國立政治大學理學院應用物理研究所
碩士論文

Graduate Institute of Applied Physics, College of Science

National ChengChi University

Master Thesis

基於 EEMD 與類神經網路建構台指期貨交易策略

A study of Trading Strategies of TAIEX Futures by

using EEMD-based Neural Network Learning Paradigms

陳原孝
Chen Yuan Hsiao

指導教授：蕭又新博士

Advisor: Shiau You Hsien, Ph.D.

中華民國 102 年 1 月
January, 2013

摘　要

　　金融市場瞬息萬變,股價漲跌似乎沒有顯著的規則,這意味著股價的行為特徵是不可精確預知和不確定的,為了在市場上增加收益和減少投資風險,研究人員不得不試圖建立一個有效預測金融市場的模型,它可以估算這種不確定性的影響,很可惜的,至今仍然沒有一個模型接近成功的。沒有成功的模型並不代表它是不存在的,相反的,研究人員需要建立更多的預測模型,以提供市場判斷的經驗法則。

　　我們使用 ARMA 與兩種不同形式的 EEMD-ANN 去對台灣加權指數期貨做預測值的精確度比較,我們比較了兩種不同的行情:趨勢與震盪。此外,預測出未來市場之價格後,我們使用 2 種交易策略去做績效測試,本研究希望能夠找到較適合使用在指數預測的預測模型。

　　另外在本文中,我們也分析影響 TAIEX 價格波動的因素,透過 EEMD,我們可以將其拆解成數具有不同物理意義的本徵模態函數(IMF),再藉由統計值選出較重要的 IMF 並分析其意義。

關鍵字:總體經驗模態、類神經網路、自回歸移動平均模型、交易策略、預測模型

Abstract

Financial market changes constantly and Stock Price Volatility (SPV) seems to be no significant rules. This means behavioral characteristic of the stock price cannot foresee and uncertain accurately. In order to increase revenue and reduce investment risk in the market, researchers had to try to establish an effective prediction model of financial markets. It can estimate the impact of this uncertainty. It's a great pity that there is not a model that is close successfully yet. That does not represent it does not exist successful model. Instead, researchers need to establish more predictive models to offer the market to judge the rule of thumb.

The forecasting results of TAIEX Index futures by ARMA Model and two types of EEMD-ANN Models were compared in two kinds of markets – trend and fluctuation. In addition, two trading strategies were tested after the future prices are forecasted. The study attempted to identify a suitable forecasting model.

Moreover, the factors for price fluctuation of TAIEX were also analyzed in the study. Through EEMD, they could be decomposed to IMFs with various physical meanings and more important IMFs were selected to be analyzed in accordance with the statistic value.

List of Contents

List of Tables

List of Figures

<u>Chapter 1 Introduction</u>

Today's world is undergoing tremendous changes. Digital virtual world made by Web have dominated human civilization in decades. Explosive increase therefore occurs in the degree of social complexity and the rate of trend changes. Some researchers consider that no one is able to predict great events or trends after the next presidential election or reformulations in Federal Reserve System in this kind of gradual complex and continual varied world. However, some researchers think that we are more capable to predict because the advanced technology make us understand the cycles of events despite the fact that life is getting complex. So far, there are more and more researchers who support the later concept. In "The Great Depression Ahead", which was written by American notable economist and predictor Harry S. Dent Jr., said: "Greater complexity leads to greater information and intelligence (including technology and now computers) that actually allow greater predictability over time from the macro to the micro arenas." Moreover, EU establishes a computer system project called "Living Earth Simulator Project." This project uses the latest super computer to analyze great amount of information in Social Networking Services and governmental statistics. They find the trends of human politics and economy, and then take steps to prevent problems like epidemic or financial crisis. Similarly, prediction on stock price will be one of the popular researching fields in financial market as well. The financial market is a complex, evolutionary, and non-linear dynamical system (Abu-Mostafa and Atiya, 1996). It is hard to discover regulations by using traditional time series predicting in high non-linear dynamical system like stock market. This is also the greatest obstacle when researchers build predicting models. The models' consequences are wrong producing or wrong deciding plenty of financial crises. Fortunately, Artificial Neural Network (ANN) provides solutions to

non-linear problems. ANN can abstract useful data among database. It is adaptively formed basing on the features presented in the data. In this study, we focus on the development of a stock market forecasting model based on artificial neural network architecture. This study constructs ARMA and two different EEMD-based ANN algorithm as a baseline model, which enables us to test the TAIEX index futures return.

ANN has been successfully applied in financial fields. Hamid et al. (2004) compared volatility forecasts from neural networks with implied volatility from S&P 500 Index futures options using the Barone-Adesi and Whaley (BAW) American futures options pricing model. Forecasts from neural networks outperform implied volatility forecasts and were not found to be significantly different from realized volatility. Implied volatility forecasts were found to be significantly different from realized volatility in two of three forecast horizons.

Lin and Yu (2009) transformed ANN prediction into a simple FK indicator for decision making trading strategy, whose profitability was evaluated against a simple buy-hold strategy. They adopted the neural network approach to analyze the Taiwan Weighted Index and the S&P 500 in the States. Consequently, they found that the trading rule based on ANNs generated higher returns than the buy-hold strategy.

Jasemi et al. (2011) presented a model to do stock market timing on the basis of the TA of Japanese Candlestick and concept of Neural Networks with astonishing results. In this approach, the network was not going to learn the candlestick lines alone or in combination, but was to present a kind of regression model whose independent variables were important clues and factors of the technical analysis patterns; and its dependent variable was the market trend in near future. In defining the independent variables two approaches were taken; one was Raw data-based and the other is Signal-based with fifteen and twenty-four variables, respectively. Experimental results,

9

in which estimated signals are compared with actual events according to real published daily data in Yahoo, finance, showed that the proposed model performs brilliantly well in emission of buy and sell signals while the first approach seems to some extent better than the second.

Yoon et al. (1991) used forth-layer Back-propagation Neural Network (BPNN) cataloging profits to companies' stock prices and predicts that whether the future performance on stock prices better or worse. He used confidence degree, economic factors, growth rate, and strategies to profit, new products, expected loss, expected profit, long-term or short-term optimistic degree, and other professional estimable variables in Input layer. He compared the consequences of this conducing way and statistic distinguishing analysis, both in-sample and out-sample results are better.

Ken-ichiKamijo and Tanigawa-Tetsuji (1990) used ANN to analyze K-line in Tokyo Stock Exchange. They analyzed triangle patterns of K-line major to find the trend of price changes. After 15 training patterns learning, they used 16 testing patterns to predict. The result in test triangle was accurately recognized in 15 out of 16 experiments. The accuracy rate was high up to 93.8%.

Douglas Wood and Bhaskar.Dasqupta (1994) used ANN, multiple regression analysis, and ARINA predicting French capital market. They found ANN's accuracy rate is over 50%, and the others are less then it.

Zheng-Hsiu Chu (2004) used multiple regression analysis, time series analysis and ANN to analyze 279 weekly trading data from 14 countries. The result was that the improved back propagation network was the best and then was multiple regression analysis in the predicting model return rate.

The preceding historical examples are all applications of ANN in financial area. Recently, ANN also developed a predicting way that is different from traditional one. It is call EMD-ANN or EEMD-ANN.

En Tzu Li (2011) found EEMD-ANN Network which was more proper to use in predicting index and she did predicted the future index price. As a result, the FK value can display a signal if to buy or sell, and confirm trading time, and make buy or sell Call-Put decisions on TAIEX options.

Lean Yu et al. (2008) used an EMD-based neural network ensemble learning paradigm for world crude oil spot price forecasting. In terms of empirical results, Yu find that across different forecasting models, for the two main crude oil prices – WTI crude oil spot price and Brent crude oil spot price – in terms of different criteria, the EMD-based neural network ensemble learning model performs the best.

Lean Yu et al. (2010) proposed an EMD-based multi-scale neural network learning paradigm to predict financial crisis events for early-warning purposes. They used EMD to decompose the important part of data. Using the neural network weights, some important intrinsic mode components were selected as the final neural network inputs and some unimportant IMFs that were of little use in mapping from inputs to output are discarded. Using these selected IMFs, a neural network learning paradigm was used to predict future financial crisis events, based upon some historical data. Experimental results revealed that the proposed multi scale neural network learning paradigm could significantly improve the generalization performance relative to conventional neural networks.

Tsai Yu Ching (2011) applied the Ensemble Empirical Mode Decomposition (EEMD) based Back-propagation Neural Network (BPNN) learning paradigm to two different topics for forecasting: the hourly electricity consumption in National Chengchi University and the historical daily gold price. They used moving-window method in the prediction process. Good accuracy was revealed after using the method and strategy in this prediction. Moreover, they used the ensemble average method, and compared the results with the data produced without applying the ensemble

average method. By using the ensemble average, the outcome was more precise with smaller errors. It resulted from the procedure of finding minimum error function in the BPNN training.

The purpose of this study is to compare the accuracies of ARMA and EEMD-ANN Models and establish trading systems in accordance with the forecasting values. It was appropriate to define a clear price trend which allows us establish trading systems to enter or exit positions. We input the IMFs which is the effect of TAIEX futures to a back-propagation neural network (BPNN) to predict a clear price. The stock index spot should be considered while trading the index futures since they are actually the spot in the future. So, we discussed factors (which affecting index) as well.
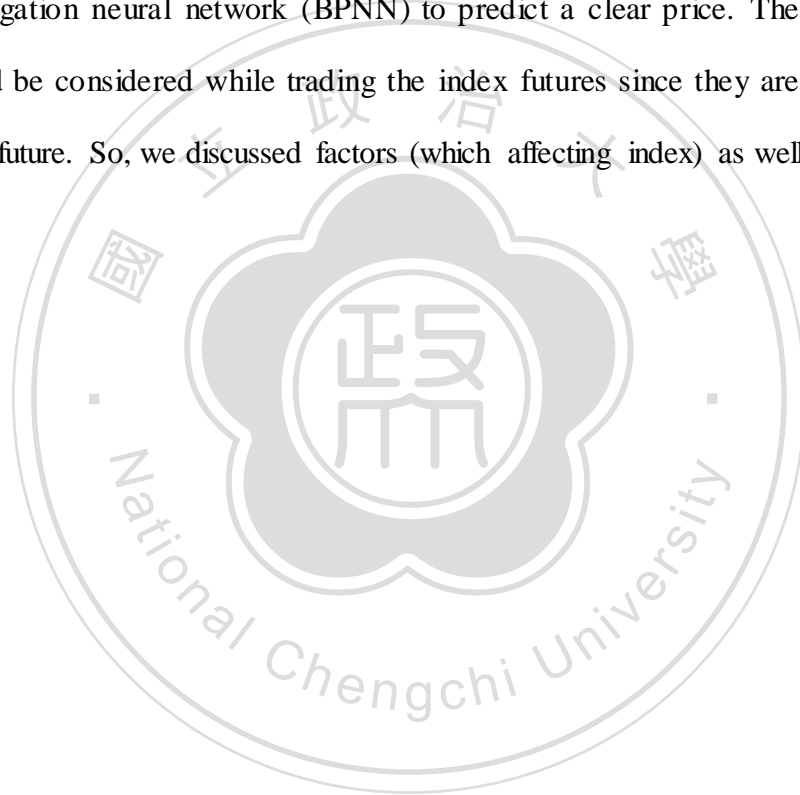
Chapter2 Methodology

2.1 Empirical Mode Decomposition (EMD)

Empirical Mode Decomposition (EMD, Huang et al., 1998) is a novel method analyzing non-linear, non-stationary data.

This method is based on local time scale, extracting intrinsic mode function from original data. It is applied by decomposing oscillation one by one of different time scale in the original data, producing a series of oscillatory mode from high frequency to low frequency. Each scale represents an IMF. It makes IMF having obvious physical meanings. Each IMF means a component of time scale oscillation in original data; while residue shows trend of them.

Each IMF has to match the following two definitions:

1. The number of total local extrema (including local maxima and local minima) and the number of zero-crossings must either equal or differ by at most one.

2. At any point, the mean value of the "upper envelope" (determined by the local maxima) and the "lower envelope" (determined by the local minima) is zero.

The IMF decomposed by two definitions above is approximate mono-components and signals which are almost orthogonal. However, there are a lot of natural data containing various frequencies. EMD uses a shifting process in order to find definition-matching IMF, symmetrize waveform, and ridding riding wave.

Use EMD to decompose the data. The input data must satisfy the following three conditions:

1. There must be two extremes of the data: a maximum and a minimum.

2. Data characteristic time scale is defined as the time difference between the two extremes.

3. If the data contains no extreme but inflection points, we do once or more times of

differential with the signal until finding the extreme. The final results can be obtained from integral component differential.

Base on the establishment of the above conditions. The EMD sifting process is as follows:

(1) Used a cubic spline lines to connect all maxima of the upper envelope $e_{t,max}$ by the local maxima of the data X_t and local minimum value; The minimum value is connected to form the lower envelope $e_{t,min}$

(2) The maximum value of the envelope plus minimal value envelope; Take the average to obtain the mean envelope

$$m_{t,k} = \frac{e_{t,max} - e_{t,min}}{2}$$

(3) The original data X_t by subtracting the mean envelope obtained component $h_{t,k}$, we get

$$h_{t,k} = X_t - m_{t,k}$$

Judgment obtained component meets the IMF two definitions. If not, then $h_{t,k}$ is repeated as the input. Step 1~3 continues shifting and filtering iterative procedure when the obtained component satisfied with two definitions of IMF.

Let $h_{t,k}$ express into $c_{t,n}$. It's expressed as follows:

$$c_{t,n} = h_{t,k}$$
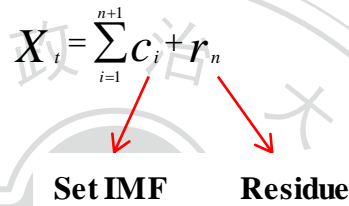
Let the original signal X_t deduct IMF component, and then we can be obtained a residual $r_{t,n.}$ It's expressed as follows:

$$r_{t,n} = X_t - c_{t,n}$$

And so on, IMF definition can be satisfied in order. Signal, $C_{t,1}, C_{t,2}, C_{t,3}, \ldots C_n,$ frequency gradually decreasing.

$$\begin{cases} X_t - c_{t,1} \\ r_{t,1} - c_{t,2} = r_{t,2} \\ r_{t,2} - c_{t,3} = r_{t,3} \\ \vdots \\ r_{t,n-1} - c_n = r_n \end{cases}$$

r_n is a monotonic function of the process. Until IMF can no longer be separated, the entire EMD decomposition process can be considered complete. r_n is trend term in EMD decomposition process, and the original signal can be expressed as the sum of the IMF and the trend term. It's expressed as follow:

$$X_t = \sum_{i=1}^{n+1} c_i + r_n$$

Set IMF   **Residue**

The shifting process automatically generates the value of the zero reference line for each IMF. The data for a long-time trend term will be separated to r_n independently. In order to avoid too much filtering to make IMF loss the physical meaning of the original single-component signal, therefore joined stopping criterion to end the shifting process, called the Standard deviation, SD values.

$$0.2 \leq SD = \sum_{t=0}^{T} \frac{(h_{t,k-1} - h_{t,k})^2}{h_{t,k-1}^2} \leq 0.3$$

SD is the amount of change between the two iterations of the component as a stop criterion. To stop filtering iterations action SD ranged from 0.2 to 0.3, and completed once the IMF's calculations. We can get the IMF by the high frequency to low frequency upon completion of the shifting process. Each IMF has physical meaning.

```
                    ┌─────────────────────┐
                    │   Original data X_t  │
                    └──────────┬──────────┘
                               │
       ┌───────────────────────┼────────────── Cubic spline ──◄───────────┐
       │                       ▼                                           │
       │            ┌──────────────────────────┐                          │
       │            │ Upper envelope e_{t,max} │                          │
       │            │ & Lower envelope e_{t,min}│                         │
       │            └────────────┬─────────────┘                          │
       │                         ▼                              ┌─────────┐
       │            ┌──────────────────────────┐                │ k=k+1   │
       │            │  Mean of envelope m_{t,k} │                └────▲────┘
       │            └────────────┬─────────────┘                     │
  ◆─────────────◆                ▼                                    │
  │ k=0         │   ┌──────────────────────────┐                     │
  │ n=n+1       │   │  h_{t,k}=X_t-m_{t,k}      │                     │
  │ X_t= r_{t,n}│   └────────────┬─────────────┘                     │
  ◆─────────────◆                ▼                         No         │
       ▲                      ◆───────◆ ──────────────────────────────┘
       │                      │  IMF  │
       │                      ◆───────◆
       │                         │ Yes
       │                         ▼
       │            ┌──────────────────────────┐
       │            │     C_{t,n}=h_{t,k}       │
       │            └────────────┬─────────────┘
       │                         ▼
       │            ┌──────────────────────────┐
       │            │ Residue r_{t,n}=X_t-c_{t,n}│
       │            └────────────┬─────────────┘
       │    No                   ▼              Yes    ┌──────────────────────────┐
       └──────┤ Residue (monotonic function) ├────────►│  IMF C_1,C_2,C_3,...R    │
              └──────────────────────────────┘         └──────────────────────────┘
```
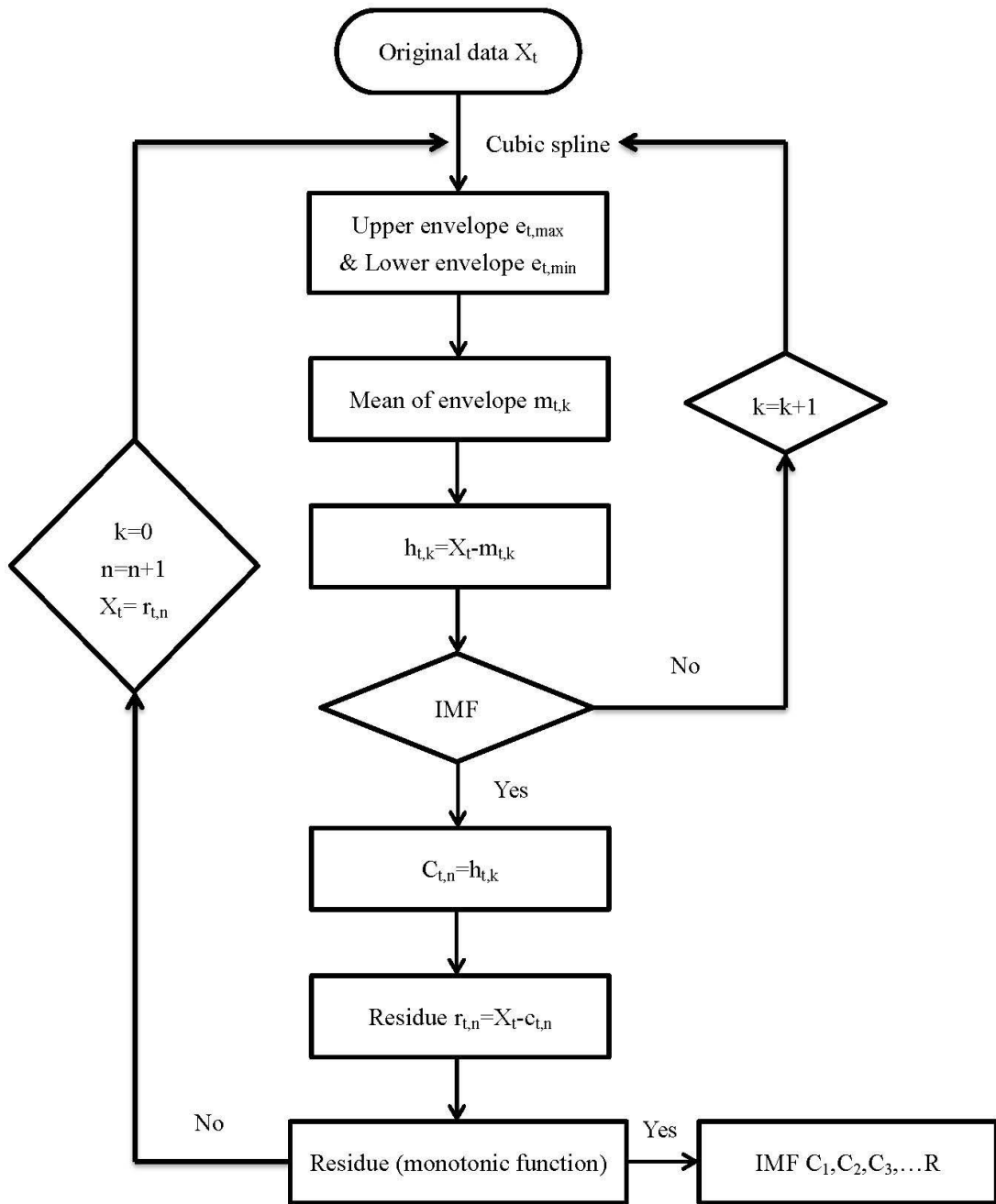
Figure 2.1 the Flow Chart of EMD

2.2 Ensemble Empirical Mode Decomposition (EEMD)

It is really a very practical method that the EMD for non-linear and non-stationary information. But there will be mode mixing problem in the process of empirical mode decomposition mixer wave. That is to say that there will be different signals mixed in the same intrinsic mode functions (IMF), or the same time scale signals appear in different intrinsic mode functions (IMF). This will reduce the physical significance of the IMF. For mixer wave, Wu and Huang (2004) proposed the Ensemble Empirical Mode Decomposition (EEMD) to overcome this problem. The EEMD modal characteristics of white noise contain various frequency scales, adding random white noise decomposition performs an overall before pre-decomposition signal. In other words, the original signal added random white noise, do EMD decomposition first, and then averaging the results of the decomposition. So, the white noise uniformly distributed in each component. The original signal will decompose to the appropriate frequency of component. Repeat the above steps, which are able to get more in line with the actual conditions of the real results.

