國立交通大學

資訊管理研究所

碩 士 論 文

隔夜效應之發現於台灣股價指數期貨之研究

A Study of Overnight Effect Mining on Taiwan Futures Market

研 究 生：李佩真

指導教授：陳安斌　教授

中 華 民 國 九 十 三 年 六 月

隔夜效應之發現於台灣股價指數期貨之研究

A Study of Overnight Effect Mining on Taiwan Futures Market

研 究 生：李佩真　　　　Student：Pei-Chen Lee

指導教授：陳安斌　　　　Advisor：An-Pin Chen

國 立 交 通 大 學

資 訊 管 理 研 究 所

碩 士 論 文

A Thesis

Submitted to Institute of Information Management

College of Management

National Chiao Tung University

in partial Fulfillment of the Requirements

for the Degree of

Master

in

Information Management

June 2004

Hsinchu, Taiwan, Republic of China

中華民國九十三年六月

隔夜效應之發現於台灣股價指數期貨之研究

學生：李佩真　　　　　　　　　　　　　　　　　　指導教授：陳安斌博士

國立交通大學資訊管理研究所碩士班

摘　　要

股市是數以千計的市場投資者買賣供需行為的結果，而每位投資者所做的買賣決策都受到各種因素的影響，例如基本面、總體經濟面及消息面等，即市場所有的資訊都會反映在價格上。這些資訊的影響都是連續不斷的，然而交易時間的限制使得股價在時間軸上為不連續的區間。非交易時間的資訊累積無法立即反映在股價的變動上，投資者只能在隔日開盤後針對非交易時間所獲得的資訊進行持股部位的調整，這是為何開盤時的交易活動較為積極的原因之一。然而開盤期間的股價的變化並沒有受到應有的重視，開盤後前十分鐘內的交易活動往往不在被許多日內交易的研究樣本中。本研究針對此隔夜效應現象進行分析，研究結果證實隔夜效應對日內交易的顯著性。

透過股價指數期貨的實驗，證實了隔夜效應的顯著性之後，本研究嘗試利用類神經網路對日內交易進行模擬，模擬結果再度肯定了隔夜效應的重要性，也因此兩個交易日之間所累積的資訊同時反映在開盤時的現象被肯定為對日內交易具價值性。

關鍵字：隔夜效應、日內交易、類神經網路

A Study of Overnight Effect Mining on Taiwan Futures Market

Student：Pei-Chen Lee　　　　Advisor：Dr. An-Pin Chen

Institute of Information Management

National Chiao Tung University

ABSTRACT

The term, overnight effect, originates from the restriction on trading hours. The influences to stock prices such as fundamental factors and macroeconomic factors are all continuous. However, due to restriction on trading hours, the generation function of stock prices is discontinuous in time-axis. Various information arrived during overnight period, but could not be responded immediately. Short-term investors could adjust their holding stocks in terms of the overnight information only after opening of the next trading day. Therefore, the influences during non-trading period would only reflect on stock prices at the opening period of the next trading day. This is also the reason why stock returns at beginning trading period are usually with high volatilities and volumes.

By experiments on stock index futures, this paper firstly investigates the phenomenon between two trading hours. After the significance of overnight effect is confirmed, the observed overnight effect would be used to analyze the following intraday trading activities by neural network. This paper concludes that information during non-cash period is valuable and should be studied.

Keywords: Overnight Effect, Intraday Trading, Neural Network

誌　　　　謝

僅將這份最值得感動的榮耀與驕傲，獻給我最深愛的父母！

兩年來指導教授陳安斌博士的悉心教導，使我接觸到人工智慧與財務金融結合的學術領域，以及為人處世的道理，本研究能順利完成，首先要感謝的當然是我的指導教授。而怡璋學長在論文撰寫的過程中不厭其煩的指導，確實引導我以更審慎的角度來思索本論文的完整性，銘感在心。

論文口試期間，左杰官老師所提出的許多寶貴意見，也讓我受益匪淺。還有陳鴻基老師與羅濟群老師的不吝指教與建議，使得本研究更臻完備，在此致上最誠摯的謝意。

還有同為實驗室的這群龐大的夥伴們，大家一起學習成長、切磋琢磨，以及彼此間不斷的加油打氣，將是我生命中值得懷念的回憶。我親愛的父母與家人的鼓勵與包容，更是這兩年期間支持我成長的最大動力。

謝謝你們！

Contents

Table Contents

Figure Contents

Notations

$Info_{overnight}$: Price innovation for Taiwan Stock Index Futures during non-cash period.

$Info_{intraday}$: Price innovation for Taiwan Stock Index Futures during intraday

$Return_t^{cc}$: Returns between Close at day *t-1* and Close at day *t*

$Return_t^{oo}$: Returns between Open at day *t-1* and Open at day *t*

$Return_{t-1}^{oc}$: Returns between Open at day *t-1* and Close at day *t-1*

$Return_t^{overnight}$: Returns between Close at day *t-1* and 10 minutes after Open at day *t*

VIF : Variance Inflation Factors

D : Durbin-Watson statistic

1. Introduction

1.1 Background

Wide empirical evidence documents that stock market is more active at the beginning period of the trading hours. Trading volume, mean return, price volatility, and number of transactions, are usually more active at the opening period. Several studies have indicated the higher trading activities at opening period are due to the flow of new information. Gerety and Mulherin [16] have found that trading volumes cluster around the opening period and the closing period because of the market participants' craving for exchanging the risk of holding their positions overnight. They also confirmed the instinct that the accumulated overnight information results in trading activities at the opening period. Investors would rebalance their holding portfolios while optimized at the previous closing period because when the market reopened, their positions would not be optimal. The adjustment makes market activities higher at the opening period.

Many researches have been focused on analysis of stock intraday returns. Whatever methodologies they introduced, most of them rejected the returns at the beginning period after opening. Because of high return volatilities and trading volumes, the returns at the beginning period are labeled as abnormal returns and, therefore, excluded. Generally, the reason to explain the high return volatilities and trading volumes is the information accumulated overnight may be reflected simultaneously at the opening. If the initial returns did not play an influential part in the results of intraday analysis, they would not be rejected by most of the researches. However, if the initial returns do play an influential part in intraday analysis, their relation becomes a main phenomenon and an interesting topic in this thesis.

The phenomenon, overnight effect, originates from the restriction of trading hours. Investors make trading decisions according to different information, such as fundamental analysis, macroeconomic analysis, news analysis and so on. By observing average number of information arrival, Berry and Howe [5] stated clearly the public information flow started to increase before opening period and turned to peak after the market has closed. The influences of the public information flow are time-relevant and continuous. However, because of the restriction on trading hours, market prices are discontinuous sections in time-axis. Reflecting the effect of all factors immediately on market prices during non-trading hours is impracticable. Although the various information during trading hours is more sizable than that during non-cash period, the latter is still worthy of studying. The reason is the public information arrived during trading period could be reflected soon while that arrived during non-trading period could not be responded directly. Investors would adjust their investment to reasonable positions in terms of accumulated overnight information. Therefore, the stock prices change after adjustment could be considered as the most acceptable view to the following trading activities. Therefore, such an important period to describe investors' expectation about the following trading activities should not be ignored by most of the intraday analysis.

