svm	1.697	22.797	4.538	8.955	4.076	6.583	3.437	12.778
LSSVM	2.092	**3.949**	**2.469**	**2.787**	**1.914**	3.949	2.159	3.562
WL_db_lssvm	2.779	7.718	3.604	5.343	2.301	7.718	2.894	6.926
WL_sym_lssvm	3.198	6.910	3.930	4.582	2.585	6.910	3.238	6.134
Garch	6.320	20.694	7.950	16.485	6.543	12.690	6.938	16.623
WL_db_garch	8.072	24.522	6.721	12.546	4.517	7.193	6.437	14.753
WL_sym_garch	3.200	20.580	3.328	10.622	2.447	6.150	2.992	12.451

In Table VIII, the average MAPE4 and MAPE20 of the 60 results in each model is displayed and LSSVM gives the best result because 4 out of 6 MAPE values are the lowest. The improvement of MAPE accuracy in wavelet functions only happen in GARCH model. The average of the MAPE4 and MAPE20 in Shanghai Composite Index is 9.188, average volatility is 4.443 average kurtosis is 3.429 and average skewness is 0.988. The average of the MAPE4 and MAPE20 in Hang Seng Index is 6.313, average volatility is 3.646, average kurtosis is 2.427 and average skewness is 0.411. The average of the MAPE4 and MAPE20 in Dow Jones Index is 4.932, average volatility is 0.822, average kurtosis is 2.175 and average skewness is -0.241. This is an amazing finding that the lower the values in volatility, kurtosis and skewness, the more accurate the forecasting result (less value in MAPE). In fact, volatility, kurtosis and skewness values are in descending order from Dow Jones to

Shanghai Indices. It is obvious that Dow Jones outperforms the other indices in this exercise as it has the least MAPE value 4.932. From the literature and the definition of EMH in the book (Bodie, Kane, & Marcus, Investments, 2005), the market is more predictable if it is EMH which means the MAPE value should be smaller. From the above experiment, Dow Jones MAPE4 and MAPE20 average MAPE is 28.01% less than HSI and 86.31% less than Shanghai Composite Index. From the statistical benchmark point of view, Dow Jones Index kurtosis value is 2.175 lowest among the 3 indices. It is less than 3 and should behave like the normal distribution and it also has the lowest skewness value -0.241 which has a slightly negative skew with a long tail on the left side. Hang Seng Index kurtosis is still within value 3 and its skewness value is 0.411 which is positive skew and a long tail at the right side. Hang Seng Index is similar to Dow Jones Index except the skewness is bigger which mean the tail is also thicker. Shanghai Composite Index kurtosis is higher than 3 which define it is not a normal distribution. Its skewness value is 0.988 which is a positive skew with a long tail on the right and it is even thicker than Hang Seng Index. This confirms our speculation that strong-form EMH market should get a better result in the above models. Dow Jones Index and Hang Seng Index average MAPE4 and MAPE20 values are very close suggesting that the US and Hong Kong security market are closely related most likely due to Hong Kong currency peg with the US. The US and Hang Seng Index have large average MAPE difference 4.257 and 2.857 to Shanghai composite index. It is interesting to point out in the famous financial tsunami crisis in 2008, the property of the datasets from

2007 to 2009 had slightly changed in the order of volatility. Shanghai Composite Index has less volatility value 3.521 than Hang Seng Index 4. However, the volatility of the Dow Jones Index still remains the lowest 0.392. Even China was not affected by the crisis due to its closed economic structure while the US was severely impacted by it but the stock market behaviour was not. It followed the US and China stock market trend. After the 1997 financial crisis, the Hang Seng Index constituent stocks included many Chinese stocks in it. Thus, the stock market of US has been impacting on China market. Concerning the above 6 algorithms, each of them has a MAPE4 forecast window for 2007 to 2009 in order to forecast 60 sliding windows or time frames. Likewise, the MAPE20 forecasts windows for 2003 to 2009 in order to forecast 60 sliding windows or time frames. Thus, it is an average forecasting result with a 60 MAPE4 and MAPE20 sliding windows for the corresponding period. It can capture all the important events and reflected all the characteristics of the markets. The average MAPE4 and MAPE20 MAPE of each market in Table VIII is a true representation of the forecasting pattern of the 3 markets. Compared to the result in section 3.1 from Table III to Table VI, they both indicated a similar conclusion that MAPE value from Dow Jones Index is the best.

In general, the improvement of accuracy using wavelet function also only happens in GARCH models. The degree of accuracy in GARCH and its wavelet function are poor compared with that of SVR and LSSVM. As explained in our data section, the pre-processing data method in GARCH cannot use windowize method and it is very likely why its result is so poor. The strength of GARCH is its flexible adaptation of the dynamics of volatilities and its ease of estimation when compared to other models. It is a return-based model but it might neglect the important intraday information. E.g. when today's closing price is equal to last day's closing price, the price return will be zero, but the price variation during today might be volatile. (Li H. Q., Forecasting Financial Volatility using Intraday Information, 2010) explained the model is not able to capture the information. Despite the renowned reputation in GARCH and previous work on the successful application of GARCH with wavelet based kernel to financial time series, our experiment cannot repeat the same result. However, the effect of wavelet based kernel is still a major contributing factor in the overall result in GARCH model. Perhaps another type of GARCH model should be employed to achieve a better result. This will be in our future work and not the scope of this research. In this section, the focus is to compare and identify the fundamental factors that cause the difference in different models and markets. We simply provide the best model for the above exercises based on our findings. The following table is a matching of which model is better in a given input data.

Table IX Three Markets Performance Max and Min Difference

Index	Sh Composite	Sh Composite	Hang Seng	Hang Seng	Dow Jones	Dow Jones
Data sets range	2007 to 2009	2003 to 2009	2007 to 2009	2003 to 2009	2007 to 2009	2003 to 2009
Forecast Horizon	MAPE4	MAPE20	MAPE4	MAPE20	MAPE4	MAPE20
SVR	7.342	19.078	21.176	32.105	18.679	21.268
WL_db_svm	6.228	95.153	15.793	44.205	14.734	28.063
WL_sym_svm	8.266	91.076	23.396	42.931	15.505	38.824
LSSVM	9.160	18.449	7.676	9.720	12.840	18.449
WL_db_lssvm	8.148	24.504	14.627	29.549	12.927	24.504
WL_sym_lssvm	8.513	17.538	18.080	21.668	15.072	17.538
Garch	16.070	51.949	27.157	69.700	20.388	74.682
WL_db_garch	19.685	96.633	26.019	66.376	18.026	39.089
WL_sym_garch	6.735	81.815	13.722	56.916	14.955	32.631

Table IX shows the difference between the maximum and minimum MAPE of the 60 results. This is crucial when selecting which model to use in forecasting. Remember these results are from the extreme volatile period caused by financial tsunami. Combining Table VIII and XX, Shanghai composite index in SVR model has the best average 1.3755 and the least difference 7.3422 in 4-day forecast. It is very likely that China stock market is still a close market and the impact of financial tsunami is small. In HSI, LSSVM model has the best average 2.4693 and least difference 7.6761 for 4-day and best average 2.7868 and least difference 9.7202. It should be noted that SVR has the best average 2.8785 and least difference 21.174 for 4-day and best average 4.1385 and least difference 32.1048 which is second to LSSVM in terms of accuracy. As far as the objective is concerned, we need to find out which is the best model for HSI forecast. From Table VIII and Table IX, it is obvious the choice is LSSVM but points to SVR. As XX is from the most current data while Table VIII and Table IX are average MAPE values throughout a fixed forecasting horizon, our final recommendation is SVR. Despite a bigger difference value, it has the smallest MAPE 0.4037. XX illustrates the accuracy of the forecasting tool while Table VIII and Table IX show the average forecasting ability during a fixed horizon. For the second choice, LSSVM is a good candidate for financial advisors for their decision making.

Table X Descriptive Statistics for Three Stock Indices during 2007 to 2009

Returns	SH Composite Index			Hang Seng Index			Dow Jones Index		
	Statistics	p-value	h-value	Statistics	p-value	h-value	Statistics	p-value	h-value
Mean	-0.0567			-0.0393			-0.1006		
Variance	6.1035			7.8655			4.2002		
Skewness	-0.0332			0.1709			0.1807		
Kurtosis	4.1061			6.1697			7.1703		
Normality	24.9141	0	1	206.2428	0	1	355.5439	0	1
Q(6)	6.0892	0.4133	0	5.427	0.4903	0	29.6717	0	1
Q(6)*	13.2112	0.0398	1	191.5078	0	1	195.1023	0	1
ARCH(6)	11.7167	0.0686	**0**	96.4186	0	1	112.366	**0**	1

Table XI Descriptive Statistics for Three Stock Indices during 2003 to 2009

Returns	SH Composite Index			Hang Seng Index			Dow Jones Index		
	Statistics	p-value	h-value	Statistics	p-value	h-value	Statistics	p-value	h-value
Mean	0.0452			0.0385			-0.0065		
Variance	3.536			3.2458			1.7012		
Skewness	-0.2169			0.0918			0.0575		
Kurtosis	5.999			12.3643			14.7956		
Normality	553.6119	0	1	5289	0	1	8390	0	1
Q(6)	19.0444	0.0041	1	9.8543	0.1309	0	63.3866	0	1
Q(6)*	128.4139	0	1	852.7444	0	1	839.8699	0	1
ARCH(6)	83.7537	0	1	366.6877	0	1	412.3289	**0**	1

Notes : Normality is the Bera-Jarque(1981) normality test;Q(6) is the Ljung-Box Q test at 6 order for raw returns;Q(6)*is LB Q test for squared returns;ARCH(6) Is Engle's (1982) LM test for ARCH effect.

Table X and Table XI report the summary of the descriptive statistics for various stock indices during the two periods based on log-return analysis. If skewness is negative, it shifts to the left and vice versa. If it is a normal distribution, kurtosis is 3. When kurtosis is greater than 3, it is more outlier-prone than normal distribution and vice versa. When normality h = 1, it is a normal distribution. When Q(6) h = 1, the statistic of raw returns indicates significant autocorrelation. When Q(6)* h = 1, the statistic of squared raw returns indicates significant correlation. When ARCH(6) h = 1, ARCH result shows significant evidence in support of GARCH effects (i.e. heteroscedasticity). Shanghai composite series has the lowest kurtosis value 4.1061 in 2007 to 2009, others are typically characterized by excessive kurtosis more than 5 and asymmetry. It can be concluded that the above series are characterized by heteroscedasticity and time-varying autocorrelation; therefore, GARCH class models should fit the dataset. However, it is only a statistical conclusion on whether a dataset is suitable for GARCH model or not. Also, it is clearly from the statistical point of view all the above datasets will not behave like normal distribution as all of their kurtosis value is more than 3. As seen from Figure 9, Figure 10, Table X and Table XI all series exhibit more variability, skewness, kurtosis and volatility clustering such that nonlinear asymmetric EGARCH model should fit it more accurately. In section 3.1.3.1Table III, all the result are generated from the EGARCH model. The statistical indicator here is not useful as the result from Table

VIII and Table IX clearly demonstrated that GARCH model is not good in the above forecasting experiment.

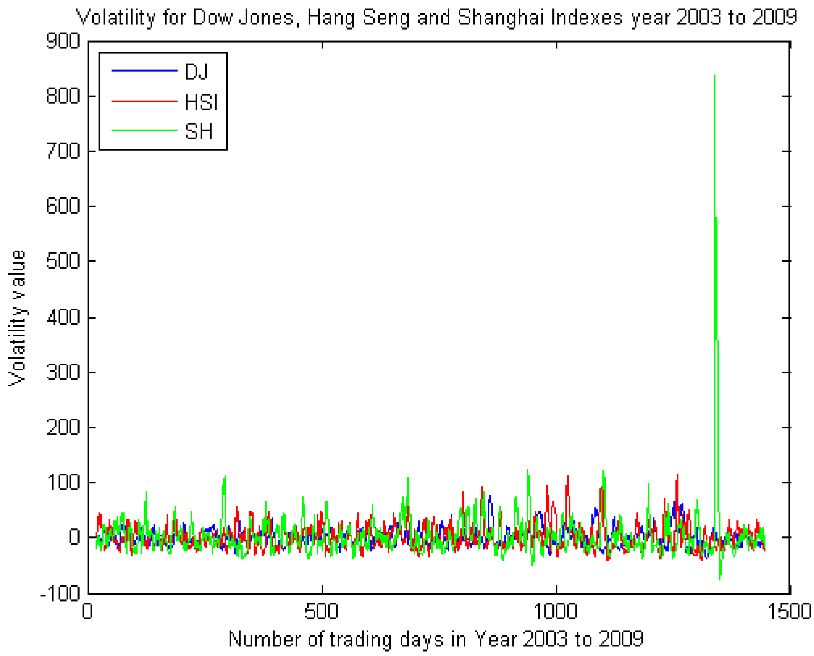

Figure 14 Volatility of The Three Markets from 2003 to 2009

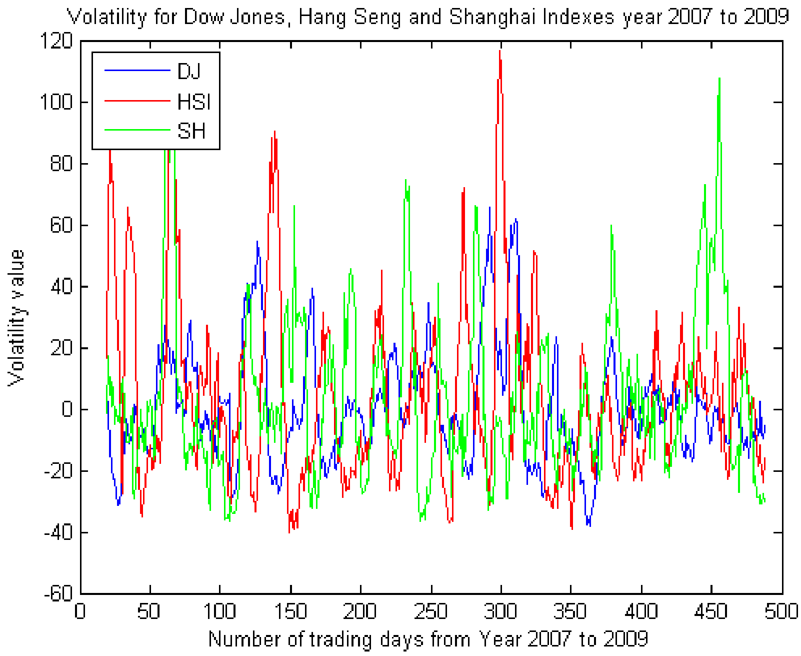

Figure 15 Volatility of The Three Markets from 2007 to 2009

Table XII Three Stock Indices Volatility Range and Average

Index	Range	2006	20026	2007	20037	20079	20039
Dow Jones	Max	46.522	72.977	131.386	131.729	65.726	77.071
	Min	-29.196	-36.009	-31.669	-36.009	-38.410	-38.410
	Average	0.661	1.000	6.167	1.829	0.391	1.252
Hang Seng	Max	83.889	83.889	111.539	111.539	116.347	116.347
	Min	-34.028	-34.855	-35.116	-35.116	-40.099	-40.099
	Average	3.057	1.475	5.891	3.170	4.355	2.937
Shanghai	Max	107.728	113.068	124.091	124.091	113.061	837.527
	Min	-30.911	-46.634	-48.695	-48.695	-36.586	-74.053
	Average	7.897	2.920	3.848	3.814	3.521	5.365

Table XIII Three Stock Indices Kurtosis and Skewness

		2003 to 2009	2007 to 2009
Dow Jones	Kurtosis	2.816	1.534
Hang Seng	Kurtosis	2.994	1.859
Shanghai	Kurtosis	3.646	3.212
Dow Jones	Skewness	-0.170	-0.311
Hang Seng	Skewness	0.928	-0.106
Shanghai	Skewness	1.299	0.676

Figure 14 and Figure 15 are the corresponding volatility value of this section datasets. Table XII depicts all the details of the volatility of the datasets for the section 3.1 and section 3.2.

3.2.3 Conclusions

Based on EMH, we have tested our models with 3 markets. The winner is LSSVM model as it produces the best MAPE4 and MAPE20 with value 2.86 and can perform equally well in the 3 markets. To continue our previous work using artificial intelligence in financial forecast, we have illustrated that it is possible to get MAPE forecast value under 2 from our experiment. The accuracy for long-term forecast in our case, i.e. 20-day or one month is always difficult but our results have demonstrated that it is still possible to get MAPE under 2. We believe this is a significant improvement and a very useful tool in financial time series analysis. The decision maker can rely on our models to analyse the market trend or benchmark for investment portfolio. From our experiment, it is a tedious task to search for the right parameters for our models and so far there is no simple solution to the above problem. The science of forecasting is still relying on trial and error approach. However, our experiments have provided a consistent approach which is to optimize the parameters using the recent historical data. The disadvantage could be time-consuming but it seems the ends justify the means if the objective is achieved.

The wavelet-based models in our experiments did not produce the best result as we would have expected. It is necessary to point out that there is no ground to compare the above result with Chapter 3 Section 3.1 as they are using different data set range. Also, it is too primitive at this stage to conclude wavelet-based model is weaker than the ordinary model as it is highly dependent on the data and the parameter setting. We may never know which is better unless we exhaust all the resource to test all available data set and wavelet-base algorithm which is beyond the scope of this research. One important finding in this section is the statistical inference on the trend or pattern of the data set has a significant influence on the choice of model. This means from the statistical point of view, the identification of the data set is an important step to choose the right model.

For example, the dataset which exhibits normal distribution behaviour ARIMA or GARCH model can be easily deployed with fruitful result. For the time being, we believe our models are sufficient to handle the current market demand even under extreme condition such as the financial tsunami. The limitation of wavelet-based forecasting is the difficulty in analysing the wavelet. To choose the right algorithm to analyse the wavelet is the key to success but it is not easy as there are many wavelet forms to choose from. This will make the forecasting difficult and the result not promising.

The study on forecasting based on EMH is rare in the literature review. (Liu Y. X., Discussing on Trend to Efficient Market Hypothesis of Securities and Futures Market, 2009) and (Timmermann & Granger, Efficient Market Hypothesis and Forecasting, 2004) are one of the few but there is no indication or benchmark on the correlation between EMH and the forecasting result. The research on the correlation between fuel prices and electricity market price by (Zhe, Zhao, & Sanderson, The Efficient Market Hypothesis and Electricity Market Efficiency Test, 2005) can only demonstrate the existence of the historical correlation but not the forecasting prices of both fuel and electricity market price. Recently research on financial time series forecasting using EMH has been neglected mainly because it is surprisingly difficult to test and considerable care has to be exercised in empirical tests. It is necessary to point out that this research only used the market dataset set which satisfies the EMH definition such as Dow Jones Industrial Average Index. It does not attempt to proof if it is EMH or not, which is beyond the scope of this research. The above result demonstrates a strong correlation between the forecasting result and the actual market behaviour of EMH, which is the contribution of this paper. An EMH weak form market such as Shanghai Composite has the largest average MAPE values from 3.021 to 16.297 in the above experiment. An EMH strong form market such as Dow Jones Industrial Average has 6 out of 9 best average MAPE value in the above experiment. Hence, the experiment has proofed strong form EMH has a better result than the weak form EMH using different AI forecasting technique.

3.3 Chart Pattern and Least Square Support Vector Machine in Stock Forecasting

3.3.1 Introduction

The analysis of Chart Pattern has been conducted for many decades. It is an important investment tool which has been employed by investors before the dawn of computer age. According to the book encyclopaedia of stock patterns by (Bulkowski, Encyclopedia of Chart Patterns, 2005) there are 53 patterns. The famous head and shoulder pattern is the key market trend indicator on the immediate future of the stock price. There is over 80% that the market will revert when the upward head and shoulder pattern is recognized. The predictability of stock pattern has been supported by many investors. In fact, the market news is full of forecasting trend based on pattern analysis. Under the classical definition, it is controversial to spot a particular stock pattern like head and shoulder as there is no pattern that can perfectly match the definition. From the investment point of view, there are two types of analysis − namely fundamental and technical. Pattern recognition is classified as technical analysis. All these approaches require expert knowledge which can only be obtained through intensive training. There are many discussions on how to employ fundamental analysis but very little in technical analysis - stock pattern recognition. It is probably because there is no systematic method to deter pattern. Fundamental analysis is based on financial and economic data, price/return movement of the stock and there has been an enormous amount of data accumulated throughout the past decades and it is relatively easy to obtain that in public domain. The recent development of Artificial Intelligence has drastically improved the efficiency in fundamental analysis. To a certain extent, it can also help to identify stock pattern recognition. However, this is not easy especially all the algorithms are developed under the guideline of the experts. But the guideline is not clean cut and sometimes it could be ambiguous. Unlike fundamental analysis which can provide an equation to develop an algorithm, the same cannot be applied in stock pattern recognition algorithm. One of the reasons is the interpretation of each pattern which is not a science but rather an art. For example, in the upward head and shoulder pattern, the general pattern is shoulder, head and shoulder. However, there is no strict definition of how far a head and a shoulder should be departed or how much a head should be higher than a shoulder. There is no strict rule on the minimum difference of the two shoulders. On top of that, the above classification is getting more and more difficult as there are so many stocks in different countries. The information from the stock pattern could be many but only a few can benefit from it. This model attempts to build a pattern recognition algorithm based on the book by (Bulkowski, Encyclopedia of Chart Patterns, 2005) to deter the pattern using the historical data of HK equity.

In the above section 3.2, the robustness and steady performance of SVM and LS-SVM have inspired this paper to explore the hybrid algorithm combining chart pattern and LS-SVM. In order to seek out a better forecasting model to increase the predictability of the above models, chart pattern is another approach to explore. If the pattern of the data set can be identified, it would definitely help the forecasting model to improve the accuracy. Hence, the following section continues the work of section 3.2.

3.3.2 Chart Pattern-based Algorithms

Seven forecasting algorithms have been developed by the candidate under the guidance by (Bulkowski, Encyclopedia of Chart Patterns, 2005) in order to test the predictability of the chart pattern. 20 years of historical data of 74 Hong Kong stocks have been scrutinized to detect an obivous pattern. Once it is discovered, 1 year of historical data of that stock prior to the detected pattern will be inputted into the LS-SVM model to forecast the future close value. The objective is to get an accuracy of mean absolute percentage error (MAPE) value below 2. As each algorithm is unique and based on the characteristic of the corresponding chart pattern, the outcome of each algorithm will be a reflection of the accuracy of that algorithm. The algorithms are the combination of chart pattern recognition and forecasting which is rare in the literature review. The comparison of each algorithm will be explained in the following sections.

3.3.2.1 Data

74 stocks from Hong Kong Stock Exchange have been selected for this experiment which covered the blue chips and red chips stock. These 74 stocks data with 20 years of historical data were downloaded from finance.yahoo.com. Only 7 chart patterns which have been classified into 2 categories in Section 2.5 are selected and the algorithm together with the identification guideline has been shown in the Appendixes. Each algorithm is applied to these 74 datasets to train up the model to seek out the possible chart pattern matching.

3.3.3 Empirical Modeling

3.3.3.1 Forecasting Model and Parameters

Our previous work (Lai & Liu, Support Vector Machine and Least Square Support Vector Machine Stock Forecasting Models, 2013) have demonstrated the advantage of employing SVM and LS-SVM in forecasting equities indices of US, Hong and China market. The robustness and steady performance of SVM and LS-SVM in (Zhang & Shen, Stock Yield Forecast based on LS-SVM in Bayesian inference, 2009) have inspired our research to explore the hybrid algorithm combining chart pattern and LS-SVM. From section 3.2.2.3 Table VI, the performance of LSSVM has encouraged the candidate to continue the research in this forecasting technique. Despite the fact that there are many discussions in the literature on chart pattern discovery using different methodologies such as (Lee, Liaw, & Hsu, Investment decision making by using fuzzy candlestick pattern and genetic algorithm, 2011) and, we believe the advantages of these algorithms are the ability to first discover the chart pattern and perform the forecasting later. The design of the algorithms ensures the chart pattern is discovered according to the literature guideline.

The approach is to select a time frame of 120 days transaction and divide it into 4 sections. High and low of each section were discovered within the 30 records and totally there were 4 highs and lows in 4 different sections. As per the chart identification guideline from (Bulkowski, Encyclopedia of Chart Patterns, 2005) these highs and lows within the sections are the key indicators if that section can be classified as a particular chart pattern. After a chart pattern is identified, one year of historical stock records dated back from the time frame discovery are used as training data to forecast the next 5-day and 22-day forecast horizons using LS-SVM. To seek chart pattern, high and low of the historical records are used but in the forecasting model, only close value is used which predicts the 5-day and 22-day future close values. LS-SVM model is selected because comparing to SVR, it is easier to operate without too many parameters to tune in order to get the best result which has been discovered in our previous work.

The 7 forecasting algorithms the candidate developed are Chart 2, Chart 3, Chart 24, Chart 26, Chart 47, Chart 48 and Chart 49. As the name implied, each algorithm targets a particular chart pattern, Chart 2 algorithm is designed to seek out pattern which exhibits the behaviour of Chart 2 as per Figure 3. Despite the facts these 7 algorithms are under the guideline of the book (Bulkowski, Encyclopedia of Chart Patterns, 2005), the algorithms are unique and specially designed to fit the Hong Kong stock market. The time horizon in the algorithm is set to 120 days which is roughly according to the stock market behaviour of the Hong Kong stock market. The flexibility of the algorithms is high as there are no parameters to input.

3.3.3.2 Empirical Result

The following results are based on the 7 algorithms.