The statistical know that adding white noise caused the error between the original signals. After a large number of statistical Wu et al. Empirical formula is given:

$$\varepsilon_n = \frac{\varepsilon}{\sqrt{n}}$$

n is represents the number of ensemble members; ε is the amplitude of the added noise; ε_n is the final standard deviation of error defined as the difference between input data and the corresponding IMFs (standard deviation of error).

In the implementation, we will modulation the two parameters ε_n and N. As ε_n increased, errors also become large. Other hand, sometimes the N value is insufficient and apt to cause the signal of the same frequency range decomposed into two signals. So it is very important to choose the appropriate averaging time. The number of the

ensemble numbers N is always set to 100 and the εn always set to 0.1 or 0.2 (Zhang et al., 2008) based on experience.

The following is EEMD decomposing process:

1. Add a white Gaussian noise $\varepsilon_{t,i}$ to the original data x_t to create a new data. That is

$$y_{t,i} = x_t + \varepsilon_{t,i}$$

x_t is original data , $\varepsilon_{t,i}$ are different realizations of white Gaussian noise.

2. Use EMD to decompose each $y_{t,i}$. You can get n IMF which can write as $c_{t,ij}$ and a residue $r_{t,i}$. $c_{t,ij}$ means the number j IMF after adding i white noise.

3. Use ensemble average to the above IMFs. You can get the final IMF , which is:

$$c_{t,j} = \frac{1}{N}\sum_{i}^{N} c_{t,ij}$$

$c_{t,j}$ means the j IMF after using EEMD decomposition to original signal.

Through the above three steps, we already can get IMF more accurately and meaningful. However, in applying, IMFs vary every time after EEMD decomposition. The reason is that EEMD adds different white Gaussian noise randomly. With the differences of ε_n's number, the results vary to each other. Even one selects the same parameter, the result is different. For example, ε_n simulates as 0.1, N simulates as 100, the results are still different from each other. But after doing averages many times, the interferences of white Gaussian noise will turn small, which is near the real answer.

In this study, we use EEMD to analyze TAIEX index, and get a series of meaningful IMF. We are trying to explain these affecting reasons.

2.3. Artificial Neural Networks (ANNs)

Artificial Neural Networks (ANNs) is an Artificial Intelligence Machine using mathematical method. And via computer's rapid calculating ability, it makes computer have ability of predicting. It can simulate non-linear function. After several times of simulations, ANN becomes a complex functional form. Therefore, it is suitable for solving complex questions of non-linear system. Besides, one doesn't have to presupposition before using ANN. You can start analyzing just preparing enough history data. For instance, if one knows the factor of Fundamental information or Technical Information, he can use ANN to start stock index prediction. After many experiments proves, ANN can be applied on many field which old computer system cannot reaches, just as Face Recognition (Rowley et al., 1998)、Financial forecasting (Gately, E., 1996)…etc. In addition, ANN has potential ability of predicting time scale models and Nonparametric estimated as well (Kuan and White, 1994).

This section gives an introduction to basic neural network architectures and learning rules. An ANN's all Neurons are layered by their functions. Three layers in general: input layer, hidden layer, and output layer. Each layer is linked by foregoing order.

1. Input layer: use to represent input variable data. Neurons' number are depends on question. In this study, we use IMF decomposed by EEMD as input data.

2. Hidden layer: use to represent the inter-effects between input data. In many research results and engineering simulation all shows hidden layer doesn't need above two layers (Hush and Horne, 1993) , pointed out network which use two layers in hidden layer, each layer only has a few neurons. It can be replaced by a network with a great amount of neurons in a layer. Moreover, the amount of neurons has to use trail route to decide its best number. ANN

cannot depict the system properly with too little number of neurons. Too large number of neurons will provide over-fitting problem (I. Kaastra, M. Boyd, 1996).

3. Output layer: use to represent output data. In this study, we predict TAIEX futures price. Therefore, the neuron's number is 1 in output layer.

The basic principle of ANN is using weight to connect each layer's neurons. If there are two neurons is stimulated at once by a connection, the strength of this connection will increase. If the two neurons in a connection is not stimulated simultaneously, e.g. one has response, the other doesn't, then the strength of this connection will weaken or vanish (Stent et al., 1973). Each input data will reach hidden layer after weighted accumulation. Through Transfer function there will be a value. The processing flow is represented in Figure2.2.
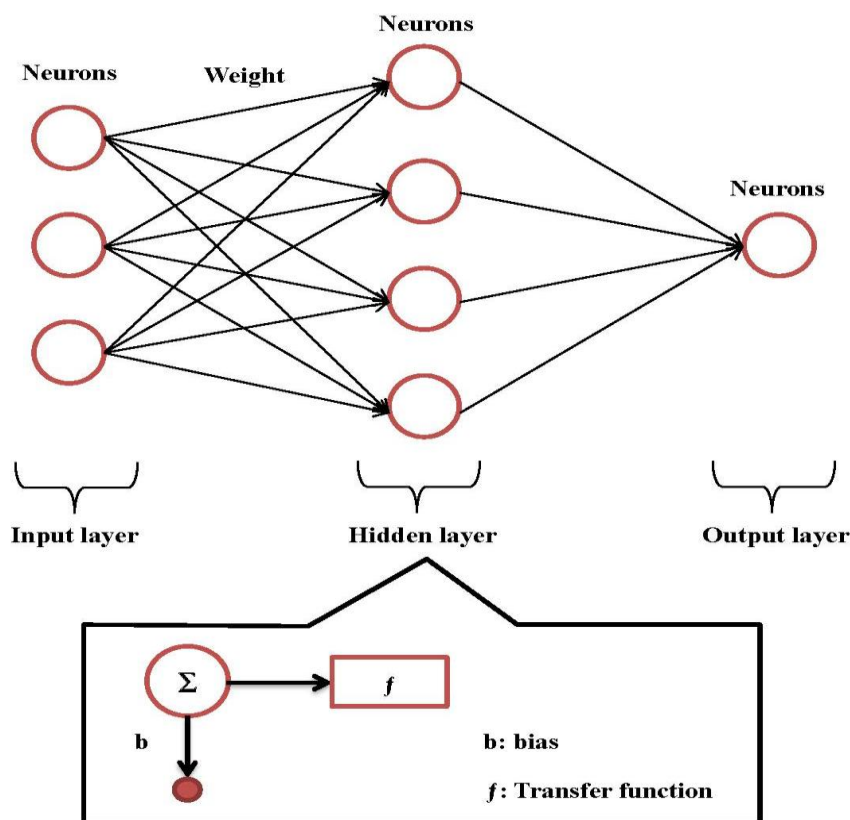

Figure 2.2 Three-Layer Neural Networks

Transfer functions are mathematical formulas that determine the output of processing neurons. The selection of transfer function will affect the whole network's trait. Therefore, networks having different characteristic have to select different transfer functions. Three kinds of transfer functions which are usually used in Back-Propagation Neural Network in this paper is: hyperbolic function, log-sigmoid function and linear function.

$$Hip(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

$$Sig(x) = \frac{1}{1 + e^x}$$

$$Lin(x) = x$$

Since financial markets are nonlinear, nonlinear transfer functions are more favored. So far, In ANN applying, the most popular one is Back-Propagation Neural Network (Rumelhart et al., 1986). It adopts the Widrow-Hoff learning rule (i.e. least mean squared (LMS) rule) (Hagan et al., 1996). The mean square error function could be presented as:

$$F(M) = \frac{1}{N} \sum_{i=1}^{N} [e_i(M)]^2 = \frac{1}{N} \sum_{i=1}^{N} [t_i - Y_i(M)]^2$$

$Y_i(M)$ is the final output value, t_i is the target value, and $e_i(M)$ is the error between the values $Y_i(M)$ and t_i. The processing flow is represented in Figure2.3.
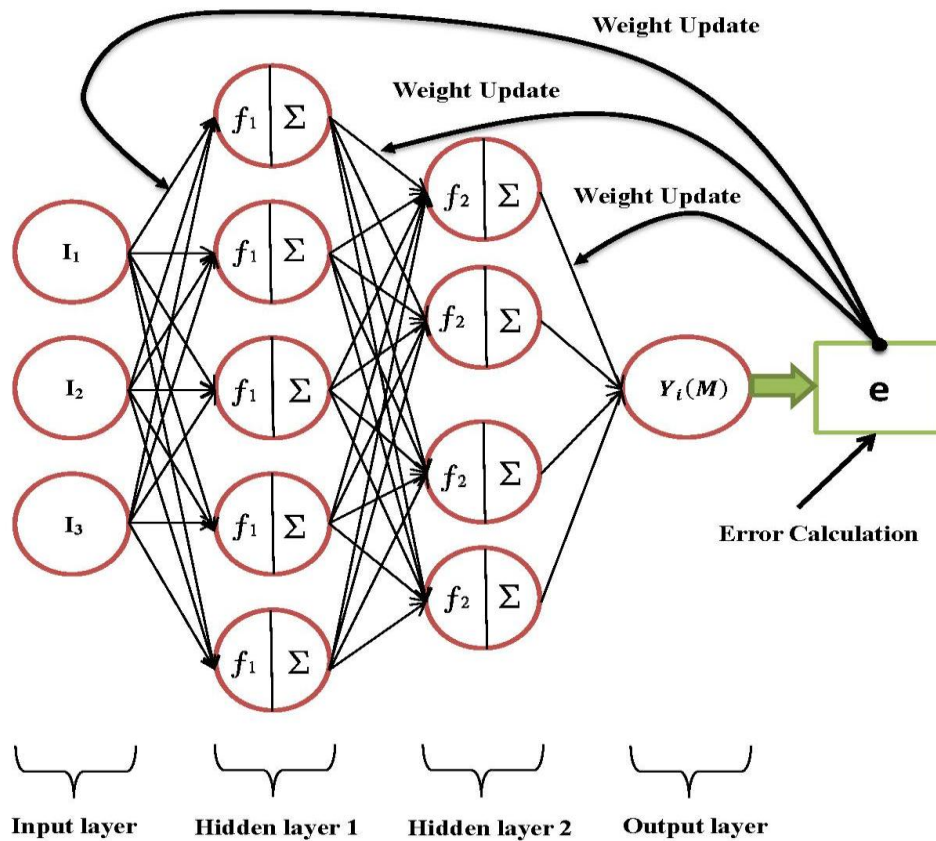
Figure 2.3 the Structure Chart of BPNN's Training Progress

The neurons here have two kinds of inputs, which are input value "I" and bias value "b". "I" is the real input data. In this study, our input values are IMFs after EEMD decomposition. They have to multiply the weights which connect hidden layer. "b" is value which directly input in neurons. It is used to apply non-leaner Threshold Value. Therefore, the formula of ANN dynamical system which has M layers can be expressed as: for $1 \leq i \leq N_m$, $1 \leq j \leq N_m$, $1 \leq m \leq M$

$$Y_i^m = f\left(\sum_{j=1}^{N} w_{ij}^m Y_j^{m-1} + b_i^m\right) = f\left(n_i^m\right)$$

Each ANN contains N_m neuron and the iteration calculation is able to reflect one layer of the process unit in the network. For instance, the neuron i on the layer m accepts all the processed and weighted inputs from (m-1) layer and then adds external

22

input of b_i^m to produce n_i^m. Y_i^m can therefore be obtained after a non-linear transfer function and finally, the value of $Y(M)$ is the output of this network.

The network shown under had j inputs, i neurons in the first hidden layer, k neurons in the second hidden layer, one neuron in the output layer. Transfer function we choose *Hyperbolic Function, Sigmoid Function and Purlin Function*.

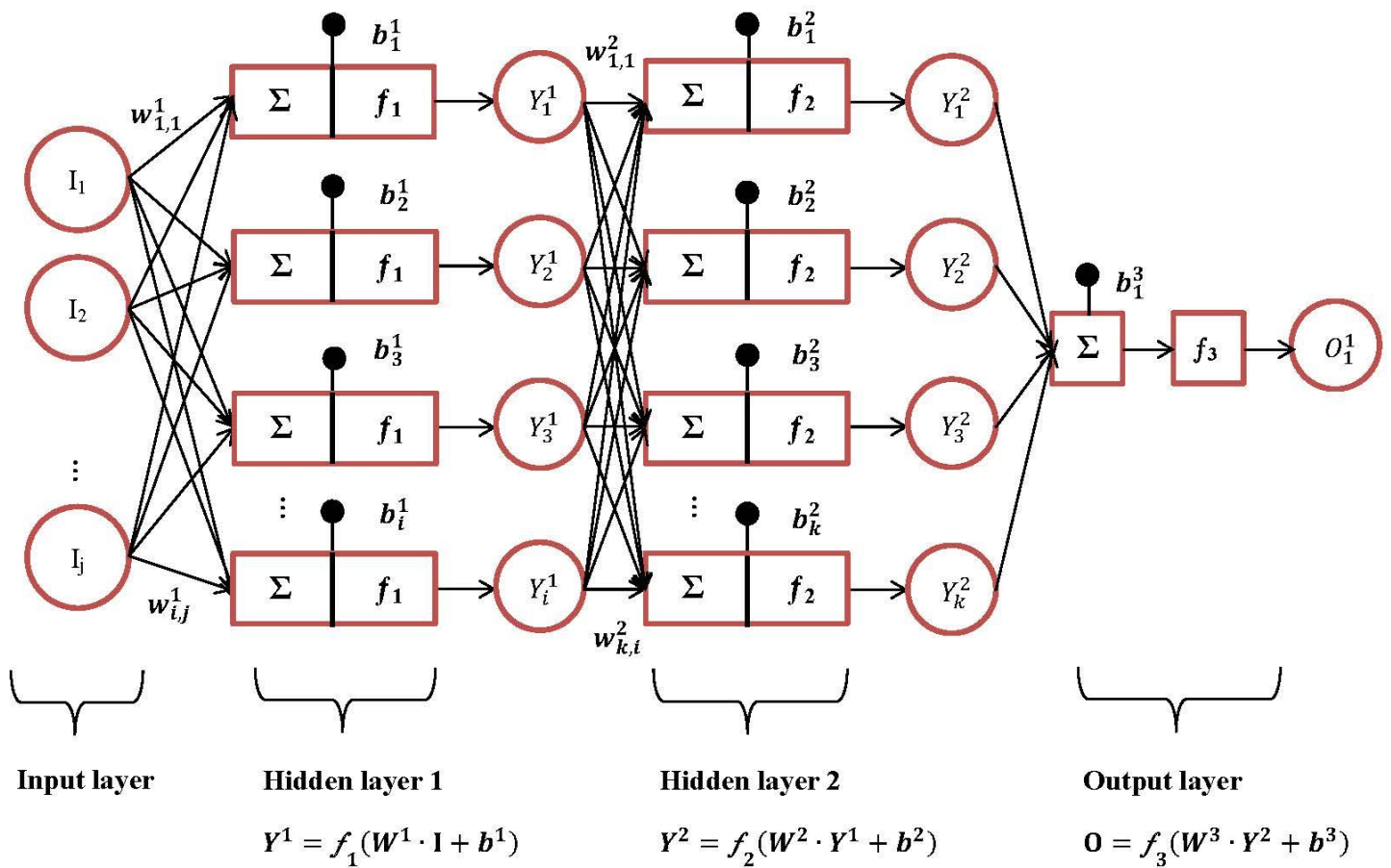| Input layer	Hidden layer 1	Hidden layer 2	Output layer
	$Y^1 = f_1(W^1 \cdot I + b^1)$	$Y^2 = f_2(W^2 \cdot Y^1 + b^2)$	$O = f_3(W^3 \cdot Y^2 + b^3)$

Figure 2.4 A Fully Connected with Four-Layer Feed-forward Network

2.3 EEMD-Based Neural Network Learning paradigms

Artificial neural networks have now been applied on multiple territories and show great forecasting results. One of the most important things is to improve the accuracy rate of ANN forecasting. The traditional artificial neural networks use Cross-validation. However, the result of the neural network learning by single series may be enough. Many researchers have done a lot of extensive research, such as Wavelet-based ANN (Martin F.et al., 2003), Fuzzy BP (Nayak, P.C. et al., 2004), and EMD-based ANN (Lean Yu et al., 2008). EEMD adopted in the study decomposes and screens the primary information of the stock price to several IMFs. Such method can stabilize the data, screen out noisy data, and leave the representative IMFs so that we could regard these IMFs as the input variables of ANN. In general, the EEMD-based neural network learning paradigm contains following steps:

(1) Adding different ε to the original time series data

(2) The input data set and target data set are extracted from the decomposed IMFs and original data.

(3) Put IMFs and residual as the input variables into ANN.

(4) Dividing the database into two sets: the training set and the testing and validation set.

(5) Creating an ANN structure with two hidden layers.

(6) The input data and target data will be co

(7) The number of iteration has infinite epochs.

(8) Reverse the processing of result to get predicts.

2.4 ARMA model

ARMA model is one of the most frequently used models in statistics. As the multi-functional random forecasting model, ARMA model are combined with two models.

1. Autoregressive model (AR):

In AR (P), the current observation value is produced by the weighted average of past observation values and current random error as the formula below:

$$X_t = c + \varepsilon_t + \sum_{i=1}^{p} \varphi_i X_{t-i}$$

2. Moving-average model (MA):

In MA (q), the observation value Yt is the weighted average of random errors in the past q periods.

$$X_t = \mu + \varepsilon_t + \sum_{i=1}^{q} \theta_i \epsilon_{t-i}$$

3. Autoregressive–moving-average model ARMA (p,q) :

Stationary random process has both the qualities of moving average and autogression. Therefore both models should be used.

$$Xt = c + \varepsilon_t + \sum_{i=1}^{p} \varphi_i X_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}$$

In this study, ARMA model is used to forecast TAXIE and establish trading strategy. The accuracy rate and forecasting efficacy of ARMA model is also compared with the results produced by EEMD-ANN models.