Lots of papers on Taiwan Stock Index Futures intraday trading analysis are focused. Nevertheless, these researches just mention the trading activities of trading hours, and few comment the period between two trading hours. This thesis assumes that the gap between Close price in the previous trading day and opening prices in this trading day shows the overnight information. To investigate the information accumulated during non-cash period and the impact of overnight information on the following intraday trading activities is the purpose of this thesis.

1.2 Expected Contribution

The higher market activities at the beginning period after opening have been already documented by many researches, and the reason is considered as reflection of information accumulated overnight. Although the initial returns are usually ignored by most of the intraday researches, to confirm the importance of overnight information to intraday trading activities is the intention of this thesis. In other words, to show the significance of overnight information to intraday returns is the main purpose of this thesis. After that, a different intraday analysis, considering overnight information, would be simulated to affirm the importance of overnight effect again by artificial intelligence. Finally, giving better understanding about the intraday trading activities is expected simulation results.

In addition, some investment strategies are suggested. However, employing overnight effect to avoid risks or speculate is the future works.

1.3 Thesis Organization

The contents of this thesis could be divided into the following parts. A theoretical study on existence of overnight effect would be discussed first. The relation exists between intraday returns and overnight returns would be explored by a regression model. After overnight effect is determined, intraday data analysis that studying overnight effect would be simulated by neural network.

The remainder of this study is organized as Figure 1. Chapter 2 describes literature reviews. The prior relative researches would be reviewed in this chapter. All methodologies applied in this thesis are also discussed in Chapter 2. Chapter 3 details the experiment design. The significance of overnight effect to intraday trading activities would be shown in this chapter by modeling regression formulas. Chapter 4 explains simulation steps. How overnight information could be used to analyze intraday trading activities would be discussed in this chapter. After simulation, the empirical results are summarized and discussed in Chapter 5. Chapter 6 concludes this thesis, and some future works will be proposed.

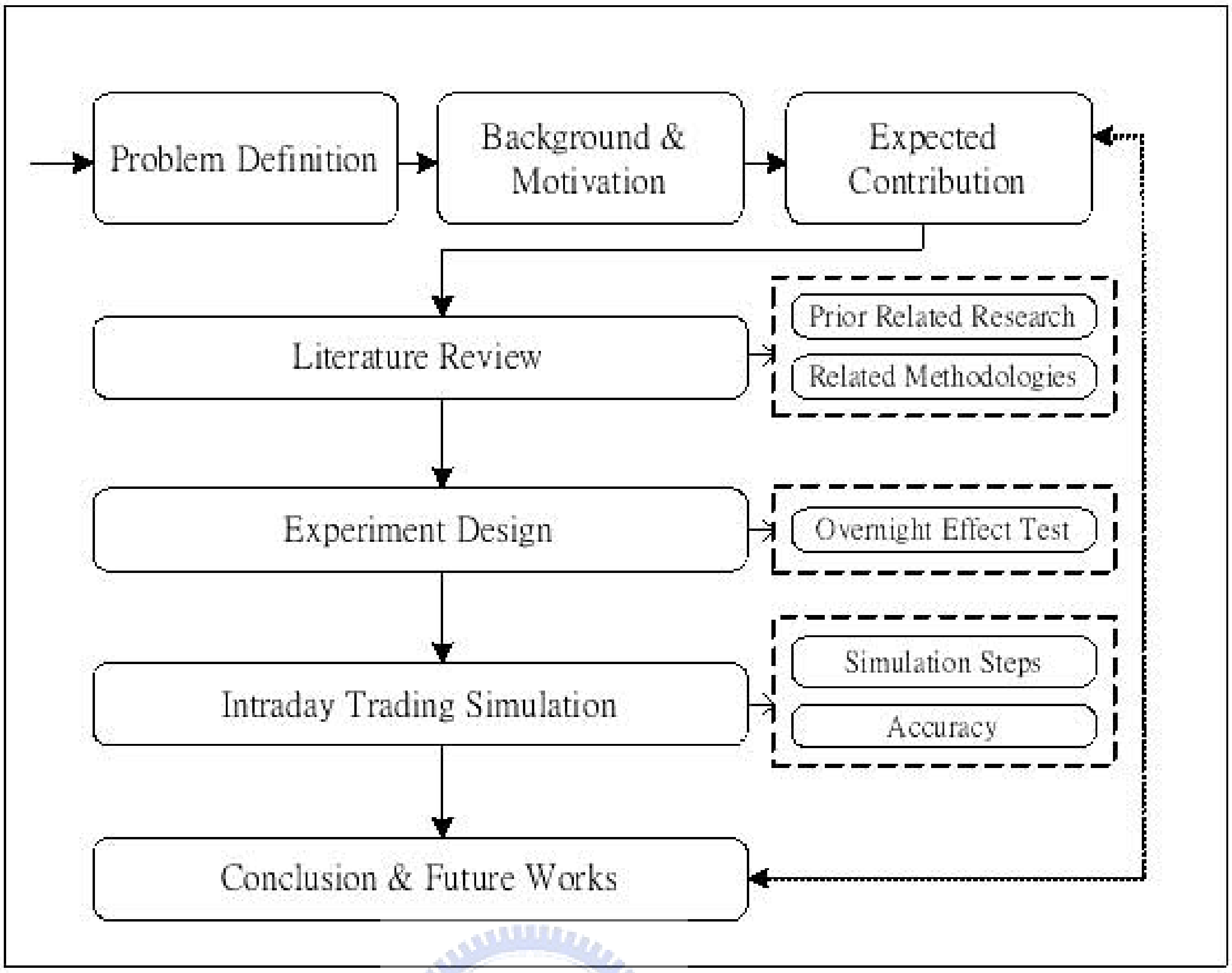


Figure 1-1. Framework

2. Literature Review

Before this research, the subject of predictability of stock market is wondered. This issue has been debated for such a long period. A lot of papers have focused on this topic. Some of them insisted on the non-predictability of stock market, however, recently more and more papers have indicated the predictability of stock market. After the issue of predictability, this chapter would discuss more about the phenomenon, overnight effect. Several theories adapted in this thesis such as neural network and candlestick analysis would also be reviewed in this chapter.

2.1 Stock Market Behavior

There have been an increasing number of researches on the predictability of stock market returns. For example, Chen, Roll and Ross [9] in 1986, Campbell and Shiller in 1988 [6], Balvers, Cosimano and McDonald [4] in 1990, and so on. They have demonstrated that stock returns are not unpredictable by using linear models. Stock returns were discovered predictable by public information on financial and economic variables such as dividend yields, interest rates, inflation rates and so on. While linear predictability was shown, the non-linear property was also found by many studies such as Gencay [15], and Abhyankar, Copeland and Wong [1]. Nowadays, it has become more acceptable that most of the financial and economic variables are non-linear. Abhyankar, Copeland and Wong [1] tested the nonlinearity in the world's four most important stock returns by several tests. The presence of nonlinear dependence was confirmed. Gencay [15] investigated the linear and non-linear predictability of stock market returns in daily Dow Jones Industrial Average Index from January 1963 to June 1988. The moving average rules with a band between the short and the long averages were used to create trading signals. Finally, Gencay [15] concluded that strong evidence of non-linear predictability was indicated in the stock market returns by employing the buy and sell signals generated from the moving average rules.