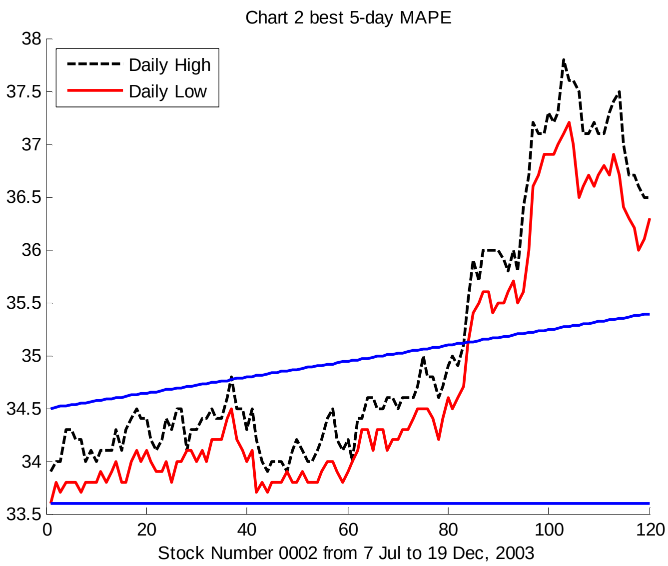

Figure 16 Best MAPE5 Result

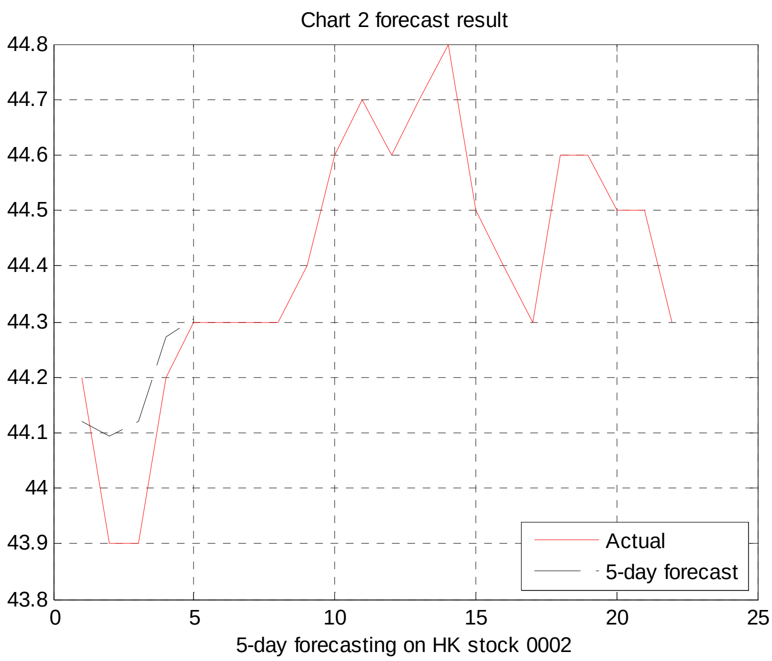

Figure 17 Chart 2 Forecast Result

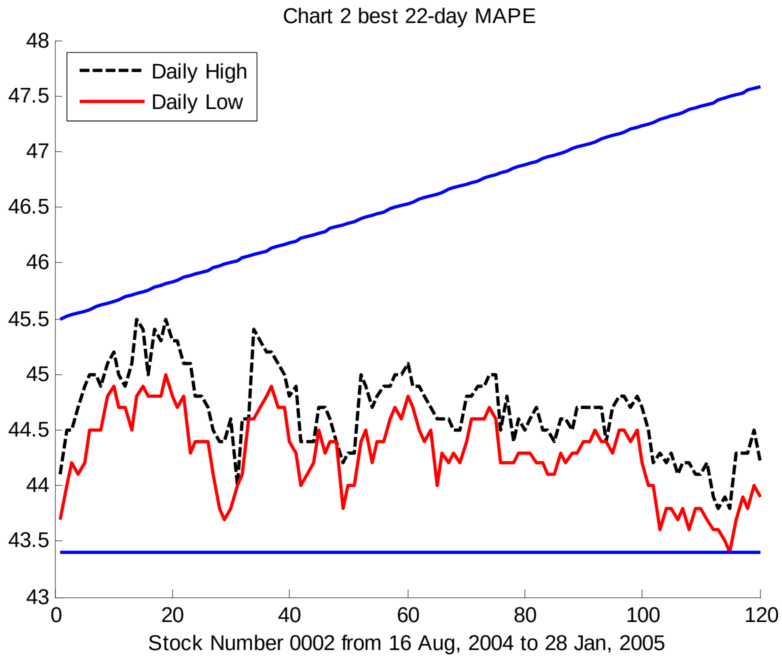

Figure 18 Chart 2 Best MAPE22 Result

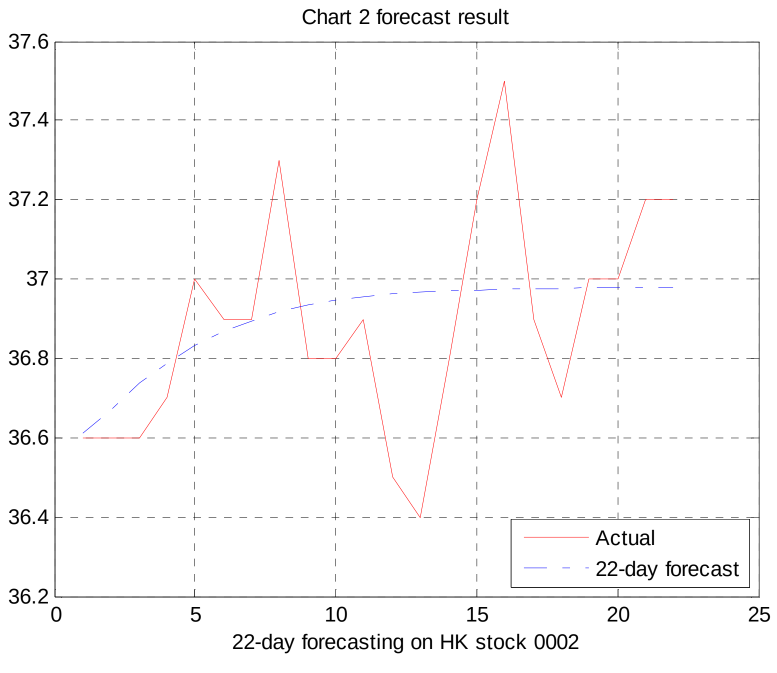

Figure 19 Chart 2 Forecast Result

Figure 16 to Figure 19 demonstrated the 120 records time frame that Chart 2 pattern is identified and the immediate forecasting result used one-year prior record as training. There are 23 out of 74 stocks that have been detected in the Chart 2 pattern. The best MAPE value for 5-day is 0.22 and 22-day is 0.28 they all come from stock no. 0002. 21 out of 23 stocks which fit the Chart 2 pattern recognition have MAPE value under 2 in 5-day forecast. 16 out of 23 stocks which fit the Chart 2 pattern recognition have MAPE value under 2 in 2-day forecast. As a summary, there is 91.3% in 5-day forecast horizon and 69.57% in 22-day forecast horizon that the predicted MAPE value is under 2 in Chart 2 pattern.

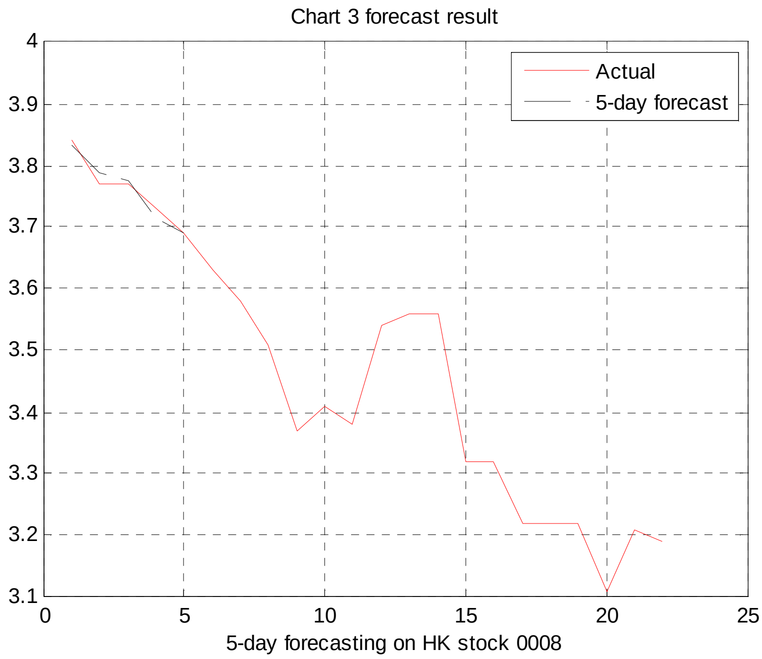

Figure 20 Chart 3 Forecast Result

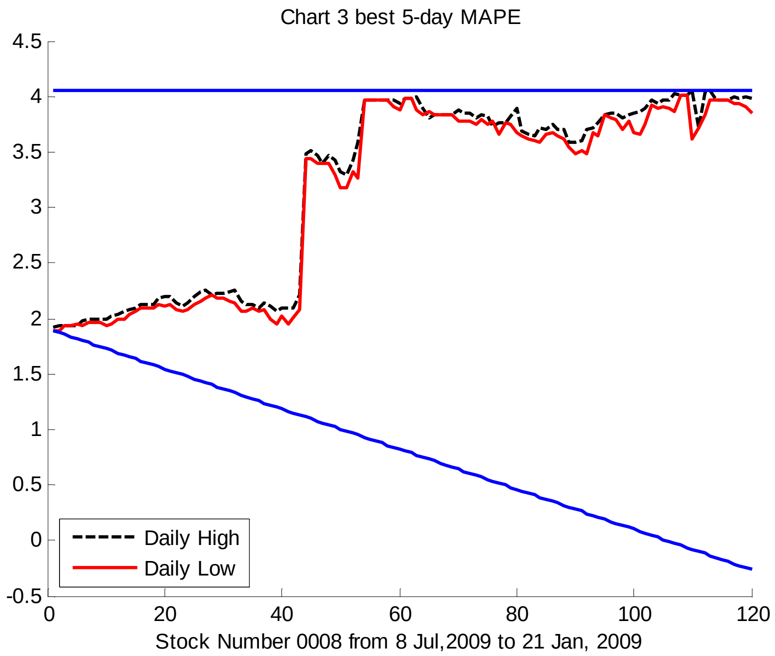

Figure 21 Chart 3 Best MAPE5 Result

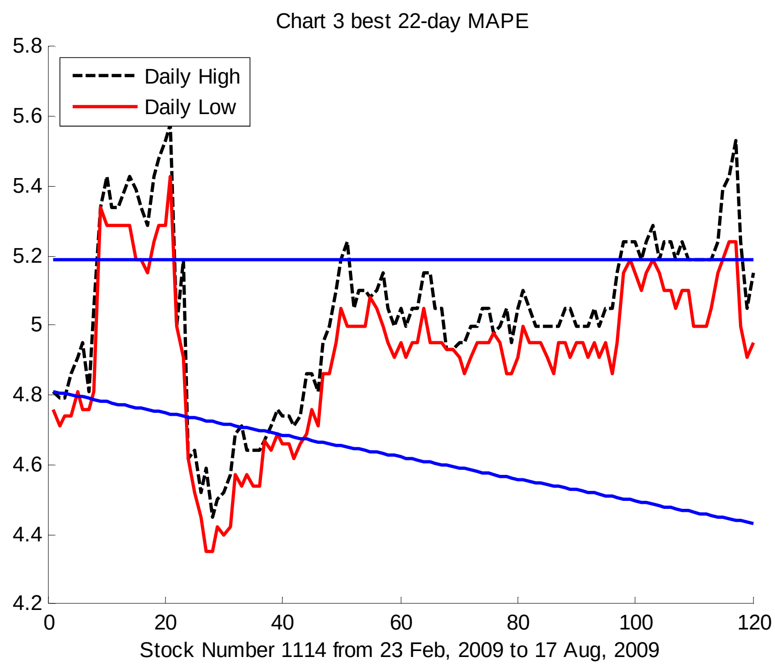

Figure 22 Chart 3 Best MAPE22 Result

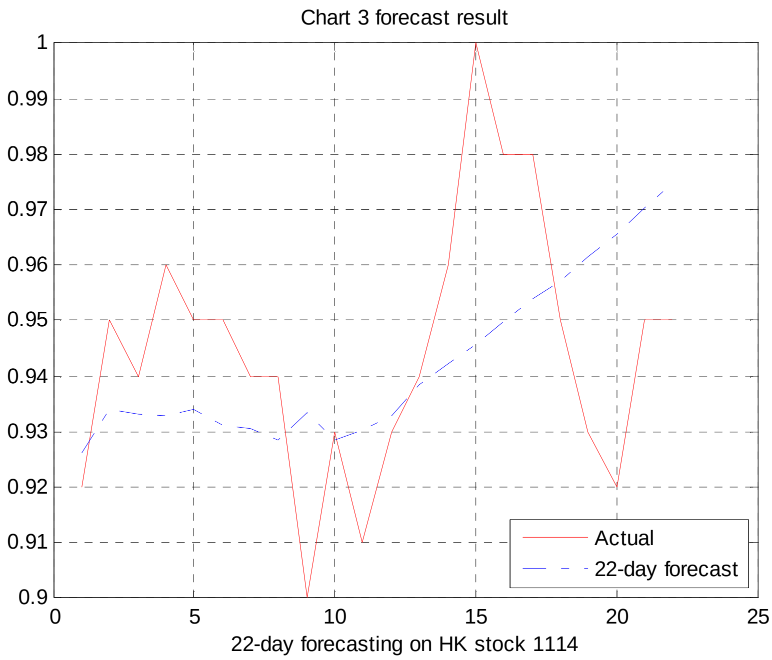

Figure 23 Chart 3 Best MAPE22 Result

Figure 20 to Figure 23 demonstrated the 120 records time frame that Chart 3 pattern is identified and the immediate forecasting result used one-year prior record as training. There are 60 out of 74 stocks that have been detected in the Chart 3 pattern. The best MAPE value for 5-day is 0.25 and 22-day is 2.05 which come from stock no. 0008 and stock no. 1114 respectively. 25 out of 60 stocks which fit the Chart 3 pattern recognition have MAPE value under 2 in 5-day forecast but none in 22-day forecast. As a summary, there is 41.67% that the predicted MAPE value is under 2 in 5-day forecast horizon.

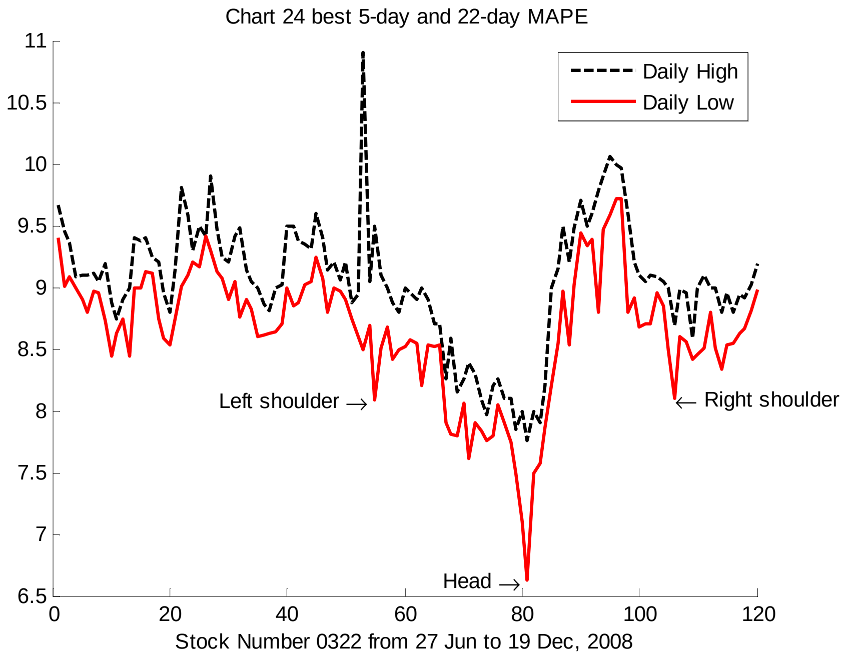

Figure 24 Chart 24 MAPE5 and MAPE22 Result

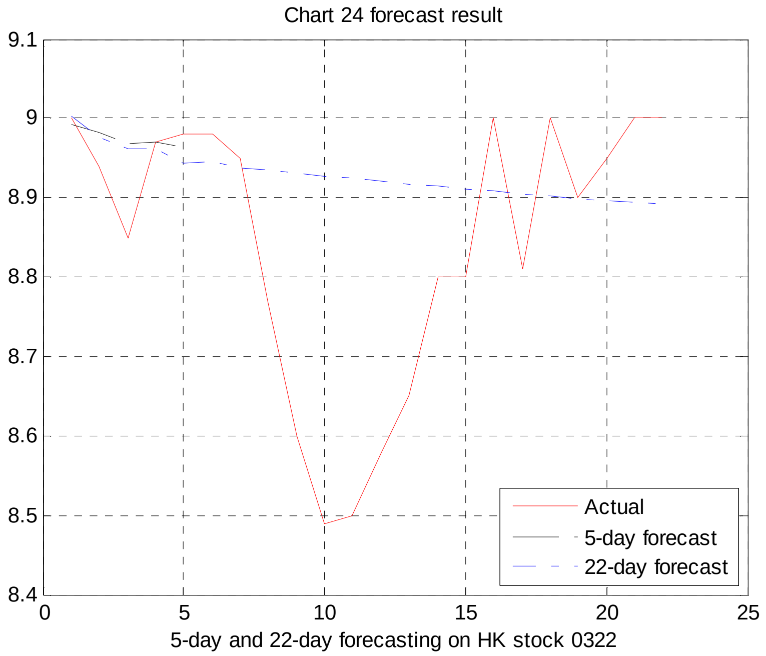

Figure 25 Chart 24 Forecast Result

Figure 24 and Figure 25 demonstrated the 120 records time frame that Chart 24 pattern is identified and the immediate forecasting result used one-year prior record as training. There are 23 out of 74 stocks that have been detected in the Chart 24 pattern. The best MAPE value for 5-day is 0.22 and 22-day is 0.28 they all come from stock no. 0002. 21 out of 23 stocks which fit the Chart 2 pattern recognition have MAPE value under 2 in 5-day forecast. 16 out of 23 stocks which fit the Chart 24 pattern recognition have MAPE value under 2 in 2-day forecast. As a summary, there is 91.3% in 5-day forecast horizon and 69.57% in 22-day forecast horizon that the predicted MAPE value is under 2 in Chart 24 pattern.

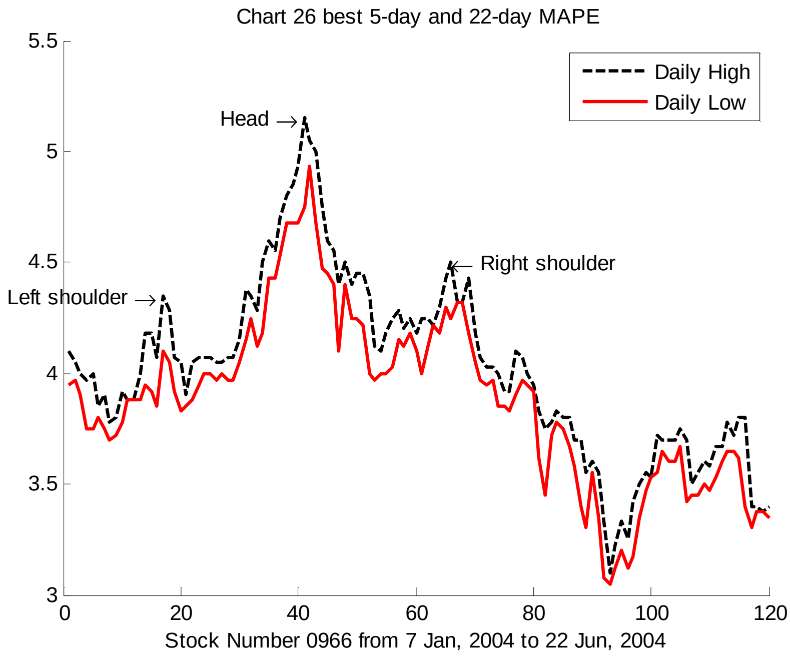

Figure 26 Chart 26 Best MAPE5 and MAPE22 Result

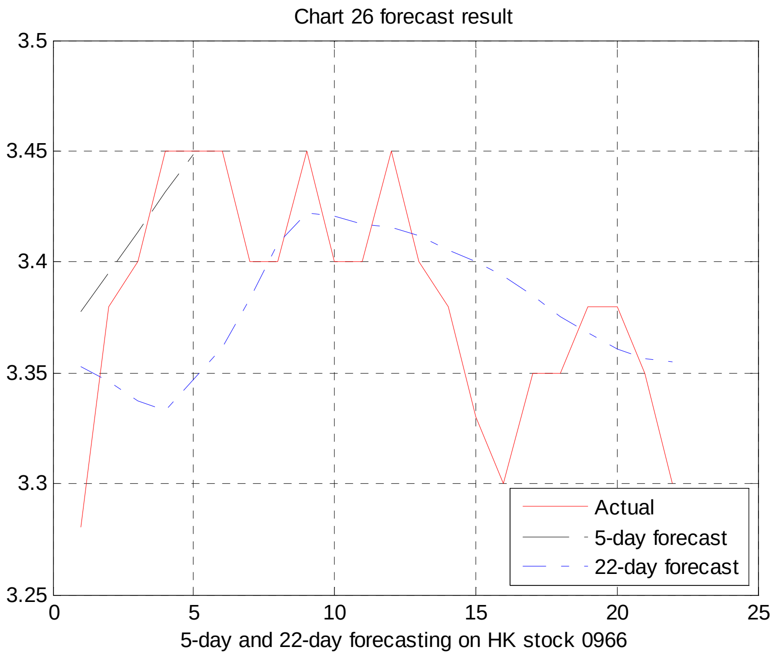

Figure 27 Chart 26 Forecast Result

Figure 26 and Figure 27 demonstrated the 120 records time frame that Chart 26 pattern is identified and the immediate forecasting result used one-year prior record as training. There are 8 out of 74 stocks that have been detected in the Chart 26 pattern. The best MAPE value for 5-day is 0.88 and 22-day is 1.29 they all come from stock no. 0966. 4 out of 8 stocks which fit the Chart 26 pattern recognition have MAPE value under 2 in 5-day forecast. 2 out of 8 stocks which fit the Chart 26 pattern recognition have MAPE value under 2 in 2-day forecast. As a summary, there is 50% in 5-day forecast horizon and 25% in 22-day forecast horizon that the predicted MAPE value is under 2 in Chart 26 pattern. It is a famous head and shoulder top pattern but the result is not promising. Only 8 records out of 74 stocks with 12 years of historical records can be identified as Chart 26 pattern and the forecasting result is only within the average.