Chapter 3 Experimental Details

3.1 Data description

Index Futures

Index futures, as one of the various commodity futures, can be categorized as product and finance in accordance with the nature of contracts in the international market. The sub-category of product includes: agricultural products, metals, energies and so forth while finance includes interest rate, exchange rate, and stock market index. One of the most known items above is the stock market index such as Japanese Nikkei 225, American S&P 500, Dow Jones Industrial Average Index, and TAIEX Index. The stock market index is generated from the weighted average or mean value of a holding of stock. For instance, the TAIEX Index is the weighted average while Japanese Nikkei 225 is general from mean value. The stock index spot should be considered while trading the index futures since they are actually the spot in the future. If not, the arbitrage trade cannot be succeeded.

The main functions of index futures are hedging, arbitrage, and speculation. As for most investors, the purpose of stock trading is hedging (back spread) and one of the most crucial advantages of index futures is the financial leverage. Stock trading requires the screening capability to avoid earning points but losing top spot. The trading of index futures can simplify the target so that investor only needs to understand the trend of market to make a good trade. The traders of TAIEX futures must refer to the stock index spot provided by the information system otherwise the future market maybe disorientated. Therefore, as for the TAIEX Index, there will be two different index prices in the market – the index spot and the index futures.

It can be told that TAIEX Futures have become one of the significant investing targets for legal entities or individual investors from the growth of trading volume.

Statistics from Taiwan Future Exchange show that legal entities have grown greatly in the recent years which suggest that the understandings of futures and options have been furthered hence more and more investors attempt to trade futures and options for hedging and arbitrage which cause the steady increase of trading volumes.

Table 3.1 the Trading Volumes from 2000 to 2005

Year	2000	2001	2002	2003	2004	2005
Volume	1,339,907	2,844,709	4,132,040	6,514,691	8,861,278	6,917,375

Table 3.2 the Trading Volumes from 2006 to 2011

Year	2006	2007	2008	2009	2010	2011
Volume	9,914,999	11,813,150	19,819,775	24,625,062	25,332,827	30,611,932

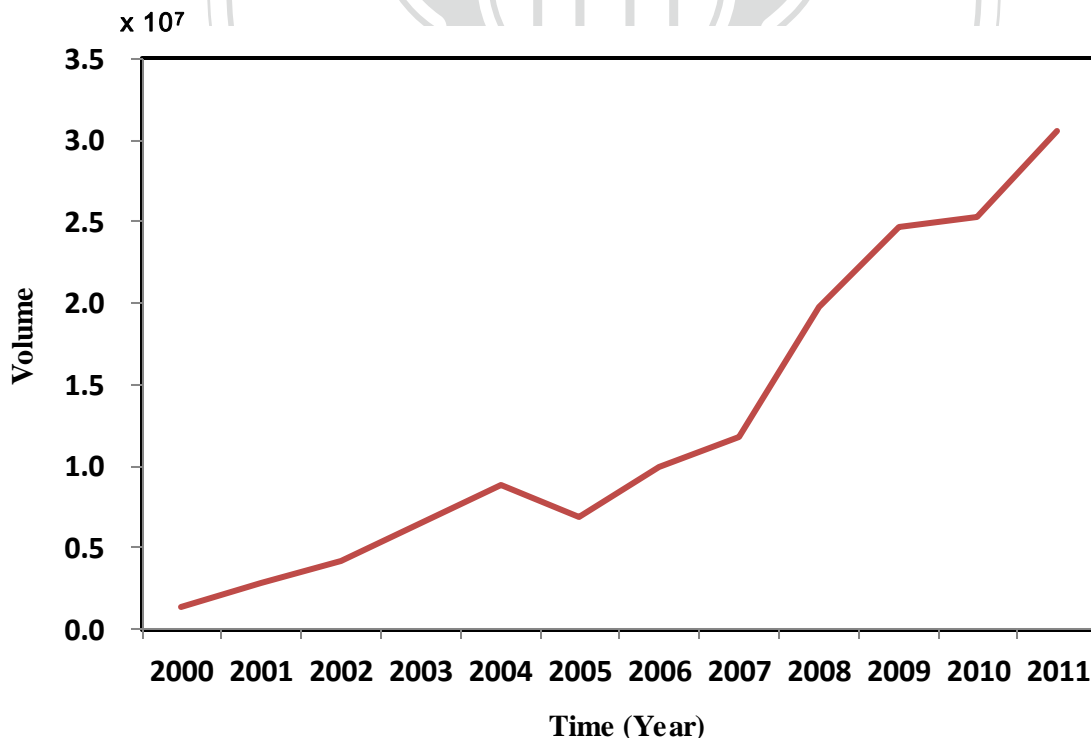

Figure 3.1 the Booming Figure of Trading Volumes during 2000~2011

The contract form of TAIEX futures is as follow:

Table 3.3 the Specification for TAIEX Futures

Item	Description
Underlying Index	Taiwan Stock Exchange Capitalization Weighted Stock Index (TAIEX)
Ticker Symbol	TX
Delivery Months	Spot month, the next calendar month, and the next three quarterly months
Last Trading Day	The third Wednesday of the delivery month of each contract
Trading Hours	1. 08:45AM-1:45PM Taiwan time Monday through Friday of the regular business days of the Taiwan Stock Exchange 2. 08:45AM-1:30PM on the last trading day for the delivery month contract
Contract Size	NTD 200 x per index point
Minimum Price Fluctuation	One index point (NTD 200)
Daily Price Limit	+/- 7% of previous day's settlement price
Margin	1. The initial and maintenance margin levels as well as the collection measures prescribed by the FCM to its customers shall not be less than those required by TAIFEX. 2. The initial margin and maintenance margin announced by TAIFEX shall be based on the clearing margin calculated according to TAIFEX's Criteria and Collecting Methods Regarding Clearing Margins, plus a percentage prescribed by TAIFEX.
Daily Settlement Price	The daily settlement price is the volume weighted average price, which is calculated by dividing the value of trades by the volume within the last one minute, or as otherwise determined by TAIFEX according to the Trading Rules.
Final Settlement Day	The same day as the last trading day
Final Settlement Price	The average price of the underlying index disclosed within the last 30 minutes prior to the close of trading on the final settlement day. Method used to calculate final settlement price.
Settlement	Cash settlement
Position Limit	Combined with the calculation of MTX position limit (on a pro rata basis of 1:4 contract size) These position limits are not applicable to omnibus accounts.

Source: Taiwan Future Exchange

The data attempted to forecast in the study were the historical data of daily TAIEX futures closing price. The data collection was made from 2010/1/4 to 2010/12/31, total of 250 business days. The Figure 3.2 of data collection is as follow:
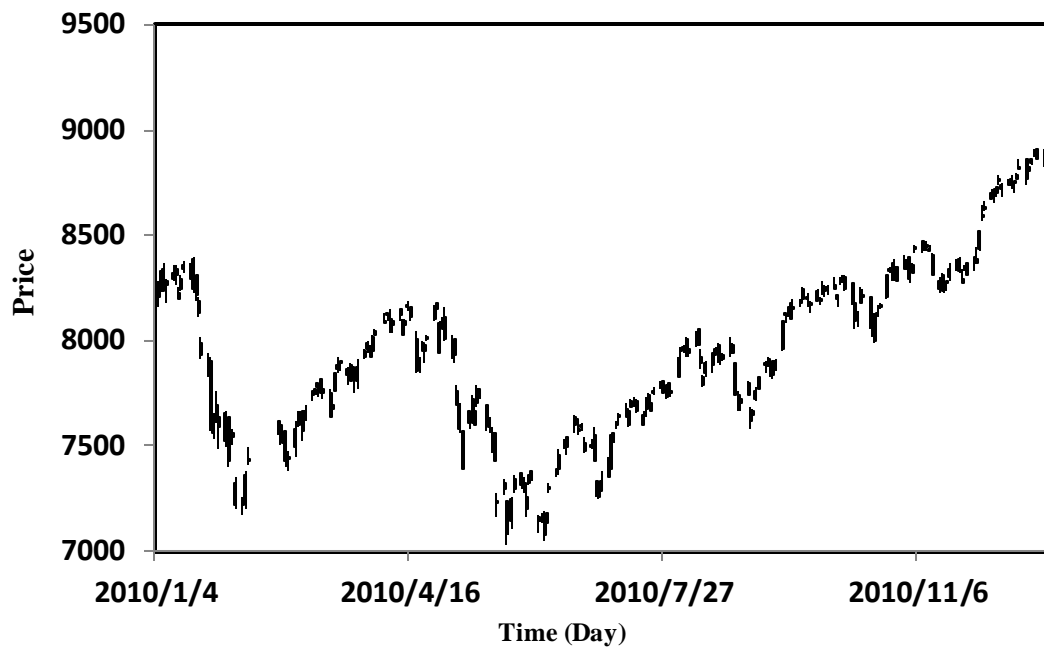

Figure 3.2 the TAIEX Futures during 2010

3.2 The Operation of Price

The reason why the market mechanism is so charming is that we always believe as long as we know the price fluctuation patterns, no loss will be made. Yet, it is not and the uncertainty of market is even higher than the fluctuation of price, which include:

1. The uncertainty of price fluctuation

Will the stock price go up or down tomorrow? Three models were used to forecast the direction of stock price of the next day in this study.

2. The uncertainty of the fluctuation range

Assume the forecasted direction is correct, we may profit only a bit before our position ends or we may profit greatly yet the comeback of price in the end causes the profit. Both cases are caused by the uncertainty of fluctuation range.

29

3. The uncertainty of the fluctuation pattern

Will the price go up and then down? When will the market callback? Will the price go directly or zigzag up? Will the callback be slow or acute? The answers to these questions remain uncertain and investors can only use trading strategy to protect their positions.

4. The uncertainty of unexpected event

Random interference, occasional events, and unexpected events are commonly occur in the market and may cause violent fluctuation at the moment. Investors must adopt trading strategies to avoid severe loss by preventing the impacts from violent fluctuation on their positions.

These uncertainties are complicate and changeful combinations therefore it is truly difficult to forecast the market. Especially, the financial operation difficulty has increased since the snowball impacts of subprime mortgage crisis in the U.S. reached on Europe and Asia after the bankruptcy of Lehman Brothers Holding Inc. as well as the global financial crisis initiated by the financial fail of Iceland government. The market ardently demands a relatively accurate forecasting system to prevent grand loss. The study compared the forecasting accuracies of three models and established the trading strategies in accordance with price forecasted in order to provide a decent forecasting model for the market.

3.3 Statistical measures

We use the following statistical measures to analyze the IMFs and residues. The measures used to analyze IMFs are mean periods, Pearson correlation coefficient, and power percentage. These measures are presented as follows:

Mean period

The mean period which is used to see the cycle of an IMF is calculated by the inverse of mean frequency. The mean frequency is the average of ─instantaneous frequency. To calculate the instantaneous frequency, we apply HHT to the extracted IMFs. In the following paragraph we will briefly introduce the HHT, proposed by Huang et al. (1998).

For any arbitrary time-series data set X(t), we can always have its Hilbert-transform Y(t) as

$$Y(t) = \frac{1}{\pi} \text{p.v.} \int_{-\infty}^{\infty} \frac{X(t')}{t - t'} dt'$$

The p.v. of this formula indicates the Cauchy principal value.

As the X(t) and Y(t) are the corresponding real part and imaginary part ,then we use X(t) and Y(t) to form the complex conjugate pair, and get an analytic signal Z(t) as

$$Z(t) = X(t) + iY(t) = a(t)e^{i\theta(t)}$$

Where

$$a(t) = \sqrt{X^2(t) + Y^2(t)}$$

And

$$\theta(t) = \arctan\left(\frac{Y(t)}{X(t)}\right)$$

Which a(t) represents the time-series amplitudes, and θ(t) represents the time-series phase. Since X(t) and Y(t) are both instantaneous values, θ(t) can indicate the changing value of the phase between two continue points. Now we can define the instantaneous frequency of Hilbert transform through the phase:

$$f(t) = \frac{w(t)}{2\pi} = \frac{1}{2\pi}\frac{d\theta(t)}{dt}$$

Therefore, the mean frequency F of an IMF can be presented as:

$$F = \frac{1}{N}\sum_{t=1}^{N} f(t)$$

And the mean period T will be

$$T = \frac{1}{F}$$

Notice: We didn't calculate the mean period of residue because it is a monotonic function, so in this study we ignored the mean period of residue.

Pearson correlation coefficient

Correlation coefficient can be used to describe the linear relationship between two variables. That is to use a number to indicate the relationship between the two variables and the direction of this relationship (positive or negative). Covariance can be used in finding out the correlation between two random variables. Cov_{XY} indicates the covariance of X and Y that is multiplying the distance of each group and adding up. It's expressed as follows:

$$Cov_{XY} = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{N - 1}$$

The value would fall between positive and negative infinite. When the value is a positive number, Y would increase when X increase while if the value is a negative number, Y would decrease as X increase. However, the degree of correlation remains unknown therefore correlation coefficient can tell more about the correlation between X and Y. The formula of correlation coefficient is as follow:

$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum(X_i - \bar{X})^2 * \sum(Y_i - \bar{Y})^2}}$$

The σ is the standard deviation and the value will fall between -1 to +1.

Power percentage

Power percentage is a measure based on variance for detecting the weight of an IMF on the original data. A higher value of power percentage indicates a stronger weight an IMF is. The power percentage is defined as follows:

$$\text{Power percentage}(\%) = \frac{\text{VAR(IMF)}}{\text{VAR(original data)}} * 100\%$$

3.4 Significant IMFs

The study starts with the analysis of basic factors that determine the fluctuation of stock price. First, the data forecasted are TAIEX Futures. Futures are the stocks in the future, therefore the factors of TAIEX Index must be identified before forecasting TAIEX Futures and the indication of IMFs can be clearly defined through the statistical measures.

As mentioned before, the EEMD method proposed by Huang can decompose time-series data into several IMFs which have its own physical meaning. Since there are lots of IMFs decomposed from the data (TAIEX Index), we must to sort out the significant IMFs which are more meaningful to analyze.

In section 3.3, the statistical measures "power percentage" is used to determine the significant IMFs, and the correlation coefficients "Pearson coefficient" is used to compare the correlation between the significant components with the original data. Moreover, the statistical measure "mean period" can be indicated IMFs own specific meanings.

First of all, the factors of price fluctuation range widely and complicatedly. Generally, the fluctuation of price is determined by the demands and supplies. There are investors choosing to sell out, there are ones who choose to buy in, and this is so-called trading behaviors. In TAIEX Index, the organizational investors hold the

dominant power on the market. Hence, the study attempted to identify the effects of organizational investors on TAIEX Index through the trading value of foreign and other investors.

The main approach for studying the fluctuation of stock prices can be categorized as follows:

1. Basic knowledge: analyze the basic situation of the macro-economy and the industries. The macro-economy reflects the integral operation efficacy and establishes the foundation for the further development of the enterprise. Therefore both macro-economy and the industries are closely related with the stock prices. The basic knowledge of the industry includes, financial status, profiting…etc. The second riches entrepreneur, Warren E. Buffett believes he can make investment judgment in accordance with only basic knowledge.

2. Technicality: refer to the technical index, trend, and K-charts that reflect the fluctuation of price. The technical analysis only considers the real pricing behaviors in the market or of the financial tools. It is believed that "the history repeats itself" and mass statistic data are used to forecast the direction of trends. The study also adopts the moving average indictors in the technical index to assist trading.

3. Unexpected situation: when unexpected event occurs, the stock price may have severe fluctuations.

From the database of Taiwan Stock Exchange Corporation, only the data of Trading Value of Foreign & Other Investor from April, 2004 to July, 2012 were available therefore the prices of TAIEX Index from April, 2004 to July, 2012 were analyzed. There are five IMFs and one residue exacted from the monthly TAIEX Index data by EEMD. The original data and the results of decomposition are shown in Figure 3.3& 3.4, and the statistical measures are listed in Table 3.4.
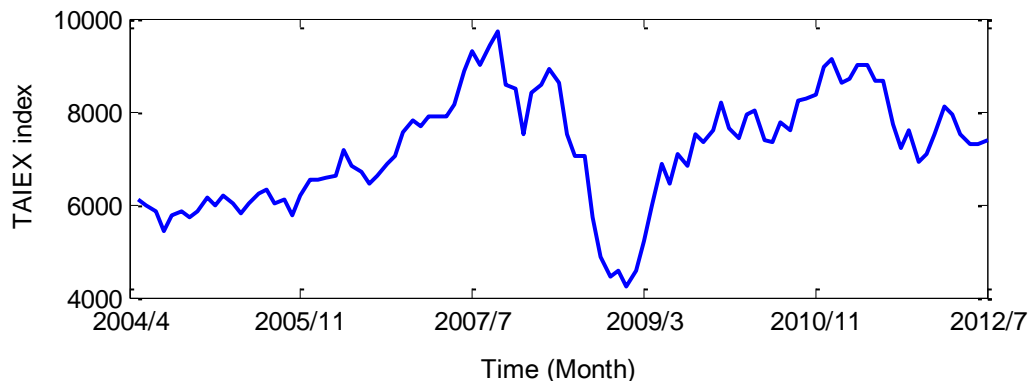
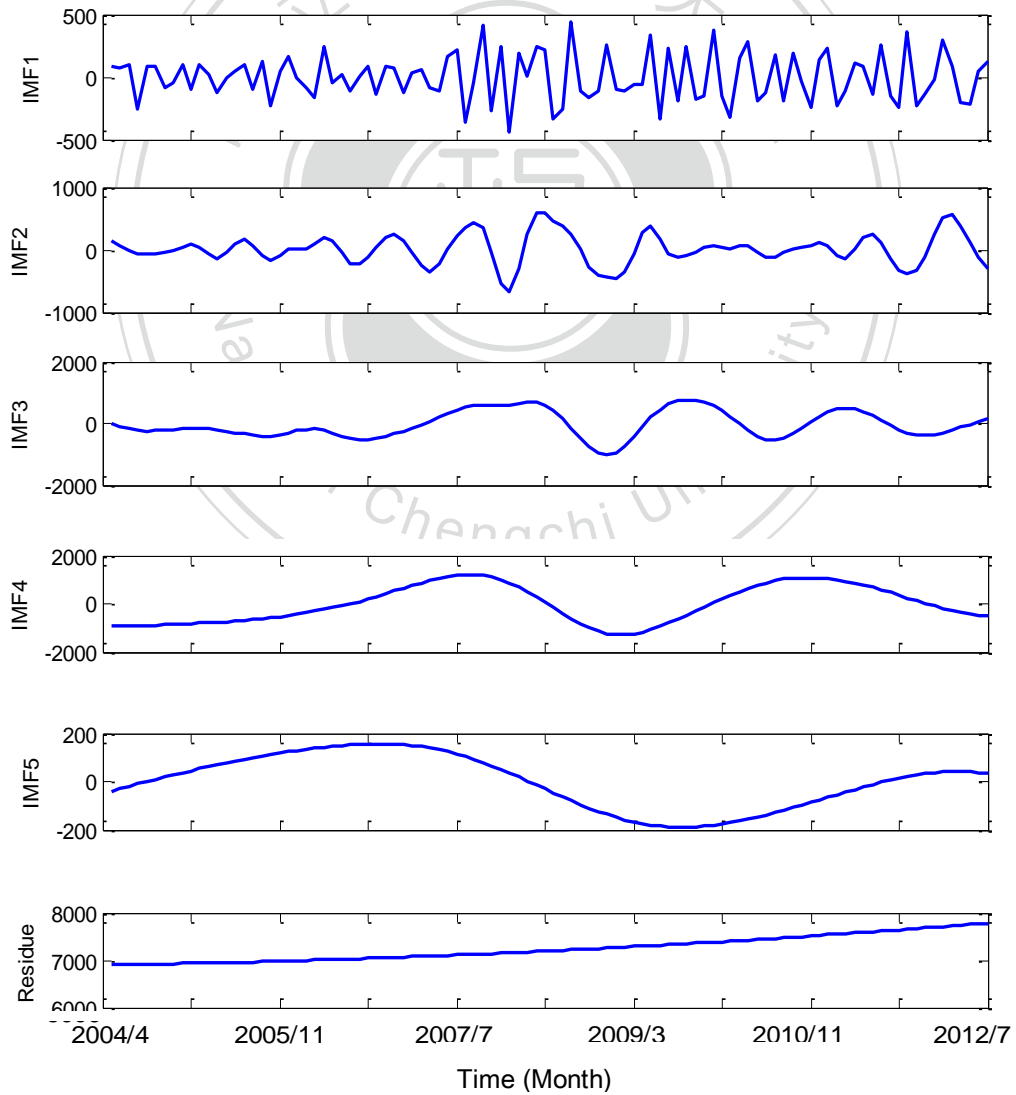Figure 3.3 the TAIEX Index from 2004 April to 2012 July