The predictability is still displayed by a wild variety of papers. For instance, Lo, Mamaysky and Wang [29] defined some technical patterns in advance and then applied nonparametric kernel regression to explore these technical patterns recognition of numerous U.S. stocks from 1962 to 1996. In this way, the effectiveness of technical analysis was evaluated. The unconditional empirical distribution of daily stock returns was compared with the conditional empirical distribution of daily stock returns, such as conditioned on head-and-shoulders or double-bottoms that they already defined in advance. Lo, Mamaysky and Wang [29] stated clearly that several technical indicators do contain incremental information and may have some practical value over the 31-year sample period.

Table 2-1 lists the prior related researches about linear predictability and Table 2-2 lists the prior related researches about non-linear predictability. Therefore, this thesis is based on the concept of predictability. After predictability, the core issue, overnight information, would be detailed in the next section.

Table 2-1. Prior related researches on linear predictability of stock returns

Authors	Year	Contribution
Chen, Roll and Ross	1986	This paper examined the relations between several macroeconomic variables and stock returns. For instance, the spread between long and short interest rates, industrial production rate and spread between high-grade and low-grade bonds. These variables have been found significant in explaining stock returns.

Campbell and Shiller	1988	This paper indicated that a long moving average of real earnings is a powerful predictor of stock returns. In other words, present value of future real dividends could be predicted by long historical average real earnings.
Balvers, Cosimano and McDonald	1990	The model proposed in this paper specified factors in arbitrage pricing. The insight of predictability phenomena could also be available from this model.

Table 2-2. Prior related researches on non-linear predictability of stock returns

Authors	Year	Contribution
Ramazan Gencay	1996	Daily Dow Jones Industrial Average Index from January 1963 to June 1988 was used to investigate the linear and non-linear predictability of stock market returns. The moving average rules with a band between the short and the long averages were used to create trading signals. Finally, this paper concluded that strong evidence of non-linear predictability was indicated in the stock market returns by employing the buy and sell signals generated form the moving average rules.
A Abhyankar, L S Copeland and W Wong	1997	This article tested the nonlinearity in the world's four most important stock returns by several tests. The presence of nonlinear dependence is confirmed.

Andrew W. Lo, Harry Mamaysky and Jiang Wang	2000	This paper employed kernel regression to discover the technical patterns, which were defined in advance. Lo, Mamaysky and Wang stated clearly that several technical indicators indeed contain incremental information and might have some practical value over the 31-year sample period

2.2 Overnight Information

The term, overnight information, originates from the restriction of trading hours. In 1990, Stoll and Whaley [35] wanted to explore the relation between the price shifts in index futures and in stocks and model the temporal association. It was intended for overnight returns to be excluded in any of the series. Hence, the first two 5-minute returns of each stock index series were not included because the index values at beginning of the trading day might be computed by the Close price of the previous trading day if some stocks were not traded yet. However, the Close prices were not correct reflections of opening values. Therefore, the initial returns were defined as noise and deleted. Until all stocks within the index had been traded, the index returns were accurate reactions to opening values.

Stoll and Whaley's [35] statement was also referenced by Darrat, Rahman and Zhong [11] in 2003. They studied the contemporaneous correlation and the lead–lag relation between trading volume and return volatility in all the Dow Jones industrial average (DJIA) stocks, and also compared them respectively under the MDH (mixture distribution hypothesis) and SIAH (sequential information arrival hypothesis) hypothesis. Similar to Stoll and Whaley, Darrat, Rahman and Zhong [11] also excluded overnight returns. The first two 5-minute returns were rejected again because of the average time for stocks to be traded in the S&P 500 Index reported by Stoll and Whaley [35]. The average time expired between the market opening and an opening transaction is around 5–7 minutes. Prices during the period were influenced by the stale Close price in the previous trading day.

Anderson et al. [2] also deleted the initial returns. Because the prices at initial period of trading day were adjusted in terms of overnight information. Returns at this interval were higher than any other 5-minutes returns. The first 5-minutes return, hence, was considered an unusual 5-minutes return and then deleted from the intraday data set. Lee and Mathur [25] also documented the first 5-min return because the unusually high volatility.

All the above intraday researches are collected in the following table.

Table 2-3. Prior related researches on intraday analysis

Authors	Year	Reasons
Hans R. Stoll and Robert E. Whaley	1990	At beginning of the trading day, the index values might be computed by the Close price of the previous trading day if any stocks were not traded yet. However, the Close prices were not accurate reflections of opening values. Therefore, the initial returns were defined as noise and deleted.
Torben G. Anderson, Tim Bollerslev	1997	The prices at initial period of trading day were adjusted in terms of overnight information. Returns at this interval were higher than any other 5-minutes returns. The first 5-minutes return, hence, was considered not a usual 5-minutes return and deleted from the intraday data set.
Ali F. Darrat, Shafiqur Rahman and Maosen Zhong	2003	Prices during the initial interval could be affected by the stale closing price of the previous day. Therefore, the first two 5-minutes returns were excluded

Torben G. Anderson, Tim Bollerslev, Jun Cai	2000	The initial returns largely reflected the adjustment to information accumulated overnight and demonstrate higher average return volatility. Therefore, the unusual returns were deleted.

The reason why these intraday researches always turned down the first one or two 5-minutes returns is wondered. Whatever reasons they rejected the initial returns, they all expressed the phenomenon due to discontinuous trading sections. If the phenomenon is not important, it is impossible for most researches to exclude the initial returns. However, if the phenomenon is important, the relation between initial returns and intraday returns is interesting topic. How to make them integrate with each other in time-axis is the point in this thesis.

Chan et al. [7] discussed overnight information and intraday trading behavior in 2000. Several cross-listed stocks on the NYSE (New York Stock Exchange) were selected to investigate how overnight price movements in local markets affected the trading activity of foreign stocks on the NYSE. The overnight period was defined as non-trading hours of NYSE. The conclusion was that local price movements influenced not only the opening returns but also their first 30-min returns. Local price movements were positively related to price variability of foreign stocks. They also had stronger influence at the NYSE open and weaker afterward. Therefore, the importance of overnight information was determined between NYSE cross-listed stocks and their local market.

Information content during non-cash trading period is also confirmed by the following research. Lee and Chen [27] explored the lead-lag relationship between the futures market during the non-cash trading period and the cash market during its opening period, and showed the futures prices in non-cash trading period contain valuable information in forecasting the opening cash price index. The obtained leading futures and cash market closing index in the previous trading day were then used to predict the following opening cash market price by back-propagation network. This paper also concluded that neural network model outperforms commonly discussed GARCH model.