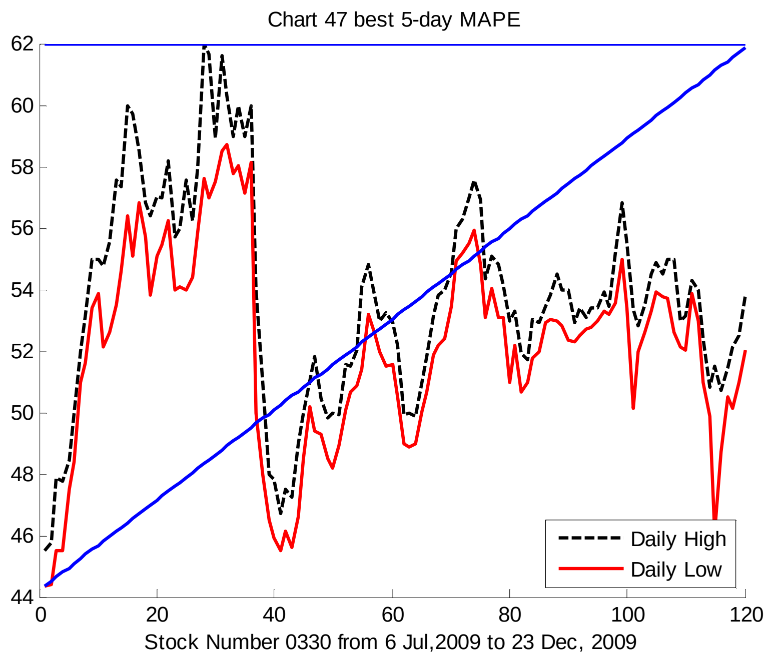

Figure 28 Chart 47 Best MAPE5 Result

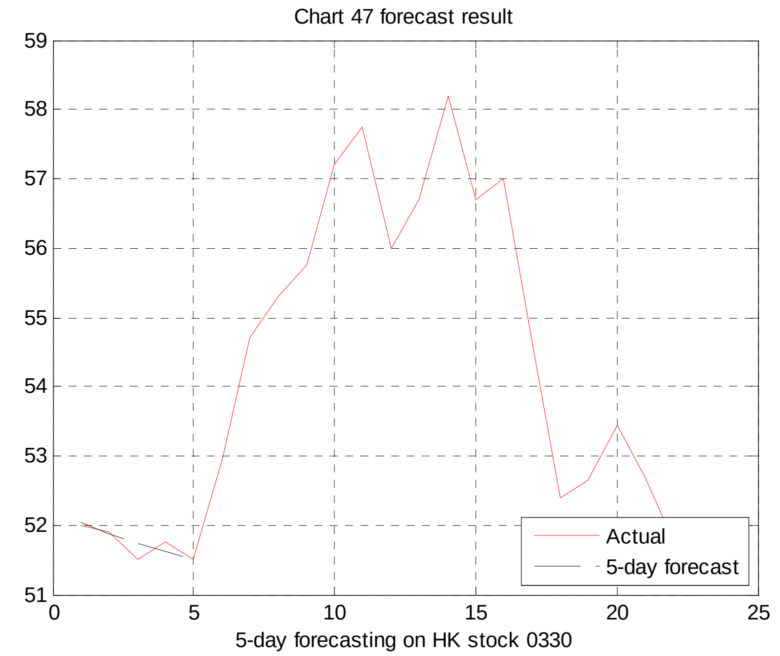

Figure 29 Chart 47 Forecast Result

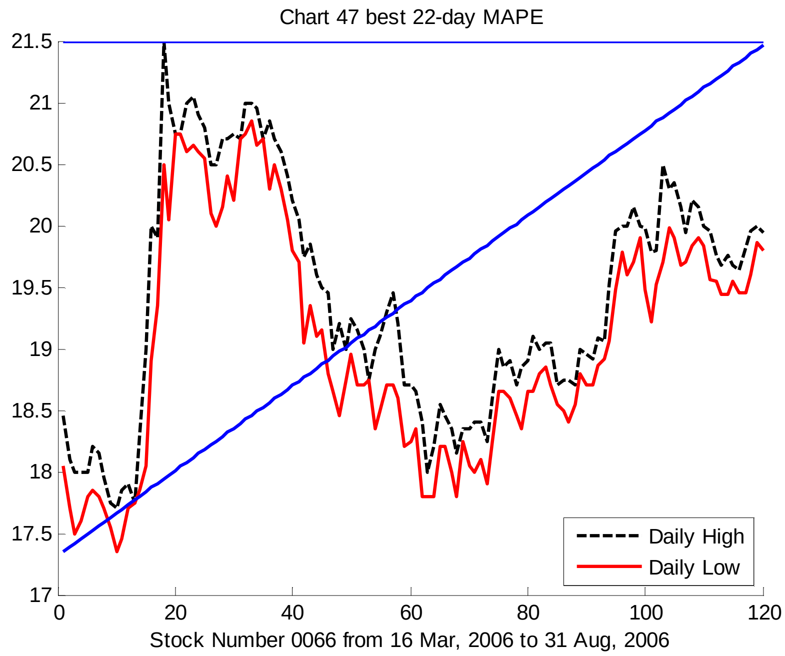

Figure 30 Chart 47 Best MAPE22 Result

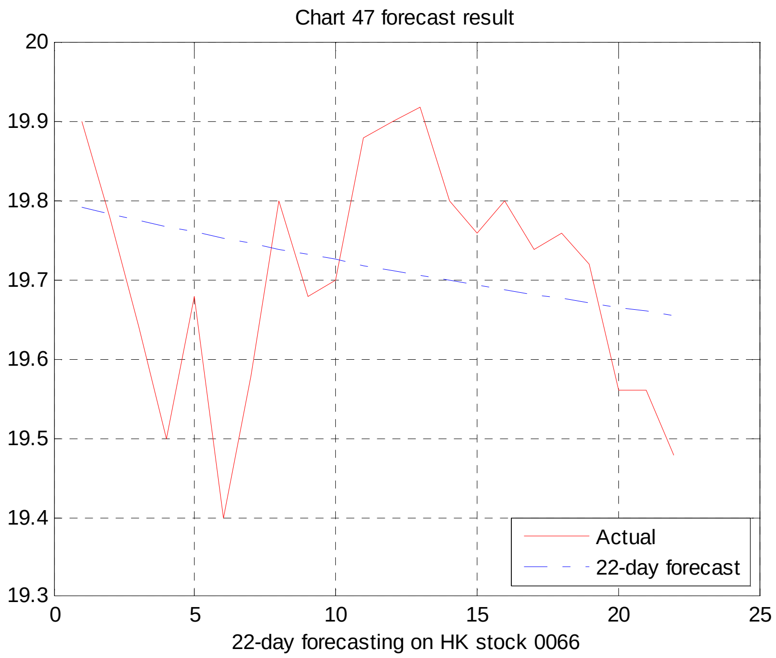

Figure 31 Chart 47 Forecast Result

Figure 28 to Figure 31 demonstrated the 120-record time frame that Chart 47 pattern is identified and the immediate forecasting result used one-year prior record as training. There are 72 out of 74 stocks that have detected the Chart 47 pattern. The best MAPE value for 5-day is 0.16 and 22-day is 0.62 they come from stock no. 0330 and 0066 respectively. 54 out of 72 stocks which fit the Chart 47 pattern recognition have MAPE value under 2 in 5-day forecast. 23 out of 72 stocks which fit the Chart 47 pattern recognition have MAPE value under 2 in 2-day forecast. As a summary, there is 75% in 5-day forecast horizon and 31.94% in 22-day forecast horizon that the predicted MAPE value is under 2 in Chart 47 pattern. This algorithm identified the most stocks and produce the best result.

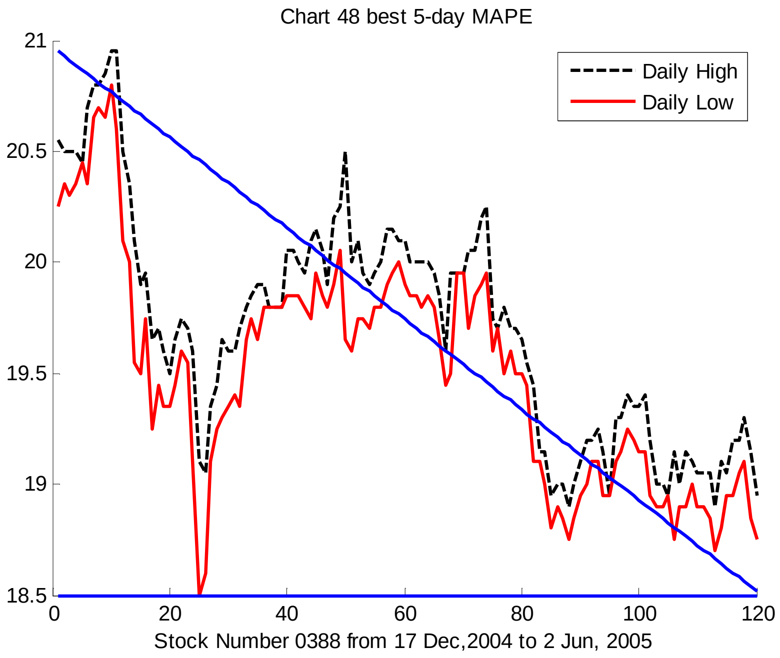

Figure 32 Chart 48 Best MAPE5 Result

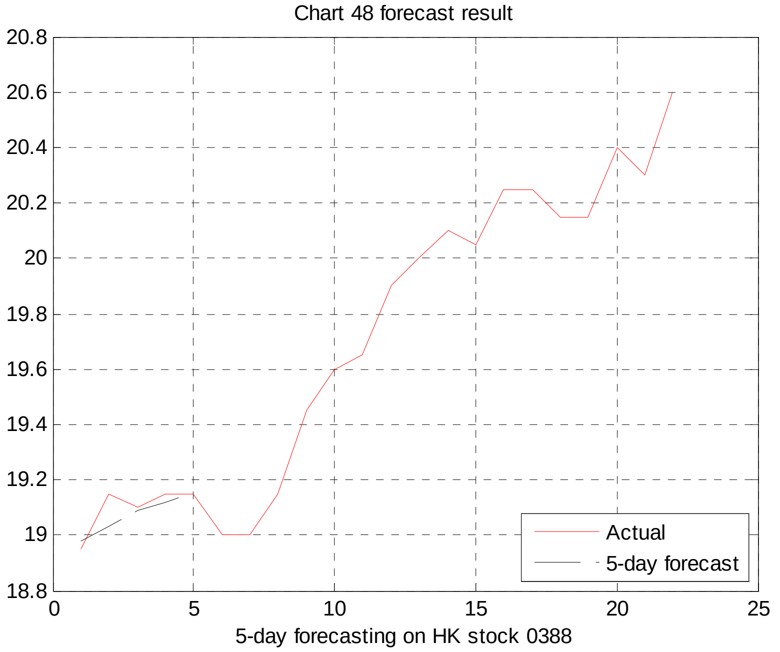

Figure 33 Chart 48 Forecast Result

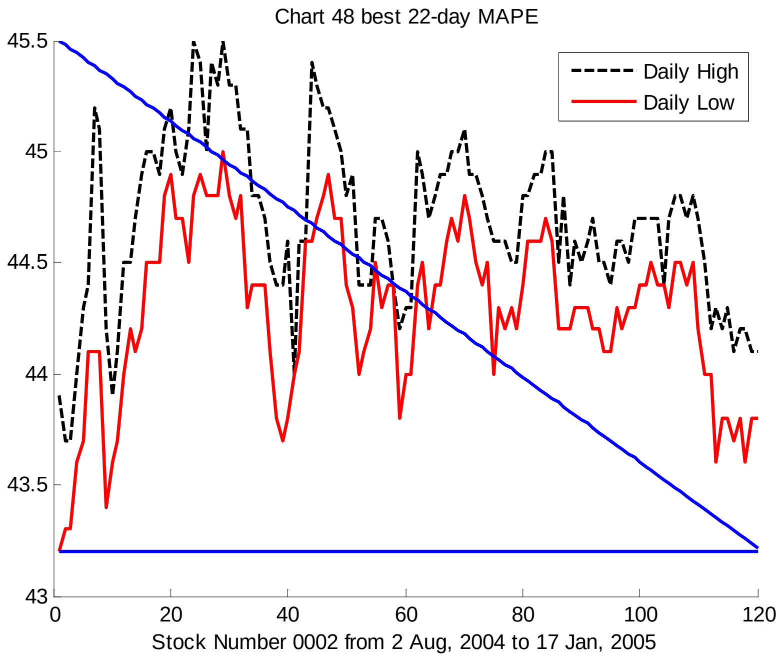

Figure 34 Chart 48 Best MAPE22 Result

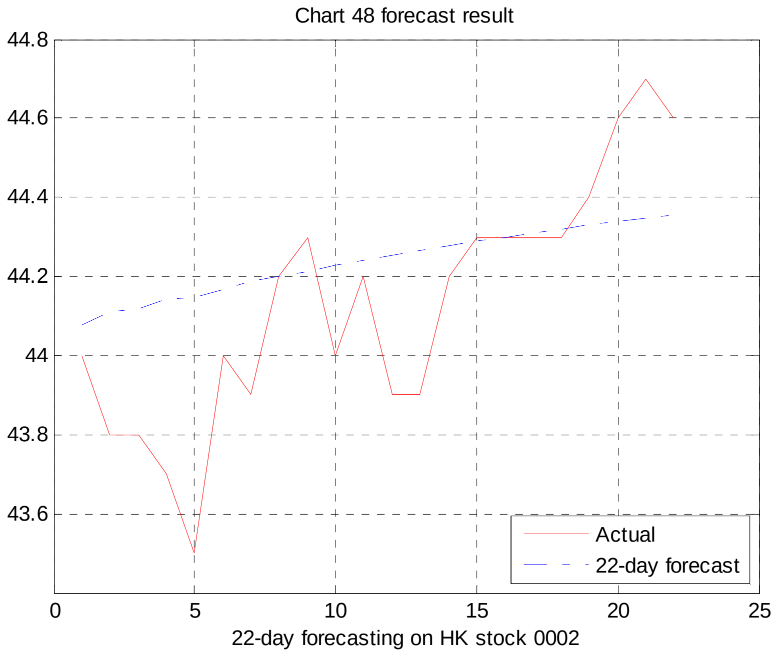

Figure 35 Chart 48 Forecast Result

Figure 32 to Figure 35 demonstrated the 120 records time frame that Chart 48 pattern is identified and the immediate forecasting result used one-year prior record as training. There are 30 out of 74 stocks that have detected the Chart 48 pattern. The best MAPE value for 5-day is 0.2 and 22-day is 0.45 they come from stock no. 0388 and 0002 respectively. 24 out of 30 stocks which fit the Chart 48 pattern recognition have MAPE value under 2 in 5-day forecast. 11 out of 30 stocks which fit the Chart 48 pattern recognition have MAPE value under 2 in 2-day forecast. As a summary, there is 80% in 5-day forecast horizon and 36.67% in 22-day forecast horizon that the predicted MAPE value is under 2 in Chart 48 pattern.

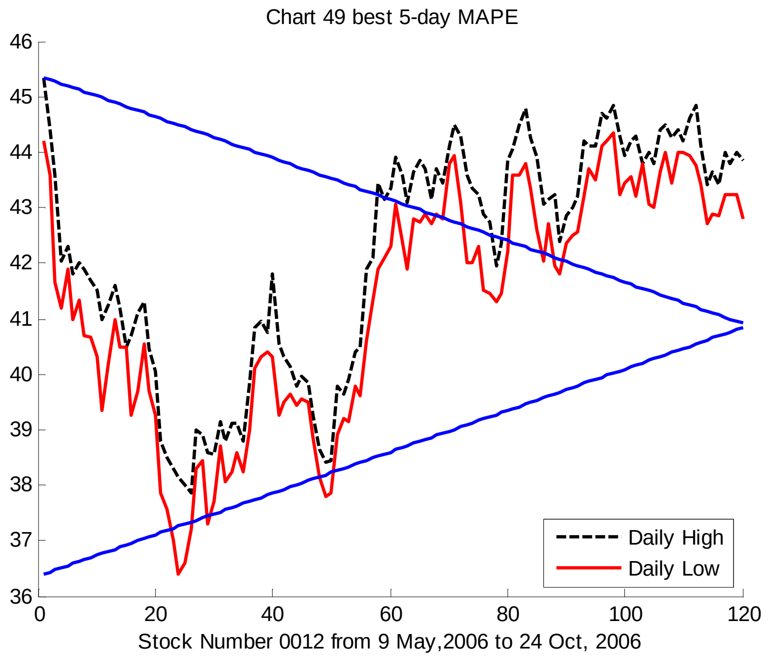

Figure 36 Chart 49 Best MAPE5 Result

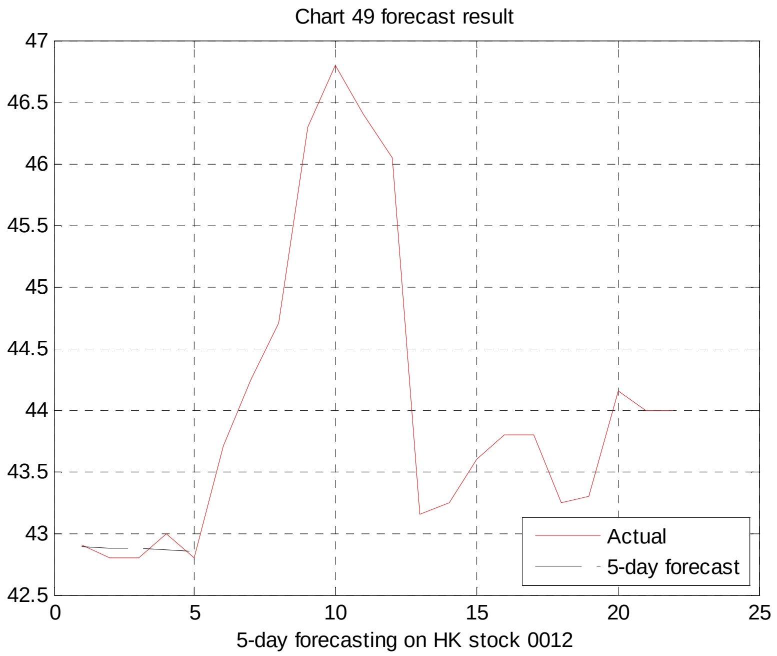

Figure 37 Chart 49 Forecast Result

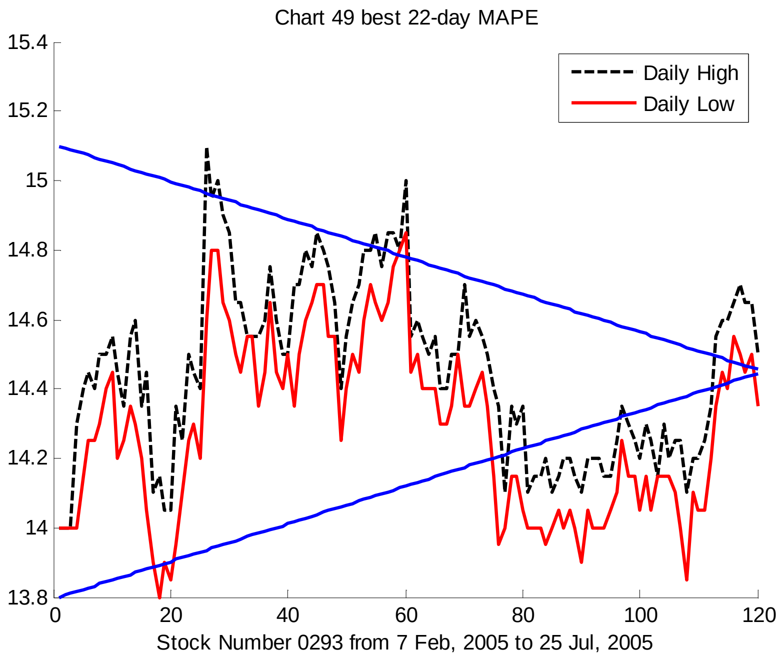

Figure 38 Chart 49 Best MAPE22 Result

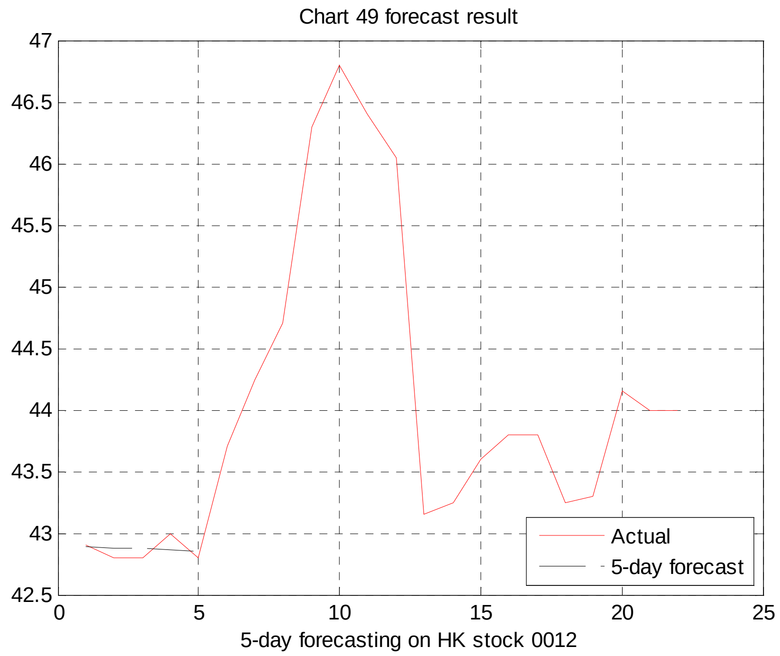

Figure 39 Chart 49 Forecast Result

Figure 36 to Figure 39 demonstrated the 120 records time frame that Chart 49 pattern is identified and the immediate forecasting result used one-year prior record as training. There are 46 out of 74 stocks that have detected the Chart 49 pattern. The best MAPE value for 5-day is 0.16 and 22-day is 0.54 they come from stock no. 0012 and 0293 respectively. 35 out of 46 stocks which fit the Chart 49 pattern recognition have MAPE value under 2 in 5-day forecast. 22 out of 46 stocks which fit the Chart 49 pattern recognition have MAPE value under 2 in 2-day forecast. As a summary, there is 76.09% in 5-day forecast horizon and 47.83% in 22-day forecast horizon that the predicted MAPE value is under 2 in Chart 49 pattern.

Table XIV Number of Patterns Discovered

Chart No.	Discover	5-day < 2 MAPE	22-day < 2 MAPE
2	33	24	13
3	60	25	0
24	13	5	2
26	8	4	2
47	72	54	23
48	30	24	11
49	46	35	22

Table XV Best MAPE Forecasting Result in Each Pattern

Chart No.	Stock Code	5-day < 2 MAPE	Stock Code	22-day < 2 MAPE
2	2	0.298	2	**0.556**
3	8	0.253	1114	2.058
24	322	0.415	322	1.564
26	966	0.879	966	1.286
47	330	0.164	66	0.616
48	388	0.199	2	0.452
49	12	**0.163**	293	0.535

Table XVI Worst MAPE Forecasting Result in Each Pattern

Chart No.	Stock Code	5-day < 2 MAPE	Stock Code	22-day < 2 MAPE
2	700	6.736	700	12.172
3	267	8.889	353	23.026
24	64	7.724	65	8.055
26	966	14.036	966	10.183
47	2600	18.473	2600	30.620
48	9	3.780	388	9.951
49	2600	**19.618**	1898	**36.010**

Table XVII Average MAPE Forecasting Result in Each Pattern

Chart No.	MAPE5	MAPE22
2	1.696	3.072
3	3.128	7.534
24	3.562	4.562
26	3.935	4.935
47	1.940	4.447
48	1.135	3.045
49	2.060	4.605
Average	2.494	4.600

From Table XV, the best result in this experiment is Chart 49 5-day MAPE 0.163 Stock No. 0012 and Chart 2 22-day MAPE 0.556 Stock No. 0002. But it also has the worst MAPE value from Table XVI. The best performance chart pattern from Table XVII is 48 with average MAPE 1.135. This is a very important finding that matches the expectation of this objective to search below 2 MAPE value. In fact, chart 47 and 48 have achieved this goal in 5-day horizon. Chart 49 is also very close. The famous head and shoulder, both top and bottom patterns are not as good as the others. In most cases, they are worse than the other chart pattern forecasting result. One of the reasons is the occurence of head and shoulder pattern is low as indicated in Table XV, only 13 and 8 patterns are discovered out of 74 stocks in Chart 24 and Chart 26 respectively. The other possible reason is that technical analysis has become very popular and these famous patterns are easily spotted by many experts. As a result, the market could over-react once the news is available and affect the market normal trend. Despite the fact Chart 24 and Chart 26 are not as good as the other patterns forecast result, it is still a good forecasting benchmark using the above algorithm as the average.

3.3.4 Conclusions

Compared to Chapter 3 Sections 3.1 and 3.2, the 7 algorithms are relatively easy to manipulate as there are no parameters to tune. The end user has nothing to operate but get the result very easily. Having said so, it should not underestimate the difficulty in implementing the algorithms as the definition of each chart pattern is getting blurred and the window frame is getting narrower. The above experiment has a window frame of 30 days to demonstrate the effect and in accordance with the textbook definition but in reality it could be shorter. In day trade operator, the window frame is in minutes instead of hours. It is a pattern which can get the results very quickly and easily but it is very difficult to manage. From Table XV, 14 of the best MAPE value has 13 of them under 2 and 10 them under 1. Our objective in the beginning of this research has set the MAPE value to 2. It can be regarded as a significant improvement compared to the result in Sections 3.1 and 3.2. The results from these 7 algorithms are much better than the previous models. From Table XVII, the average MAPE5 is 2.494 which is better than the SVR model MAPE5 2.585. However, it is only a rough comparison as the result from Table VII and from Table V is not the same dataset or environment. The dataset from the forecasting result in Table XVII is selected by the chart pattern algorithm while the dataset in Table V has no pre-screening process. However, it is worthy to point out Chart pattern algorithm has an advantage in this case.

The future of chart pattern identification will be heavily relying on computer-generated patterns and domain expert knowledge gained by us. In other words, the more computer resources, the higher accuracy you will get. In fact, many banking institutes already have allocated very heavy investment on IT infrastructure to get an edge in this very competitive digital war game so that the algorithm can spot the market pattern earlier than the competitors. In addition, the LSSVM only uses the close value of the stock as input feature and this methodology will continue due to computerization of program trading. Waves of buy programs and sell programs generated by algorithm trading would move the market irregularly. Electronic trading will react so fast that the volume will be less important and hence the key performance index is price only. The experiment is limited to 74 stocks and 7 out of 54 patterns but it should be extended to all stocks not restricted to one market but many markets worldwide with all spotted patterns.

It has been shown that the hybrid algorithm combining chart pattern and LS-SVM can produce a very promising result. Compared to our previous work (Lai & Liu, Stock forecasting using Support Vector Machine, 2010), this algorithm offers a methodology to get the better MAPE forecasting value by screening the chart pattern first before forecasting. The findings have also included the most common pattern like head and shoulder which is not as good as the other patterns in terms of forecasting result. The application is wide - not only for financial forecast but also other scientific fields with obvious graphical patterns in data representation. This section is part of the overall strategy in dealing with forecasting problem. Section 3.1 to Section 3.2 have employed statistical benchmark such as kurtosis and skewness to seek out the useful model in different markets. This section has an important discovery to this research which is the property of the datasets. The analysis of the chart pattern is similar to the analysis of the stock market such as Dow Jones, Hang Seng and Shanghai. It has been pointed out in Table II and Table XIII of Section 3.1 and Section 3.2 there is a certain relationship of the forecasting outcome and the kurtosis and skewness value. They both are the characteristics of the datasets defined by the statistical benchmark. The reason to revisit these values is its significance in the definition of pattern from the statistical point of view. In summary, the meaning of these values such as kurtosis is 0 means the pattern is normal distribution when the skewness value is 0. The limitation of chart pattern is the dependency of the appearance of the pattern from the literature. This means there is no forecasting without the spotted chart pattern. The application of this methodology is very limit. In the following section, normal distribution pattern has been thoroughly discussed on the effect of the forecasting result.