Figure 3.4 the IMF1~ IMF5 and the Residue Decomposed from TAIEX Index

35

Table 3.4 the Statistical Measures of the Components Decomposed from TAIEX
Index

	Mean period(month)	Pearson coefficient	power percentage
IMF1	3.16	0.2	2.65%
IMF2	8.04	0.39	4.73%
IMF3	22.52	0.66	16.88%
IMF4	50.52	0.89	34.35%
IMF5	71.65	0.08	0.83%
Residue		0.44	3.90%
Sum			63.33%

IMF3 and IMF4 are considered the most important. The power of percentage of
IMF4 is about 34% and the latter is about 17%. In addition, the mean period of IMF3
is approximately two years while it is four years for IMF4 which suggests, the two
years and four years loops of TAIEX Index are significant. However, from the
perspectives of corporate profits, there are quarterly and annual reports. The influence
of financial reports are expected to be seen in the data therefore IMF1 is seen as high
frequency term, IMF2 and IMF3 are combined as mid frequency term, while IMF4
and IMF5 are the low frequency term based on the mean period. The four terms
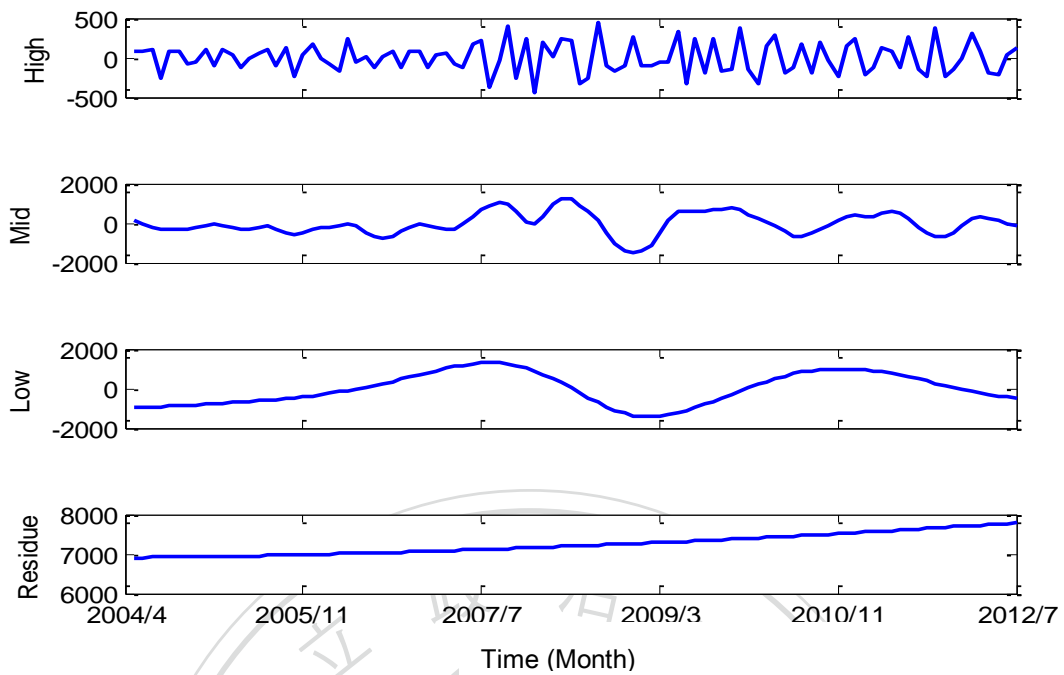combined from the TAIEX Index components are shown in Figure 3.5.

Figure 3.5 the Four Terms Combined from the TAIEX Index Components

Table 3.5 the Statistical Measures of each term Combined from the TAIEX Index Components

	Mean period(month)	Pearson coefficient	power percentage
High frequency term	3.21	0.21	2.46%
Mid frequency term	13.46	0.72	20.48%
Low frequency term	50.47	0.88	39.82%
Trend		0.41	4.67%
Sum			67.73%

In Table 3.5, it is shown that the combined IMF has the mean period of approximately three months in the high frequency term, one year in the mid frequency term, and four years in the low frequency term. The influential orders from high to low are low frequency, mid frequency, trend, and high frequency on TAIEX Index. Each frequency and trend is significant and analyzed.

High frequency term

It is considered that high frequency term is greatly related to quarterly reports. The organization would also adjust their investments in accordance with quarterly reports. To analyze the correlations between TAIEX Index and organizations, the Trading Value of Foreign & Other Investors was also decomposed as five IMFs and one residue by EEMD. The Trading Value of Foreign & Other Investors of decomposition are shown in Figure 3.6
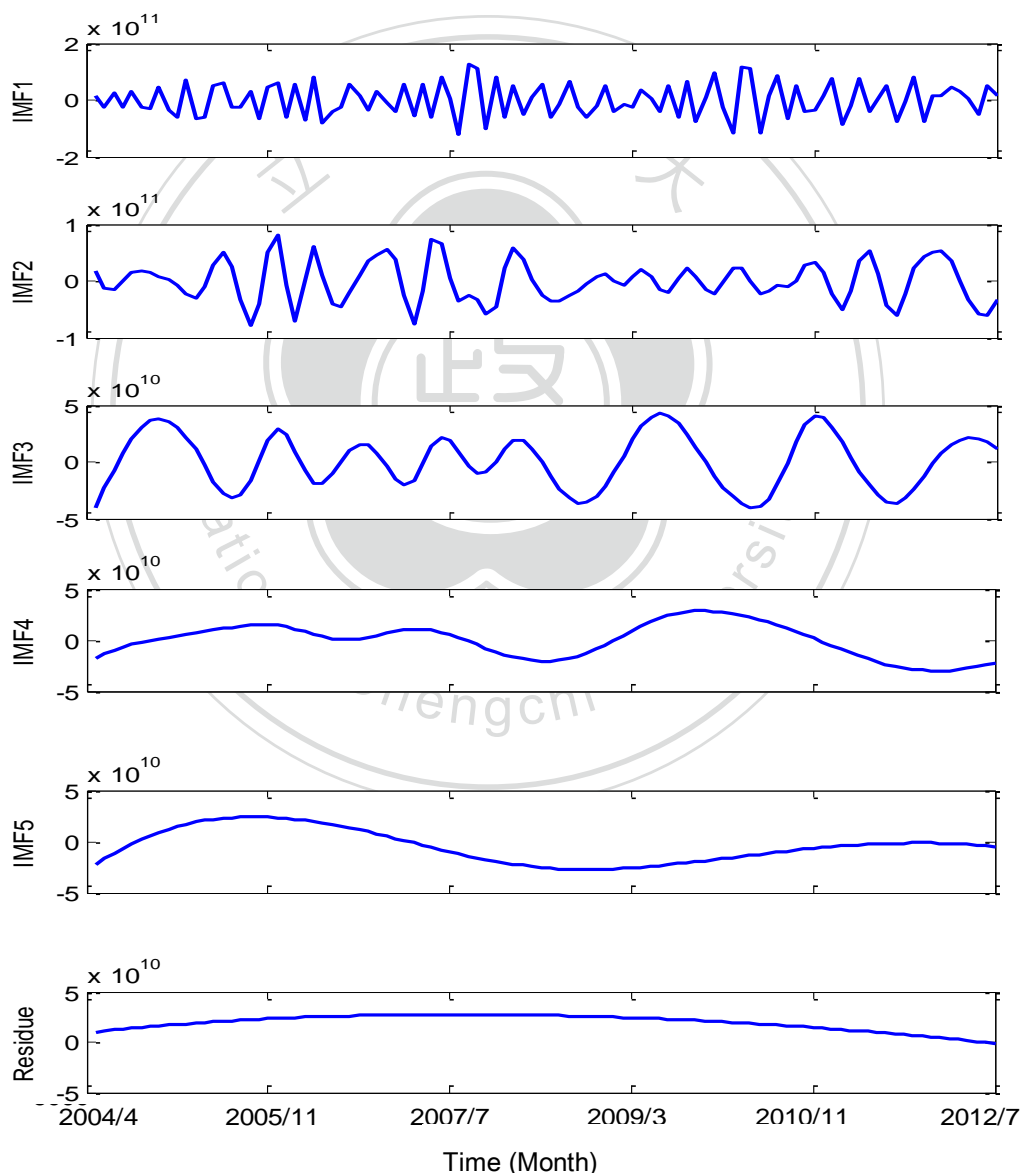

Figure 3.6 the IMF1~ IMF5 and the Residue Decomposed from the Trading Value of Foreign & Other Investors

Table 3.6 the Statistical Measures of each term from the Trading Value of Foreign & Other Investors

	Mean period(month)	Pearson coefficient	power percentage
IMF1	3.00	0.74	50.06%
IMF2	6.77	0.57	17.72%
IMF3	12.77	0.37	8.05%
IMF4	36.58	0.33	4.02%
IMF5	64.48	0.24	3.77%
Residue		0.16	0.97%
Sum			84.59%

From the table 3.6, IMF1 are the most important and the powerful percentage is around 50%. The mean period is a three months loop. Pearson correlation between the high frequency term of TAIEX Index and IMF1 were analyzed and shown in Table 3.7. The results suggested, there was a 0.75 positive correlation. After comparing with actual data, such trend is also reflected on the relative high point and low point of high frequency term as well as the IMF1. Figure 3.7 shows the relative high point and low point. It is concluded, the organization would adjust their position in accordance with the profit of the corporation and the adjustment of position would in turn have effects on the stock market.

Nonetheless, there is no significant correlation between high frequency term and original data (0.21) which suggests, the trading behaviors of organizations would not have effects on the price fluctuation of TAIEX Index for the long term.
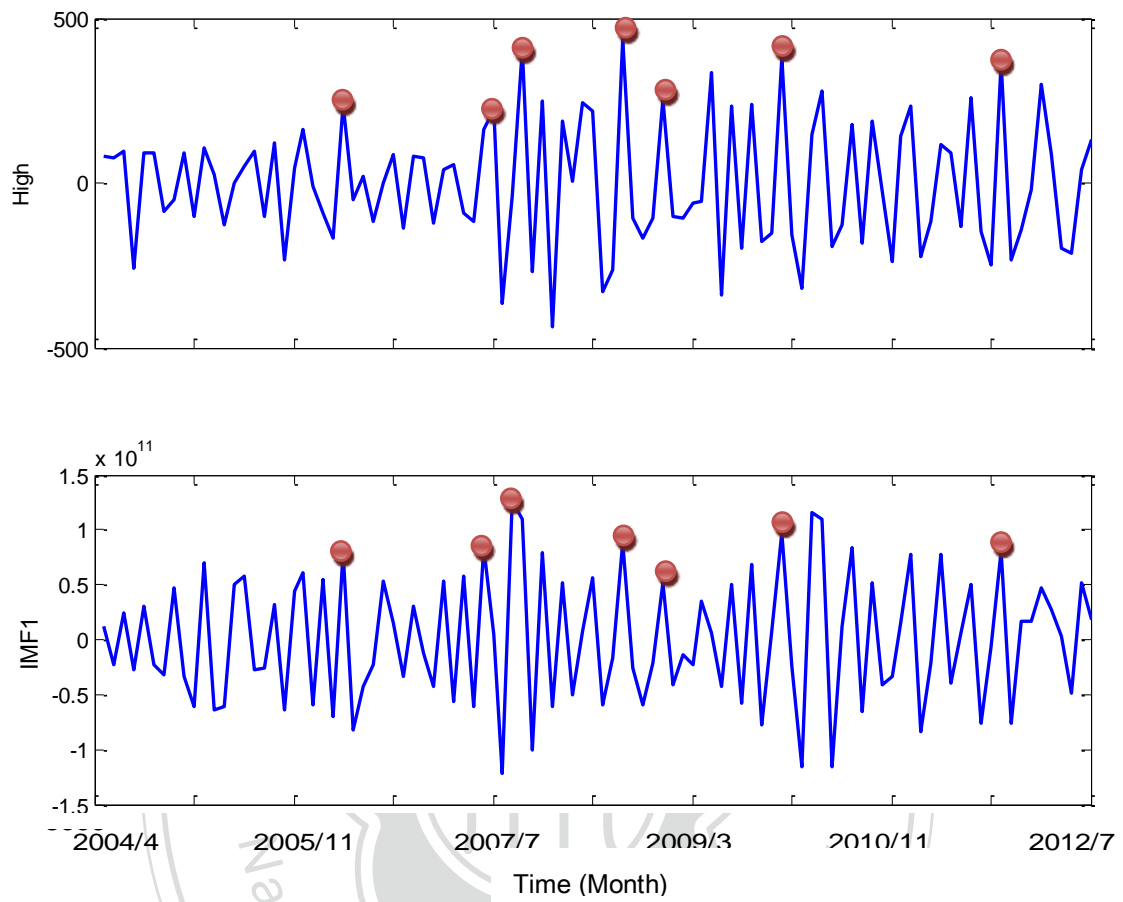
Figure 3.7 the Relative High Point and Low Point

Table 3.7 The Correlation Coefficients Compared the Trading Value of Foreign &
Other Investors to TAIEX Index in terms of High Frequency

Matches of TAIEX Index	Pearson correlation
High Frequency	0.75

Mid frequency term

There is significant correlation between the mid frequency term and original data (0.72). The power percentage is 20% as the second most influential factor for TAIEX Index. The meanings and phenomenon of mid frequency are discussed as follows. The mean period of mid frequency term is around one year, therefore it is considered to be related to the annual reports released by the corporations. The annual reports indicate the annual profits for the corporations. The long-term investors would adjust their position in accordance with the annual reports. In addition, it is shown in many literatures, there are abnormal effects on the stock market and the most noted is January Effect.

"The January Effect might be the granddaddy of all calendar anomalies, but it is not the only one. For inexplicable reasons, stocks generally do much better in the first half of the month than the second half, do well before holidays, and plunge in the month of September. Furthermore, they do exceptionally well between Christmas and New Year's Day, and until very recently, they have soared on the last trading day of December, which is actually the day that has launched the January Effect. Why these anomalies occur is not well understood, and whether they will continue to be significant in the future is an open question. But their discovery has put economists on the spot. No longer can researchers be so certain that the stock market is thoroughly unpredictable and impossible to beat. " (Stocks for the Long Run, 4e, Jeremy J.Siegel) Generally, investors may sell losing stocks to avoid taxes and make another purchase in the next January. Furthermore, the sell-off in December that causes the drop of price also induce buy in and therefore the stock price tends to go up in January. "If taxes are a factor, however, they cannot be the only one, for the January Effect holds in countries that do not have a capital gains tax. Japan did not tax capital gains for individual investors until 1989, but the January Effect existed before then.

Furthermore, capital gains were not taxed in Canada before 1972, and yet there was a January Effect in that country as well." (Stocks for the Long Run, 4e, Jeremy J.Siegel)

In addition, there are other explanation for January Effects. "There are other potential explanations for the January Effect. Workers often receive extra income, such as from bonuses and other forms of compensation, at year-end." (Stocks for the Long Run, 4e, Jeremy J.Siegel) However, as January Effects attract more attentions, investors now are inclined to use it as the operational strategy so the effects are weakened significantly. Liu (2010) argued that January Effect can be found in the stock market in Taiwan and United Kingdom. The observations on mid frequency term also show that the prices rose in January, 2005, 2006, 2007, 2009, 2011, and 2012. The reason why there was no January Effect in 2008 and 2010 was due to the global financial crisis and European sovereign-debt crisis. Figure 3.8 shows the January Effect.
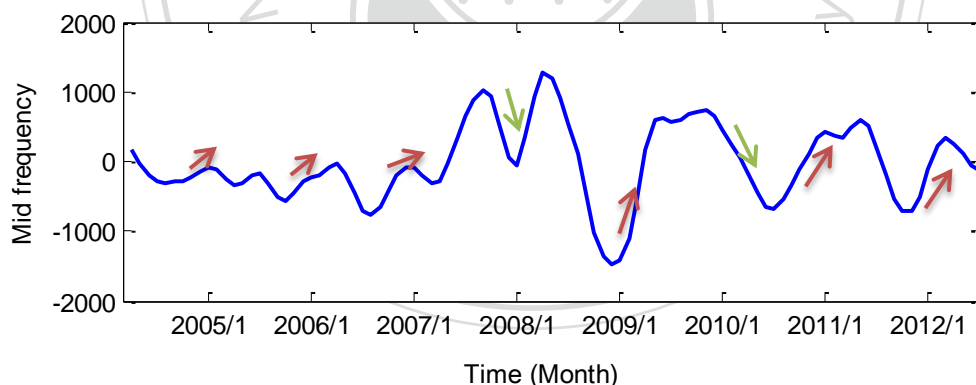

Figure 3.8 the January Effect

The incidents that affect TAIEX Index such as, the Subprime mortgage crisis in 2007, economic crisis in 2008, European sovereign-debt crisis in 2010, enterprise review in 2011, and the utility rise as well as taxation of stock income in 2012, can all be reflected in the mid frequency term. Table 3.8 and Figure 3.9 shows mid-frequency

term can reflect important financial incidents.