So many researches have shown clearly the importance of overnight information. However, the information collected during non-cash period is not studied as much as deserved. Therefore, the purpose of this thesis is to explore the unknown knowledge between two trading hours so that intraday analysis could be more precise.

2.3 Candlestick Analysis

Candlestick charts were proposed by Japanese rice traders in the 17th century. It is a graphical way to display the different combinations of High, Low, Open and Close prices at the same trading day. Each combination is a different visualization of stock price movement. It may represent a different prediction about the following trading activities. A candlestick chart includes a rectangle, real body, and two shadow lines. The real body is the difference between Open price and Close price. The difference between Open price and Close price, real body, provides a way to measure the trend and the extent of intraday prices movement. The real body is black if Open price is higher than Close price while white if Open price is lower than Close price. Hence, white candlestick reveals downtrend and black candlestick reveals uptrend. The two shadow lines, upper shadow line and lower shadow line, respectively deliver the information about High price and Low price. The difference between High price and Low price indicates the scope of intraday trading activity and a degree of volatility.

There had been many researches based on Open, High, Low and Close price. Some of them attempted to investigate the information content about these four prices so that the others could exploit the affirmation about candlestick line to develop some application systems about analysis on stock prices.

Fiessa and MacDonald [14] intended to investigate the information content of these four prices, respectively Open, High, Low and Close, by using range and cointegration methods. They argued that the analysis of High and Low prices could be informative if the following statements are realized. One was that High and Low prices could reveal some information about shifts in supply and demand structure; the other was changing order flows is key factor while determining market prices. They also indicated that Curcio and Goodhart [10], DeGrauwe and Decupere [12] have done some academic work about support and resistance to support the first statement. The second statement is proven by Menkhoff [33]. Menkhoff emphasized the relation between order flow analysis and expectation process formation of foreign exchange market participants. After this stage, Fiessa and MacDonald [14] attempted to explore the dynamic and structural relationships about High, Low and Close prices. This paper concluded that a stable structural relationship between High, Low and Close prices does exist. The technical analysis of High, Low and Close prices played an important role while learning Granger causality in high frequency exchange rates.

Lam [24] investigated the relationships between Open, High, Low and Close prices by linear regression, adaptive recursive estimation and nonlinear neural network. This paper concluded that based on the prior information of Open price, these prices, High, Low and Close prices, could be forecasted better.

What Fiessa and MacDonald [14] stated or Lam [24] concluded is all about the importance of Open, High, Low and Close prices. There are still many researches focusing on the information content of candlestick analysis. Most of them confirm these four prices are informative than other prices while describing intraday trading activities. This is also the reason why these four prices are adapted in this thesis.

Some researches also used the relations between these four prices to forecast the price trend in the future. For instance, Lee and Jo [26] applied candlestick analysis to develop an expert system to help investors to get the trading signals. They divided stock price pattern into 5 groups, including rising, falling, neutral, trend-continuous and trend-reversal. Each group has its own corresponding patterns, which defined in advance such as bigblack, bullish harami, dark cloud cover and so on. The expert system was built and tested according to the defined rules. The average success rate to judge the buying or selling timing was 72%. However, Lee and Jo [26] indicated the constraint on this expert system, the lack of an automatic machine learning. To solve this limitation, a neural network model is chosen in this thesis because its ability in self-learning.

According to these above researches, it is believed that Open, High, Low and Close prices reveal more information than any other prices during intraday trading activities. This is also the reason why candlestick analysis is used in this thesis.

2.4 Neural Network

Neural network, a part of artificial intelligence, is an information-processing paradigm inspired by the densely interconnected, parallel structure of human brains. The key element of this paradigm is the novel structure of the information processing system. Artificial neural networks learn by example just as human beings. Learning in biological systems involves adjustments to the synaptic connections that exist between the neurons. An artificial neural network is configured for a specific application through a learning process. For example, data classification, data clustering, pattern recognition and so on.

The earliest kind of neural network is a single-layer perceptron network proposed by McCulloch and Pitts [32] in 1943. The inputs are fed directly to the outputs via a series of weights. The inputs and outputs of McCulloch-Pitts neuron are only binary and the activation function of McCulloch-Pitts neuron is always the unit step function.

MP model was followed by Hebb learning in 1949. This is the first learning law. In 1957, Rosenblatt proposed the concept of perceptron. In 1960, Widrow and Hoff presented Widrow-Hoff learning rule and also invented ADALINE (ADAptive LINear Element). All that mentioned above are important development in neural network.

In 1969, a famous monograph which entitled Perceptrons by Marvin Minsky and Seymour Papert showed that it was impossible for a single layer perceptron network to learn an XOR function. After that, neural network was paid little attention to because of the disadvantage proven by Marvin Minsky and Seymour Papert.

However, there were still a lot of researches focusing on neural network. For instance, the back-propagation algorithm was discovered by Webos in 1974. ART, Adaptive Resonance Theory, was developed by Carprnter and Grossberg in 1980. Kohonen proposed SOM [21], Self-Organizing Maps, and LVQ, Learning Vector Quantization. These are all critical development in neural network history.

The key element of the neural network paradigm to be differentiated form other methods of artificial intelligence is self-learning. Neural network is also well known for its ability to process nonlinear problem.

Besides the strong evidence of non-linear predictability of stock returns found by Gençay [15], Qi and Maddala [36] also demonstrated that the neural network model could improve upon the linear regression model in terms of predictability, however, not in terms of profitability.

Another example is indicated by Kanas and Yannopoulos [17]. A linear and a nonlinear artificial neural network model are compared with each other while forecasting monthly Dow Jones returns and Financial Times indices. For these two indices, the empirical results showed clearly that artificial neural network predicts more accurately than linear model while out-of-sample data is used. This conclusion also supported that the relation between stock returns and fundamentals is nonlinear. Since stock returns are non-linear predictability, neural network is considered practicable to predict stock returns well and, hence, is chosen as a useful tool to analyze intraday returns.

Neural network has been widely applied in several financial areas, such as exchange rate prediction, bankruptcy forecasting and stock market prediction [37][28][30].

For exchange rate prediction, Yao and Tan [37] expressed the applicability of neural network model to forecast foreign exchange rates. Foreign exchange rates between American Dollar and five other major currencies, Japanese Yen, Deutsch Mark, British Pound, Swiss Franc and Australian Dollar are predicted by using neural networks. Beneficial forecasts could be achieved without any use of extensive market data or knowledge while using neural network.

The subject of bankruptcy is also an important issue in the finance area and is concern of investors, managers, auditors and other finance personnel. Leshno and Spector [28] applied various neural network models to predict bankruptcy cases and also made a comparison among them. Accuracy of neural network model exceeded the other classical discriminate analysis model.

The implied volatility, usually calculated using the Black-Scholes model, is such a popular issue in stock market prediction. Malliaris and Salchenberger [30] used neural network to forecast the S&P 100 implied volatility. The past volatilities and other options market factors were used to develop a neural network model to forecast future volatility. The performance of this network showed itself as a valuable and predictive tool.