3.4 Support Vector Regression with Lévy Distribution Kernel

3.4.1 Introduction

Normal distribution is widely used in the financial model (Chen, Hardle, & Jeong, Forecasting Volatility with SVM-Based GARCH Model, 2010). In the famous Black-Scholes details in the book (Bodie, Kane, & Marcus, Investments, 2005) formula in Section 2.6 for call option price, the stock price is based on Brownian motion which follows the normal distribution. Robert Merton and Myron Scholes won the Nobel Prize in Economic Sciences in 1997 because of this workable option-pricing model. This famous formula also led to the collapse of Long Term Capital Management Incorporation which made a 5 Trillion US dollar hole in financial history. Subsequently, it sparked the financial crisis in 1997 which hit all Asian markets drastically. Yet, it is still widely used by options market participants even after the notorious financial tsunami in 2007. It is arguable whether the financial tsunami has anything to do with the assumption that market behaves like the normal distribution. But the connection of the assumption to the first financial crisis has been widely recognized particularly the Black-Scholes formula stock price is simulated under the normal distribution. On the other hand, Mandelbrot (1963) has observed that logarithm of relative price changes on a financial and commodities markets exhibit a long tailed distribution. His conclusion was that Brownian motion in exp(Bt) should be replaced by symmetric-stable Lévy motion with index $\alpha < 2$. This yields a pure-jump stock-price process. Roughly speaking, one may envisage this process as changing its values only by jumps. Normal distributions are α-stable distributions with $\alpha = 2$, so Mandelbrot's model may be seen as a complement to the Osborne (1959) or Samuelson (1965) model.

In the above section 3.3, there are many Chart Pattern analyses in the literature review and our contribution in that section would not be significant but our approach has pointed to the right direction and has improvement compared to that of section 3.1 and section 3.2. The chart pattern algorithm has a significant disadvantage. It can only forecast the future event if and only if a Chart pattern is discovered. The application of Chart Pattern forecasting will be limited by this restriction. But our work is not in vain as it is obvious the dataset pattern is crucial in forecasting. Another approach is to fit the historical data into a known distribution function. Usually, the Normal distribution is assumed but we would like to investigate the Lévy distribution approach. Our motivation is based on the paper, Process in Finance: Theory, Numerics, and Empirical Facts by (Raible, Lévy Processes in Finance: Theory, Numerics, and Empirical Facts, 2001), which illustrated that the adoption of Lévy distribution instead of Normal distribution has many advantages.

In the following section, it explains the discovery that the historical stock movement does not follow normal distribution; rather, it follows heavy tail like distribution such as Lévy distribution according to literature review (Tu, Clyde, & Wolpert, Lévy Adaptive Regression Kernels, 2007). The reason to employ normal distribution in stock movement is its simplicity and the characteristics of normal distribution. It is a continuous distribution with conjugate family, moment of generating function and the famous central limit theorem that a large number of independent random variables, each with a well-defined mean and well-defined variance, will be approximately normally distributed. Lévy distribution is one of the few distributions which is stable and that has probability density functions which can be expressed analytically, the others being the normal distribution and the Cauchy distribution. All three are special cases of the stable distributions that do not generally have a probability density function which can be expressed analytically.

The following work objective is to extend the work of (Lai & Liu, Stock forecasting using Support Vector Machine, 2010) to develop a consistent approach in stock forecasting using SVR so that a platform to compare different forecasting tools can be established. The performances of SVR in predicting stock prices in the Hang Seng Index (HSI), Dow Jones Industrial Index (DJ) and Shanghai Composite Index (SH) over a 5-day and 22-day horizons respectively have been examined. The experiments are carried out in MATLAB R2011 environment with algorithms developed by the candidate.

3.4.2 Normal and Lévy Distribution

The probability density function (pdf) of the distribution is used to simulate the stock movement in financial forecasting. To use pdf in (4) of the SVR, it must satisfy the Mercer condition and it should be positive definite. Normal distribution is the good choice for reasons mentioned above with the following pdf Equation 54 and Figure 40

$$f(\mathrm{x}) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(\mathrm{x}-\mu)^2}{2\delta^2}}$$

Equation 54

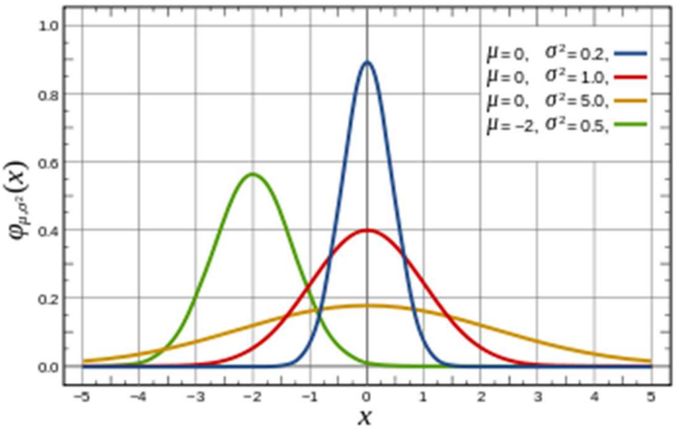

Figure 40 Normal Distribution Probability Density Function

As one of the few stable distributions, Lévy distribution also satisfies Mercer Condition with the following pdf Equation 45 and Figure 41

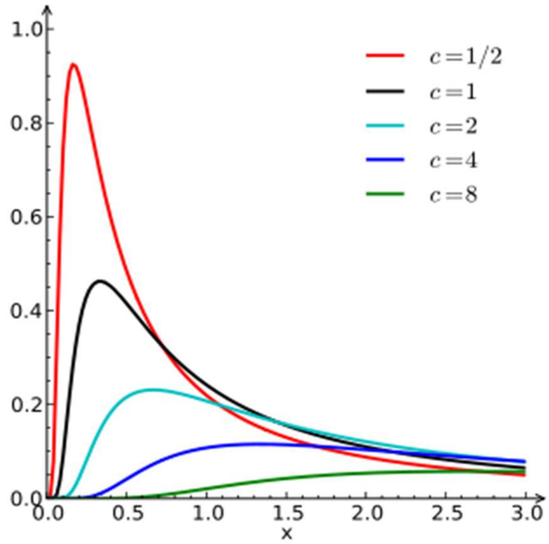

Figure 41 Lévy Probability Density Function

$$f(x;\mu,c)=\sqrt{\frac{c}{2\pi}}\,\frac{e^{-\frac{c}{2(x-\mu)}}}{(x-\mu)^{3/2}}$$

Equation 55

3.4.3 Experiments and results

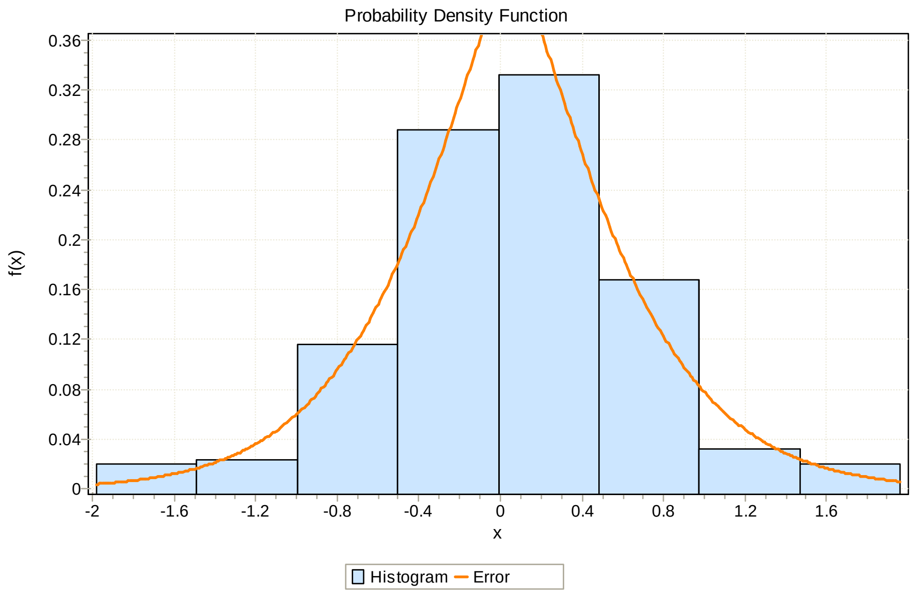

Figure 42 Dow Jones 2006 Data Set

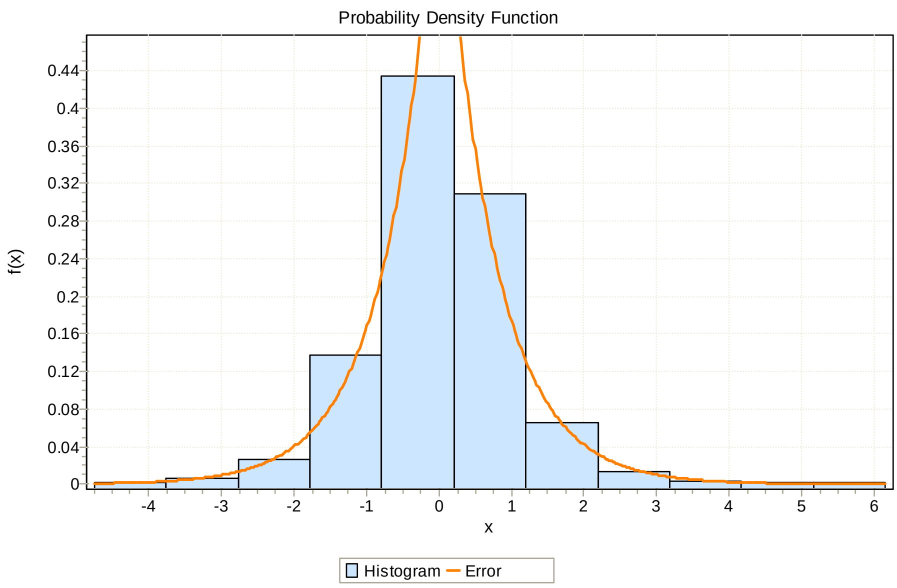

Figure 43 Dow Jones 2002 to 2006 Data Set

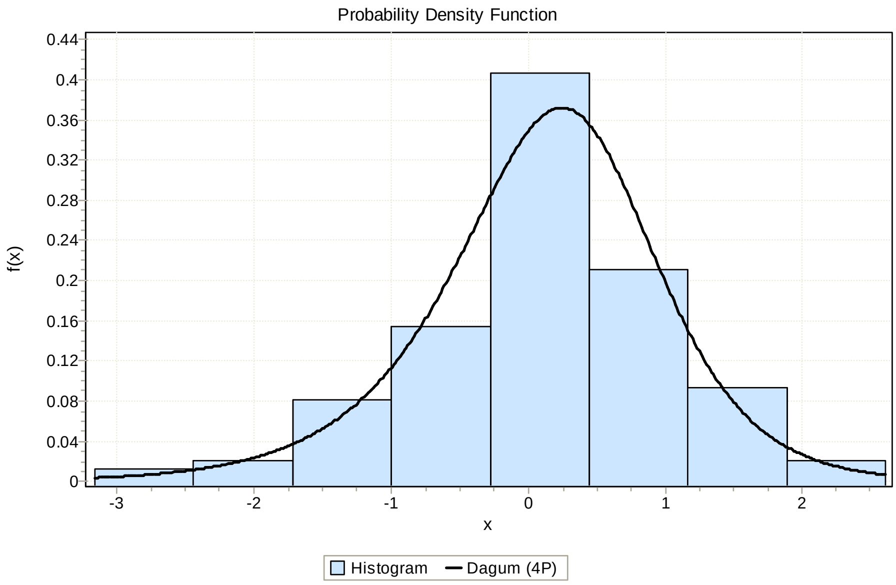

Figure 44 Hang Seng Index 2006 Data Set

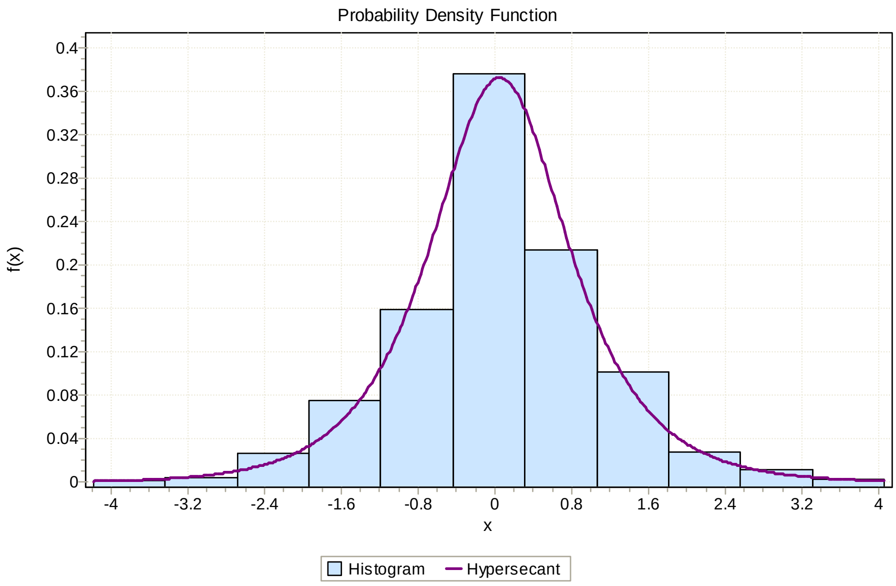

Figure 45 Hang Seng Index 2002 to 2006 Data Set

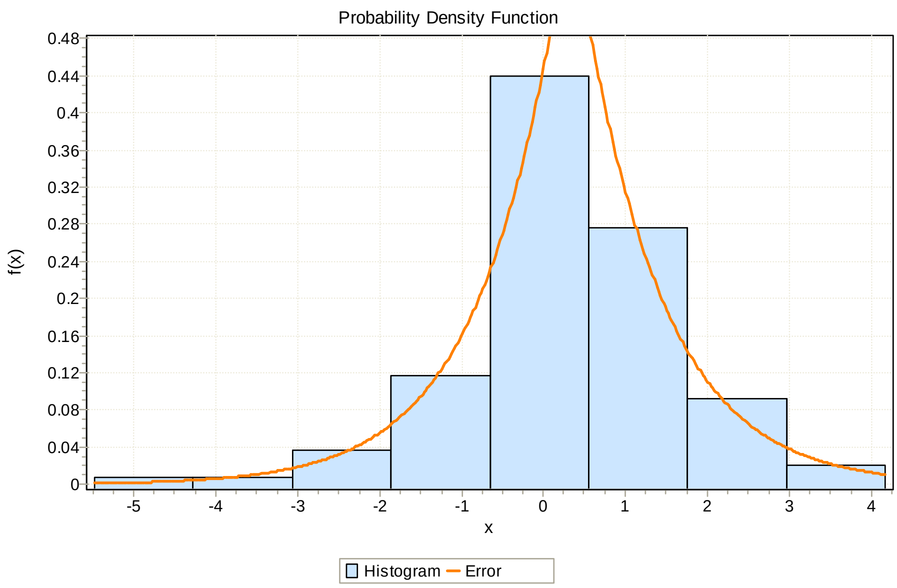

Figure 46 Shanghai Composite Index 2006 Data Set

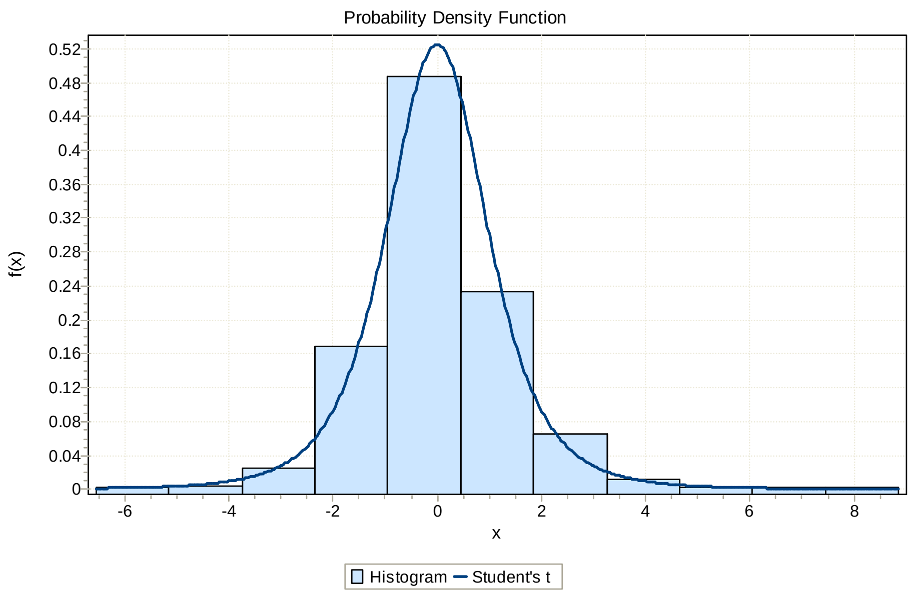

Figure 47 Shanghai Composite Index 2002 to 2006 Data Set

Figure 42 to Figure 47 are the histograms of the log return for the dataset 2006 and 2002-6 of the three different markets. Despite the fact that they look like normal comparing to the heavy tail distribution, all of the above histograms have kurtosis more than 3 which means they are not normally distributed. The difference among Figure 42 to Figure 47 compared with the normal distribution curve is the heavy tail behaviour.

As explained in the introduction, SVR maps data into higher dimensional space using kernel function. In (4) RBF kernel function is $K(x_{k,}x_l) = \varphi(x_k)^T \varphi(x_l)$. The formula in that kernel function is $e^{(-gamma*(x_k - x_l)^2)}$ which is a normal pdf. Another kernel using Lévy pdf with the formula $\frac{e^{(-gamma*(x_k - x_l))}}{(x_k - x_l)^{3/2}}$ has been included in SVR model. Table XVIII is the result from SVR using normal distribution kernel. The higher the C, the less the percentage of error. The g which stands for the gamma in the kernel function also has different settings. The value of g is to adjust the kernel function and the higher the value the greater the mapping dimension is. The parameters C and g were based on our previous work (Lai & Liu, Stock forecasting using Support Vector Machine, 2010).

Table XVIII Prediction Result of SVR Using Normal Distribution

Market	Volatility	Kurtosis	Skewness	MAPE5	SMAPE5
DJ2002_6	1.00	2.360	0.763	0.7	1.22
DJ2006	0.66	2.821	-0.387	0.38	1.22
DJ2003_7	1.83	2.080	-0.308	1.14	2.86
DJ2007	6.17	2.710	0.289	1.94	4.48
HSI2002_6	1.47	2.254	0.612	2.71	5.58
HSI2006	3.06	2.333	0.326	2.71	5.58
HSI2003_7	3.17	2.198	0.774	0.94	5.52
HSI2007	5.90	4.042	1.092	1.75	5.07
SH2002_6	2.92	3.428	0.741	7.32	18.23
SH2006	7.90	5.273	0.766	9.74	19.24
SH2003_7	3.81	1.742	-0.016	1.85	8.6
SH2007	3.83	5.092	1.830	1.85	8.53
Overall average of MAPE value					4.965

Short term forecast for 2007 with 5-days horizon considers the 2006 dataset only as input. The lowest MAPE result is 0.38 in DOW Index. For 2007 with 5-day forecast horizon using a dataset from 2002 to 2006, the lowest result is 0.70 also in DOW Index. Long term forecast for 2007 with 22-day horizon considers the 2006 dataset only as input, the lowest result is 1.22 in Dow Index. For 2007 with 22-day forecast horizon using a dataset from 2002 to 2006, the lowest result is 1.22 in DOW Index.

Short term forecast for 2008 with 5-days horizon considers the 2007 dataset only as input. The lowest MAPE result is 1.75 in Hang Seng Index. For 2008 with 5-day forecast horizon using a dataset from 2003 to 2007, the lowest result is 0.94 also in Hang Seng Index. Long term forecast for 2008 with 22-day horizon considers the 2007 dataset only as input, the lowest result is 4.48 in Dow Jones Index. For 2008 with 22-day forecast horizon using a dataset from 2003 to 2007, the lowest result is 2.86 also in Dow Jones Index. In Table XVIII, there are 10 MAPE values out of 24 lower than 2.

SVR requires much effort to tune the parameters c and g in order to get a better result. It is rather difficult to determine which setting is correct. We discovered from the experiments that the parameter c has to be set with the range between 1,000 and 8000 in order to produce meaningful results and the c value must be more than 500, please refer to

(Lai & Liu, Stock forecasting using Support Vector Machine, 2010) for details.

Table XIX Prediction Result of SVR Using Lévy Distribution

Market	Volatility	Kurtosis	Skewness	MAPE5	SMAPE5
DJ2002_6	1.00	2.360	0.763	2.69	3.29
DJ2006	**0.66**	2.821	-0.387	0.36	0.98
DJ2003_7	1.83	2.080	-0.308	1.24	3.21
DJ2007	6.17	2.710	0.289	1.94	4.48
HSI2002_6	1.47	2.254	0.612	3.22	3.3
HSI2006	3.06	2.333	0.326	7.13	7.15
HSI2003_7	3.17	2.198	0.774	1.15	5.36
HSI2007	5.90	4.042	1.092	1.6	5.45
SH2002_6	2.92	3.428	0.741	9.88	12.07
SH2006	7.90	5.273	0.766	4.14	6.49
SH2003_7	3.81	**1.742**	**-0.016**	21.7	28.76
SH2007	3.83	5.092	1.830	1.98	6.71
Overall average of MAPE value					6.012

Table XIX demonstrates the improvement using Lévy distribution in the 3 markets. There are 2 in DOW market better than the result in Table XVIII, 2 in Hang Seng and 4 in Shanghai. It seems the more efficient the market, the less Lévy distribution kernel can improve the forecasting result. The significant improvement using Lévy distribution in Shanghai market is remarkable. There are 9 out of 24 results having an improvement and the Shanghai market SH2006 was improved from 19.24 to 6.49 using Lévy distribution kernel. It is a weak-form EMH market and it is not likely to follow normal distribution movment in its price trend. One of the advantages of SVR is the mapping of data into higher dimension using kernel and avoid finding the distribution parameters of the original data. The kurtosis of the log return dataset DJ 2006, DJ2007, SH2006, SH2002-6 and SH2007 are 4.24, 4.56, 5.6, 7.03 and 4.95 respectively. After the dataset is mapped into higher dimension using normal distribution, the kurtosis of the log return dataset DJ 2006, DJ2007, SH2006, SH2002-6 and SH2007 are 6.25, 10.63, 21.92, 117.23 and 6.36 respectively.After the dataset is mapped into higher dimension using Lévy distribution kernel, the log return kurtosis reduces to 4.09, 4.54, 6.17, 7.94 and 4.35 respectively. It seems the closer the kurtosis value of the higher dimension data to the original dataset, the better the forecasting result. The significant improvement is from SH2006 MAPE22 which is 12.75 less than the Normal distribution SVR. It is the same for SH2006 MAPE5 which is 5.6 less than the Normal distribution kernel in SVR. From

Table XIX, SH2006 has the highest volatility value 7.9, kurtosis 5.273 and the skewness 0.766 ranked the 4th highest. For SH2002_6, its MAPE22 value is 6.16 less than the Normal distribution kernel SVR. The volatility value is 2.92 which ranked the 8th highest, the kurtosis value is the 4th highest and the skewness value is 0.741 which ranked the 6th highest. This is another evidence that the higher the volatility and kurtosis value which means the market is more fluctuating and volatile, the more accurate is the use of Lévy distribution kernel in SVR. The four highest skewness values in HSI2003_7, HSI2007, SH2006 and SH2007 are 0.774, 1.092, 0.766 and 1.83 respectively. Lévy distribution kernel in SVR has improvement in these four dataset. The fat tails effect is the pattern that

Lévy distribution identified and this experiment has proved it is successful based on the result.

3.4.4 Conclusions

The potential to use Lévy distribution in SVR model has been demonstrated with convincing evidence and promising future application. The improvement in forecasting accuracy is significant in the 3 markets, especially in Shanghai market. We believe this is a contribution in forecasting methodology as there is no such finding yet in the literature review. Our finding shows that there is a strong correlation between MAPE value with the volatility as well as the skewness value.

It is important to point out that the experiments have demonstrated promising forecasting result in strong form EMH market such as DOW in the USA especially using SVR model. 6 out of 8 forecasting values have MAPE under 2 which is the objective of this research. For HSI Hong Kong market, the performance is similar to DOW despite it is a semi-strong EMH market but it can still provide a good decision-making platform for investment. For weak form EMH SH in China, the forecasting ability is obviously poor. This is because the volatility of the market strongly affects the forecasting power. In the future, it is necessary first to forecast the volatility of the market and then develop a hybrid SVR model to input the volatility attribute into it in order to improve the performance.

Attempt to fit historical data into known distribution function is not novel. Yet, there is no perfect fit especially in financial time series data. From the theoretical point of view, there are more than hundred of distribution functions and more will be discovered due to big data era. The parameters (such as μ location and c >0 in Lévy distribution) and characteristics (such as probability density function, cumulative density function or moment generating function also in Lévy distribution) in each distribution has a very rigid statistical definition. It is impossible that a historical data will fall into a particular distribution function. MATLAB or EasyFit software (Mathwave Technologies, 2004) have offered simple fitting to their database distribution function. However, there are many discreprencies in the parameters and characteristics. Thus, those sofware programs are just assumed it will fit into a certain distribution which is similar to many research approaches. Statisticians have their very important role in the history of forecasting as mentioned in Chapter 2 Section 2.4. The fundamental assumption that many data may fall into normal distribution function has a very long root from these people. Our finding has challenges if this assumption is a necessary and sufficient to support the forecasting theory. There are a few more distribution functions that can be employed as SVR kernel such as Hyperbolic Sceant, Student's t and Laplace distribution which the candidate has already tested but not included in the study. Lévy distribution has outperformed the others as far as our research is concerned. It is obvious from

Table XIX that the overall MAPE of the Lévy distribution is not good as it is only 6.012 much higher than the normal distribution 4.965.