Table 3.8 Incidents

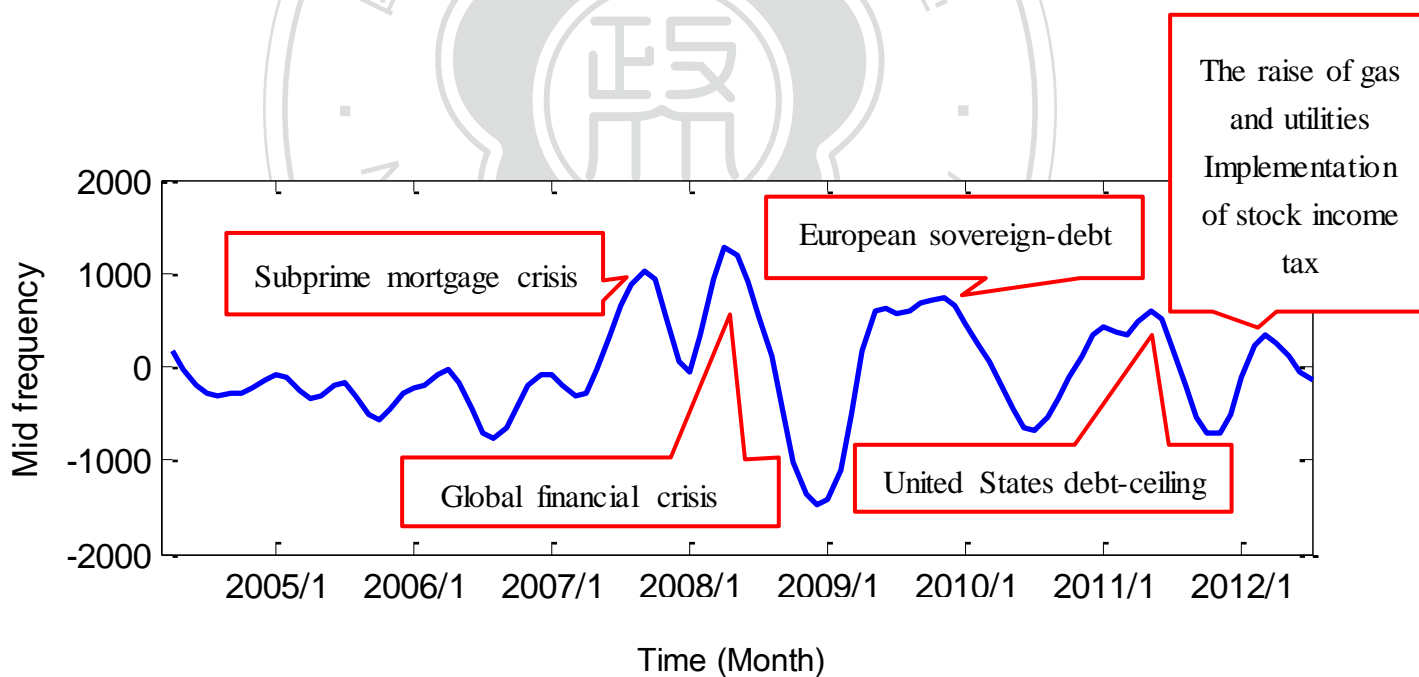| Event	Time	Price
(1) Subprime mortgage crisis	Oct-07	9859→7384
(2) Global financial crisis	May-08	9309→3955
(3) European sovereign-debt crisis	Jan-10	8395→7080
(4) United States debt-ceiling crisis	Aug-11	8707→6744
(5) The raise of gas and utilities Implementation of stock income tax	Apr-12	8121→6857

Figure 3.9 the Big Events

Low frequency term

There is significant correlation between the low frequency term and original data (0.88). The power percentage is 40% as the most crucial factor of TAIEX Index. The meanings and phenomenon of low frequency are discussed as follows. The low frequency term is considered related to the booming cycle of TAIEX Index.

The booming cycle is also known as economic cycle or business cycle, indicating the circulation trend of recession, depression, recovery, and prosperity of GDP. Business circulation is shown in Figure 3.10 the trend Y indicates Real GDP that is, GDP would increase with trend Y through time even if there is no business circulation. The economics have long-term growth and the growth pace would be determined by saving rate, population growth, and development of technology. The Real GDP is fluctuated when the business circulation occurs.
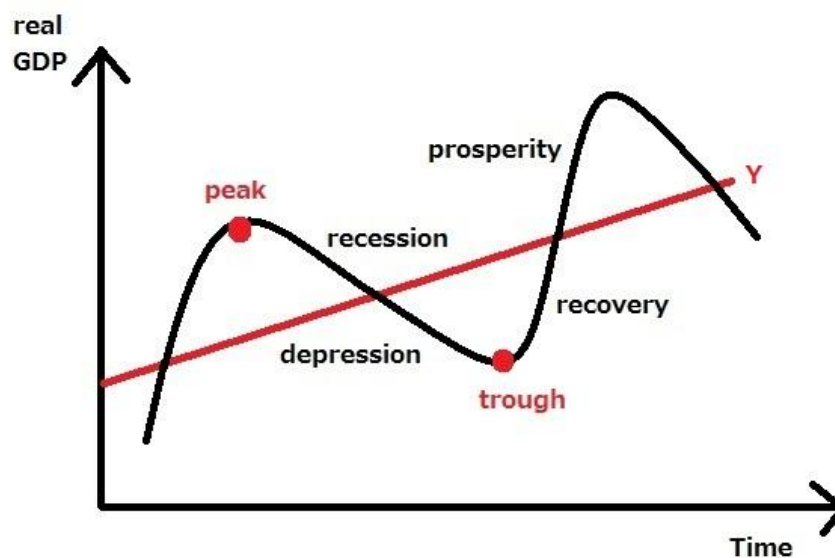

Figure 3.10 Business Cycle

One of the purposes of studies on business circulation is to identify the turning point to prepare for the future economics in advance. Council for Economic Planning

and Development (ECPD) would determine the economic status on the basis of macroeconomics. The leading indicator was chose to make comparisons with low frequency term and according to ECPD, the indicator is consisted with seven factors as follows: Index of export orders、Monetary aggregates, M1B、Stock prices index、Index of producer's inventory、Average monthly overtime in industry & services、Building permits (including only housing, mercantile, business and service, industry warehousing)、SEMI book-to-bill ratio.

The Figure 3.11 and 3.12 shows the comparison of low frequency term and leading indicator. The Table 3.9 shows the correlation coefficient is 0.98 which suggests the low frequency term is related to booming loop.
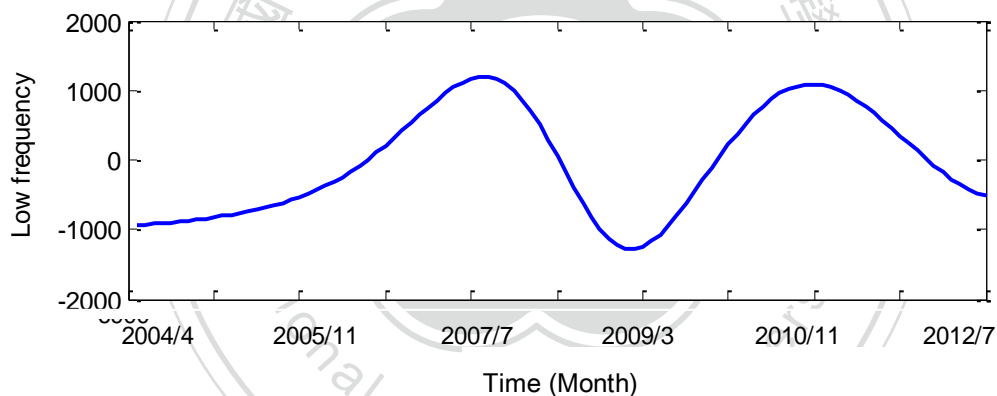

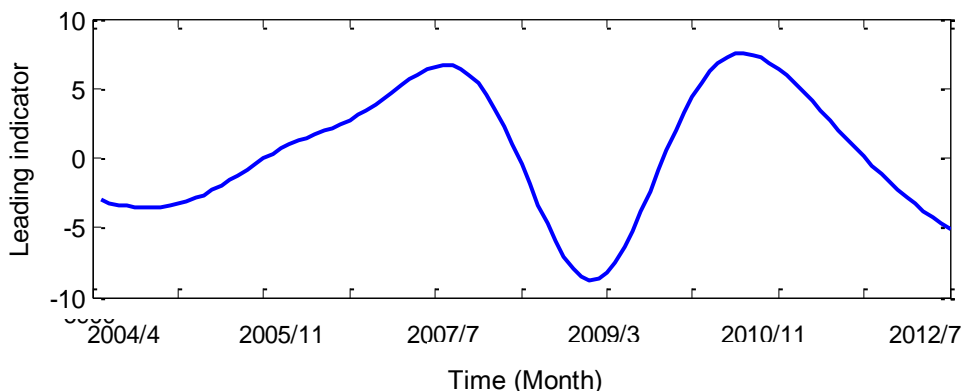Figure 3.11 Low Frequency term from 2004 April to 2012 July

Figure 3.12 Leading indicator from 2004 April to 2012 July

Table 3.9 the Correlation Coefficients Compared Leading indicator with TAIEX Index in terms of Low Frequency

Matches of TAIEX Index	Pearson correlation
Low Frequency	0.98

In addition, the mean period is around four years therefore statistics show the business circulates every four years in the U.S. The reason may be the presidential election in every four years but how the election impacts stock market? "The Federal Reserve and U.S. government tend to be more simulative in their monetary and fiscal policies building up to the presidential elections to support the present administration or party—and then have to deal with the excesses by use of tightening and cutbacks into the midterm elections, when there is less at stake. Investors also tend to get optimistic in the first year of a presidency, with so many promises having been made during the election, and then realize that not much will change by the second year. There may be other influences that we have not yet identified, but this is another consistent cycle." (The Great Depression Ahead, Harry S. Dent Jr.)

Furthermore, some suggest corporations adjust their inventory every four years therefore it may also serve as one of the causes of four year circulation (Kitchin, 1923). Council for Economic Planning and Development in Taiwan points out, the economics in Taiwan has circulated 10 times and the duration for each time as 56 months.

Trend

The long-term trend of index is considered to be influenced by the Real GDP. GDP refers to the market value of all officially recognized final goods and services produced within a country in a given period of time and is generally used to assess the economic performance of a certain country. As shown in Figure 3.10, the economic activities may be fluctuated in short terms yet long term speaking, the accumulation of capital and development of technology would increase GDP. The Figure 3.13 shows the comparison of real GDP and trend and it is observed that they have the same trend.
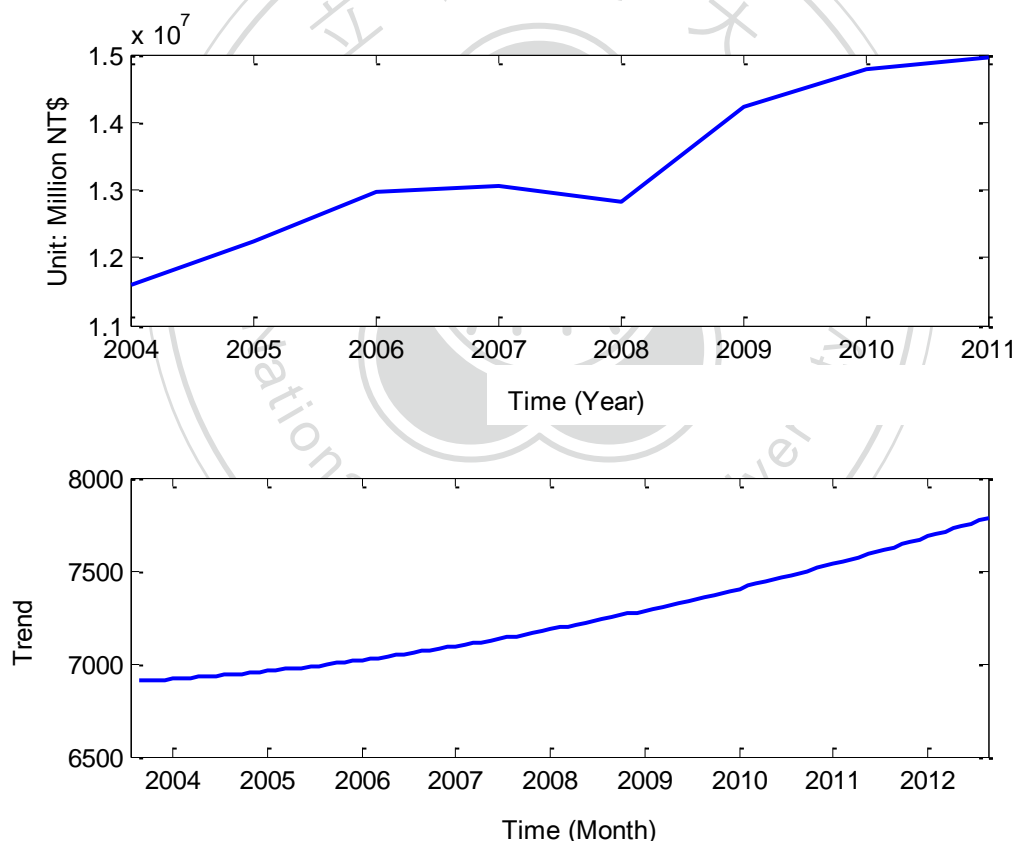

Figure 3.13 the Comparisons between GDP and Trend from 2004 to 2012

We also compares with GDP and Trend from 1991 to 2011. Long term speaking, the Figure 3.13 shows the comparison of low frequency term and leading indicator

and the Table 3.10 shows the correlation coefficient is 0.98 which suggests, trend and real GDP are closely related no matter whether for long term or short term. According to Table 3.5, the correlation coefficient between trend and real GDP is 0.41 and the power percentage is 4% which indicates trend has less effect on the TAIEX Index number from 2004 to 2012 and the most important is business circulations represented by low frequency term.
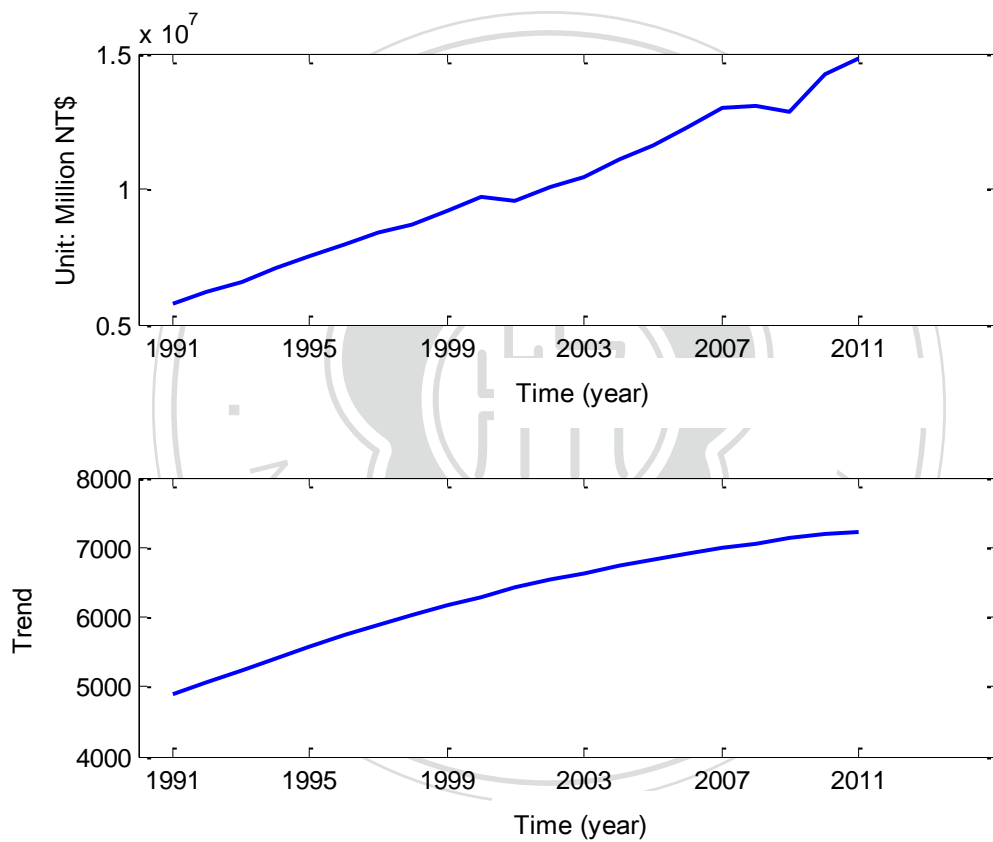

Figure 3.14 Comparisons between GDP and Trend from 1991 to 2011

Table 3.10 the Correlation Coefficients Compared Real GDP with TAIEX index in terms of Trend

Matches of TAIEX Index	Pearson correlation
Trend	0.98

3.4. Experiment design

In the experiment, we input the closing price of 60 days to train to model forecasting the closing price the next day and conduct a successive forecasting of stock price. Three models were used in the study – ARMA model and two different types of EEMD-ANN model. The forecasting accuracies of three models were compared. The forecasting strategy of EEMD-ANN model includes three steps as follows:

Step1: extract data

Except for there were only 13 business days in February, 2010 (the Chinese New Year came across), the business days in other months were 21-23 days with total of 250 business days. The closing prices of 60 days previous to the forecasting month were used as the beginning input and *Moving Window* was used to forecast each business day in the month. For instance, the closing prices from 2010/6/8 to 2010/8/31 are the inputs for the forecasting of the closing price of 2010/9/1.

Step2: EEMD and ANN training

After inputting the TAIEX Futures, the new data – IMFs were generated from the decomposition made by EEMD and IMFs were viewed as the factor of TAIEX Futures which were used as input data to construct EEMD-ANN Model 1 and EEMD-ANN Model 2.

a. EEMD-ANN Model 1

n+1 sets of IMFs (n sets of IMFs and residue) were produced by decomposing the TAIEX Futures with EEMD, every set of IMF was then input into the corresponding BPN models respectively. Each BPN model was distinguished by their target data. When IMF_m (m=1, 2…n+1) are input in BPN_m, The IMF_m discarded the first point was regarded as the target data. For BPN training, the inputs and target data were put into a BPN network and the discarded last points of IMFs were used as the input for

this trained network to make forecasting. In other words, the IMF of next minute was forecasted and the TAIEX Future could be retrieved by adding each IMF.
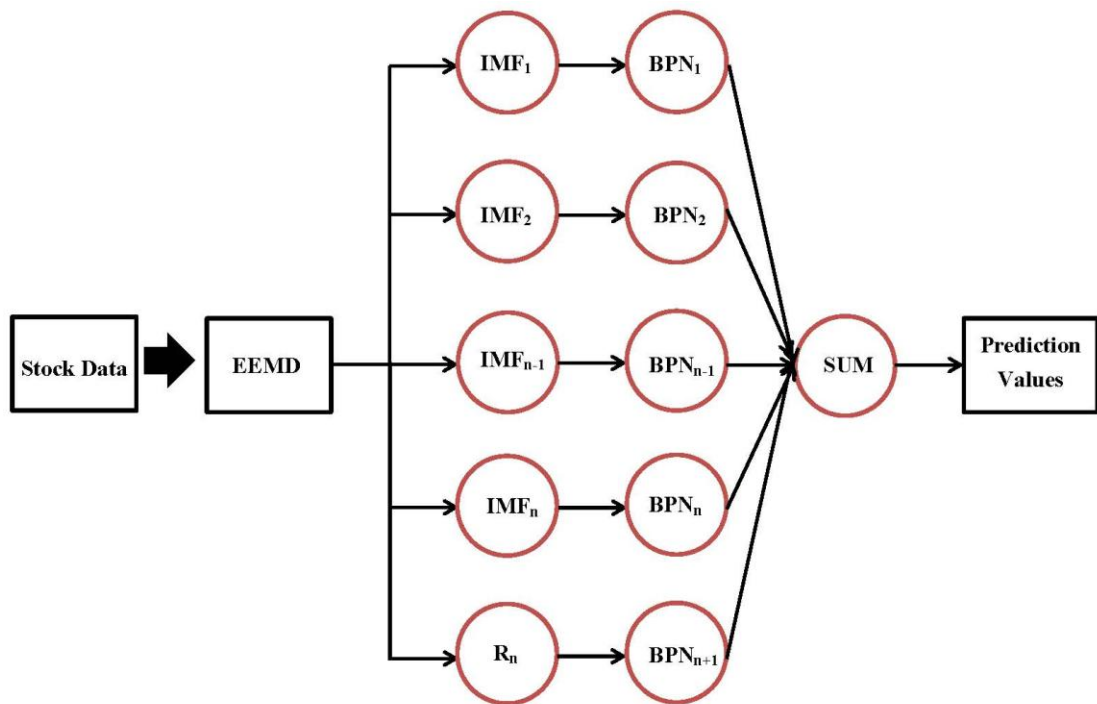