2.5 Learning Vector Quantization

There are several fields in artificial neural network such as classification, clustering and mapping. Learning Vector Quantization is one of classification methods. It is a competitive learning algorithm and considered as a supervised version of the Self-Organising Map (SOM) algorithm by Kohonen [19][20] in 1988.

The basic principle of LVQ is to find the codebook vector that is closest to the input vector. For instance, if input vector and the nearest codebook vector belong to the same class, then the nearest codebook vector is moved toward input vector. On the contrary, then the nearest codebook vector is moved away from input vector.

The algorithm of LVQ is listed as follows：

I. Codebook and learning rate initialization.

II. Do step 3-7 until stopping condition is satisfied.

III. For each input vector x，repeat steps 4-5

IV. Find the nearest codebook m_c so that $\|x - m_c\|$ is a minimum distance.

V. Update the nearest codebook m_c according to follows ：

If x and m_c belong to the same class, $m_c(t+1) = m_c(t) + \alpha(t)[(x(t) - m_c(t))]$

If x and m_c belong to different classes, $m_c(t+1) = m_c(t) - \alpha(t)[(x(t) - m_c(t))]$

VI. Reduce learning rate

VII. Test stopping condition. It may be a fixed number of iterations or the learning rate reaching a sufficiently small value.

$\alpha(t)$ is learning rate and should be smaller than 1 and larger than 0, $0<\alpha(t)<1$. However, Kohonen suggested in the LVQ1 algorithm, α should be initially smaller than 0.1.

Actually, α could be constant or reduce with time. In LVQ_PAK (The Learning Vector Quantization Program Package) developed by LVQ programming team of the Helsinki University of technology, α decreases linearly in time.

A novel learning vector quantization (LVQ2.1) is proposed [20][22][23] in order to deal with the overlapping situation. The classification decision is the same with that of LVQ1. However, the difference is that if the distance between input vector and winner (m_i) is almost the same as that between input vector and the runner-up (m_j), then both m_i and m_j should be updated simultaneously. In other words, LVQ2.1 deals with an input that falls into the window that is the overlapping area between the nearest class and next-to- nearest class. The window is defined as follows.

$$\min\left(\frac{d_i}{d_j},\frac{d_j}{d_i}\right) > s, s = \frac{1-\omega}{1+\omega}$$

d_i and d_j are the Euclidean distances of input vector x from m_i and m_j. In LVQ2.1, the window width is recommended as 0.2-0.3.

The learning algorithm is given by

If x and m_i belong to different classes, $m_i(t+1) = m_i(t) - \alpha(t)[(x(t) - m_i(t))]$

If x and m_j belong to the same class, $m_j(t+1) = m_j(t) + \alpha(t)[(x(t) - m_j(t))]$

(Note x must fall into the window so that LVQ2.1 algorithm could be used.)

LVQ3 is an improved vision of LVQ 2.1. To deal with the situation if x, m_i and m_j belong to the same class, LVQ3 is expended. The learning algorithm is given by

If x and m_i belong to different classes, $m_i(t+1) = m_i(t) - \alpha(t)[(x(t) - m_i(t))]$

If x and m_j belong to the same class, $m_j(t+1) = m_j(t) + \alpha(t)[(x(t) - m_j(t))]$

If x, m_i and m_j belong to the same class, $m_k(t+1) = m_k(t) + \varepsilon\alpha(t)[(x(t) - m_k(t))]$

$$(k \in \{i, j\})$$

ε is found applicable between 0.1 and 0.5 in a series of experiments. However, the optimal value depends on the window size. LVQ3 is a stable algorithm because the optimal placement of codebook does not change in the long run while it does change in LVQ2.1.

Kangas and Kohonen [18] indicated that accuracies of these three options, the LVQ1, the LVQ2.1 and the LVQ3 [11], almost the same. The LVQ1 and LVQ3 are stable. In LVQ1, learning rate could reach almost optimal result at a quick convergence. Therefore, LVQ1 is chosen as neural network model in this research.

3. Overnight Effect

3.1. Definition

What is overnight effect? It is a phenomenon that all information accumulated during overnight period would be reflected on the beginning period after opening. Overnight period is defined as non-cash period. Information produced during overnight period is defined as overnight information. Briefly, overnight effect is the phenomenon between two trading days.

The term, overnight effect, originates from the restriction on trading hours. The influences to stock price such as fundamental factors and macroeconomic factors are all continuous. However, due to restriction on trading hours, the generation function of stock prices is discontinuous in time-axis. Information arrived during overnight period could not be responded immediately. Investors could adjust their holding stocks in terms of the overnight information only after opening of the next trading day. Therefore, the influences happened during non-trading period would only reflected at the opening period of the next trading day. This is also the reason why returns at beginning period after opening are usually with high volatilities and volumes.

In this research, all factors are assumed as functions of time and continuous time relative data. The following formulas express the relations between continuous effect of all factors and discontinuous trading sections.

$$Stock\,price = Real\,price + Effect\ price \tag{1}$$

$$Effect\ (fundmental\ ) = f_1(t) \tag{2}$$

$$\vdots$$

$$Effect(macroeconomics) = f_2(t) \tag{3}$$

Formula (1) explains the value of stocks. Stock price is affected not only by its real value but also various effects such as formula (2) and formula (3). $f_1(t)$ and $f_2(t)$ are different continuous functions of time. As formula (2) and formula (3) show, fundamental factor affects stock prices as function 1 while macroeconomic factor affects stock prices as function 2. There are still various kinds of formulas as these two. Each of them represents the influence of a different factor.

Figure 3-1 describes that these trading rules affect the stock prices continuously. Because of the restrictions on trading hours, the generation function of market prices is discontinuous in time-axis as Figure 3-2 illustrated. All these two figures want to emphasize is the conception of overnight effect. It comes from the inconsistence between continuous influences and discrete trading hours. The phenomenon between two trading periods is the key point in this thesis.