As pointed out earlier there are 9 out of 24 better results using Lévy distribution kernel which means there are 15 worse results compared to Normal distribution kernel. Despite the fact that there are more worst results than better results, this kernel is still worthy to check and carry on with another approach. Improving the SVR mechanism is not easy but this Lévy distribution has definitely improved the accuracy under specific data characteristics. It seems for financial time series, it is necessary to use different approaches to seek out the best result. However, the above SVR algorithm is not good at handling high dimensional data especially in big data environment. The limitation of the above algorithms is that Lévy distribution could be significant but not necessary contribute to all financial time series pattern. There are many that do not conform to Lévy distribution and the application therefore is not universal. The difficulties in applying Lévy Distribution kernel in SVR is the mapping mechanism using kernel method is unknown in the literature review and by assumption there is no need to. As the data is mapped into a higher dimension which could be infinite, it is very difficult to control the result after the mapping.

3.5 Extreme Learning Machine

3.5.1 Introduction

ELM has been highly praised as the path towards human brain alike learning by Guang-Bin Huang (2013) (Huang & Chen, Enhanced random search based incremental extreme learning machine, 2008) (Huang G. B., What are Extreme Learning Machines? Filling the Gap between Frank Rosenblatt's Dream and John von Neumann's Puzzle, 2015). The classical NN has lost its attention in the research domain mainly because of the difficulty in handling parametric techniques especially when it is approximating complex nonlinear mappings directly from the input samples. SLFNs has been investigated to be the feasible solution to compensate for this problem. There are two types of SLFNs network architectures investigated in this research. The first one is SLFNs with additive hidden nodes and the second is (RBF) networks in (ELM). The SLFNs is best employed in online applications such as (Bosman, Iacca, & Wortche, Online Extreme Learning on Fixed-Point Sensor Networks, 2013) or (Janakiraman, Nguyen, & Assanis). In batch learning, each new data set must be used to re-train the learning parameters such as learning rate, number of learning epochs, stopping criteria and other predefined parameters which takes long computational power while online SLFNs are much shorter. In the first SLFNs architectures with additive hidden nodes, the back-propagation (BP) algorithm is basically a batch learning algorithm and it is the backbone of this architecture.

From Sections 3.1 to 3.4 in Chapter 3, various forecasting algorithms have been studied and some of them have achieved the objective of under 2 in MAPE value. However, all the above algorithms are based on mature forecasting techniques which have been more than two decades of history. ELM is relatively a new technique but with strong NN background and similarity to SVM. The research in financial time series forecasting is incomplete without the investigation of ELM. With the simpler hidden nodes structure combined with back-propagation network, it is expected that ELM should be able to generate a better result than the models described above.

In this model, two architectures are investigated using the same dataset as described in Section 3.6.1. The objective is to use ELM as a forecasting platform to compare different forecasting tools. The performances of SVR in predicting stock prices in the Hang Seng Index (HSI), Dow Jones Industrial Index (DJ) and Shanghai Composite Index (SH) over a 4-day and 20-day horizons respectively have been examined. The experiments are carried out in MATLAB R2011 environment with the algorithm developed by the Guang-Bin Huang.

3.5.2 Empirical Modeling

The SLFNs network functions with n hidden nodes are expressed in the following.

$$f_n(x) = \sum_{i=1}^{n} \beta_i g_i(x) = \sum_{i=1}^{n} \beta_i G(x, a_i, b_i), \qquad a_i \in C^d,$$

$$x \in C^d, \; b_i \in C, \quad \beta_i \in C,$$

Equation 56

where g_i or $G(x,a_i,b_i)$ denotes the output function of the i^{th} hidden node and β_i is the (output) weight of the connection between the i^{th} hidden node and the output node.

While additive hidden nodes are calculated as follows:-

$$g_i(x) = g(a_i \times x + b_i),\ \ a_i \in R^d,\ \ \ b_i \in R,$$

Equation 57

where g is the activation function of hidden nodes,

And RBF hidden nodes are calculated as follows:-

$$g_i(x) = g(b_i \| x - a_i \|),\ \ a_i \in R^d,\ \ \ b_i \in R.$$

Equation 58

3.5.3 Empirical Results

The results of the ELM model are compared with that of ARIMA and MA models.

Table XX ELM Prediction Result

Market	Volatility	Kurtosis	Skewness	MAPE5	MAPE22
				Linear	Linear
DJ2002_6	1.00	2.360	0.763	3.555	6.691
DJ2006	**0.66**	2.821	-0.387	3.443	6.373
DJ2003_7	1.83	2.080	-0.308	4.281	7.467
DJ2007	6.17	2.710	0.289	4.335	6.861
HSI2002_6	1.47	2.254	0.612	1.721	10.438
HSI2006	3.06	2.333	0.326	1.612	9.961
HSI2003_7	3.17	2.198	0.774	2.065	11.683
HSI2007	5.90	4.042	1.092	1.861	10.986
SH2002_6	2.92	3.428	0.741	1.836	7.061
SH2006	7.90	5.273	0.766	1.541	6.722
SH2003_7	3.81	1.742	-0.016	1.934	7.716
SH2007	3.83	5.092	1.830	1.844	7.515
Average value on 3 markets on MAPE5 and MAPE22					5.396

Table XX has indicated clearly that the RBF kernel has no effect on improving the accuracy of the forecasting. The average MAPE value is 5.396. Unlike the result using Support Vector Machines in Chapter 3 Section 3.1 where RBF kernel has the best performance, RBF kernel did not improve the result. Even in NN model, the use of kernel is much better than the linear model. One of the possible explanations of the difference is RBF kernel is better with more hidden layers structure while SLFNs is a single hidden layer structure which the kernel has very limited effect on the adaptation process. In general, ELM model has very good performance on Shanghai complex index and Hang Seng Index as almost all MAPE5 value is under 2. This is an interesting finding as SVR models have better performance in Dow Jones index. The overall MAPE average of ELM ranked only the third with reference to Table VI. NN is 5.463 while ELM is 5.396 but it would not be fair to judge that ELM is better than NN. As pointed out in the above section, it is also important to judge the complexity of the model. The running time of all models is very fast and ELM is also very easy to implement, unlike SVR model as the hidden nodes/neurons do not need to be iteratively turned in wide types of neural networks and learning modes in ELM. Here, it is necessary to point out the advantage of the structure of ELM which can facilitate a huge amount of data in the model, therefore is an ideal choice in a big data environment. In Chapter 4, a much larger dataset than the above 4 data sets will be employed to compare with ELM.

Tables XXI – XXII give results tested on ARIMA and MA models.

Table XXI ARIMA and MA Model MAPE5 Results

Market	Volatility	Kurtosis	Skewness	MAPE5	MAPE5	MAPE5
				ARIMA	ARIMA	MA
				(0,1,1)	(1,1,0)	
DJ2002_6	1.00	2.360	0.763	0.238	0.244	0.258
DJ2006	**0.66**	2.821	-0.387	**0.230**	**0.231**	**0.258**
DJ2003_7	1.83	2.080	-0.308	2.204	2.207	2.371
DJ2007	6.17	2.710	0.289	2.196	2.196	2.371
HSI2002_6	1.47	2.254	0.612	0.990	0.991	0.972
HSI2006	3.06	2.333	0.326	0.989	0.989	0.972
HSI2003_7	3.17	2.198	0.774	2.196	2.197	1.657
HSI2007	5.90	4.042	1.092	2.176	2.176	1.657
SH2002_6	2.92	3.428	0.741	2.224	2.227	2.957
SH2006	7.90	5.273	0.766	2.162	2.168	2.957
SH2003_7	3.81	**1.742**	**-0.016**	1.235	1.235	1.083
SH2007	3.83	5.092	1.830	1.227	1.227	1.083
average				1.506	1.507	1.550

The best MAPE5 result is from model ARIMA(0,1,1) which is 0.230 while the SMAPE5 is 0.115. However, ARIMA(0,1,1), ARIMA(1,1,0) and MA models all have very similar MAPE5 results. Moving average model is just a simple average of the last two transaction day values. It is not covered in the Literature Review Chapter 2 as it is a very simple model which is self-explanatory. The short-term 5-day financial time series forecasting is quite accurate using 2 days moving average forecasting method as the smoothing effect is lesser. This is because the 5-day fluctuation is normally not very bumpy except in very special financial circumstance like 2008 financial tsunami crisis. It is interesting to point out the best MAPE value is in DJ2006 which has the lowest volatility value 0.66.

Table XXII AIRMA and MA Model MAPE22 Results

Market	Volatility	Kurtosis	Skewness	MAPE22	MAPE22	MAPE22
				ARIMA	ARIMA	MA
				(0,1,1)	(1,1,0)	
DJ2002_6	1.00	2.360	0.763	**0.530**	**0.532**	**0.504**
DJ2006	**0.66**	2.821	-0.387	0.545	0.545	0.504
DJ2003_7	1.83	2.080	-0.308	5.716	5.717	5.878
DJ2007	6.17	2.710	0.289	5.709	5.709	5.878
HSI2002_6	1.47	2.254	0.612	1.737	1.738	1.712
HSI2006	3.06	2.333	0.326	1.735	1.735	1.712
HSI2003_7	3.17	2.198	0.774	9.165	9.165	8.569
HSI2007	5.90	4.042	1.092	9.142	9.142	8.569
SH2002_6	2.92	3.428	0.741	4.767	4.769	6.003
SH2006	7.90	5.273	0.766	4.484	4.500	6.003
SH2003_7	3.81	**1.742**	**-0.016**	6.538	6.538	6.595
SH2007	3.83	5.092	1.830	6.538	6.538	6.595
Average				4.717	4.719	4.877

Table XXII shows the best MAPE22 result is from MA model. ARIMA and MA models are different from NN and SVR models. The MAPE5 value and the first 5 values of MAPE22 are the same in ARIMA and MA models as they are just based on the last values to generate the future value. In NN and SVR, the model uses the last 22 days value to validate the training model in order to generate the future 22 values. GARCH model is similar to ARIMA model. The best MAPE22 value in all the models is in DJ200-6 which has the second lowest volatility value 1.00.

Table XXIII ARIMA and MA Model Prediction Results

Market	MAPE	MAPE	MAPE
All	ARIMA (0,1,1)	ARIMA (1,1,0)	MA
Average	3.111	3.113	3.213

Table XXIII shows the best result is from ARIMA(0,1,1) with average MAPE value 3.111. This value will be used as a benchmark to compare with the other models results in Chapter 4 using Conditional Restricted Boltzman Machine. In fact, financial time series is not affected by seasonal factor like clothing or food market industry. There is no indication that Summer financial activity is lower than Spring or vice versa. The global effect on financial time series is strong as the developed world economy moves more or less at the same pace. That is why Hang Seng Index will follow the Dow Jones Average Index movement. However, the global trend may move at the same pace but not necessary at exactly at the same time except the big events like financial tsunami.

Each market is different despite the influence of other markets and the effect of the global trend may be different too. Hence, it is still worthwhile to study the individual market and able to forecast the future event. ARIMA(1,1,0) model has more or less the same result as the ARIMA(0,1,1) which is 3.113. The difference is not significant to differentiate which one is better. ARIMA(0,1,1) is equivalent to MA(1,1) model while ARIMA(1,1,0) is equivalent to AR(1,1) model.

3.5.4 Conclusions

ELM model is the latest neural network forecasting tool which has great potential to develop such as (Cao, Huang, & Sun, Optimization-Based Extreme Learning Machine with Multi-kernel Learning Approach for Classification, 2014) and (Atsawaraungsuk, Horata, Sunat, Chiewchanwattana, & Musigawan, Evolutionary Circular Extreme Learning Machine, 2013). In the paper (Huang G. B., What are Extreme Learning Machines? Filling the Gap between Frank Rosenblatt's Dream and John von Neumann's Puzzle, 2015), it has been clearly pointed out the fundamental advantage of ELM is that there is no need to iteratively tuned in wide types of neural networks and learning models the hidden nodes/neurons. As a consequence, ELM has a faster computational speed than any of the models in this thesis. It is by far the most efficient and easiest to handle the algorithms and can handle quite a large amount of data. Nevertheless, the RBF kernel in ELM did not improve the accuracy of the above models but there are 11 more networks such as Threshold networks (Huang, et al., Can threshold networks be trained directly?, 2006), fully complex neural networks (Huang, Saratchandran, & Sundararajan, Fully complex extreme learning machine, 2005) or Fourier series (Huang, Chen, & Siew, Universal Approximation Using Incremental Constructive Feedforward Networks with Random Hidden Nodes, 2006) that can be incorporated into ELM to improve the accuracy. More datasets with different applications other than financial time series forecast have to be conducted in order to find out the reason. Only Close value of the 3 market indices is inputted into the ELM model which is the same as ARIMA GARCH and MA models. ELM average MAPE value is 5.396 which is not as good as ARIMA model 3.111. They all use the same input parameter but adopt a very different approach in forecasting. ARIMA has very strong statistical background using Yule-walker equation to calculate the auto-correlation function while ELM is a relatively new model but it rapidly gathers all the attention due to its theoretical complement to Support Vector Machine. The average MAPE value of SVR model is 4.833 and is better than ELM 5.396 value. This is intriguing as both using partition theory to distinguish or classify the data in order to select the crucial points to represent the whole data set. Later, these points are used to forecast the result. Again, it is not easy to judge if ELM is better than SVR based on the above experiment. As there are parameters to tune, ELM is definitely much better than SVR in terms of usage. The limitation of ELM is not deep learning and it does not gather each layer of the network to perform the forecasting.

Box-Jenkins model from George Box and Gwilym Jenkins is a famous time series analysis algorithm using autoregressive moving average ARMA or ARIMA to fit the historical record. Once a formula is discovered to fit the historical data, it will be projected into the future events. Here, it should point out that ARIMA so far has the best performance in forecasting the 3 market Indices value particularly in a 5-day horizon but it is not very practical for long term forecasting. ARIMA is difficult to achieve long-term prediction as per (Nkayam, Ata, & Oka, Predicting time series of individual trends with resolution adaptive ARIMA, 2013), this is partly due to the ARIMA employed Yule-Walker equation to estimate parameters by replacing the theoretical covariance with estimated value. For long term forecasting, it is very difficult to fit the parameters into an equation. Therefore, Resolution Adaptive ARIMA in (Nkayam, Ata, & Oka, Predicting time series of individual trends with resolution adaptive ARIMA, 2013) is developed to increase the long-term prediction ability. SVR selects support vector from the dataset to represent the whole data even though it depends on the spread of the data, it is still possible for SVR to perform long-term forecasting. ARIMA model uses all the data to search for the best fit and in terms of the fitness to historical data, ARIMA is better for stationary data and short-term forecast than SVR as the MAPE5 of SVR is 2.585 while MAPE5 of ARIMA is 1.506 and all the MAPE5 value in SVR are higher than ARIMA model. The more data to fit the historical data, the better the result. However, ARIMA can only view the data in a simple plane and it is not easy to classify which data is more important while SVR can map the data into hyperplane in order to select some points to represent the pattern. In this respect, SVR methodology is much better than ARIMA. Judging from the result, the average MAPE in ARIMA is 3.111 while SVR is 4.833. We cannot simply say ARIMA is much better than SVR. The average ARIMA MAPE22 value is 4.717 while SVR is 8.109. But there are 4 datasets in which MAPE22 value in SVR is better than ARIMA. The strength of SVR comes from its ability to classify while ARIMA is strong in regression. Both algorithms enjoy high reputation in its own field from the literature review. The history of these two algorithms is relatively new no more than 30 years compared to the classic statistical theories which could be more than 250 years old such as Bayseian theory. It is sufficient to say that up to this point ARIMA has produced by far the best result (lowest MAPE5 value 0.23) from the above experiments and it is a very good means to forecast Dow Jones Average Index. The limitation of ARIMA is not suitable for high dimension data. It is a statistical conceptual model which is not designed to handle high dimension data.

Chapter 4 New Approach Using Conditional Restricted Boltzmann Machine

4.1 Introduction

With the dawn of big data environment, there are many new techniques to handle such vast volume of data. The MNIST database (Mixed National Institute of Standards and Technology database) is a large database of handwritten digits that is commonly used for training various image processing systems. The database contains 60,000 training images and 10,000 testing images. This is a very near-human performance. In the paper presented by (Hinton, Osindero, & Teh, A Fast Learning Algorithm for Deep Belief Nets, 2005), the generalization performance of the network has a 1.25% error on the 10,000-digit official test set. This beats the 1.5% achieved by the best backpropagation nets when they are not handcrafted for this particular application. It is also slightly better than the 1.4% errors reported by Decoste and Schoelkopf (2002) for support vector machines on the same tasks. In the program conducted by Ruslan Salakhutdinov and Geoff Hinton using Restricted Boltzmann Machine (RBM) in deep learning training environment, the error rate is 1.5%. It is not the focus of this research to discuss the research development of MNIST dataset but to point out the Restricted Boltzmann Machine has obtained a remarkable success. Inspired by this success, Conditional Restricted Boltzmann Machine (CRBM) which has the ability to forecast future event is selected for this research as part of a further investigation for the domain of this research – forecasting problem. In (Cai & Lin, Forecasting High Dimensional Volatility Using Conditional Restricted Boltzmann Machine on GPU, 2012), the application of CRBM in high dimensional data to forecast the volatility of multivariate asset return is a typical example of how this tool can handle high dimensional data. In this section CRBM is selected as a candidate for a forecasting model is not only because of its ability and success in handling big data but the conceptual difference between CRBM and all other forecasting models should provide a different perspective in the research domain. Forecasting using big data is an unexplored territory and many literature reviews in forecasting only selected some low dimension segment of data for forecasting. It is a reasonable deduction that high dimensional big data should provide a wider and deeper spectrum of hidden information for the right forecasting tool to exploit.

The classic neural network or supporting vector machine theory on classification is based on Euclidean distance method. The shorter the distance, the more cluster the data. The mapping of the dataset into higher dimension will facilitate to draw a hyperplane to separate the data but it is still under Euclidean distance concept. RBM has a different approach. It assigns low energy based on probability concept to those connections that are relevant to each other while high energy to those that are irrelevant. RBM also combines the deep learning theory which only trains one layer at a time to obtain proper information. The conceptual details of deep learning theory can be found on the IPSAM Summer School 2012 (Research, 2012). From the conceptual point of view, Euclidean distance methodology is limited to finding the centre point to calculate the distance. Distance cannot be calculated without a point of reference. Here, which centre point is the most relevant is very crucial to the success of the algorithm. In RBM, probability concept is applied throughout the whole dataset. The centre point is irrelevant as each point will be calculated independently without reference to any point using an energy-based model with probability distribution calculation method.

Conditional RBM or (CRBM) is similar to RBM but many non-conditional RBM algorithms are not applicable to train CRBM. The Equation 44 which defines a joint probability over v and h of RBM and Equation 52 which defines probability distribution v and u of CRBM have indicated that CRBM models the distribution p(v|u) by using an RBM to model v and using u to dynamically determine the biases or weights of that RBM. By definition, RBM neurons must form a bipartite graph: a pair of nodes from each of the two groups of units, commonly referred to as the "visible" and "hidden" units respectively, may have symmetric connection between them, and there are no connections between nodes within a group. Conditioning vector u is to determine increments to the visible and hidden biases of the RBM and therefore a unique distribution p(v|u) is generated. CRMB are generally used for forecasting (Cai & Lin, Forecasting High Dimensional Volatility Using

Conditional Restricted Boltzmann Machine on GPU, 2012), (Minh, Larochelle, & Hinton, Conditional Restricted Boltzmann Machines for Structured Output Prediction, 2010) and (Taylor, Sigal, Fleet, & Hinton, Dynamical Binary Latent Variable Models for 3D Human Pose Tracking, 2010). RBMs have found applications in dimensionality reduction (Hinton & Salakhutdinov, Reducing the Dimensionality of Data with Neural Networks, 2006), classification, (Teh & Hinton, 2001) and (Larochelle & Bengio, Classification using discriminative restricted Boltzmann machines (PDF, 2008) collaborative filtering (Salakhutdinov, Hinton, & Minh, Restricted Boltzmann machines for collaborative filtering., 2007), feature learning (Coates, Lee, & Ng,). An analysis of single-layer networks in unsupervised feature learning, 2011) and topic modelling (Salakhutdinov & Hinton, Replicated softmax: an undirected topic model, 2011)

4.2 Empirical Modeling

In Chapter 3 forecasting models, the output value is either predicted by using 4 attributes Open, High, Low and Close or just the close value in the case of ARIMA or GARCH model. The historical data consist of around 1200 records and hence the data matrix is about 1200x4. The famous Black-Scholes formula uses only one value or attribute for stock price prediction and yet it is still used in many financial markets for option trading. It is not our research to challenge or even validate the accuracy of the Black-Scholes formula despite its controversial assumption that the market behaves like a normal distribution. There is a lot of research using different approaches instead of the Black-Scholes formula but it has not yet gained world-wide recognition in either academic research or in the financial institution. We would like to offer a different approach in financial time series analysis which is using more attributes other than one stock but stocks with similar activity in open markets such as the index constituents of Dow Jones or Hang Seng Index. From (L., 2012), the market interdependence is an important aspect which must incorporate into the forecasting domain. In the beginning of this research, it has been pointed out clearly that financial tsunami has a great impact on the financial market on a global scale. To what extent it affects each market is beyond the scope of this research. However, several markets' interdependence can be investigated with proper forecasting model. It is due to this theory that the candidate contemplates more attributes input into the model can discover the interdependence. As a result, the better the forecasting result.

Since the research focuses on financial time series forecasting, CRBM is employed because it works on the miniature big financial time series data which by definition is a continuous data. The program for CRBM was provided by Graham Taylor, Geoffrey Hinton and Sam Roweis (Taylor, Hinton, & Roweis, Modeling Human Motion Using Binary Latent Variables, 2006) which is originally designed to forecast robot movements. The candidate modified the algorithm in order to apply financial time series data into it. It can handle up to 120 dimensions of data with 60000 records.

It is not the intention of this research to study the characteristics of Human Motion as it is a big topic. However, it is interesting to borrow the concept of style translation process which is transforming an input motion into a new style while preserving its original content. Human Motion is a coordination of head, hand, body and foot movement. To a certain extent, financial time series is also a coordination of different markets their movement at the same time. In (L., 2012), it has been pointed out the breadth and depth of financial markets' interdependence has been blamed for their domino effects during recent financial crises. Modern economic and investment theories already showed this feature.

The technical specification of the model is as follows which is based on (Taylor, Hinton, & Roweis, Modeling Human Motion Using Binary Latent Variables, 2006). This program trains a Conditional Restricted Boltzmann Machine in which visible, Gaussian-distributed inputs are connected to hidden, binary, stochastic feature detectors using symmetrically weighted connections. Learning is done with 1-step Contrastive Divergence. Directed connections are present, from the past configurations of the visible units to the current visible units, and the past configurations of the visible units to the current hidden units. It is a 2-step deep learning procedure to return the visible units and hidden units weight factor. These factors will be used to generate the future value of the model. A frame is selected from the original dataset as a point of reference to generate 5-day or 22-day forecasting values. The following is a brief description of the algorithm.

As described before, RBM used energy-based learning method a details description can be found in (LeCun, Chopra, Hadsell, Ranzato, & Huang, A Tutorial on Energy-Based Learning, 2006)**.** The dependency between variables in Energy-Based Model (EBM) is achieved by associating a scalar energy with each configuration of the variables. The observed configurations of the variables are given lower energies while the unobserved configurations are given high energies. An energy function in called "negative parabolic log likelihood function" is setting the value of the observed variable and finding values of the remaining value that minimize the energy. This means for any setting of the hidden units, the visible unit is defined by the following equation.

$$-\log p(v,h) = \sum_i \frac{(v_i - c_i)^2}{2\sigma_i^2} - \sum_j b_j h_j - \sum_{i,j} \frac{v}{\sigma_i} h_j w_{ij} + Const$$

Equation 59

The σ_i is the standard deviation of the Gaussian noise for visible unit i.

The main advantage of using this undirected, "energy-based" model rather than a directed "belief net" is that inference is very easy because the hidden units become conditionally independent when the states of the visible units are observed. The conditional distributions (assuming $\sigma_i = 1$) are:

$$p(h_j = 1 | v) = f(b_j + \sum_i v_i w_{ij})$$

Equation 60

$$p(v_i \mid h) = N(c_i + \sum_j h_i w_{ij}, 1)$$

Equation 61

where f(·) is the logistic function, N(μ, V) is a Gaussian, bj and ci are the "biases" of hidden unit

j and visible unit i respectively, and w_{ij} is the symmetric weight between them.

Maximum likelihood learning is slow in an RBM but learning still works well if we approximately

follow the gradient of another function called the contrastive divergence. The learning rule is:

$$\Delta w_{ij} \propto < v_i h_j >_{data} - < v_i h_j >_{recon}$$

Equation 62

where the first expectation (over hidden unit activations) is with respect to the data distribution and the second expectation is with respect to the distribution of "reconstructed" data. The reconstructions are generated by starting a Markov chain at the data distribution, updating all the hidden units in parallel by sampling Equation 60 and then updating all the visible units in parallel by sampling Equation 61. For both expectations, the states of the hidden units are conditional on the states of the visible units, not *vice versa*. The learning rule for the hidden biases is just a simplified version of Equation 62:

$$\Delta b_j \propto <h_j>_{data} - <h_j>_{recon}$$

Equation 63

So far the RBM has not incorporated any temporal information. Temporal dependencies can be modelled by treating the visible variables in the previous time slice as additional fixed inputs. Fortunately, this does not complicate the inference. Two types of directed connections are added : autoregressive connections from the past n configurations (time steps) of the visible units to the current visible configuration, and connections from the past m visible to the current hidden configuration. The addition of these directed connections turns the RBM into a conditional RBM (CRBM).