Figure 3.15 Flow Chart of EEMD-ANN Model 1

b. EEMD-ANN Model 2

Similarly, the procedure in model 2 started with decomposing TAIEX Futures by EEMD, n+1 sets of IMFs were generated while the different part of model 2 from model 1 was that all the IMF_m (m=1,2…n+1) were put into BPN at once and data trained were TAIEX Futures without being decomposing. The TAIEX Futures discarded at the starting point was regarded as the target data. In the training of BPN, inputs and target data were put into BPN network and the discarded last point of IMF_m (m=1,2…n+1) were used as inputs for the trained network to make forecasting. The point represented the forecasted value of TAIEX Futures.
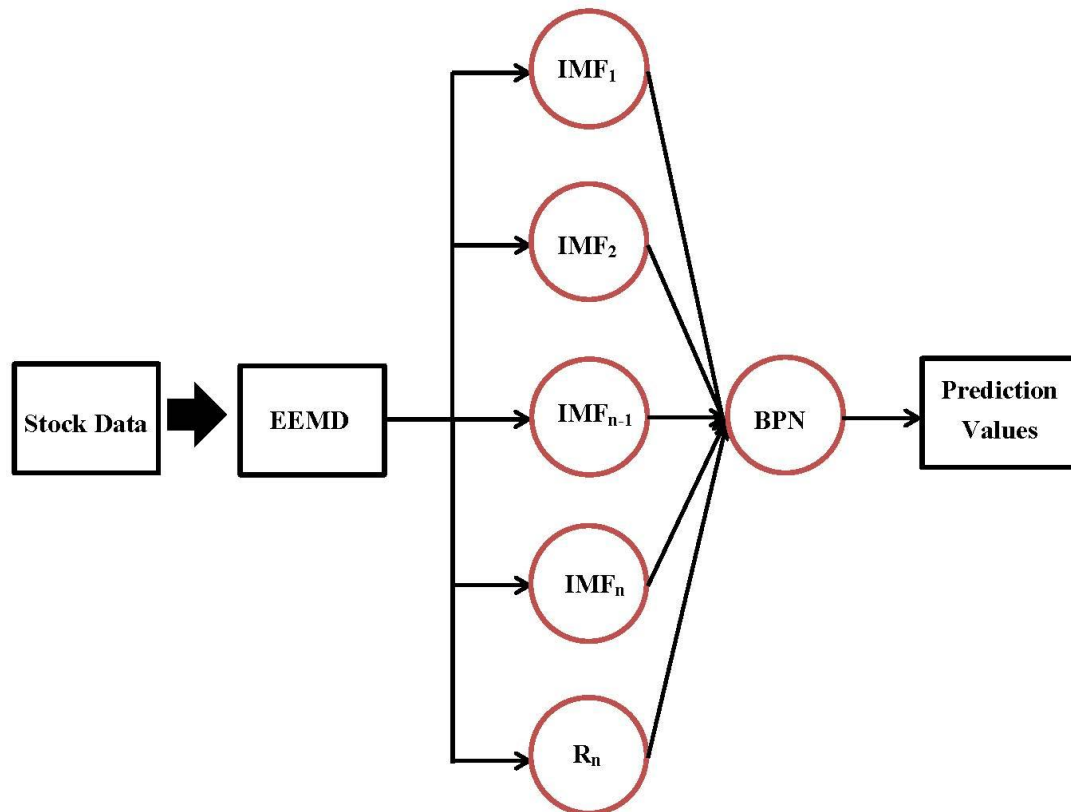
Figure 3.14 Flow chart of EEMD-ANN Model 2

Step3: Moving window and forecasting

The models used moving window concept to forecast the price of next minute. That was, the forecasted value could be the reference for trading in one minute. While the actual value of closing price was collected, it could be used to forecast the value of next day and so forth. The three steps repeated every day in 2010.

Figure 3.15 Training Network by Moving Window Process (Model 1)

Figure 3.16 Training Network by Moving Window Process (Model 2)
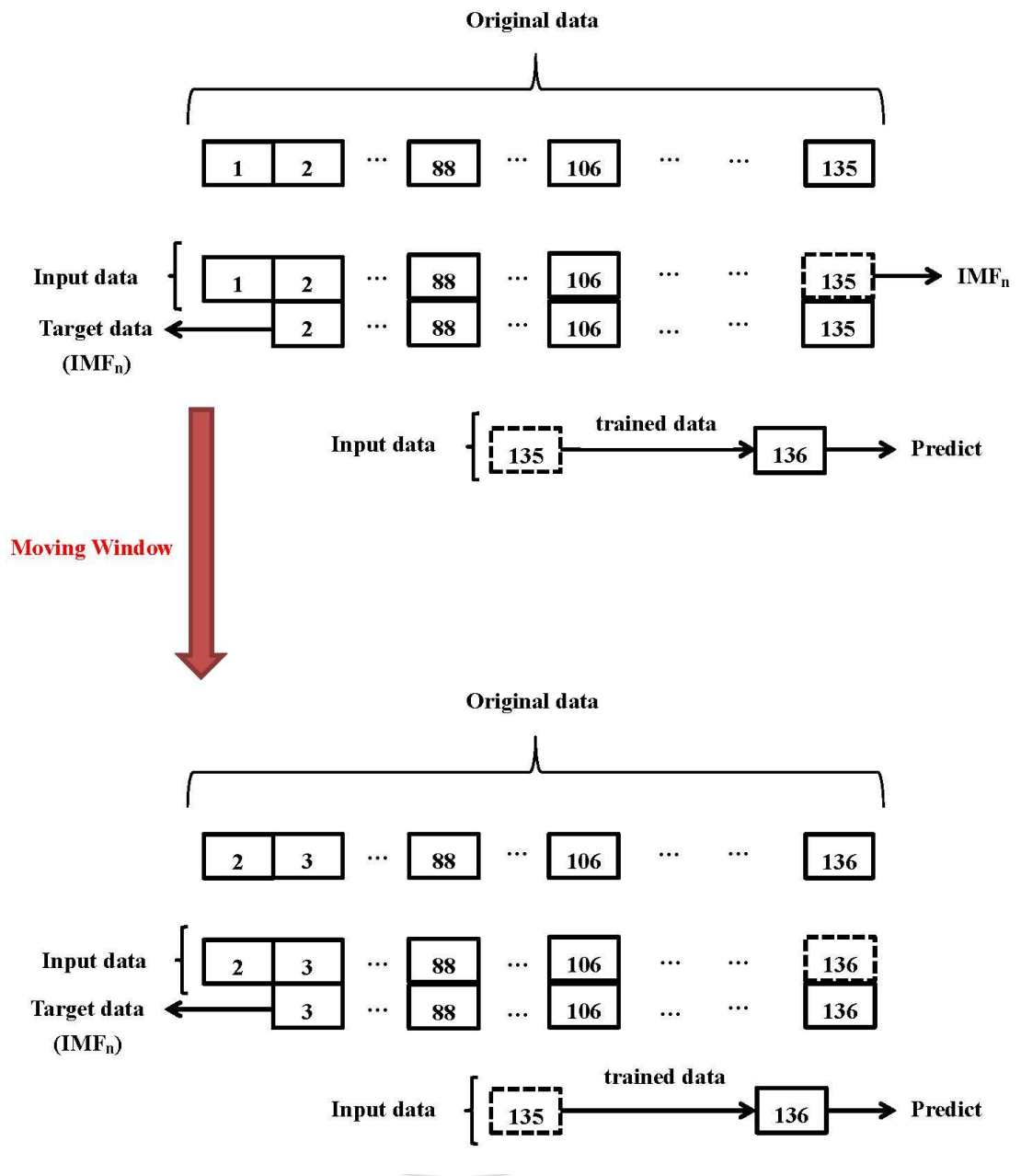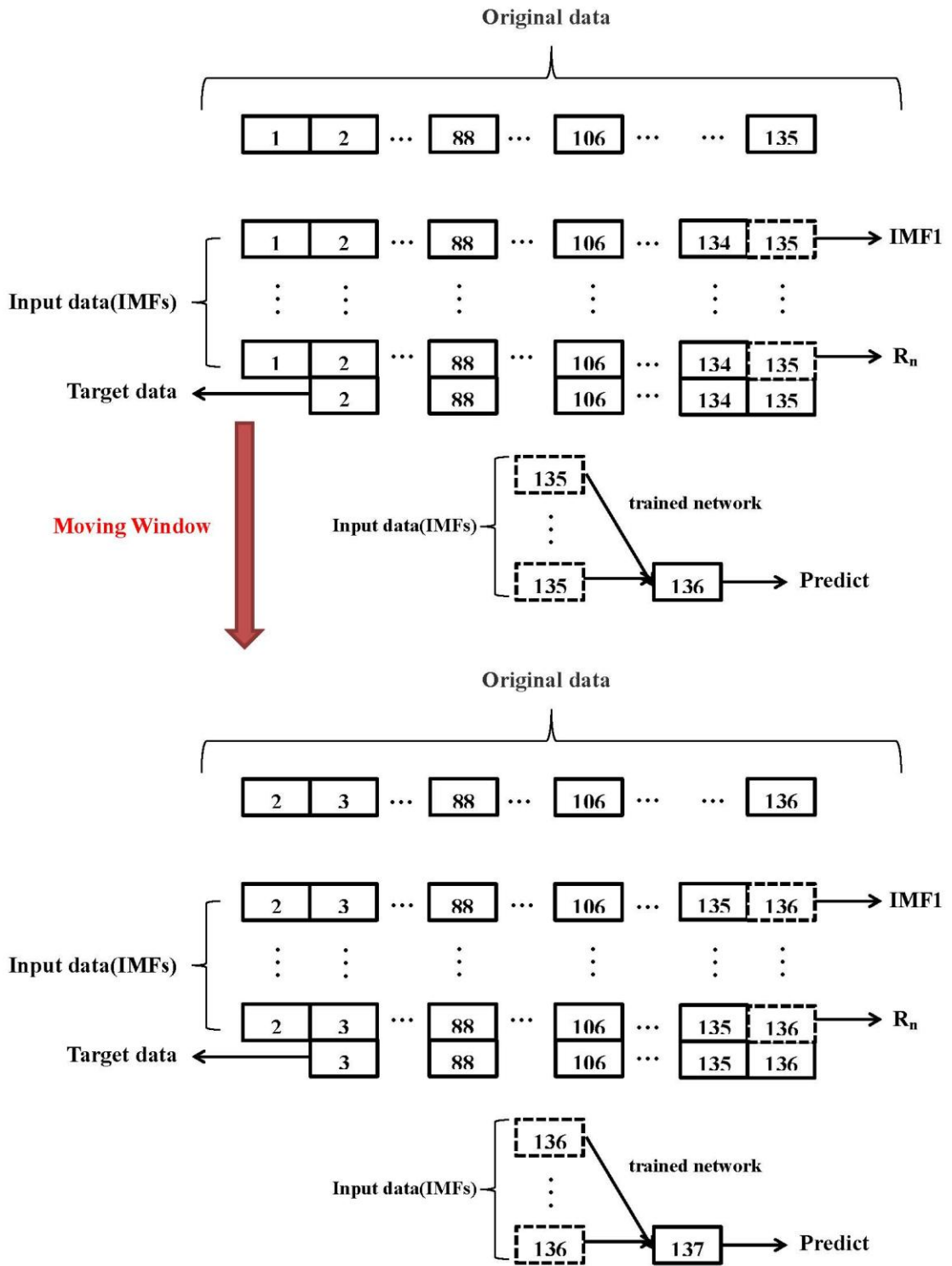
3.4. Performance

Forecasting results

The stock price fluctuates in a certain range every day therefore the forecasted value must fall within the range of fluctuation to be accurate. One of the most common charts in the market is called candlestick chart or K-chart which composed with opening price, closing price, highest price, and lowest price. Generally, the red line represents the rise of stock price (closing price higher than opening price) and green line represents the fall of stock price (closing price lower than opening price). Zone between opening and closing is call real body while the highest and lowest prices extended above and below from real body are called shadows. The trade is good and the forecast is accurate when the forecasted price falls between the highest and lowest prices of the minute which is very important because the forecasted value can only be used to establish the trading strategies when the trading is good. An example of K-chart is demonstrated as Figure 3.19:
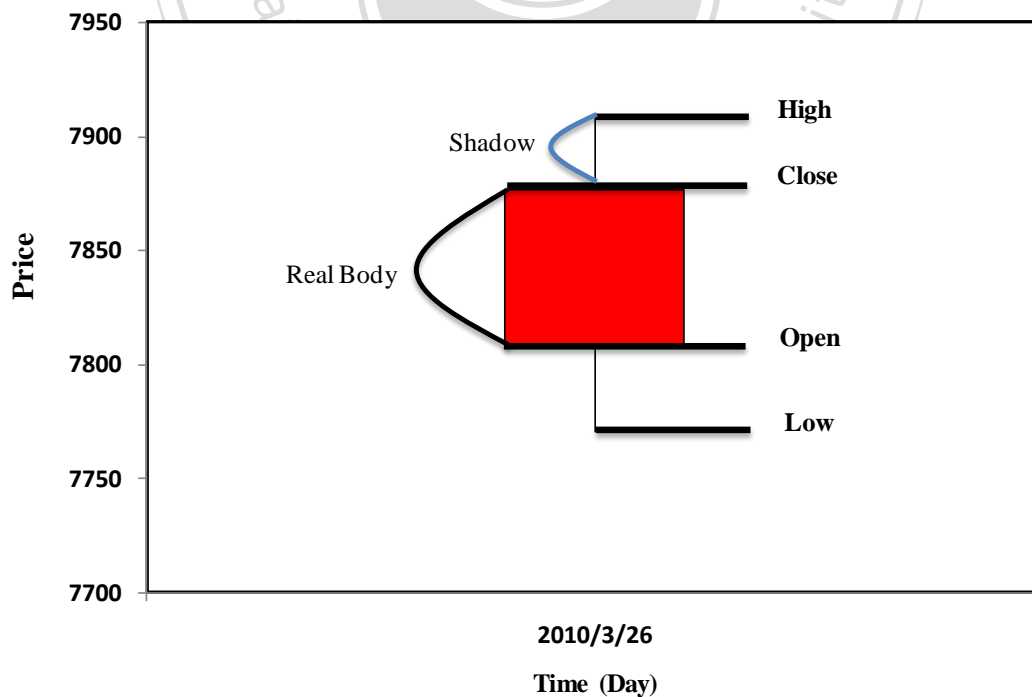
Figure 3.19 the Shape of a Candlestick

For example, the data of opening, closing, highest, and lowest prices are as follows:

Table 3.11 the Example of TAIEX Futures

Time	Open	High	Low	Close
2010/12/27	8867	8908	8829	8892
2010/12/28	8919	8919	8840	8871

When the closing price at 2010/12/7 is 8892, the TAIEX Future forecasted in next day will fall between 8919 and 8840 for the forecast to be accurate and the stock can be traded successfully as shown in the Figure 3.20:
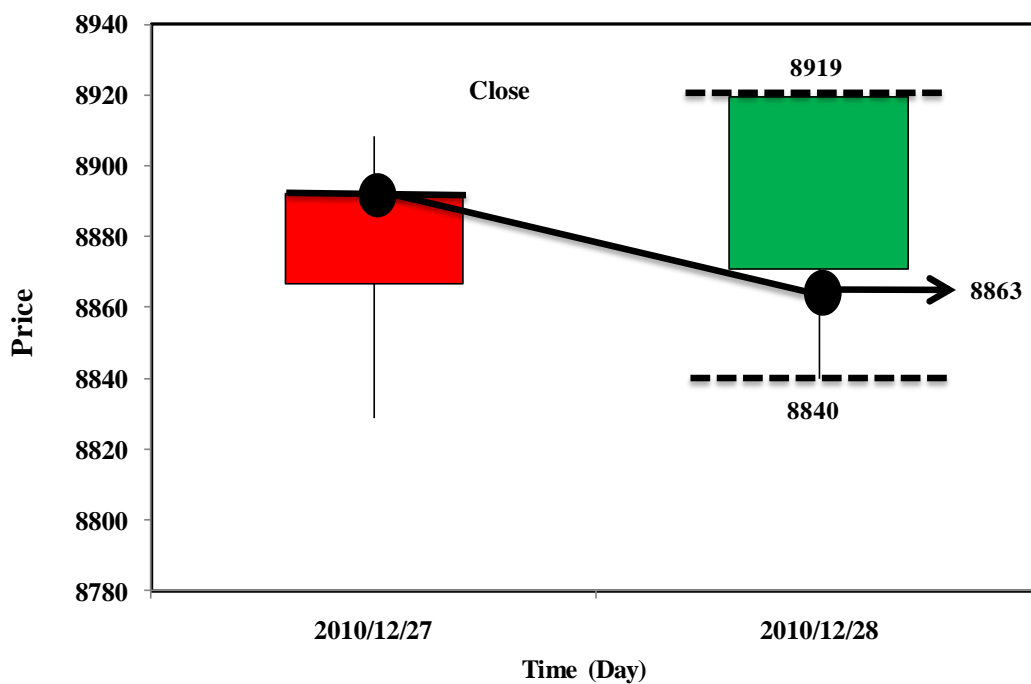
Figure 3.20 Trade Success

On the other hand, when the forecasted price falls outside of the range between highest and lowest prices – assuming the forecasted price is 8803 – it means the forecast is inaccurate and the trade is unable to be proceeded.
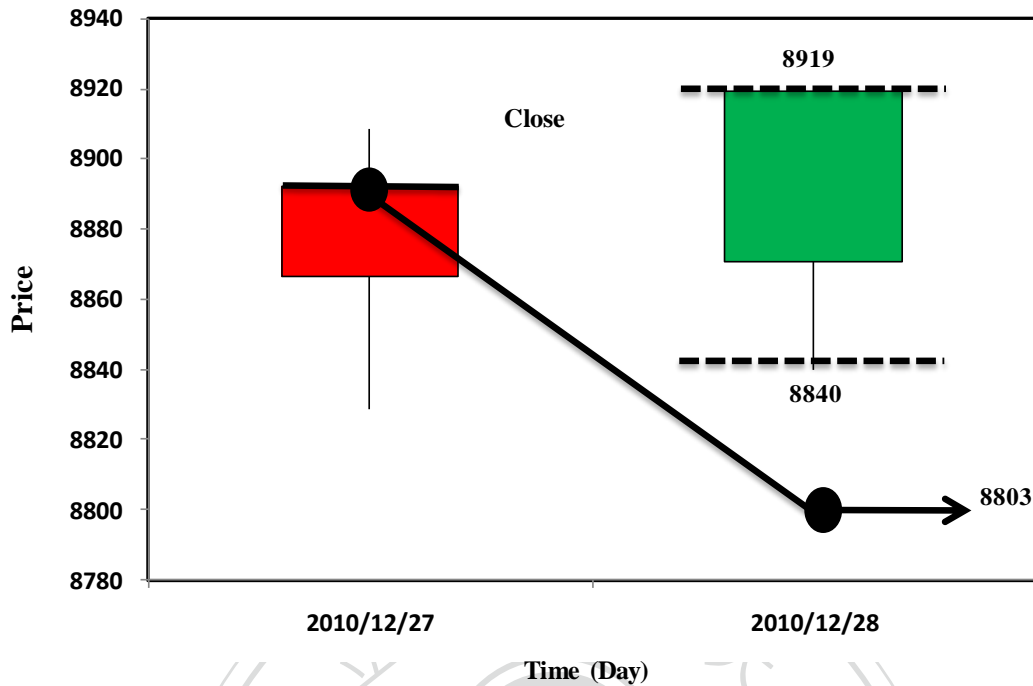
Figure 3.21 Trade Failure

By using the methods mentioned above, the accuracies of two models were examined by the formula shown below:

$$\textbf{Accuracy}\,(\%) = \frac{\text{accurate forecasts}}{\text{total forecasts}} * 100\%$$

Generally, there are two quotations in the stock market – quotation fluctuation and quotation trend. The business days in 2010 were divided into two parts – first half from January to June and latter half from July to December. The Figure 3.22 & 3.23 shows the market fluctuated more acutely in the first half year of 2010 (quotation fluctuation) while the fluctuation from July to December were milder (quotation trend). The forecasting accuracies of three models in these two quotations were compared.