For this reason, this thesis begins with focusing on whether all influences happened during non-cash period, between 13:45 and 08:45 (period T) from previous day, would affect the opening prices of next trading day or not. Furthermore, it is suspected that the returns at the beginning period after opening of the next trading day would be the sum of the factors occurring during overnight period. The idea is formulated by formula (4).

$$Return_{Opening} = \sum_{e=1}^{n} Effect(e) = \sum_{e=1}^{n} f_e(T) \quad (4)$$

$Opening$	：open price
e	：element influencing prices
n	：number of elements
T	：time period between two trading times

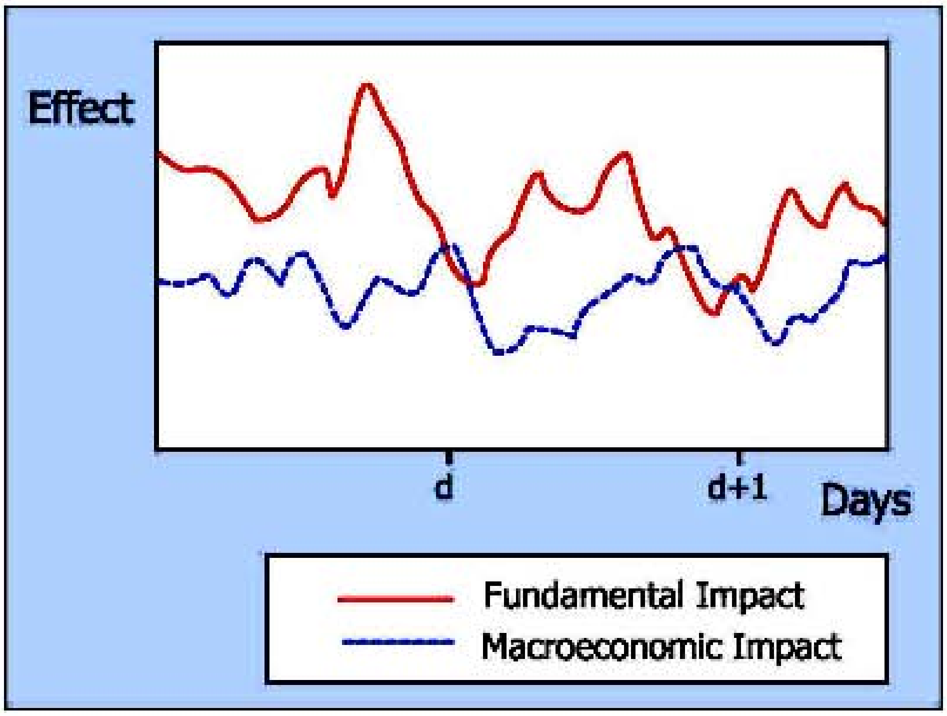


Figure 3-1. An illustration of continuous impact on market prices. The factors influence the stock prices continuously, such as fundamental factors, macroeconomic factors and etc.

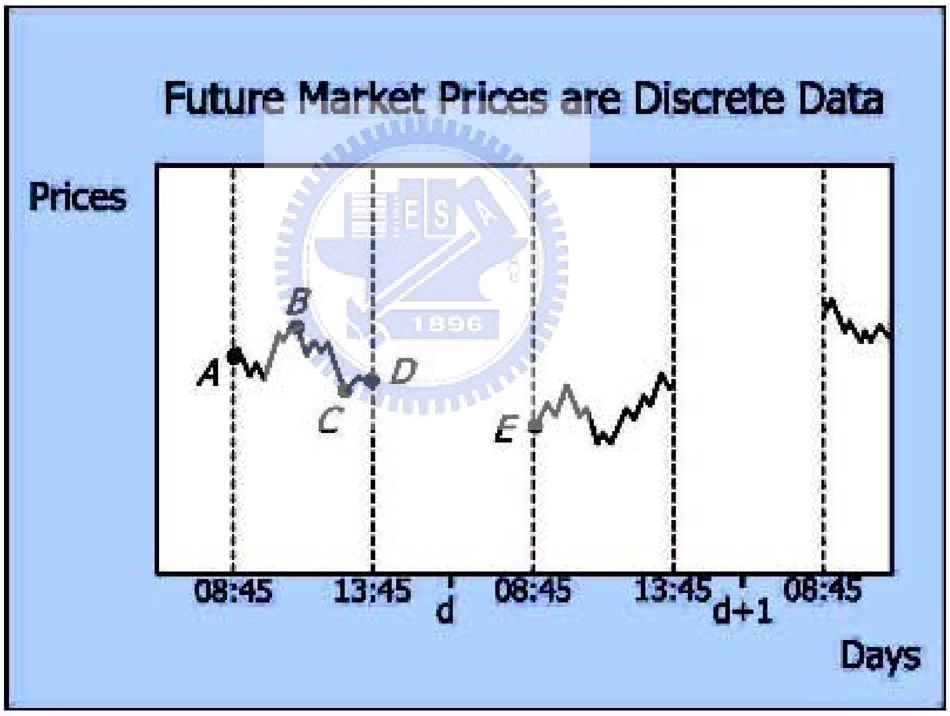


Figure 3-2. An illustration of discontinuous trading sections. The Open, High, Low, Close prices in the previous trading day are plotted as point A, B, C and D. The market prices are discrete sections in time-axis because of the restrictions on trading hours.

Overnight returns are used to be rejected by intraday analysis because they are defined as abnormal returns. However, if opening prices really reflect the most of overnight information, then overnight returns actually represent market participants' view to the following trading day after digesting information accumulated overnight. Therefore, the returns at the beginning period after opening should not be neglected. On the contrary, the overnight returns should be paid much attention to because they show the newest view of market investors to the following trading activities.

In order to have a better understanding about intraday trading analysis, the importance of overnight returns should be confirmed first. In other words, the overnight returns should be proved significant to the following intraday returns. After investigating the relation between information content of overnight returns and of intraday returns, a simulation on employing overnight information to intraday analysis would be made by neural network. The experiment architecture is illustrated as Figure 3-3.