Inference in the CRBM is not more difficult than in the standard RBM. Given the data at time t, t − 1, . . . , t − n, the hidden units at time t are conditionally independent. We can still use contrastive divergence for training the CRBM. The only change is that when we update the visible and hidden units, we implement the directed connections by treating data from previous time steps as a dynamically changing bias. The contrastive divergence learning rule for hidden biases is given in Eq. 5 and the equivalent learning rule for the temporal connections that determine the dynamically changing hidden unit biases is:

$$\Delta d_{ij}^{t-q} \propto v_i^{t-q}(<h_j^t>_{data} - <h_j^t>_{recon})$$

Equation 64

where d_{ij}^{t-q} is the log-linear parameter (weight) connecting visible unit i at time t − q to hidden unit j for q = 1..n. Similarly, the learning rule for the autoregressive connection

$$\Delta a_{ki}^{t-q} \propto v_k^{t-q}(v_i^t - < v_i^t > recon)$$

Equation 65

where a_{ki}^{t-q} is the weight from visible unit k at time t−q to visible unit i.

The origin algorithm is to model human motion and generate the motion from the captured data. As the model is able to efficiently capture complex non-linearity in the data without sophisticated pre-processing or dimensionality reduction, it is also suitable for other high-dimensional time series such as financial time series in a big data environment. The candidate was inspired by this algorithm and modified it for financial data forecasting. Human motion and financial movement are similar in nature. The head and shoulder movement is restricted by the body structure while financial movement is bounded by the economy and political environment. However, the past time steps for financial forecasting is set to 1 for visible and 1 for hidden variables. But in human motion, the past time steps m and n values are set to 3. In human motion, the updates are conditional upon the past 2 to 3 time steps in learning a model because the complexity of motions requires a few steps to gather the information. The relationship between head and shoulder has more information than financial movements between 2 equities. Hence, more detailed information is required in order to update the learning model. In financial movement, 1 past step is enough to update the training model because 2 or 3 past time steps information could be redundant as the movement is affected by external factors such as the political environment or economic situation.

4.3 Empirical Results

In order to compare with the models in Chapter 3, the previous 3 markets indices of Dow Jones, Hang Seng and Shanghai complex with the same time horizon 2006, 2002 to 2006, 2007 and 2003 to 2007 are used as inputs for the CRBM model.

Table XXIV Overall result of All Models

Methods	MAPE5			MAPE 22			Overall
	DJ	HSI	SH	DJ	HSI	SH	Average
A(011)	1.217	1.588	1.712	3.125	5.445	5.582	3.111
A(110)	1.219	1.588	1.714	3.126	5.445	5.586	3.113
MA	1.390	1.325	2.307	3.241	5.181	6.668	3.352
SVR RBF	1.040	1.525	5.190	2.473	5.020	13.753	4.833
CRBM	2.901	1.121	5.389	5.336	6.470	8.181	4.900
NN	1.203	4.219	4.533	3.054	4.154	13.035	5.033
SVR Lévy	1.556	3.276	9.425	3.026	5.504	13.847	6.106
ELM	3.972	4.450	2.914	12.399	20.644	21.466	10.974
GARCH	3.972	4.450	2.914	12.399	20.644	21.466	10.974
	MAPE4			MAPE 20			
SVR RBF	1.376	2.879	2.09	6.49	4.139	2.524	3.250
WL_db_lssvm	2.779	3.604	2.301	7.718	5.343	7.718	4.911
WL_sym_garch	1.593	1.717	3.319	7.982	10.803	18.480	7.316
WL_sym_SVR	1.224	1.870	3.801	7.619	12.265	22.399	8.196
WL_db_SVR	1.184	1.891	4.991	7.767	13.846	22.750	8.738
WL_db_garch	1.685	2.660	7.976	9.430	13.187	19.276	9.036
WL_sym_lssvm	1.776	1.801	3.624	53.361	7.441	33.574	16.930
LSSVM	11.533	19.338	24.743	12.721	15.052	22.020	17.568

In Table XXIV, CRBM ranked the fourth which is not too bad by any standard. CRBM model has an overall average MAPE value of 4.9 while the best overall average MAPE from ARIMA(0,1,1) is 3.111. The difference is 57.5% which need to improve.

CRBM is the only model that simultaneously forecast all the MAPE values in Table XXIV which attempt to use the interdependence of these 3 markets for forecasting. From this experiment, it is not successful. There could many reasons one of which could be there is only 3 markets which many not be easy to seek out the correlation. The other reason could be the historical data is not long enough only 1200 records. The last but not least reason is the dataset selected within the financial tsunami period which is too volatile to capture the interdependence information. Despite this is not satisfactory, the candidate is not discouraged and has presented another approach using this method in different dataset environment to improve the accuracy.

Table XXV Best Result of Each Model in a Specify Time Horizon

Methods	MAPE4	MAPE5	Market	Horizon	Structural difference	remarks
CRBM	NA	1.03	HSI	2002-7	12 attributes to input parameters	time step = 1
A(011)	NA	0.230	DJ	2006	yule-walker equation regression	
A(110)	NA	0.231	DJ	2006	yule-walker equation regression	
MA	NA	0.258	DJ	2006	Last 2 records average regression	
ELM	NA	1.541	SH	2006	Linear kernel	
NN	NA	0.273	DJ	2002-6	grnn network	
SVR RBF	NA	0.380	DJ	2006	rbf kernel	parameter c =1000 g = 1
SVR Lévy	NA	0.360	DJ	2006	rbf kernel	parameter c =1000 g = 1
GARCH	NA	1.280	DJ	2006	EGARCH network	transformed prices to returns
Chart pattern_LSSVM	NA	0.163	NA	NA	Pattern 49 search result	
SVR RBF	0.318	NA	DJ	2006	rbf kernel	parameter c =1000 g = 1
WL_db_SVR	0.501	NA	DJ	2006	daubechies wavelet	
WL_sym_SVR	0.354	NA	DJ	2006	symmlets wavelet	
WL_db_lssvm	0.262	NA	DJ	2002-6	daubechies wavelet	
WL_sym_lssvm	0.262	NA	DJ	2006	symmlets wavelet	
GARCH	1.040	NA	DJ	2006	EGARCH network	
WL_db_garch	0.617	NA	HSI	2007	daubechies wavelet	
WL_sym_garch	0.758	NA	SH	2007	symmlets wavelet	

Table XXV gives a summary of all the models structural difference. They are compared using the 4 datasets as described in the above section. Except CRBM, most of the forecasting models are based on literature. The candidate has selected 4 datasets around the financial tsunami period to check the models' performance. The best performance has been listed out and the corresponding conditions and criteria have been described on how to reach it. The best result is from the chart pattern LSSBM model 0.163 but it is not within the 4 datasets and the design of the algorithm is quite different from the others. However, the algorithm is only usable if chart pattern is discovered. The candidate has conducted many experiments in order to summarize the result as per the above table of which should be considered as a contribution in financial time series forecasting. Despite most of the models are from literature for many years and there have been many applications but to compare all of them using the same benchmark and obtain a remarkable result is difficult. All the above MAPE value is under 2 which has been the objective of this thesis. The above table could a guidance of which model to apply in a different market and time horizon.

In the following experiment, 9 different markets are used instead of 3. It seems the performance is much better.

Table XXVI Nine Different Markets MAPE Using CRBM

Market	MAPE5	MAPE22
Dow Jones	0.533	5.129
Hang Seng	1.594	1.966
Nikkei	1.216	4.498
Shanghai	3.095	8.096
Straits	0.979	1.269
Korea	0.488	1.598
Philippine	1.778	8.535
Bangkok	1.683	7.661
Taiwan	1.038	2.561
Overall average MAPE value		**2.984**

In Table XXVI, there are 9 indices of which 8 of them are Asian Indices plus the US major stock index. The assumption is that the Asian stock index has been influenced by US stock movement for a very long time and its impact is still in effect. From the DataStream financial data resource centre (Thomson, 2012), only 2563 trading days were recorded prior to 29 August, 2014 of the above 9 indices. Thus, the experiment can only use 2563 records and input into the model. The close value of each market is selected. In other words, the data matrix is 2563x9. Table XXVI has a remarkable improvement in its accuracy. The number of records in this experiment 2563 is more than that in Table XXIV of which it has only 1200 records. The average MAPE 2.984 is even better than the best result 3.111 from ARIMA(0,1,1). In contrast to Table XXVI which only has 3 markets Open, High, Low and Close values as input, only 9 Close values are used in Table XXVII. It seems the more dimensions, the better the CRBM model. We would like to look at more examples to examine whether the dimension is a curse or a blessing. Despite the previous argument on Shanghai Complex Index which is not a good market indicator, it is not bad when it is combined with other Asian Stock Index. As expected before, the longer the historical records and more market involve, the interdependence is easier to capture. As a result, the forecasting accuracy is higher.

So far the dataset is focused on Index value. Although stock exchange index plays a significant role in the market movement as well as a key indicator of how a particular stock will trade, it is still not as good as to directly forecast the individual stock.

Table XXVII Constituents of Thirty Stocks in Dow Jones Index

	Constituents	MAPE5	MAPE22
1	3M	0.277	1.42
2	AE	0.452	2.344
3	AT&T	1.384	1.389
4	Boeing	1.988	1.891
5	Caterpillar	1.049	3.655
6	Chevron	0.603	4.241
7	Cisco	1.374	1.889
8	Coca-Cola	1.156	5.841
9	DuPont	0.347	4.307
10	Exxon	0.527	3.063
11	GE	0.824	1.441
12	Goldman Sachs	1.394	3.859
13	Home Depot	1.386	8.009
14	Intel	0.915	1.741
15	IBM	0.405	0.742
16	J&J	0.891	4.404
17	JP Morgan	1.215	2.813
18	McDonalds	1.487	1.739
19	Merck	1.526	4.219
20	Microsoft	0.889	4.887
21	Nike	1.331	4.772
22	Pfizer	1.474	1.563
23	P&G	0.553	5.679
24	Travelers	0.506	3.182
25	United Health	3.49	4.469
26	United Tech	0.789	4.332
27	Verizon	1.738	2.474
28	Visa	0.752	1.665
29	Walmart	0.909	3.29
30	Walt Disney	1.037	1.805
Average		1.089	3.238
Overall average MAPE value			**2.163**

Table XXII has average MAPE value 2.163 which is even better than Table XXVI. Only 707 records were found in the 50 stocks of SSE 50 prior to 29 August 2014. Hence, only 707 close values of each stock are inputted into the model. In other words, the data matrix is 707x30.

The dimensionality of this data set is 30 instead of 9 in the above. As all the records are in the Dow Jones Industrial Average constituents, the close relationship is assumed but it is still amazingly to find out the accuracy is that high. Unlike the other forecasting techniques which can only forecast one value at a time, the CRBM model does forecast all the values at the same time. Each stock MAPE value is calculated independently e.g. 3M stock forecasted 5 days stock price will be compared with the actual stock price to derive the MAPE5 value. This is very efficient to use in online trading. The forecasting technique of CRBM is different than the other model as it uses the relative movement of all the stocks in one time step as an anchor to infer all the stocks movement in the next time step. While the other models deducted the historical pattern of a stock and forecast the next time step based on that pattern, CRBM has a significant structural difference compare to them.

Table XXVIII Constituents of Fifty Stocks in Hang Seng Index

No.	Constituents	MAPE5	MAPE22	No.	Constituents	MAPE5	MAPE22
1	1299 HK Equity	1.085	2.504	26	101 HK Equity	0.936	4.049
2	3988 HK Equity	1.683	2.222	27	11 HK Equity	0.69	1.551
3	3328 HK Equity	1.716	3.87	28	12 HK Equity	5.378	6.542
4	23 HK Equity	0.99	1.735	29	1044 HK Equity	0.854	3.494
5	1880 HK Equity	2.273	5.192	30	3 HK Equity	1.285	2.716
6	2388 HK Equity	0.693	3.617	31	388 HK Equity	2.218	2.455
7	293 HK Equity	3.246	1.474	32	5 HK Equity	0.388	0.981
8	1 HK Equity	2.781	7.522	33	13 HK Equity	1.083	4.938
9	939 HK Equity	1.754	3.997	34	1398 HK Equity	1.983	3.407
10	2628 HK Equity	2.972	2.447	35	135 HK Equity	0.95	6.583
11	2319 HK Equity	3.385	7.495	36	992 HK Equity	5.22	9.245
12	144 HK Equity	0.801	3.512	37	494 HK Equity	1.761	11.474
13	941 HK Equity	4.957	7.955	38	66 HK Equity	1.156	2.239
14	688 HK Equity	2.862	9.318	39	17 HK Equity	2.236	3.094
15	386 HK Equity	2.4	7.248	40	857 HK Equity	4.047	4.563
16	291 HK Equity	2.228	14.604	41	2318 HK Equity	2.422	5.376
17	1109 HK Equity	2.785	5.536	42	6 HK Equity	1.57	3.26
18	836 HK Equity	2.584	6.158	43	1928 HK Equity	8.587	21.544
19	1088 HK Equity	2.18	2.729	44	83 HK Equity	1.329	6.319
20	762 HK Equity	4.773	6.485	45	16 HK Equity	0.87	1.842
21	267 HK Equity	2.482	10.134	46	19 HK Equity	1.347	2.544
22	2 HK Equity	1.357	2.678	47	700 HK Equity	1.869	8.446
23	883 HK Equity	1.938	5.268	48	322 HK Equity	1.2	3.751
24	1199 HK Equity	4.087	7.891	49	151 HK Equity	7.451	8.416
25	27 HK Equity	5.902	22.428	50	4 HK Equity	1.993	4.547
Average						2.455	5.748
Overall average MAPE value							**4.101**

Similarly, in Table XXVIII, the average MAPE value 4.101 which is not too bad compared to Chapter 3 results. In fact, the MAPE5 average value 2.455 which is close to our target under 2. In order to compare with Table XXVII which used 707 records, the same number of records are inputted into the model. In other words, the data matrix is 707x50. The dimensionality of this data set is 50 instead of 30 in the above. Like Table XXVII all the records are in the Hang Seng Index constituents.

Table XXIX Constituents of Fifty Stocks in SSE50

No.	Constituents	MAPE5	MAPE22	No.	Constituents	MAPE5	MAPE22
1	601288	1.482	1.382	26	600332	3.171	9.139
2	600585	2.128	1.777	27	600837	2.818	2.799
3	601169	7.560	7.826	28	601688	2.943	5.500
4	601328	3.259	1.705	29	600015	2.653	1.055
5	600637	0.350	1.730	30	601398	1.633	1.094
6	601299	1.613	5.182	31	601166	2.475	2.577
7	601818	1.384	4.713	32	600111	1.197	4.259
8	601118	1.457	17.482	33	600887	6.018	3.943
9	601628	3.476	2.987	34	600010	2.491	20.128
10	600036	1.863	1.830	35	600518	1.169	3.034
11	600999	2.222	3.988	36	600519	3.398	2.376
12	600016	1.633	3.351	37	600406	5.725	6.504
13	601117	1.390	6.623	38	601336	4.221	2.430
14	601601	3.606	2.103	39	601857	1.219	1.592
15	600028	4.453	5.337	40	601318	2.679	2.444
16	601088	2.346	1.755	41	600048	3.014	5.838
17	601989	6.768	13.344	42	600104	5.365	9.462
18	601668	1.834	6.182	43	600703	1.787	8.176
19	600050	2.908	3.483	44	600031	1.835	3.921
20	600030	3.950	2.007	45	600547	0.648	3.626
21	601766	2.150	2.264	46	600196	1.170	2.188
22	601006	5.532	9.073	47	600018	4.350	12.578
23	601901	2.253	1.876	48	600832	0.172	1.624
24	600383	0.977	7.662	49	600000	2.034	1.649
25	600256	9.594	10.235	50	600089	5.531	7.690
Average						2.918	5.030
Overall average MAPE value							**3.974**

In Table XXIX, we cannot use the Shanghai Stock Exchange Composite Index which is a stock market index of all stocks (A share and B share) that are traded on the Shanghai Stock Exchange for comparison as it has 1083 constituents which are beyond the scope of this research.

Instead, SSE 50 Index is selected from the 50 largest stocks in Shanghai Stock Exchange which have the most influential impact on the market. Like Table XXVII and Table XXVIII, 707 trading days close value records on 29 August, 2014 are inputted into the model. Only 707 records were found in the 50 stocks of SSE 50 after 29 August 2014 and it is the reason to select 707 trading days. The performance of the model in SSE 50 is even better than Hang Seng Index. The overall average is 3.974 while Hang Seng Index is 4.101. It is fair to conclude the result is similar. However, the MAPE value is a different story. 2.918 comparing to 2.455 in Hang Seng Index indicates that the short term forecasting is better than SSE 50. Hence, the short-term forecast MAPE5 value ranking is Dow Jones, Hang Seng and SSE50. This is in line with Chapter 3 Section 3.2.2.2 EMH analysis and it is consistent with different data sets even though it is not the same forecasting technique. As explained in the result of Table XXIX that Shanghai complex index is not an open market and there are 1021 stocks in the index, the co-ordination of Dow Jones and Hang Seng Index with Shanghai complex index is not obvious. As Dow Jones only has 30 and Hang Seng has 50, Shanghai Complex Index could be too complicated to present a clear picture of the market. On top of that, Shanghai Complex Index also contains A and B share stocks of which B share is not normally being traded in Shanghai Stock Exchange market. The result in Table XXIX clearly indicates that SSE50 stocks are closely related and can give a clear direction on the market movement. It would not be fair to compare Table XXVI with Table XXVII to Table XXIX as it has 2563 trading days. Hence, the following Table XXX only takes 707 trading days ended on 29 August 2014 for comparison.

Table XXX Forecasting Result of The Nine Markets with Less Input Records

Market	MAPE5	MAPE22
Dow Jones	0.618	2.147
Hang Seng	1.633	2.056
Nikkei	1.809	3.304
Shanghai	3.247	5.594
Straits	0.779	1.433
Korea	0.461	0.935
Philippine	1.502	4.999
Bangkok	1.373	4.995
Taiwan	1.002	1.139
Overall average MAPE value		2.168

Here, the overall MAPE value has been improved from 2.984 to 2.168. This is an interesting finding and also a typical example that dimension is not always a curse but could be a blessing. Why 2563x9 historical records forecasting result is not as good as 707x9. A possible explanation in financial time series is that the impact of long historical records could have less impact on recent events as today's markets are interwoven with current event less than a year, few years or 10 years ago. More experiments on this theory will be demonstrated in Chapter 5. Unlike natural phenomena such as sun black spot, the longer the historical records, the better the forecasting result because the fluctuation of the data is not as volatile as the financial market and the pattern will be clearer if more historical records are available for input.

4.4 Conclusion

In this chapter, a revolutionary model has been introduced for financial time series forecasting. We believe this is the first attempt to apply Conditional Restricted Boltzmann Machine in financial time series forecasting. The potential for this model based on big data technology is very broad. Big data not only induce impacts on large enterprises but to a certain extent most of our daily life. Our data storage is no longer measured by MB but by GB and will easily go to TB or more. Technology to deal with big data in a computer environment is still restricted in the hands of a few companies like Horton. However, the benefit to acquire valuable knowledge from the big data is almost unlimited. Due to the limit of this research, we only use a very small fraction of the financial market data from Bloomberg, Data stream and Shanghai Exchange high frequent databases. This is no comparison with the real big data which is usually more than 10 Peta bytes in financial data alone. There are lots of financial information in other formats such as chat forum in text format which is also very valuable for analysis on how the market sentiment will turn into. We just demonstrated one type of data format that can be handled by latest big data analysis algorithm.

The advantage of CRBM is its ability to handle high dimension data although it is not a fast algorithm. High dimension data in this research means the number of parameter in the dataset. The structural difference between CRBM has a significant advantage over other models in the literature. The reason to input high dimension data for CRMB is to study all the multiple variables characteristic and infer the next time step based on the information acquired. This novel approach has wide application in short-term forecasting particularly in a high frequent transaction environment. In the next section, this application will be further elaborated. The model is a 2 step deep learning algorithm which cannot run on a normal desktop computer but customized computer with high computational power. In additional, the basic idea in CRBM is to seek out the co-relationship between different attributes but if there is a not a number (NaN) value in one of the attributes, the model will return NaN and will not be able to generate results. Hence, the dataset must filter out the NaN values so that the model can work properly. That is why in the about example only 707 records are selected because not all the stocks in SSE50 have more than 707 trading days records. In this section, we have demonstrated that the best forecasting result so far is from using more attributes and less historical records on the same segment of the market like Dow Jones, Hang Seng and Shanghai complex Index constituents stock. It is not our intention to seek out the most accuracy forecasting tool in financial time series as it would be impossible to define and maintain as explained in Chapter 1 Section 1.1. Like MNIST image dataset analysis, there is always room for improvement with newly improved algorithms. Instead, we would like to contribute a new methodology in financial time series analysis. In Chapter 3, there are many classical forecasting technologies introduced which are applicable in many fields but can only forecast one attribute at a time even they can input many attributes into the model like ELM. In our example, it is the future close index value of the corresponding stock index. Many models like ARIMA or GARCH only use one attribute to forecast its own future value. It seems this approach, despite its successful reputation, should consider another alternative.

Chapter 5 Performance and Evaluation on CRBM

5.1 Application of CRBM in Big Data Environment

From Chapter 4, we have demonstrated the advantage of using CRBM algorithm. So far, we have investigated the forecasting techniques of Neural Network, Support Vector Machine, Least Square Support Vector Machine, ARIMA, GARCH, Wavelet based Support Vector Machine, Wavelet based Least Square Support Vector Machine, Wavelet based GARCH, Extreme Learning Machine and Conditional Restricted Boltzmann Machine and their respectively forecasting results are illustrated from Table III to Table XXX. All these datasets are not huge as the biggest one is only 2568x9 in the above experiments. We have mentioned that CRBM is generally involved in big data and the following section attempts to employ big data environment and use the above model to analyse the data. In this Chapter, the Shanghai Stock Exchange High Frequency data from 2003 to 2013 which is provided by the library of Hong Kong Polytechnic University is analysed using CRBM model. The data is originally from GTA Information Technology Company and its name is Level-1 China Security Market Trade & Quote Research Database 2012 version. Its recording interval is ranging from 1 minute, 5 minutes, 10 minutes, 15 minutes, 30 minutes to 60 minutes. Please refer to (CSMAR, China Security Market Trade & Quote Research Database, 2013) for the data structure. The high frequency financial data has provided a unique platform to study the microscopic structure of the market movement. The total database size is 1.8 T bytes. There are many types of files in financial time series such as csv, excel, text and mat. Even in this dataset, there are 2 types csv and txt formats and it must be converted to Matlab data format MAT in order to be utilized in Matlab program. This is very computationally expensive. Unlike MNIST or NORB image dataset which is provided by a research institute in MAT format, it is very difficult to manipulate many financial time series data storage formats and converting them into MAT format is very tedious. This is quite a challenge for time series research. Not only is the data format difficult to handle, but the integrity of the data is also a problem. Datastream, Yahoo or Bloomberg are famous companies to keep track of the financial record, there are still many gaps in between due to some particular event to a company and this happens quite a lot in SSE50.

CRBM was based on RBM as described in section 4.1. The application of big data in RBM has many examples in the literature review in the last few years. But many are in the areas of optimization and classification such as MNIST or NORB datasets. Big data has given birth to the demand of new algorithms to handle such huge volume of data. At the same time, it has also miraculously broken many of the records in optimization and classification domain which has no breakthrough for many years. We, in the history of computational complexity domain, have never attained such success. It has also revived many of the classical algorithms such as the neural network backward propagation to revisit its application in big data environment and has attained unprecedented breakthrough. As one of the inventors of this algorithm Geoffrey Hinton from Canadian Institute for Advanced Research puts it "we discover this algorithm too soon without a properly environment to fully utilize its potential 30 years ago", now the destiny of this algorithm has finally revealed. The above is just a tip of iceberg as many breakthroughs in different domains will be coming. Classic theories and algorithms will be revisited to seek out its application in this new era. However, in forecasting domain, the application is still rare. One of the possible reasons is that traditional forecasting algorithm such as ARIMA or GARCH cannot be applied easily in a big data environment due to its inability to handle high dimension data which has been discussed in section 3.5.4. SVM has many applications in classification problem in big data but the SVR in forecasting domain is rare. The other possible reason is the availability of data in forecasting domain is not that many. Scientific data such as weather, natural phenomena, earthquake historical data and interstellar position data are very difficult to obtain and usually excluded from public access. Most of the time, the user must pay to access and analyse such data such Hadoop. Financial time series data are the most popular one as they encourage the investor to study but still some of the data like the dataset employed in this research must be purchased from the data warehouse. From Chapters 3 to 4, all the datasets are from the public source such as finance yahoo or the Hong Kong stock exchange. High frequent transaction financial data are not open to the public and not easy to get.

5.2 SSE50 High Frequent Financial Data Analysis

Year 2010 to 2013 SSE50 high frequent data are employed here to compare the result from previous Chapter 4. Each year, there are 12 months of records but not all stocks have records in all the months. Not all 50 stocks of SSE50 have the same volume of high frequent data at the time the data is captured as some stocks have a higher volume of transaction. One of the possible reasons is that they are suspended from trading under the security act if the daily fluctuation of the stock price is more than 10%. The other reason is it is a new stock just added into the SEE50 family. For example, there are 7 stocks which only started to have full records from 2012. Hence, there are only 43 stocks in SSE50 in 2011 and 50 stocks in SSE50. On top of that, each stock in the same year or month has different records than the other stock in the SSE50 constituent stocks. Hence, the datasets of CSMAR are filtered so that the same number of records of each stock in each year is picked.