Figure 3.22 Fluctuations from January to June
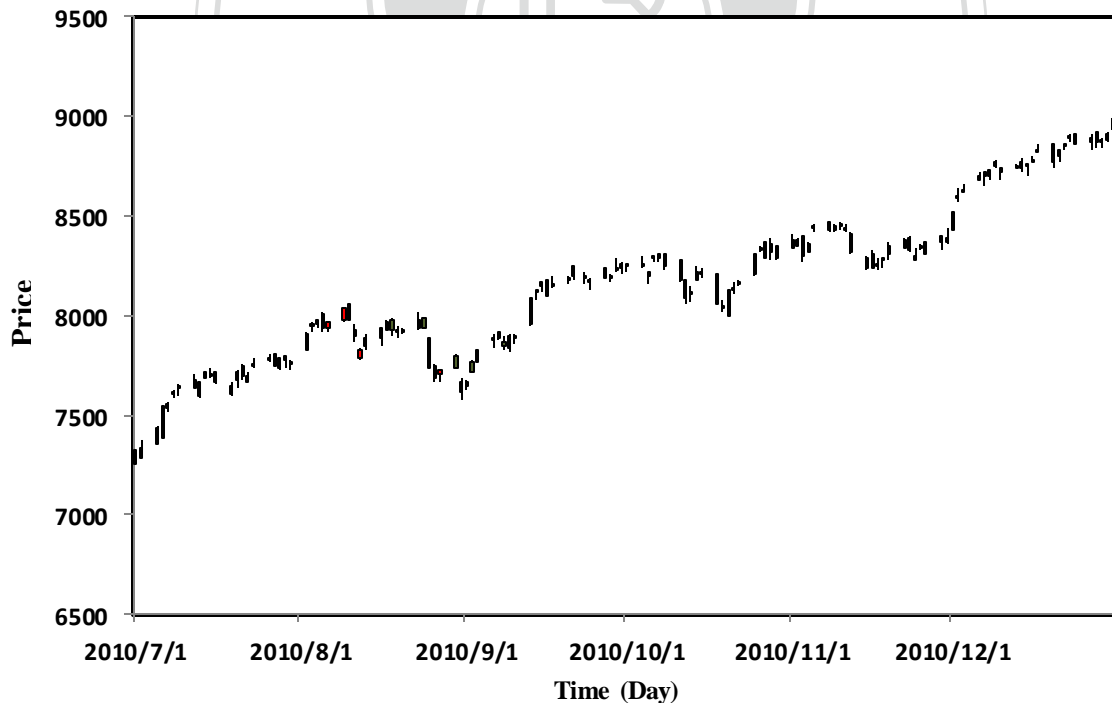
Figure 3.23 Trends from July to December

Table 3.12 demonstrates the forecasting accuracy for the first half year, it is observed that Model 1 (ε=0.3) has the accuracy are of 52% which is superior to ARMA and the other model. Table 3.13 shows the forecasting accuracy for the latter half year and Model 1 (ε=0.3) has the outstanding accuracy rate of 53% f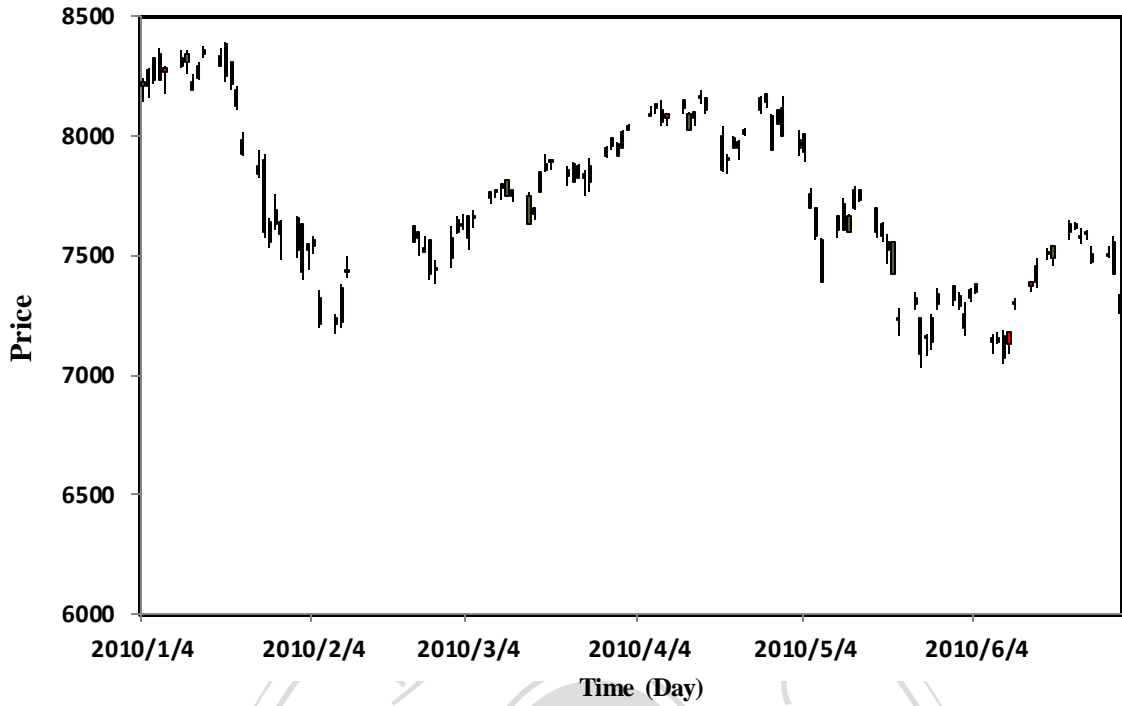rom all other models. After further examination, it is found that ARMA has accuracy rate of 46% for the first half year while the number drops to 38% in the latter half year which suggests ARMA model is more suitable for forecasting the quotation fluctuation while EEMD-ANN (Model 1) plays excellent forecasting role in both quotation fluctuation and trend.

Table 3.12 the Forecasting Accuracy for the First Half Year

Condition	ARMA	ε=0.1		ε=0.2		ε=0.3	
		Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
Month	Accuracy (%)						
Jan	55%	40%	50%	60%	60%	60%	50%
Feb	62%	46%	23%	62%	38%	62%	46%
Mar	56%	48%	35%	57%	52%	61%	43%
Apr	33%	33%	29%	48%	48%	48%	38%
May	28%	33%	38%	38%	43%	38%	38%
Jun	43%	38%	43%	38%	43%	38%	48%
Average Accuracy (%)	46%	40%	36%	51%	47%	52%	44%

Table 3.13 the Forecasting Accuracy for the Latter Half Year

Condition	ARMA	ε=0.1		ε=0.2		ε=0.3	
		Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
Month		Accuracy (%)					
Jul	59%	46%	45%	32%	59%	36%	50%
Aug	32%	50%	45%	59%	45%	50%	45%
Sep	38%	52%	48%	52%	29%	67%	33%
Oct	48%	33%	43%	43%	38%	52%	57%
Nov	41%	59%	36%	59%	32%	59%	45%
Dec	9%	39%	31%	65%	35%	52%	39%
Average Accuracy (%)	**38%**	**47%**	**41%**	**52%**	**40%**	**53%**	**45%**

The forecasting accuracy of each month was further compared. The market fluctuation from the highest to lowest prices daily and the fluctuation value is the highest price minus lowest price in the business day. The average fluctuation range is to add up the fluctuation value of every business day in the month and divide with the days.

$$\text{average fluctuation range} = \frac{\sum_1^n High_n - Low_n}{n}$$

It is defined in the study that accuracy is when forecasting price falls between the highest and lowest value of actual price. Therefore, it is easier for the model to make forecast when the fluctuation range is wide and vice versa. Table 3.14 shows the average fluctuation rage for each month, it is shown that fluctuation values are higher

in January, February, and May while the accuracy of Model 1(ε=0.2, ε=0.3) is exceptional. ARMA model also has good accuracy rate in January and February. Nonetheless, the forecast made by Model 2 misses completely. September, November, and December have relatively mild fluctuations therefore it is more difficult to product accurate forecasts. It is shown that Model 1 has same exceptional forecasting accurate rate in all three months; Model 2 has average forecasting accuracy; while ARMA model shows a miserable accurate rate. Especially in December, when the average fluctuation value is 57, EEMD-ANN Model maintains good forecasting accuracy while ARMA Model has a total miss.

In accordance with previous analysis, EEMD-ANN Model tends to produce more accurate forecasts in any quotations; EEMD-ANN Model 2 has average accurate rate in most quotations; while ARMA Model demonstrates a highly unstable forecasting accurate rate and only produces acceptable forecasting results in quotation fluctuation.

Table 3.14 Average Fluctuation Range for Each Month

Month	Average fluctuation range
Jan	121
Feb	123
Mar	78
Apr	78
May	112
Jun	77
Jul	67
Aug	74
Sep	64
Oct	76
Nov	63
Dec	57

Chapter 4 Algorithmic Trading

Algorithmic trading refers to the use of automation for trading in financial markets. Once trading opportunities are presented, the trading can be automatically executed via a computer. The use of algorithmic trading was most commonly used by large institutional investors in order to capture profit opportunities or hedge a portfolio.

In this research, we uses ARMA model and two different types of EEMD-ANN model to decide the price, we decomposed the TAIEX futures for the period of 2010 for 60 days previous to the forecasting month as the input of our test, and time delay for one day as the target. Moreover, we use two different forms of trading strategy to decide the order timing and order quantity. Algorithmic Trading can remove human emotion element in trading to order accurately.

4.1. Trading Strategy I

Correct Trend (%)

The Definition of Direction Ratio

$$\text{Correct Trend(\%)} = \frac{\text{Correct signal}}{\text{Total point}} * 100$$

The correct signal is when the stock price is forecasted to rise in the next day, and the forecast turned out to be correct. On the other hand, incorrect signal is if the stock price is forecasted to rise but it actually drops.

Table 4.1 shows the correct trend (%) of every month. Model 1(ε=0.3) demonstrates the best correct signal with correct trend (%) of 60%; Model 1(ε=0.2) has correct trend (%) of 58% as the second best; Model 1(ε=0.1) has the correct trend (%) of 57%; while ARMA Model possesses the correct trend of 52%. As for Model 2 (ε=0.1、ε=0.2、ε=0.3), the correct trends are 50%, 49%, and 51% respectively which

are less satisfactory.

Table 4.1 the Correct Trend (%) of Every Month

Condition	ARMA	ε=0.1		ε=0.2		ε=0.3	
		Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
Month		**Correct Trend** (%)					
Jan	60%	65%	55%	45%	55%	55%	55%
Feb	62%	69%	46%	62%	46%	62%	46%
Mar	43%	43%	52%	52%	52%	48%	52%
Apr	52%	57%	48%	57%	52%	48%	52%
May	52%	42%	57%	52%	57%	57%	57%
Jun	43%	57%	52%	48%	48%	52%	57%
Jul	59%	50%	55%	64%	55%	64%	59%
Aug	50%	59%	64%	64%	55%	56%	45%
Sep	62%	57%	52%	52%	38%	71%	43%
Oct	67%	57%	33%	67%	47%	62%	57%
Nov	50%	68%	56%	64%	50%	64%	59%
Dec	26%	61%	30%	74%	35%	74%	35%
Sum	**52%**	**57%**	**50%**	**58%**	**49%**	**60%**	**51%**

A simple trading strategy was tested in the study. The inputs are the daily closing price therefore the forecasting value produced by ARMA and EEMD-ANN models are the possible closing price for the next day. Hence, the trading strategy of buy-in would be executed when the forecasting value of the next day is higher than the actual closing price while the sell-out strategy will be executed when the forecasting value of the next day is lower than the actual closing price on the certain day. In this study, a "buy" signal (value = 1) will be output by the system when the forecast suggest a rise of price, a "sell" signal (value = -1) will appear when the forecasting value is lower than the actual closing price of the previous day, while a "hold" signal (value = 0) will be shown when encountering other situations. To avoid complicating the trading strategy, there will be only one time buy in or sellout. The position is hold until the

contrary signal turns out and the position will be close when the contrary signal appears. For instance, the trading chart of 2010/9 is shown Figure 4.1. The closing price of 2010/9/1 is predicted to be higher than the actual closing price of 2010/8/31 therefore a buy signal is putout. The stock was bought in during the closing of 2010/8/31 and the forecasting on 2010/9/2 maintained to be buy in therefore the position was held until the close of 2010/9/2 when the forecasting for the closing price of 2010/9/3 showed that the closing price of 2010/9/3 was lower than the actual closing price of 2010/9/2. The forecasting result is viewed as the sell signal therefore the position was sold out during the close of 2010/9/2. The trading strategy was repeated until the end of this month.
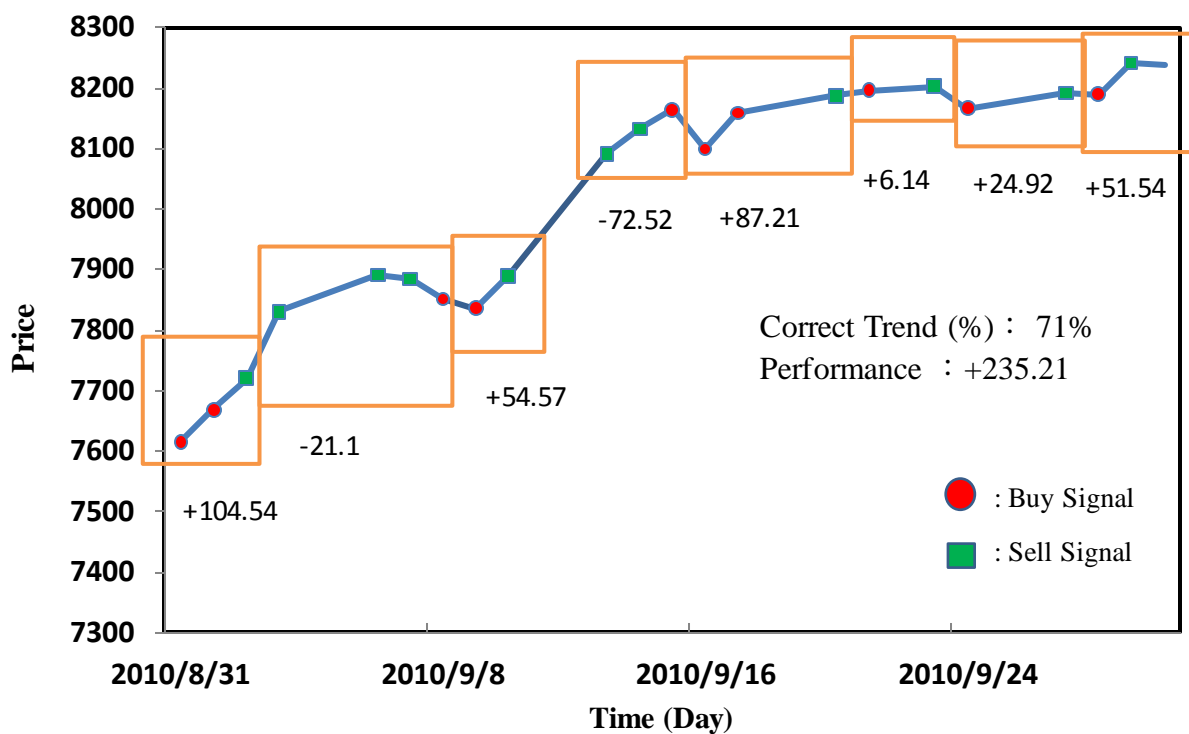

Figure 4.1 the Trading Chart of 2010 September

Performance

It is mentioned in Chapter 3 that the operation of price is determined by four types of uncertainty. The first one is the uncertainty of price fluctuation. Trading strategy I

attempts to assure whether the price will go up or down to determine the action of sell or buy. Generally, the more accurate the forecasting is, the better estimation results will be. However, it is not always the case because the results are not simply influenced by the uncertainty of price fluctuation. Assume the forecasted direction is correct, we may profit only a bit before our position ends or we may profit greatly yet the comeback of price in the end causes the profit. Both cases are caused by the uncertainty of fluctuation range. Therefore, the investing profit will be improved if the model is able to forecast the trend of price. Next, the forecasting of price trend by these three models will be discussed. The Table 4.2 demonstrates the profit situation of trading strategy 1 in 2010.

Table 4.2 the Profit of Trading Strategy I in 2010

Condition	ARMA	$\varepsilon = 0.1$		$\varepsilon = 0.2$		$\varepsilon = 0.3$	
		Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
Month		Trading Strategy I					
Jan	-123.67	-126.76	528.75	-375.04	333.06	-294.63	149.77
Feb	14.48	384.65	-336.13	239.49	-109.48	608.66	-68.27
Mar	-365.2	-28.99	193.89	399.25	65.45	305.11	-176
Apr	-200.27	340.48	-222.22	304.02	-201.88	65.5	83.36
May	198.12	-563.51	71.49	-141.28	91.28	-324.16	240.73
Jun	-64.85	72.95	160.25	-360.5	65.51	-132.68	450.25
Jul	53.31	8.29	-207.67	64.61	206.44	18.39	30.95
Aug	144.35	98.99	209.29	384.18	178.35	310.62	51.27
Sep	79.79	71.92	-78.71	-230.01	-416.97	235.21	-384.93
Oct	310.08	-83.22	-287.69	-66.04	-265.89	382.5	26.5
Nov	-141.7	245.26	86.52	94.71	-52.83	116.5	199.16
Dec	-600.02	316.25	-540.95	385.61	-504.22	434.25	-505.01
Sum	-695.58	736.31	-423.18	699	-611.18	1725.27	97.78

Usually, the higher correct trend (%) is the better profit will be in most months and vice visa. It is also observed that, the profit may vary even if the correct trends (%) are identical in the same month because of the different forecasting points of the models. For instance, the Model 1 (ε=0.3) and ARMA have the same correct trends (%) of 62% in February. Model 1 (ε=0.3) profits greatly with total amount of 608.66 while ARMA profits only 14.48. In addition, it is also found the profit of Model 2 is higher than Model 1 and ARMA Model in January and May even if its correct trend is relatively lower than the other two which suggests it possesses better trend forecasting capability. Generally, Model 1(ε=0.3) has total profit of 1725.27 in 2010 and the best ratio of correct trend (%) and the next is Model 1 (ε=0.2、ε=0.3). ARMA Model possesses the correct trend (%) of 52% which is higher than Model 2 yet its profit is less than EEMD-ANN Model 2. Such phenomenon suggests that Model 2 possesses advanced forecasting capability to correctly estimate the trend of price.