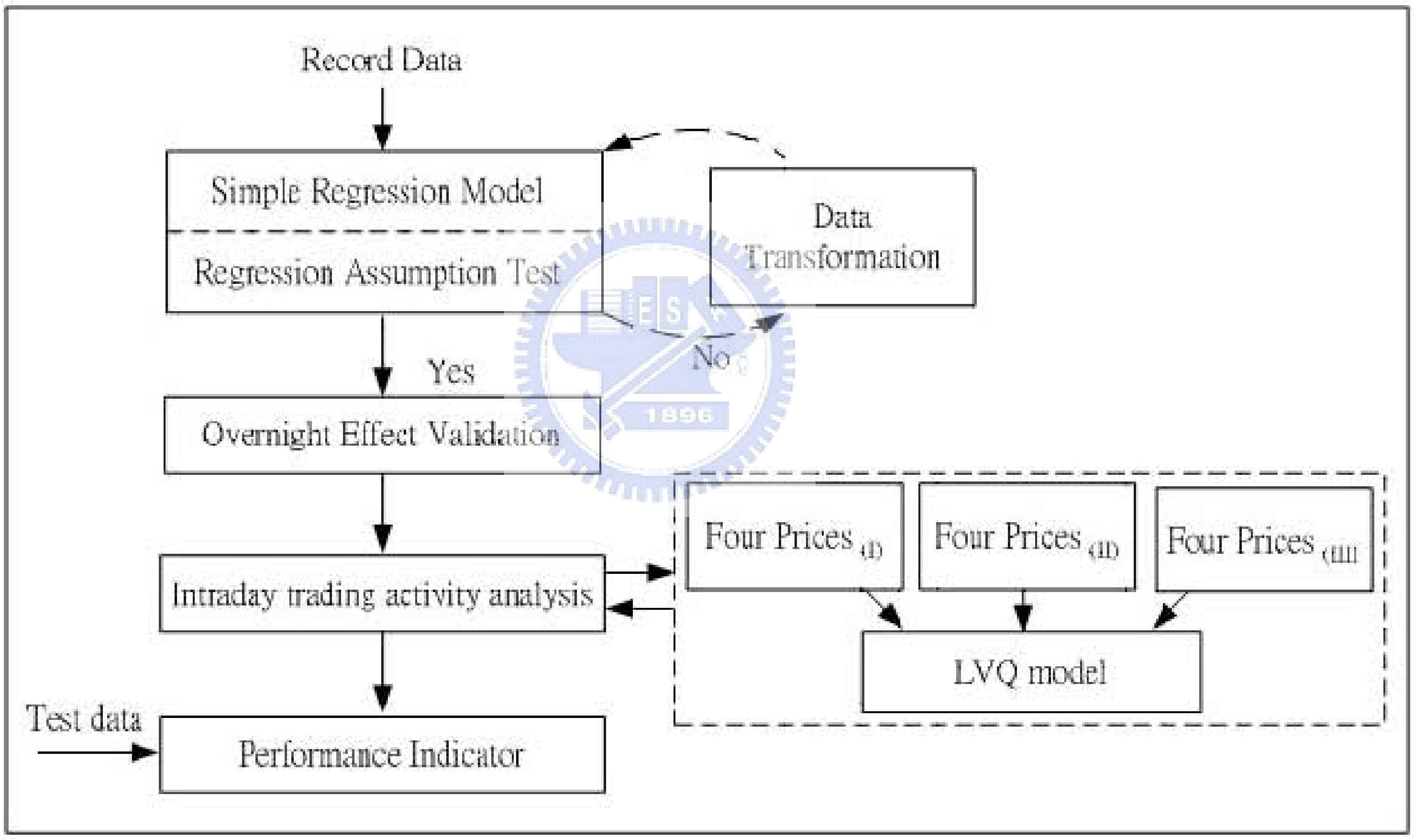


Figure 3-3. An illustration of experiment architecture.

3.2. Regression Model

According to Chan et al. [7], the following regression model is established in order to show the significance of overnight effect. In this thesis, Taiwan Stock Index Futures between 2001/04/13 and 2002/04/12 are used to determine the importance of overnight returns.

Let $Info_{overnight}$ denotes the price innovation for Taiwan Stock Index Futures during non-cash period.

Let $Info_{intraday}$ denotes the price innovation for Taiwan Stock Index Futures during intraday.

Then the effect of overnight returns on intraday return can be assessed by the regression model：

$$Return_i^{cc} = \alpha_1 + \beta_1 \times Info_{overnight} + \beta_2 \times Info_{intraday} + \varepsilon_i \tag{5}$$

It is quit difficult to quantifying $Info_{overnight}$ and $Info_{intraday}$. So it is a proper idea to replace them as concept of returns. Variable $Info_{overnight}$ could not be directly inferred from Open price and Close price at previous day because overnight information influences not only Open price but also opening period. Variable $Info_{intraday}$ also could not be directly inferred from Open price and Close price in this trading day since $Info_{overnight}$ are not directly observed just from Open price and Close price at previous day. Therefore, the overnight information and intraday information could be observed from open-to-open variable and close-to-close variable as Figure 4 illustrated.

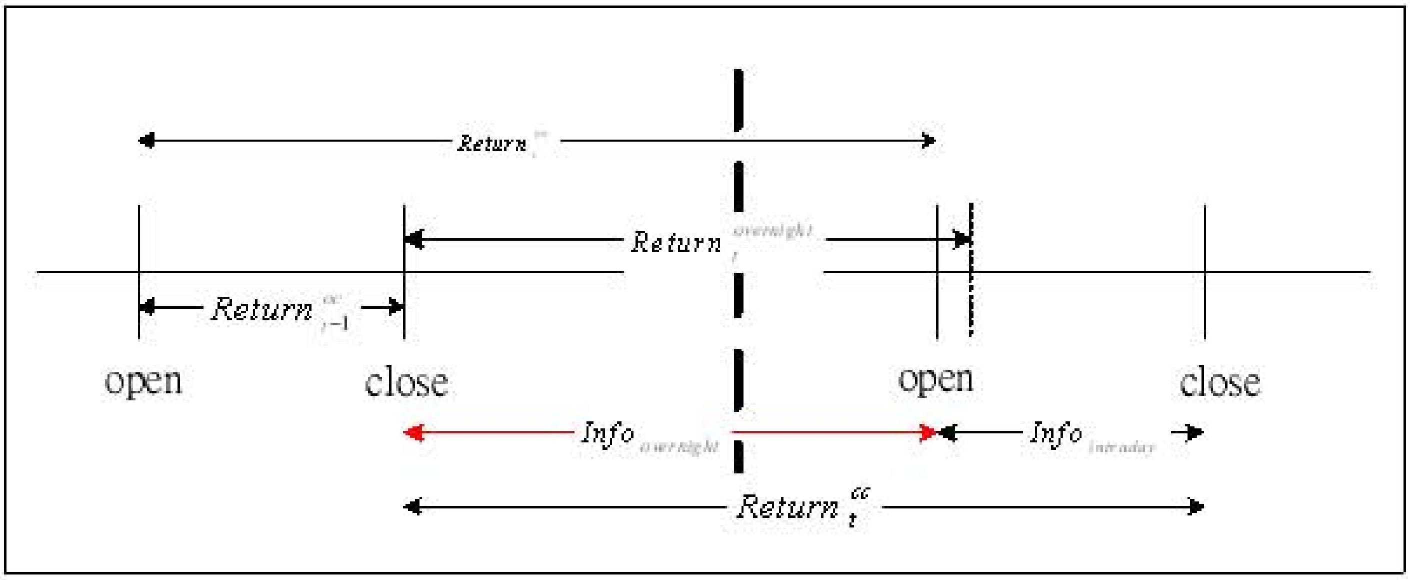


Figure 3-4. An illustration of relations between variables $Return^{cc}_t$, $Return^{oo}_t$, $Return^{oc}_{t-1}$ and $Return^{overnight}_t$

As Figure 3-4 illustrated, $Return^{oo}_t$ is composition of $Return^{oc}_{t-1}$ and $Info_{overnight}$, while $Return^{cc}_t$ is composition of $Return^{overnight}_t$ and $Info_{intraday}$. The relations could be formulated as follows.

$$Return^{oo}_t = a_1 + b_{11} \times Return^{oc}_{t-1} + b_{12} \times Info_{overnight} \tag{6}$$

$$Return^{cc}_t = a_2 + b_{21} \times Return^{overnight}_t + b_{22} \times Info_{intraday} \tag{7}$$

Valuable $Info_{overnight}$ and Valuable $Info_{intraday}$ can be captured by estimating formula (6) and formula (7). Then formula (8) and formula (9) are obtained.

$$Info_{overnight} = \frac{1}{b_{12}} Return^{oo}_t - \frac{a_1}{b_{12}} - \frac{b_{11}}{b_{12}} \times Return^{oc}_{t-1} \tag{8}$$

$$Info_{intradayt} = \frac{1}{b_{22}} Return_t^{oo} - \frac{a_2}{b_{22}} - \frac{b_{21}}{b_{22}} \times Return_{t-1}^{overnight} \quad (9)$$