The last 4 years from 2010 to 2013 are selected for this experiment because all the data files are in csv format. The rest is in txt format. We would like to test all the dataset in the same format to ensure the integrity of the data is preserved. As a result, the data structure of all the datasets in these years are as follows. In year 2013, it is 312555x50 records and 365396x43. This mean there are 50 stocks which have 312555 records in each stock and there are 43 stocks which have 365396 records in each stock. In year 2012, it is 266128x50 and 419558x43. In year 2011 is 455923x43 and year 2010 is 395517x43. There are 3 accumulated years of records, from 2012 to 2013 is 579183x50 and 784954x43, from 2011 to 2013 is 1240877x43 and from 2010 to 2013 is 1636394x43. These datasets are huge compared to the previous 2563x12. It is tedious and computational expensive algorithm to run the simulation. Without the latest computer hardware setup as depicted in Section 1.4, it is not possible to run in an ordinary laptop. The running time of each of the above setup is more than 48 hours. It is a big disadvantage compared to all the models in Chapter 3 and 4. However, the preliminary success in Chapter 4 has triggered a further investigation on the potential ability of this model. It took more than a month to produce the following result. Despite saying so, it does not mean that it is time-consuming and therefore not worthwhile nor significant. Certainly, it will be a waste of time and resource if the end does not justify the mean. Without the success of Chapter 4 using CRBM, it would be too ambitious to exploit the resource to seek out the benefit of CRBM in this dataset.

5.3 Empirical Results

Table XXXI SSE50 Forecasting Result in Year 2013

SSE50	Constituents	MAPE500	SSE50	Constituents	MAPE500
1	600000	4.5428	26	600832	7.250
2	600010	6.6320	27	600837	2.976
3	600015	5.9076	28	600887	5.066
4	600016	8.7228	29	600999	6.096
5	600018	6.8908	30	601006	3.382
6	600028	3.3252	31	601088	2.722
7	600030	3.0228	32	601117	2.700
8	600031	8.0970	33	600118	11.149
9	600036	1.9601	34	601166	4.187
10	600048	8.3414	35	601169	3.129
11	600050	2.8730	36	601288	1.727
12	600089	5.2806	37	601299	4.703
13	600104	3.1120	38	601318	5.004
14	600111	10.6813	39	601328	5.445
15	600196	10.3960	40	601336	1.415
16	600256	12.3538	41	601398	3.628
17	600332	7.6767	42	601601	2.547
18	600383	4.6453	43	601628	2.686
19	600406	3.1035	44	601668	3.551
20	600518	3.8809	45	601688	1.624
21	600519	6.7135	46	601766	4.884
22	600547	14.0277	47	601818	3.599
23	600585	3.5379	48	601857	1.625
24	600637	6.4399	49	601901	6.397
25	600703	8.0345	50	601989	4.327
Average					**5.240**

As described before, one day of the transaction could have generated more than 2,000 to 3,000 records. Therefore, we need to forecast around 500 future values or roughly half a day transaction to compare. The forecasting window of 500 is based on the percentage of the daily transaction volume. If it is 2,000 records, it is almost a quarter of a day. The SSE is open for trading from Monday to Friday. The morning session begins with centralized competitive pricing from 07:45 to 07:55 and continues with consecutive bidding from 09:30 to 11:30. This is followed by the afternoon consecutive bidding session, which starts from 13:00 to 15:00. Hence, it is roughly 4 hours per day. The forecasting window of 500 is approximately 1-hour trading details of all the stocks involved. This is very useful for online trading particularly in option related derivative as they are heavily being traded in the stock market. The fluctuation in 1 hour may not be that intensive for many stocks but the momentum could be very strong for option price. In Chapter 4, the forecasting horizon for consideration is 5-day and 22-day which corresponds to roughly 1 week and 1 month trading periods. The result from Table XXXI is very encouraging as the MAPE500 value is 5.24 compared to Table XXIX MAPE value of 3.974. Again the dataset matrix in the result of is Table XXIV 707x50 while the Table XXXI is 312555x50. The running time is 34610 seconds for 312555x50 and the Table XXXI is an average of 5 runs which is around 2 days computational time. It would be meaningless to use only 707 records in this experiment as it is less than 1 transaction day. In order to reduce the computation time, 11775 records are selected to see if the accuracy can be improved with less historical records. From experimental point of view, there is no harm trying as it is necessary to reduce the running time so that it can be applicable in an ordinary computer set up or in a mobile computing platform.

Table XXXII SSE50 in Year 2013 with 11775 Input Records

SSE50	Constituents	MAPE500	SSE50	Constituents	MAPE500
1	600000	0.234	26	600832	0.944
2	600010	0.287	27	600837	0.204
3	600015	0.636	28	600887	0.290
4	600016	0.469	29	600999	0.799
5	600018	0.190	30	601006	0.635
6	600028	0.477	31	601088	0.121
7	600030	0.464	32	601117	0.338
8	600031	0.163	33	600118	4.387
9	600036	0.817	34	601166	0.291
10	600048	0.766	35	601169	0.200
11	600050	0.188	36	601288	0.260
12	600089	0.316	37	601299	0.284
13	600104	0.226	38	601318	0.248
14	600111	0.172	39	601328	0.199
15	600196	1.086	40	601336	0.149
16	600256	0.173	41	601398	0.178
17	600332	0.577	42	601601	0.226
18	600383	2.401	43	601628	0.146
19	600406	0.163	44	601668	0.210
20	600518	0.363	45	601688	0.195
21	600519	0.344	46	601766	0.539
22	600547	0.495	47	601818	0.179
23	600585	0.193	48	601857	0.166
24	600637	0.286	49	601901	0.454
25	600703	0.585	50	601989	0.227
					0.479

With less number of historical records, the accuracy has been improved from 1.823 to 0.479 as per the result from Table XXXII and once again demonstrating that the stocks within the same region or market are very closely related. The MAPE value of 0.479 based on forecasted of next 500 transactions cannot directly be compared with the best result from ARIMA 3.111 as per Table XXIII. First, the previous one MAPE500 is most likely on the same day with volatility -0.16 while the latter MAPE22 is the next 2 transaction days forecast with volatility 3.48. In order to compare the result with another algorithm that is also suitable in the big data environment, ELM is used as a reference. The reason why only 11557 records are selected is because it is the maximum data matrix the ELM can handle. This was the testing result on an ordinary computer with 2G RAM i-3 Laptop with only 80G hard disk memory. The following table is the result of ELM.

Table XXXIII SSE50 Forecasting Using ELM

SSE50	Constituents	MAPE500	SSE50	Constituents	MAPE500
1	600000	0.169	26	600832	1.146
2	600010	0.000	27	600837	0.351
3	600015	0.277	28	600887	0.144
4	600016	0.609	29	600999	0.665
5	600018	0.258	30	601006	0.598
6	600028	0.373	31	601088	0.147
7	600030	0.520	32	601117	0.254
8	600031	0.110	33	600118	0.254
9	600036	0.501	34	601166	0.136
10	600048	0.331	35	601169	0.171
11	600050	0.195	36	601288	0.218
12	600089	0.213	37	601299	0.238
13	600104	0.172	38	601318	0.414
14	600111	0.071	39	601328	0.162
15	600196	0.431	40	601336	0.091
16	600256	0.152	41	601398	0.157
17	600332	0.152	42	601601	0.137
18	600383	0.318	43	601628	0.097
19	600406	0.127	44	601668	0.258
20	600518	0.190	45	601688	0.160
21	600519	0.645	46	601766	0.512
22	600547	0.634	47	601818	0.168
23	600585	0.154	48	601857	0.086
24	600637	0.330	49	601901	0.556
25	600703	0.299	50	601989	0.098
Average					**0.289**

With the same data matrix 11557x50, the result from 2 different forecasting models is very close. CRBM MAPE500 value is 0.386 while ELM is 0.289. Without doubt, ELM is better in CRBM in this regard but the mechanism is quite different. Each stock is forecasted individually in ELM and the remaining 49 stocks are used as a reference benchmark. In other words, it cannot simultaneously forecast all the values at once but use one value at a time in the process. Thus the running time is quick as less computational power is required.

Table XXXIV Summary Result of Year 2010 to 2013

Year	Data structure	MAPE500	Volatility	Kurtosis	Skewness
2010	395,517x43	1.513	1.6E+241	2.40	0.60
2011	455,923x43	6.808	5.1E+235	2.16	0.06
2012	419,558x43	8.429	2.6E+188	2.37	0.09
2013	365,396x43	6.111	1.1E+186	2.45	0.39
2012to13	784,954x43	15.158	1.3E+188	2.41	0.24
2011to13	1,240,877x43	18.106	1.7E+235	2.33	0.18
2010to13	1,636,394x43	25.607	4.0E+240	2.35	0.28
2010	11,775x43	0.372	0.22	3.78	0.70
2011	11,775x43	0.832	0.21	2.68	1.22
2012	11,775x43	0.119	0.35	2.26	-0.32
2013	11,775x43	0.451	-0.22	2.60	-0.01
2012	419,558x50	13.634	2.2E+188	2.39	0.11
2013	313,055x50	5.24	9.3E+185	2.42	0.37
2012to13	732,613x50	13.725	1.1E+188	2.41	0.24
2012	11,775x50	13.634	0.35	2.24	-0.34
2013	11,775x50	0.479	-0.16	2.49	0.00

Table XXXIV gives the summary of all the results from 2010 to 2013. The maximum data this model can handle is 1,636,394x43 dimension of data for the accumulated records of all 43 stocks from years

2010 to 2013 and it took at least 1 week to run 5 times of the algorithm.

It is necessary to point out in the CMAR data provided by the candidate, only 43 stocks were found with records from year 2010 to 2013. The MAPE500 result is disappointing as it is 19.26 which is way beyond our target value at 2. However, the encouraging news is all the 11775x43 and 11775x43 datasets MAPE500 value from year 2010 to 2013 are under 2 which satisfied our objective.

5.4 Conclusion

It is difficult to represent financial factors in our model. There are many financial factors such as interest rate, GDP, political environment, monetary policy, exchange rate, economic cycle, inflation rate, unemployment rate and many others. The interest rate has been relatively stable in recent years and in our experiment, there is almost no impact on the output. Comparatively, many other factors do not have any significant effect on short-term forecasting. Unlike stock data which reflects the performance and market sentiment in a particular market, all other factors are more general in nature and to pinpoint the effects on a particular stock is very difficult. Stock market sometimes refers to irrational market which cannot be represented by simple economic factors but only the stock price itself. In fact, the recent algorithm trading only focuses on price and neglect other factors, particularly in daily or short-term trading.The characteristics of big data can be described by 5 V which are

Volume, Velocity, Variety, Veracity and Value. The CRMB model in this section major addresses the volume aspect of the big data. Volume refers to the vast amount of data generated every second in the Shanghai Stock Exchange. Many of the models in the above experiments only use daily close value parameter for forecasting. The total number of parameters in the datasets is limited. High dimensional data means many parameters are included in the dataset. They are not from the same stock but are traded in the same market. In SSE50 dataset used in this experiment, there are totally 50 parameters. They are the close values of 50 stocks which vary in value every minute.

This algorithm trains CRBM in which Gaussian visible units are connected to hidden, binary, stochastic feature detectors using symmetrically weighted connections. Learning is done with 1-step Contrastive Divergence. Directed connections are present, from the past configurations of the visible units to the current visible units, and the past configurations of the visible units to the current hidden units. It is a 2-step deep learning procedure to return the weight factor of the visible units and hidden units . Hence, the depth of the CRBM model is 2.

The number of epochs is set to 2000 which can ensure the network is well trained. There is no need to proceed sequentially through the training data sequences. The updates are only conditional on the past n time steps, not the entire sequence. Instead, a mini-batches with a size of roughly 8% of the corresponding dataset is set as the testing data. This is a unique training method and the number of hidden units is set to 150. The running time of the dataset with dimensional data 1,636,394x43 was around 96 hours while the shortest 11,775x43 was around 1 hour using the computer as depicted in Chapter 1.4.

We have identified the potential of recurrent deep learning and leads to the development of CRBM model. The recurrent neural network is equipped with additional recurrent connections that have important capabilities not found in feedforward networks. It is regarded as an expressive model to deal with non-linear sequential processing tasks. The previous ARMIA or GARCH model, on the other hand, assumes that the market is linear by nature even the algorithm itself is autoregressive. Given the autoregressive nature of financial time series data, the recurrent deep learning algorithm is a perfect match for it.

The following figures demonstrated the learning ability of CRBM in the dataset SSE2012 11775x43 from 1 layer to 3 layers. The third layer is for demonstration purpose as the model only goes to 2 layers. It is obvious the deeper the learning mechanism, the more accurate the result and therefore it should be the future development of this research.

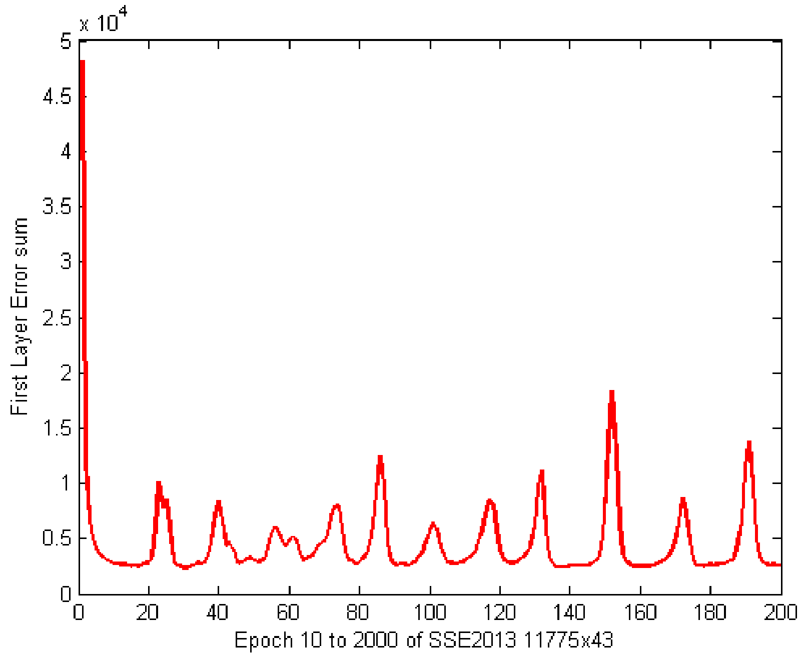

Figure 48 CRBM First Layer Learning Error

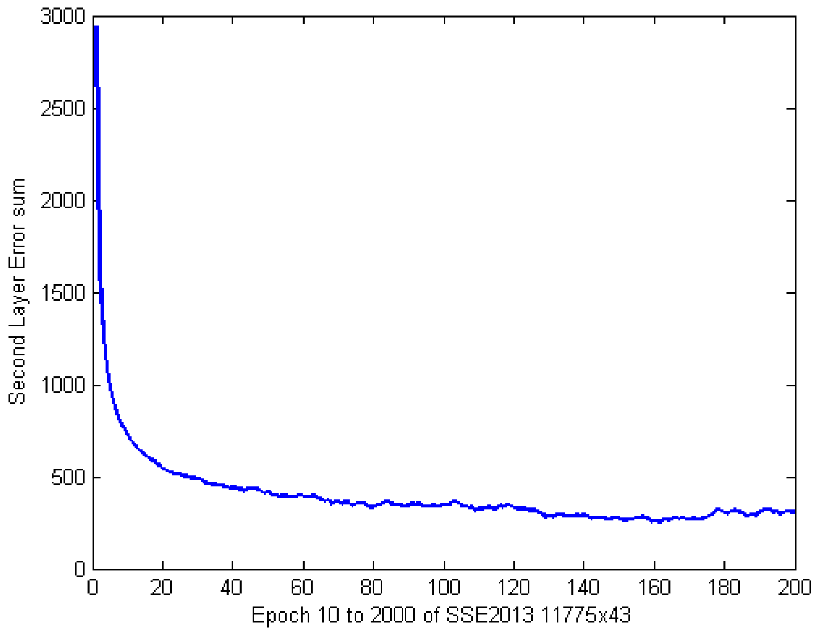

Figure 49 CRBM Second Layer Learning Error

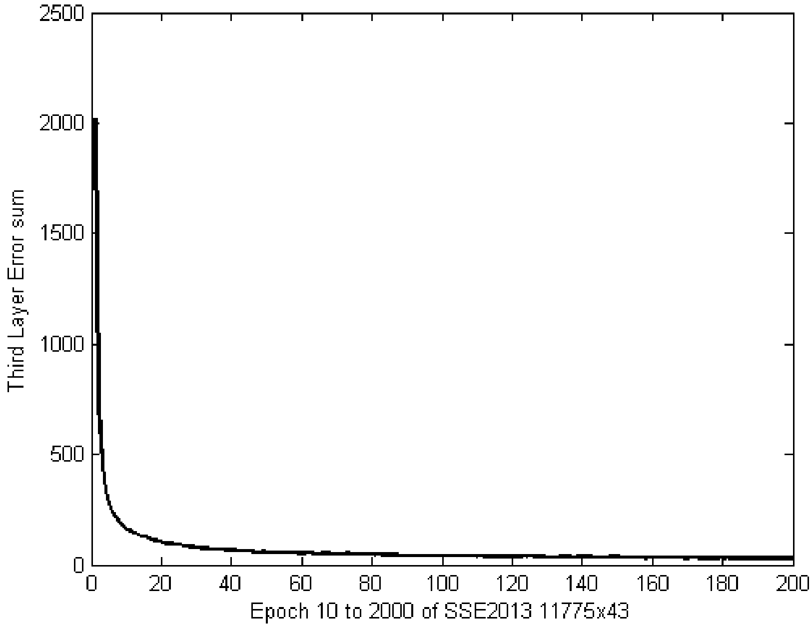

Figure 50 CRBM Third Layer Learning Error

Chapter 6 Research Conclusion and Future Development

There are many negative comments on financial time series forecasting. The most famous one is "no one is poor if it is workable". To a certain extent, it does affect our motivation in this research as it is no longer a popular subject and it would be difficult to make contribution, let alone the breakthrough, because there have been so many literature reviews and theories throughout many decades. As the Chinese saying goes, "the boat will turn straight when it reaches the end of a bridge", the big data era has inspired almost every domain in computational intelligence and it also inspired the candidate to carry on this thesis. The theoretical background on forecasting is based on statistical theory but many of them are not easy to make a breakthrough. Many of them have a very long history such as Bayesian theory which was proposed over 250 years ago and there are not many related literature reviews on forecasting. Many papers focus on the financial part but not the forecasting part since modern financial theory may include so many topics ranging from economic, politics to international business. Forecasting theory, on the other hand, is relatively covering a narrow subject with more or less statistical and artificial intelligence elements to support the theoretical background. To combine these 2 topics to form the research topic financial time series forecasting is very difficult because it is necessary to balance between the financial and forecasting theory. Many models presented in this study only investigated the forecasting ability but neglected the financial factors. This is because financial factors such as interest rate, political stability, currency rate and many others are not easily represented in a forecasting model. In fact, we have put these factors into the models but this does not affect the result nor have any influence on the models. The famous Black-Sholes model as described in Chapter 2 Section 2.6 is a good example that only the current stock price is relevant and not the others. We have not identified a model that can incorporate financial factors such as interest rate, currency rate, unemployment rate or inflation rate to affect the result of the forecasting model. Our approach, on the other hand, is to look at different markets with different financial characteristics such as EMH theory. Although the same model is applied, the result reveals the characteristics of the financial market.

Despite the negative sentiment, the importance of financial time series forecasting is very obvious in the financial institution. It is well known that many big financial institutions have many supercomputers to monitor and trade in the financial market. They have outperformed many regular investors due to their unique in-house algorithms developed by their own specialists using latest forecasting methodologies. There are so many forecasting techniques on the website of financial service providers but it is still very primitive and it does not consider the use of latest artificial intelligence models. Many financial researchers have used a very simple mean reverting technique to do financial forecasting and it serves the purpose most of the time. In Chapter 3, we have demonstrated that a simple 2-day moving average model already ranked the third among other models. This may work for a particular stock with very low volatility. However, more researches on different methods are required. In our research, we have explored many techniques that combine traditional statistical theory and the latest artificial intelligence methodology to get better results.

Our contribution in this paper is the discovery of CRBM algorithm to forecast all the constituent stocks in a market index at the same time with very good accuracy. When computational power was limited in early ages, dimension reduction is very important as it can dramatically reduce the workload and speed. However, the trade-off is the relevant information that could be lost during the reduction process. We have shown that CRBM algorithm can handle high dimensional data file which could track the relationship of all the dimensions in the dataset. Our application of the model in Index market has pointed out all the relevant stocks movement that are closely related and it cannot be proved easily without this model. The application of CRBM in the high frequent data in the above experiment has distinguished advantage in online trading as the model can provide more information about the stock movement within a very short period of time. This is similar to predict the next frame of human motion based on the previous frame but the motion is running at a very slow speed. In financial time series forecasting, this edge is the survival mode for a large volume programming transaction which only relies on the result of the data.

It is true that many financial data must be pre-processed due to different data format and storage chronological factors. On top of that, there are too many financial institutions which provide financial information but without a unified standard. Thus, the future work of this CRBM algorithm must cater to all kinds of data format from the financial ocean of data. Text format is a big challenge for this model. As a conclusion, there is no holy grail in forecasting model but only the most useful one according to our research. We have found CRMB algorithm a novel approach in financial time series forecasting with the high potential application. Unlike many algorithms such as SVR, ARIMA, wavelet-based or neural network, this algorithm does not need to tune parameters so that the model can obtain good forecasting result. Hence, it is very convenient and quick to use.

The possible future development of this algorithm could be in mobile computing. This prevailing technology enables everyone to get access to mobile data anywhere in the world. Mobile banking is a service that almost every banker must provide for the clients. It is based on mobile data management technology which has to solve the data dissemination, caching, replication in mobile environments and most importantly the location dependent data services problem. The mobile service providers have become the largest big data collectors as the number of user of mobile devices are more than computers. Baidu has launched a heat map of where Chinese travellers are heading to, coming from, and which routes are most popular during the Lunar New Year, the country's largest national holiday.

The heat map, which went online January 27 2014, was created by data Baidu gathered from smartphones through the use of the location-based Baidu Maps mobile app, which has more than 200 million registered users and receives 3.5 million position requests every day. The heat map updates every four to eight hours, showing the most popular destinations, points of origin, and travel routes. The above CRBM algorithm can be modified and installed as one of the mobile apps to forecast the next destination for travellers so that the users can avoid the peak traffic.

Appendixes

Charts Identification Guidelines

Chart 2- pattern recognition. A horizontal, or nearly so, trend line that connects the minor lows. Must have at least two distinct minor lows before drawing a trend line. An up-sloping trend line bounds the expanding price series on top. Must have at least two minor highs to create trend line.

Chart 3 – pattern recognition. A horizontal line of resistance joins the tops as a trend line. Must have at least two distinct touches (minor) highs before drawing a trend line. The expanding price series is bounded on the bottom by a down-sloping trend line. Must have at least two distinct minor lows to create a trend line.

Chart 24 – pattern recognition. A three-trough formation with the centre trough below the other two. It looks like a head-and-shoulders bust flipped upside down. The three troughs and two minor rises should appear well defined. The left and right shoulders should be opposite one another about the head, somewhat equidistant in both time and price. There are wide variations, but the formation is noticeably symmetrical about the head.

Chart 26 – pattern recognition. After an upward price trend, the formation appears as three bumps, the centre one is the tallest, resembling a bust. The two shoulders appear at about the same price level. Distance from the shoulders to the head is approximately the same. There can be wide variation in the formation's appearance, but symmetry is usually a good clue to the veracity of the formation.

Chart 47 – pattern recognition. Two price trend lines, the top one horizontal and the bottom one sloping up, form a triangle pattern. The two lines join at the triangle apex. Prices rise up to and fall away from a horizontal resistance line at least twice (two minor highs). Prices need not touch the trend line but should come reasonably close (say, within $0.15). The line need not be completely horizontal but usually is. Prices decline to and rise away from an up-sloping trend line. Prices need not touch the trend line but should come close (within $0.15). At least two trend line touches (minor lows) are required.

Chart 48 – pattern recognition. A triangular-shaped pattern bounded by two trend lines, the bottom one horizontal and the top one sloping down, that intersect at the triangle apex. A horizontal (or nearly so) base acts to support prices. Prices should touch the base at least twice (at least two minor lows that either touch or come close to the trend line).

Chart 49 – pattern recognition. Compute the formation height by subtracting the lowest low from the highest high. For upward breakouts, add the difference to the highest high or for downward breakouts, subtract the difference. Alternatively, symmetrical triangles can be halfway points in a move, so project accordingly. As consolidation, prices usually leave the triangle in the same direction as when they enter.

Reference

Addison, S. (2002). *The illustrated wavelet transform handbook.* Bristol BSI 6BE, United Kingdom: Institue of Physics Publishing Limited.

Aronson, D. (2007). *Evidence-based technical analysis.* New Jersey: John Wiley & Sons, Inc. doi:ISBN 0-470-00874-1

Asteriou, D., & Hall, S. G. (2011). *ARIMA models and the Box–Jenkins methodology.* (P. MacMillan, Ed.) Applied Econometrics.

Atsawaraungsuk, S., Horata, P., Sunat, K., Chiewchanwattana, S., & Musigawan, P. (2013). Evolutionary circular extreme learning machine. *Computer Science and Engineering Conference (ICSEC), 2013 International* (pp. 292-297). IEEE Conference Publications. doi:10.1109/ICSEC.2013.6694796

Bao, Y., Liu, Z., Guo, L., & Wang, W. (2005). Forecasting stock composite index by fuzzy support vector machines regression. *Fourth International On Machine Learning and Cybernetics,* (pp. 3535-3540). Guangzhou.

Bao, Y., Lu, Y., & Zhang, J. (2004). Forecasting stock price by SVMs regression. *Springer-Verlag*, 295-303.

Bertoin, J. (1996). *Levy processes.* Cambridge University Press.

Bjorn. (2003, May). *Questioning the inefficient market hypotheses.* Yale University. Yale: Faculty of the Graduate School of Yale University.

Bjorn, J., & Wim, S. (2010). An overview of wavelet based multiresolution analyses. South Carolina. Retrieved from cm.bel-labs.com/who/wim/paper/overview.pdf

Bloomfield, P. (1999). *Fourier analysis of time series.* John Wiley & Sons, Inc.