An example is shown in the Figure 4.2 & 4.3, Model 1 made a forecast of price dropping on 2010/2/5 therefore the short position was taken and the price in fact dropped 324 the next day so it was a great gain. However, ARMA Model signaled buy in on the day and made another wrong suggestion on the next day caused great loss.
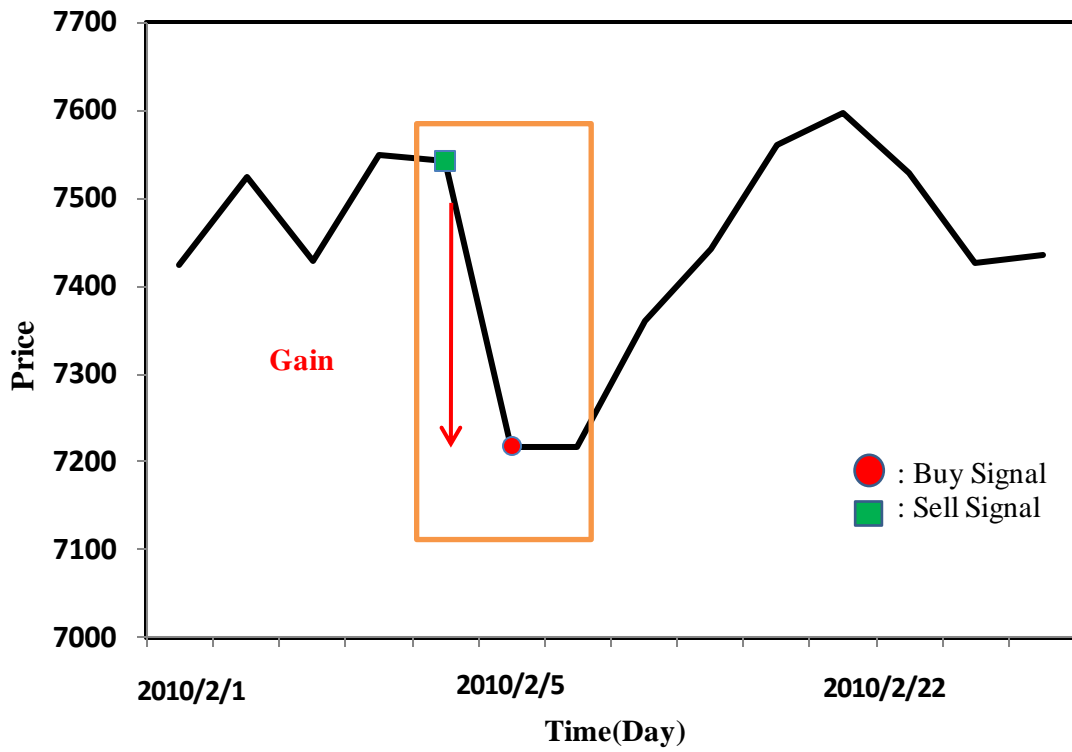
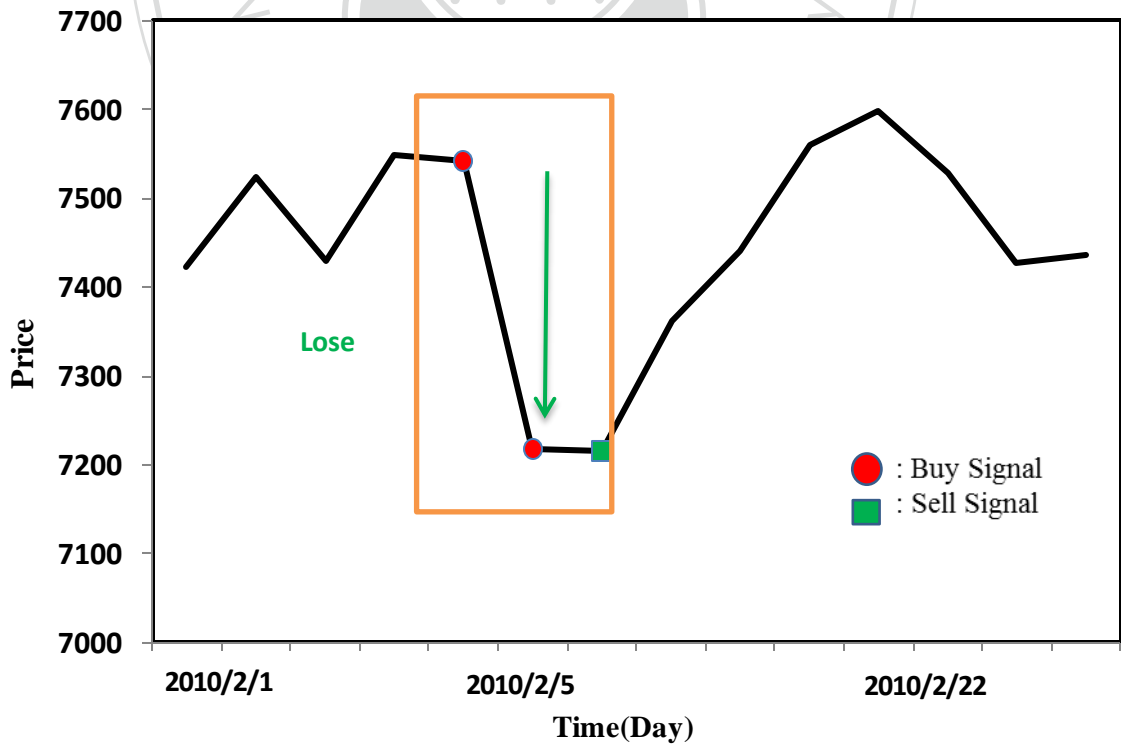Figure 4.2 Forecast of Price Dropping on 2010/2/5 (Model 1)

Figure 4.3 Forecast of Price Rising on 2010/2/5 (ARMA)

4.2. Trading Strategy II

The Moving Average Indicator and stop-loss point was added in Trading Strategy 2 to optimize the previous strategy. The main reason for using moving average indicator is to determine whether to buy low and sell high or reverse the order and sell high while Stop-loss point is to prevent the wide loss caused by false-forecasting to gain better profits.

Moving Average Indicator

Granville presents the Moving Average Indicator in 1962. It is the most popular technical indicator used to analyze time series data. The equation is expressed as follows: (for j=1 to n)

$$MA = \frac{\sum_{j}^{n} y_j}{n}$$

n is a time period. y_j is the nth day's previous closing prices.

In short, it represents the average stock price within a specific time frame. Where did the concept of "moving" comes from? If there is a new data into the sum, an old data drops out. The moving average either increases, or decreases.

We propose a definition of bull and bear markets. If the stock price is above the moving average, it is a bull market. If the stock price is below the moving average, it is a bear market. A good trading strategy is to hold long position (buy low, sell high) in the bull market and hold short position (sell high, buy low) in the bear market. It is because the stock price tends to be rise rather than drop during the bull market and the risk may rise when maintaining the short position. If the stock price happens to rise while the trading system signals sell high, the loss will be great and vice versa.
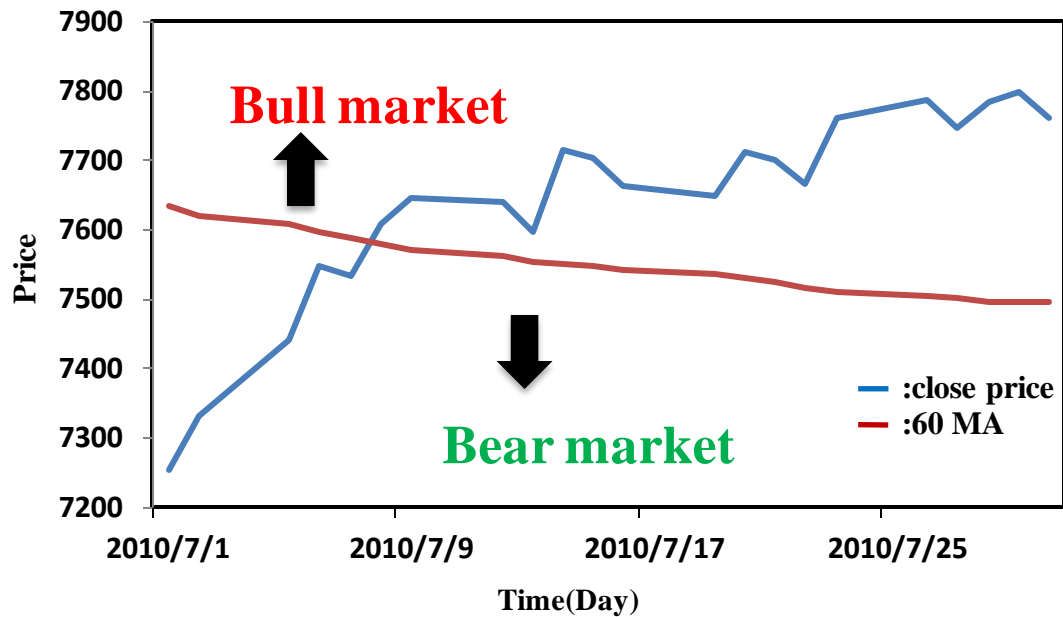
Figure 4.4 Bull and Bear Market

Stop-loss point

The setup of stop-loss point is to prevent wide loss caused by false-forecasting. There are two stop-loss points: one is when the gap of futures surpasses a certain loss value and the other is when the loss achieves a certain value during intraday. Under both circumstances, the positions are chose to be ended. In the study, the loss-stop point is set as 50 and it is because the average gap point in the stock market in Taiwan is 45. In other words, if the number is set less than 45, it may be too frequently to meet the stop-loss point therefore the stop-loss point is set as 50 which is higher than 45.

At Figure 4.5 shows the first stop-loss method. Model 2 signaled sell out in 2010/7/7 because the forecast suggest the drop of price in 2010/7/8. Therefore, short position was taken at the closing of 2010/7/7 with the price of 7534. Unfortunately, the opening price on 2010/7/8 was 7613 with rise of 169. A stop-loss strategy was immediately implemented and the dotted line in the figure below was the stop-loss
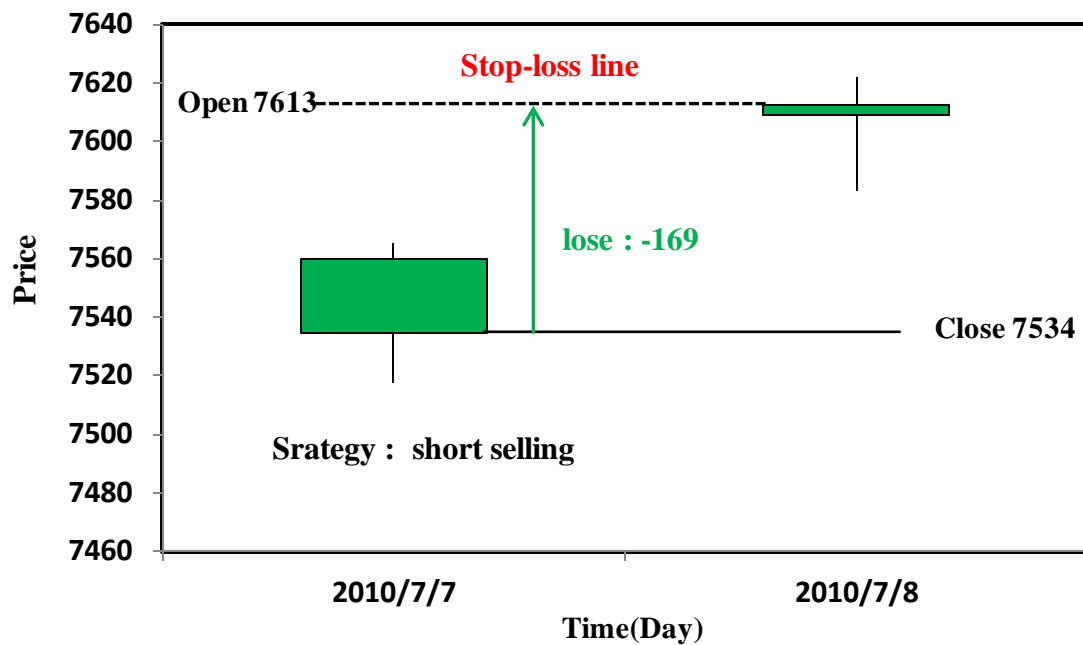
line.

Figure 4.5 the First Stop-loss Method

The second stop-loss strategy is shown in the Figure 4.6. Model 2 signed sell out on 2010/7/2 because it was forecasted to be drop on 2010/7/5. Therefore, short position was taken at the closing price of 7331 on 2010/7/2 yet the opening price of 2010/7/5 didn't reach the stop-loss value so the position was held. Nonetheless, the price rose greatly to 7381 – the stop-loss point in the day, so the position was closed to prevent further loss.
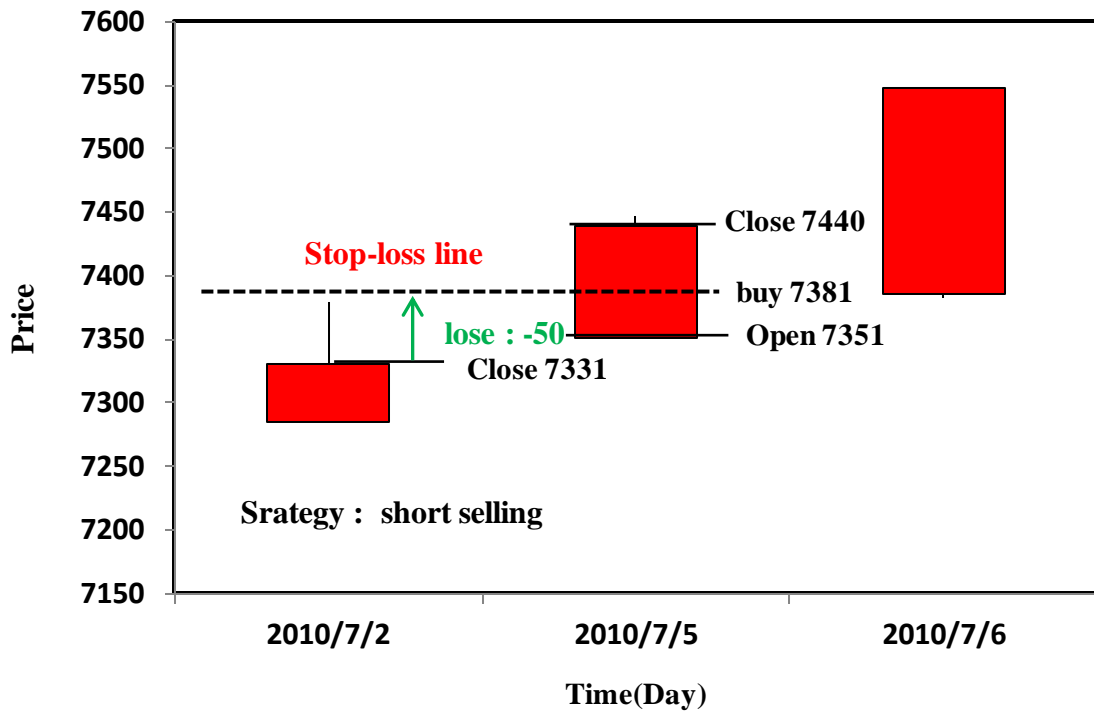
Figure 4.6 the Second Stop-loss Method

Performance

It is mentioned in Chapter 3 that the stock price may be influenced by unexpected incidents. Therefore, the trading strategies must to be able to prevent such great fluctuation caused by unexpected incidents to have better performances so that the ultimate goal of profit can be reached.

Table 4.3 the Profit of Trading Strategy Ⅱ in 2010

Condition	ARMA	ε =0.1		ε =0.2		ε =0.3	
		Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
Month	Trading Strategy II						
Jan	220.83	-8.7	47.06	5.28	-11.38	139.83	52.22
Feb	1.95	415.18	-44.06	121.02	21.25	310.52	-50
Mar	-139.29	-86.17	-45.34	22.67	59.33	79.63	-50.28
Apr	278.23	175.17	245.23	84.86	91.51	86.31	175.17
May	288.98	115.56	146.25	106.58	272.33	28.14	147.18
Jun	-44.16	184.99	38.81	63.79	238.39	190.05	168.34
Jul	-2.94	74.81	-51	-26.39	-4.28	3.84	-169.89
Aug	0	282.43	152.62	209.51	-40.25	162.52	-53.71
Sep	362.62	234.76	241.58	109.41	84.82	313.06	132.03
Oct	300.54	118.55	-63.31	334.62	-100	231.44	69.41
Nov	-58.67	150.84	171.91	282.75	57.96	130.67	137.76
Dec	0	458.84	59.07	387.46	3.79	404.24	94.77
Sum	1208.09	2116.26	898.82	1701.56	673.47	2080.25	653

Chapter 5 Conclusion

In this study, we used EEMD-base ANN learning algorithm to forecast TAIEX Futures per day. The results show the future price that forecasted by EEMD-ANN Model1 would produce better outcome than ARMA Model and EEMD-ANN Model 2 that makes overall forecasts. Two trading strategies were tested and it was found EEMD-ANN Model 1 in fact had better forecasting. Moreover, the trend forecasting ability of EEMD-ANN Model is also better than ARMA Model which could serve as a wonderful reference for future researchers.

In addition, EEMD is able to decompose the factors of TAIEX Index; it is also found that the most crucial factor of TAIEX Index is booming cycle in low frequency term, annual reports in mid frequency term, which also reflects the January Effect and incidents, while high frequency term is related to the Trading Value of Foreign & Other Investors. The organizational investors would change their position in accordance with the quarterly reports and the overall trend of TAIEX is affected by Real GDP.

References

Abu-Mostafa Y. S. and Atiya A. F., 1996. Introduction to Financial Forecasting, Applied Intelligence, 6: 205-213.

Yoon Youngohc., Swales, George, 1991. Predicting Stock Price Performance: A Neural Network Approach, Proceedings of the Twenty-Fourth Annual Hawaii International Conference on System.

Kamijo, K. and Tanigawa, T., 1993. Stock price pattern recognition: a recurrent neural network approach, in Trippi, R. and Truban, E. (eds), Neural Networks in Finance and Investing.

Kuan, C. M. White, H., 1994. Artificial neural networks: An econometric perspective, Econometric Reviews, 13, 1-91.

Wood, Douglas and Bhaskar.Dasqupta, 1994. Modelling and Index of the French Capital Market, Economic and Financial Computing, Autumn/Winter, pp.119-136

En Tzu Li, 2011. TAIEX Option Trading by using EEMD-based Neural Network Learning Paradigm. Master Thesis of Graduate Institute of Applied Physics, College of Science NCCU.

Zheng-Hsiu Chu, 2004. On Study of The Relationship between Taiwan Stock Market and The International Stock Markets. Master Thesis of Department of Statistics, College of management NCKU.

D. E. Rumelhart and J. L. McClelland, 1986. Parallel Distributed Processing：Explorations in the Microstructure of Cognition, Vol. 1, MA: MITPress.

H. A. Rowley, S. Baluja, and T. Kanade, 1998. Neural network-based face detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 20(1), January 1998.

Kaastra, I., Boyd, M., 1996. Designing a neural network for forecasting fi nancial and economic time series. Neurocomputing, 10, 215-236.

M. T. Hagan, H. B. Demuth, Neural Network Design, Thomson Learning, 1996.
Stent, G. S., 1973. A Physiological Mechanism for Hebb's Postulate of Learning. PNAS 70 (4), 997–1001.

Hush, D. and Horne, B.1993. Progress in supervised neural networks. *IEEE Signal Processing Magazine*, 10(1):8-39.

Gately, E., 1996.Neural Networks for Financial Forecasting. John Wiley, New York

Martin F. and Aguado J.A., 2003.Wavelet-based ANN approach for transmission line protection, IEEE Trans. Power Delivery, vol.18, no.4, pp.1572-1574.

Nayak, P.C., Sudheer, K.P., Rangan, D.M. and Ramasastri, K.S., 2004. Aneuro-fuzzy computing technique for modeling hydrological time series, Journal of Hydrology, 291(1-2): 52-66.

Joseph E. Granville, 1960. A Strategy of Daily Stock Market Timing for Maximum Profit. Englewood Cliffs, N. J.: Prentice-Hall, p.155.

Kitchin, Joseph., 1923. Cycles and Trends in Economic Factors, Review of Economics and Statistics, The MIT Press, 5 (1), pp. 10-16.

Harry S. Dent Jr., 2008. The Great Depression Ahead: How to Prosper in the Crash Following the Greatest Boom in History (New York: Free Press, 2008), at17-39.

Siegel, J.J., "Stocks for the Long Run", 4nd ed., New York: McGraw-Hill, 1998.

Chang-Hsu Liu, 2010. Calendar Anomalies: A Comparative Study of International Equity Markets.

Jasemi, M., Kimiagari, A. M. and Memariani, A., 2011. A modern neural network model to do stock market timing on the basis of the ancient investment technique of Japanese Candlestick. Expert Systems with Applications 38, 3884–3890.