Bodie, Z., Kane, A., & Marcus, A. J. (2005). *Investments* (Sixth Edition ed.). Mcgraw-Hill International Editiion.

Boggess, A., & Narcowich, J. F. (2009). *A first course in wavelets with fourier analysis.* New Jersey: John Wiley & Sons, Inc.

Bollerslev, T. (2007). *Glossary to ARCH (GARCH).* Retrieved from http://public.econ.duke.edu/~boller/Papers/glossary_arch.pdf

Bosman, H., Iacca, G., & Wortche, H. (2013). Online extreme learning on fixed-point sensor networks. *Data Mining Workshops (ICDMW), 2013 IEEE 13th International Conference on* (pp. 319-326). IEEE Conference Publications. doi:10.1109/ICDMW.2013.74

Box, G., & Jenkins, G. (1970). *Time series analysis: forecasting and control.* San Francisco: Holden-Day.

Bulkowski, N. T. (2005). *Encyclopedia of chart patterns.* John Wiley & Sons, Inc.

Burges, C. J. (2005). *A tutorial on support vector machines for pattern recognition.* Retrieved from http://research.microsoft.com/pubs/67119/svmtutorial.pdf

Cai, T. T., & Shen, X. T. (2011). *High dimensional data analysis.* World Scientific Publishing Co Pte Ltd.

Cai, X., & Lin, X. (2012). Forecasting high dimensional volatility using conditional restricted Boltzmann machine on GPU. *Parallel and Distributed Processing Symposium Workshops and PhD Forum (IPDPSW), 2012 IEEE 26th International* (pp. 1979-1986). IEEE Conference Publications. doi:10.1109/IPDPSW.2012.258

Cao, B., Zhan, D., & Wu, X. (2009). Application of SVM in financial research. (pp. 507-511). IEEE. doi:10.1109/CSO.2009.313

Cao, J., Lin, Z. P., & Huang, G. B. (2012). Self-Adaptive evolutionary extreme learning machine. *36*, pp. 285-305. Neural Process Lett. doi:10.1107/s11063-012-9236-y

Cao, L. J., Tay, & Francis, E. H. (2010). Support vector machine with adaptive parameters in financial time series forecasting. *Transactions on Neural Networks. 14*, pp. 1506-1518. IEEE.

Cao, L., Huang, W., & Sun, F. (2014). Optimization-based extreme learning machine with multi-kernel learning approach for classification. *Pattern Recognition (ICPR), 2014 22nd International Conference on* (pp. 3564-3569). IEEE Conference Publications. doi:10.1109/ICPR.2014.613

Chance, D. M. (2011). Derivation and interpretation of the Black-Scholes-Merton model. *Teaching Note 99-02*. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=6C35B6F03963D035F98D1BEA4186AE5A?doi=10.1.1.363.2491&rep=rep1&type=pdf

Chen, C., Hardle, W. K., & Jeong, K. (2010). Forecasting volatility with SVM-based GARCH model. *Journal of Forecasting, 29*, pp. 406-433. doi:10.1002/for.1134

Cherubini, V., Luciano, E., & Vecchiato, V. (2004). *Copula methods in finance*. John Wiley & Sons Ltd.

Clements, P. M., & Hendry, F. D. (1998). *Forecasting economic time series.* Cambridge University Press.

Coates, A., Lee, H., & Ng, A. (2011). An analysis of single-layer networks in unsupervised feature learning. *International Conference on Artificial Intelligence and Statistics (AISTATS).* (pp. 215-223). AISTATS.

Crowley, P. (2007). A guide to wavelets for economists. *Journal of Economic Surreys, 21*, pp. 207-264.

CSMAR. (2013). *China security market trade and quote research database*. Retrieved from http://www.gtadata.com.cn

Dai, W., & Lu, C. J. (2008). Financial time series forecasting using a compound model based on wavelet frame and support vector regression. *4th International Conference on Natural Computation*, (pp. 329-332).

Eickholt, J. (2006). Energy based models, restricted Boltzmann machines and deep networks. Retrieved from https://www.researchgate.net/publication/265151259_Energy_Based_Models_Restricted_Boltzmann_Machines_and_Deep_Networks

Engle, R. F. (2001). GARCH 101: The Use of ARCH/GARCH models in applied econometric. *Journal of Economic Perspectives*, *15*(4), 157-168.

Famma, E. (1991). Efficient capital markets II. *Journal of Finance, 46*(5), 1575-1616.

Fernandex, Irma, B., Zanakis, Stelios, H., & Walczak, S. (2002). Knowledge discovery techniques for predicting country investment risk. *Computers and Industrial Engineering*, *43*, pp. 787-800.

Gavrishchaka, V. V., & Banerjee, S. (2006). Support vector machine as an efficient framework for stock market volatility forecasting., (pp. 147-160). doi:10.1007/s.027-005-005-5

Gen, L., & Ma, J. (2008). Least squares support vector regression based CARRX model for stock index volatility forecasting. *2008 International Seminar on Future Information Technology and Management Engineering* (pp. 406-409). IEEE. doi:10.1109/FITME.2008.69

Gestel, T. V., Suykens, J., Baestanes, D., Lambrechts, A., Lanckriet, G. B., Vandaele, B. D., & Vandewalle, J. (2001). Financial time series prediction using least squares support vector machines within the evidence framework. *IEEE Transactions on Neural Networks* (pp. 809-821). IEEE.

Harvey, A. C. (1993). *Time series model.* Harvester Wheatsheaf.

He, Y. L., Liu, J. N., Wang, X. Z., & Hu, Y. X. (2012). Optimal bandwidth selection for re-substitution entropy estimation. *291*(8), pp. 3425-3460.

Hentschel, L. (1997). Nesting symmetric and asymmetric GARCH models. *Journal of Financial Economics, 39*, pp. 71-104.

Hinton, G. H., Osindero, S., & Teh, Y. W. (2005). A fast learning algorithm for deep belief nets. *Neural Computation. 18*, pp. 1527-1554. Massachusetts Institute of Technology.

Hinton, G., & Salakhutdinov, R. (2006). Reducing the dimensionality of data with neural networks. *Science, 313*, pp. 504-507. doi:10.1126/science.1127647.

Hoshmand, A. R. (2010). *Business forecasting a practical approach.* New York: Routledge Taylor & Francis Group. doi:10:0-415-98855-1

Hu, Y. X., & Liu, J. N. (2013). Application of feature weighted support vector regression using grey correlation degree to stock price forecasting. *Neural Computing and Applications, 22*(1), pp. 143-152.

Huang, C., Huang, L. L., & Hang, T. T. (2012). Financial time series forecasting based on wavelet kernel support vector machine. *Natural Computation (ICNC), 2012 Eighth International Conference on* (pp. 79-83). IEEE Conference Publications. doi:10.1109/ICNC.2012.6234569

Huang, G. B. (2015). What are extreme learning machines? Filling the gap between Frank Rosenblatt's Dream and John von Neumann's puzzle. *Cognitive Computation, 7*(3), pp. 263-278. doi:10.1007/S12559-015-9333-0

Huang, G. B., & Chen, L. (2007). Convex incremental extreme learning machine. *Neurocomputing, 70*, pp. 3056-3062.

Huang, G. B., & Chen, L. (2008). Enhanced random search based Incremental extreme learning machine. *Neurocomputing, 71*(16-18), pp. 3460-3468.

Huang, G. B., Chen, L., & Siew, C. K. (2006). Universal approximation using incremental constructive feedforward networks with random hidden nodes. *17 no.4*, pp. 879-892. IEEE Transactions on Neural Networks.

Huang, G. B., Saratchandran, P., & Sundararajan, N. (2005). Fully complex extreme learning machine. *Neurocomputing, 68*, pp. 306-314.

Huang, G. B., Zhu, Q. Y., & Siew, C. K. (2006). Extreme learning machine: theory and applications. *Neurocomputing, 70* (1), pp. 489-501.

Huang, G. B., Zhu, Q. Y., Mao, K. Z., Siew, C. K., Saratchandran, P., & Sundararajan, N. (2006). Can threshold networks be trained directly? *Circuits and Systems II: Express Briefs, IEEE Transactions on, 53*(3), 187-191. doi:10.1109/TCSII.2005.857540

Huang, G., Chen, L., & Siew, C. K. (2006). Universal approximation using incremental constructive feedforward networks with random hidden nodes. *IEEE Transactions on Neural Networks* (pp. 879-892). IEEE.

Huang, S. C., & Wu, T. K. (2008). Wavelet-based relevance vector machines for stock index forecasting. *2006 International Joint Conference on Neural Networks*, (pp. 603-609). Vancouver.

Huang, W., Nakamori, Y., & Wang, S. Y. (2004). Forecasting stock market movement direction with support vector machine. *Computers and Operations Research, 32*, pp. 2513-2522. doi:10.1016/j.for.2004.03.016

Hui, C. L., & Ng, S. F. (2006). A new approach for prediction of sewing performance of fabrics in apparel manufacturing using artificial neural networks. *The Journal of The Textile Institute, 96*(6), pp. 401-405. doi:10.1533/joti.2005.0101

Hyun, J. S., Dong, H. E., & Sung, S. K. (2005). One class support vector machines - An application in machine fault detection and classification. *Computers and Industrial Engineering*, (pp. 395-408).

Ince, H., & Trafalis, T. B. (2004). Kernel principal component analysis and support vector machines for stock price prediction. (pp. 2053-2058). IEEE.

Ingo, S., & Christmann, A. (2008). *Support vector machine.* Springer.

Janakiraman, V., Nguyen, X., & Assanis, D. (2013). A Lyapunov based stable online learning algorithm for nonlinear dynamical systems using extreme learning machines. *Neural Networks (IJCNN),* *The 2013 International Joint Conference* (pp. 1-8). IEEE Conference Publications. doi:10.1109/IJCNN.2013.7090813

Kim, K. J. (2004). Toward global optimization of case-based reasoning systems for financial forecasting. *Applied Intelligence, 21*, pp. 239-249.

Kong, K. M., Wong, H. Y., Lee, S., & Liu, J. (2009). Financial trend forecasting with fuzz chaotic oscillatroy-based neural networks (CONN). *FUZZY-IEEE 2009* (pp. 1947-1952). IEEE Conference Publications. doi:10.1109/FUZZY.2009.5277326

Lai, L. K., & Liu, J. (2013). Chart pattern and least square support vector machine in stock forecasting. *International Journal of Advanced Engineering Applications*, pp. 45-55.

Lai, L. K., & Liu, J. N. (2008). A neural network and CBR-based model for allocating work in progress. *5th International Conference on Information Technology and Applications*, (pp. 533-538).

Lai, L. K., & Liu, J. N. (2009). A neural network and CBR-based model sewing minute value. *IEEE 2009 Internatioal Joint Conference on Neural Networks* (pp. 1696-1701). Atlanta, Georgia: IEEE. doi:10.1109/IJCNN.2009.5178803

Lai, L. K., & Liu, J. N. (2010). Stock forecasting using support vector machine. *Journal of Machine Learning and Cybernetics*, (pp. 1607-1614). Qingdao. doi:10.1109/ICMLC.2010.558099

Lai, L. K., & Liu, J. N. (2012). WIPA:Neural network and case based reasoning models for allocating work in progress. *Journal of Intelligent Manufacturing*, 409-421. doi:10.1007/s 10845-010-0379-2, 23,3

Lai, L. K., & Liu, J. N. (2013). Support vector machine and least square support vector machine stock forecasting models. *Horizon Research Publishing Corporation*, pp. 1-18.

Lai, L. K., & Liu, J. N. (2013). Support vector regression with Levy distribution kernel for stock forecasting. *International Conference on Advanced Computer Science and Electronics Information*, (pp. 654-657). Beijing. doi:102.8331/icacsei.2013.16

Lai, L. K., Hu, R., & Liu, J. N. (2012). A weighted support vector data description based on rough neighborhood approximation. *2012 IEEE 12th International Conference on Data Mining Workshops*, (pp. 635-642).

Larochelle, H., & Bengio, Y. (2008). Classification using discriminative restricted Boltzmann machines. *Proceedings of The 25th International Conference on Machine Learning* (p. 536). ICML. doi:10.1145/1390156.1390224. ISBN 9781605582054. edit

LeCun, Y., Chopra, S., Hadsell, R., Ranzato, M., & Huang, F. (2006). A tutorial on energy-based learning. *The Courant Institute of Mathematical Sciences, New York University*. New York: MIT Press, 2006. Retrieved from http://web.stanford.edu/class/cs379c/archive/2012/suggested_reading_list/documents/LeCunetal06.pdf

Lee, C., Liaw, Y., & Hsu, L. (2011). Investment decision making by using fuzzy candlestick pattern and genetic algorithm. *IEEE Conference Publications* (pp. 2696-2711). IEEE Conference Publications. doi:10.1109/FUZZY.2011.6007707

Lee, R., & Liu, J. N. (2001). iJade stock predictor - An intelligent multi-agent based time series stock prediction system. *2nd Asia-Pacific Conference on IAT* (pp. 495-499). World Scientific. doi:10.114219789812811042_0067

Lee, R., & You, J. (2001). iJade weatherman - A multiagent fuzzy-neuro network based weather prediction system. *Systems, Man, and Cybernetics, Part C: Applications and Review, 34*(3), pp. 369-377.

Lee, S. H., Kim, H., Jang, H., & Lim, J. S. (2008). Forecasting short-term KOSPI time series based on NEWFM. *Advanced Language Processing and Web Information Technology, 2008. ALPIT '08. International Conference on* (pp. 303-307). IEEE Publication. doi:10.1109/ALPIT.2008.2

Leigh, W., Frohlich, C., Hornik, S., Purvis, P., & Roberts, T. (2008). Trading with a stock chart heuristic. *Systems, Man and Cybernetics, Part A: Systems and Humans. 38 Issue 1,* pp. 93-104. IEEE Conference and Journal Publisher. doi:10.1109/TSMCA.2007.909508

Leroy, S. (1973). Risk aversion and the Martingale property of stock prices. *International Economic Review. 14 (2)*, pp. 436-446. International Economic Review.

Leuven, K. U. (2011). *Katholieke Universiteit Leuven*. Retrieved from http://www.esat.kuleuven.be/sista/lssvmlab

Li, H. Q. (2010). Forecasting financial volatility using intraday information. *Advanced Management Science (ICAMS), 2010 IEEE International Conference on* (pp. 518-521). IEEE. doi:10.1109/ICAMS.2010.5553185

Li, Y. F. (2008). *Research on stock value investment based on artificial intelligence.* Harbin: Harbin Institute of Technology.

Liang, N. Y., Huang, G. B., Saratchandran, P., & Sundararajan, N. (2006). A fast and accurate online sequential learning algorithm for feedforward networks. *IEEE Transactions on Neural Networks, 17 No.6*, pp. 1411-1423.

Lin, C. J., & Chang, C. C. (2001). *LIBSVM*. Retrieved from http://www.csie.ntu.edu.tw/cjlin/libsvm

Lin, C. J., Chen, B. J., & Chang, M. W. (2001). Load forecasting using support vector machines: A study on ENUNITE competition 2001. *ENUNITE Competition 2001* (pp. 1-7). Department of Computer Science and Information Engineering, National Taiwan University.

Liu, L. (2012). Essays on financial market interdependence. *167*. Lund: The Department of Economics Lund University. Retrieved from http://lup.lub.lu.se/record/2437256

Liu, Y. X. (2009). Discussing on trend to efficient market hypothesis of securities and futures market. *Information Management,* *Innovation Management and Industrial Engineering, 2009 International Conference on. 3,* pp. 149-152. IEEE Conference Publications. doi:10.1109/ICIII.2009.346

Lucas, R. E. (1978). Asset prices in an exchange economy. *Econometrica, 46*(6), pp. 1429-1445.

Lv, J., & Zhang, Y. (2011). Application of least squares support vector machine in futures price forecasting. (pp. 169-172). IEEE Conference Publications. doi:10.1109/ICECTECH.2011.5941825

Mager, J., Paasche, U., & Sick, B. (2008). Forecasting financial time series with support vector machines based on dynamic kernels. *IEEE Conference on Soft Computing in Industrial Application* (pp. 252-257). IEEE.

Majumdar, A., Majumdar, P. K., & Sarkar, B. (2006). An investigation on yarn engineering using artificial neural networks. *The Journal of The Textile Institute, 97*(5), pp. 429-434. doi:10.1533/jot.2006.0266

Malkiel, B. G. (2003). The efficient market hypothesis and its critics. *Economic Perspective, 17*(1). Retrieved from http://www.jstor.org/stable/3216840

Malyscheff, A. M., Trafalis, T. B., & Raman, S. (2002). From support vector machine learning to the determination of the minimum enclosing zone. *Computers and Industrial Engineering, 42*(1), pp. 59-74.

Minh, V., Larochelle, H., & Hinton, G. E. (2010). Conditional restricted Boltzmann machines for structured output prediction. *CIFAR 2010*. University of Toronto. Retrieved from https://www.cs.toronto.edu/~vmnih/docs/uai_crbms.pdf

Mitsdorffler, R., & Diederich, J. (2008). Prediction of first-day returns of initial public offering in the US stock market using rule extration from support vector machines. *Studies in Computational Intelligence*, (pp. 185-203).

Montgomery, D. C., Johnson, L. A., & Gardiner, J. S. (1990). *Forecasting and time series analysis.* McGraw Hill.

Moreira, J. M., Jorge, A. M., Soares, C., & Sousa, J. F. (2006). Improving SVM-linear pedictions using CART for eample slection. *Foundations of Intelligent Systems, 4203*, pp. 632-641.

Nkayam, H., Ata, S., & Oka, I. (2013). Predicting time series of individual trends with resolution adaptive ARIMA. *Measurements and Networking Proceedings (M&N), 2013 IEEE International Workshop on* (pp. 143-148). IEEE Conference Publications. doi:10.1109/IWMN.2013.6663793

NN3, A. N. (2007). *Artificial Neural Netowrk and computational Intelligence Forecasting Competition*. Retrieved from http://www.neural-forecasting-competition.com/NN3/

Olson, D., & Mossman, C. (2001). Cross-correlations and predictability of stock returns. *Journal of Forecasting, 20,* 145-160.

Pasila, F., Ronni, S., & Wijaya, L. H. (2008). Long-term forecasting in financial stock market using accelerated LMA on neuro-fuzzy structure and additional fuzzy C-means clustering for optimizing the GMFs. *International Joint Conference on Neural Networks*, (pp. 3960-3965).

Posedel, P. (2006). Analysis of the exchange rate and pricing foreign currency options on the Coration market : The NGARCH models as an alternative to the Black-Scholes model. *Financial Theory and Practice*, *30 (4),* pp. 347-368.

Raible, S. (2001). *Levy processes in finance: Theory, numerics, and empirical facts.* Institut fur Mathematische Stochastick Albert-Ludwigs-Universitat Freiburg.

Research, C. I. (2012, July 9). *Graduate Summer School:Deep Learning, Feature Learning July 9-27, 2012*. Retrieved from http://www.ipam.ucla.edu/programs/summer-schools/graduate-summer-school-deep-learning-feature-learning/

Rua, A. (2010). A wavelet approach for factor-augmented forecasting. *Journal of Forecasting, 30*(7), pp. 666-678. doi:10.1002/for.1200

Salakhutdinov, R., & Hinton, G. (2010). *A better way to pretrain deep Boltzmann machines.* University of Toronto.

Salakhutdinov, R., & Hinton, G. (2011). *Replicated softmax: an undirected topic model.* Neural Information Processing Systems 23.

Salakhutdinov, R., Hinton, G., & Minh, A. (2007). Restricted Boltzmann machines for collaborative filtering. *Proceedings of the 24th International Conference on Machine Learning - ICML '07.* ICML. doi:10.1145/1273496.1273596.

Sapankevych, N. I., & Sankar, R. (2009). Time series prediction using support vector machines : A survey. *Computational Intelligence Magazine*, (pp. 25-38).

Shen, W., Zhang, Y., & Ma, X. (2009). Stock return forecast with LS-SVM and particle swarm optimization. *2009 International Conference on Business Intelligence and Financial Engineering* (pp. 143-147). IEEE. doi:10.1109/BIFE.2009.42

Sivakumar, P., & Mohandas, V. P. (2009). Modeling and predicting stock returns using the ARFIMA-FIGARCH a case study on Indian stock data. *2009 World Congress on Nature and Biologically Inspired Computing*, (pp. 896-901).

Sui, X. S., Qi, Z. Y., Yu, D. R., Hu, Q. H., & Zhao, H. (2007). A novel feature selection approach using classification complexity for SVM of stock market trend prediction. *2007 International Conference on Management Science and Engineering,* (pp. 1654-1659). Harbin.

Suyken, Johan, A. K., Gestel, T., Brabanter, V., Jos, D., Moor, . . . Vandewalle, J. (2002). *Least Squares Support Vector Machines.* World Scientific Publish Co. Pte. Ltd.

Taylor, G. W. (2006). Retrieved from http:;//www.cs.toronto.edu/~gwtaylor/publications/nips2006nhmubly

Taylor, G., Hinton, G. E., & Roweis, S. (2006). *Modeling human motion using binary latent variables.* Toronto: University of Toronto.

Taylor, G., Sigal, L., Fleet, D., & Hinton, E. (2010). *Dynamical binary latent variable models for 3D human pose tracking.* Toronto: CIFAR.

Technologies, M. (2004). Retrieved from www.mathwave.com/easyfit-distribution-fitting.html

Teh, Y., & Hinton, G. (2001). Rate-coded restricted Boltzmann machines for face recogniton. *In Advances in Neural Information Processing Systems, 3*, pp. 908-914.

Thomason, M. (1999). The practitioner methods and tools. *Journal of Computer Intelligent in Finance, 7 No.3*, 36-45.

Thomason, M. (1999). The practitioner methods and tools. *Journal Computer Intelligent, 7 No.3,* 36-45.

Thomson. (2012). *Thomson Reuters*. Retrieved 2012, from http://www.thomsonreuters.com/en/prdoucts-services/financial/investment-management/datastream-professional.html

Tian, W. J., & Wang, M. Y. (2009). Modeling and forecasting method based on support vector regression. *2009 second Intertnational Conference on Future Information Technology and Management Engineering*, (pp. 183-186).

Timmermann, A., & Granger, C. (2004). Efficient market hypothesis and forecasting. *International Journal of Forecasting, 20*, 15-27.

Tu, C., Clyde, M., & Wolpert, R. L. (2007). *Levy adaptive regression kernels.* Duke University Department of Statistical Science.

Vapnik, V., Golowich, S. E., & Smola, A. (1997). Support Vector method for function approximation, regression estimation, and signal processing. *Advances in Neural information Processing Systems*, pp. 281-287.

Wang, D. L., & Li, Y. J. (2009). A novel nonlinear RBF neural Nnetwork ensemble model for financial time series forecasting. *Advanced Computational Intelligence (IWACI), 2010 Third International Workshop on* (pp. 86-90). Suzhou: IEEE.

Wang, S. W., Zhao, Y., & Zhgn, X. F. (2011). Application of support vector machine and least squares vector machine to freight volume forecast. (pp. 104-107). IEEE Conference Publications. doi:10.1109/RSETE.2011.5964227

Wang, W., Zhao, H., Li, Q., & Liu, Z. (2009). A novel hybrid intelligent model for financial time series forecasting and its application. *2009 International Conference on Business Intelligence and Financial Engineering*, (pp. 279-282).

Wong, A. (2002). *Prediction of clothing sensory comfort using Neural Networks and fuzzy logic.* PhD thesis, Hong Kong Polytechnic University, Engeering, Hong Kong.

Wong, A. (2010). *Winner of bull and bears market.* China People's University Publisher. doi:ISBN 978-7-300-12491-9

Worasucheep, C. (2008). Trading index mutual funds with evolutionary forecasting. *Congress on Evolutionary Computation* (pp. 430-435). IEEE.

Wunder, G., & Mayo, H. (1995). Study supports efficient maket hypothesis. *Journal of Financial Planning, 8*(3), p. 128.8.

Yang, H., King, I., & Chan, L. (2002). Non-Fixed and asymmetrical margin approach to stock market prediction using support vector regression. *9th International Conference on Neural Information,* (pp. 1398-1402). Fukushima.

Yang, Y., Su, M. C., Zhou, Y., & He, C. C. (2011). Support vector machine based forecasting of the contract prices of stock index futures. *Business Intelligence and Financial Engineering (BIFE), 2011 Fourth International Conference on* (pp. 49-53). IEEE Conference Publications. doi:10.1109/BIFE.2011.113

Zhai, Y. A., Hsu, S. K., & Halgamuge. (2007). Combining news and technical indicators in daily stock price trends prediction. *4th International Symposium on Neural Networks* (pp. 1087-1096). Berlin: Springer. doi:10.1007/978-3-540-72395-0_132

Zhang, Y., & Shen, W. (2009). Stock yield forecast based on LS-SVM in bayesian inference. *2009 ETP International Conference on Future Computer and Communication* (pp. 8-11). IEEE. doi:10/10-0FCC.2009.34

Zhang, Z. (2001). Is China a weak-form efficient market hypothesis market. *Chinese Financial Research, 3*, pp. 34-40.

Zhe, L., Zhao, Y. D., & Sanderson, P. (2005). The efficient market hypothesis and electricity market efficiency test. *2005 IEEE/PES Transmission and Distribution Conference andExhibition:Asia and Pacific Dalian, China* (pp. 1-7). Dalian: IEEE publisher.

Zhou, J. G., & Tian, J. M. (2007). Predicting corporate financial distress based on rough sets and wavelet support vector machine. *2007 International Conference on Wavelet Analysis and Pattern Recognition*, (pp. 602-607). Bejing.

Zhu, Y., Zhang, Y., & Pan, Y. (2013). Large scale restricted boltzmann machines on single GPU. *Big Data, 2013 IEEE International Conference on* (pp. 169-174). IEEE.