Substituting these two variables, $Info_{overnight}$ and $Info_{intradayt}$, into formula (5), this regression model can be rewritten as following equation:

$$Return_t^{overnight} = \alpha^* + \beta_1^* \times Return_t^{cc} + \beta_2^* \times Return_t^{oo} + \beta_3^* \times Return_{t-1}^{oc} \quad (10)$$

where $\alpha^* = -\frac{\beta_2 a_2}{\beta_2 b_{21}} - -\frac{\beta_1 a_1 b_{22}}{\beta_2 b_{12} b_{21}}, \beta_1^* = \frac{b_{22} - \beta_2}{\beta_2 b_{21}}, \beta_2^* = \frac{\beta_1 b_{22}}{\beta_2 b_{12} b_{21}}, \ \beta_3^* = -\frac{\beta_1 b_{11} b_{22}}{\beta_2 b_{12} b_{21}}.$

Since Variable $Info_{overnight}$ could not be directly observed, how $Return_t^{overnight}$ could be estimated? According to Stoll and Whaley, $Return_t^{overnight}$ includes Open price and the first two 5-minute returns. This method is also adopted by Darrat, Rahman and Zhong [11]. Although this method is aimed at stock market, this thesis focuses on Taiwan Stock Market Index Future. However, Taiwan Stock Market Index Future is derivative financial instrument and its target is Taiwan weighted stock index. Hence, it is necessary for futures to be fully reflected until the stock market reflected all overnight information. The term, overnight return, is defined as return at the first ten minutes. Figure 3-4 illustrates overnight return by the dash line. In this way, the issue how to measure $Return_t^{overnight}$ is solved and formula (10) could be estimated.

Regression coefficients are estimated subject to the constraints according to formula (10) by SPSS 11.0. The method for Independents is stepwise. Stepwise regression is one method to find the best explanatory variables. According to partial correlations, explanatory variable is entered in turn till no significantly increasing explanatory power.

Table 3-1 reports regression results. As expected, the β_1^*, β_2^* and β_3^* coefficients are high for the overnight return with large t-statistics. This indicates that most of the overnight information is connected to opening prices. For Taiwan stock market index future, estimates of β_1^*, β_2^* and β_3^* are all significant for the overnight period. Table 3-1 is able to conclude the relations between these variables. The adjusted R^2 is 0.744 in this regression model.

Table 3-1: Regression results of overnight effect by SPSS 11.0

Variables	Coefficient	t statistic
$Return_t^{cc}$	0.222 (β_1^*)	5.547
$Return_t^{oo}$	1.667 (β_2^*)	18.793
$Return_{t-1}^{cc}$	-1.534 (β_3^*)	17.221

3.3. Regression Model Assumptions Tests

There are several assumptions while using linear regression model, including multicollinearity, auto-correlation and normal distribution. In order to test the data set could be modeled by linear regression model or not, following statistic tests are made.

I. Multicollinearity

When it comes to regression model, the most important topic is multicollinearity. Multicollinearity in regression models represents strong correlations between independent variables and would enlarge the variances of the parameter estimates. This may lead to no significance of individual independent variables and may cause incorrect signs and magnitudes of regression coefficient estimates. Consequently, multicollinearity would result in wrong conclusions about the regression model.

Two common procedures, tolerance and VIF (Variance Inflation Factors), are used to diagnose the existence of multicollinearity. Tolerance and VIF could be computed according to the following two equations.

$$VIF(\hat{\beta}_j) = \frac{1}{1-R^2} \tag{11}$$

$$Tolerance(\hat{\beta}_j) = \frac{1}{VIF} = 1 - R^2 \tag{12}$$

It is considered that the higher VIF or the lower tolerance index, the higher probability of multicollinearity. If values of VIF exceed 10 (or tolerance < 0.1), they are often regarded as presence of multicollinearity. Table 2 shows the values of VIF about these three variables listed in formula (10).

Table 3-2: Collinearity statistics

Variables	Tolerance	VIF
$Return_t^{rr}$	0.807	1.240
$Return_t^{oo}$	0.164	6.113
$Return_{t-1}^{oc}$	0.162	6.169

Values of VIF are all smaller than 10, hence, it could be concluded as no multicollinearity exists in this regression model.

II. Auto-correlation

The tests for auto-correlation are concerned with the dependence between error terms in a series. If the error terms are related to each other, then it is considered as violation of the linear regression assumptions. Auto-correlation could be detected by using Durbin-Watson Test. The Durbin-Watson test statistic is designed for detecting errors that follow a first-order autoregressive process. The D-W statistic is computed by the following equation.

$$D = \frac{\sum_{t=2}^{n} (e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^{\,2}} \tag{13}$$

If D-W statistic is less than 4-dU and larger than dU, it is considered as no auto-correlation in this series. The relations between D-W statistic, dL and dU are illustrated as Figure 3-5.

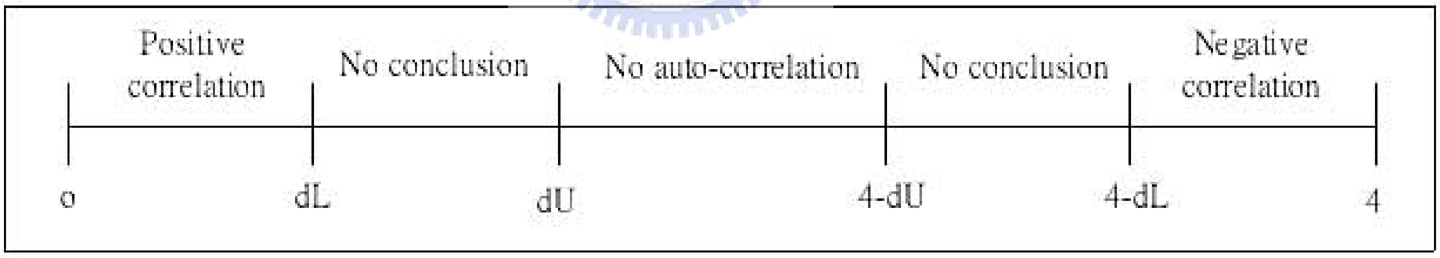


Figure 3-5. Durbin-Watson test

The value of dU could be checked in Durbin-Watson test table. According to Durbin-Watson test table, the value of dL and dU are obtained. The value of dL is 1.61 and the value of dU is 1.74. The D-W statistic is 1.992. Therefore, it is concluded that there is no auto-correlation in this regression model.

III. Normal distribution

The linear regression model is only valid under the assumption that the error terms are distributed as normal distribution. This assumption could be verified by looking at the plot residuals. In this thesis, the normal P-P plot of SPSS is used to test the normality assumption. If the normality assumption is not violated, points will cluster around a straight line.

It is clearly illustrated in Figure 3-6 that the points do cluster around the straight line. Therefore, the normality assumption is not violated in this data set.

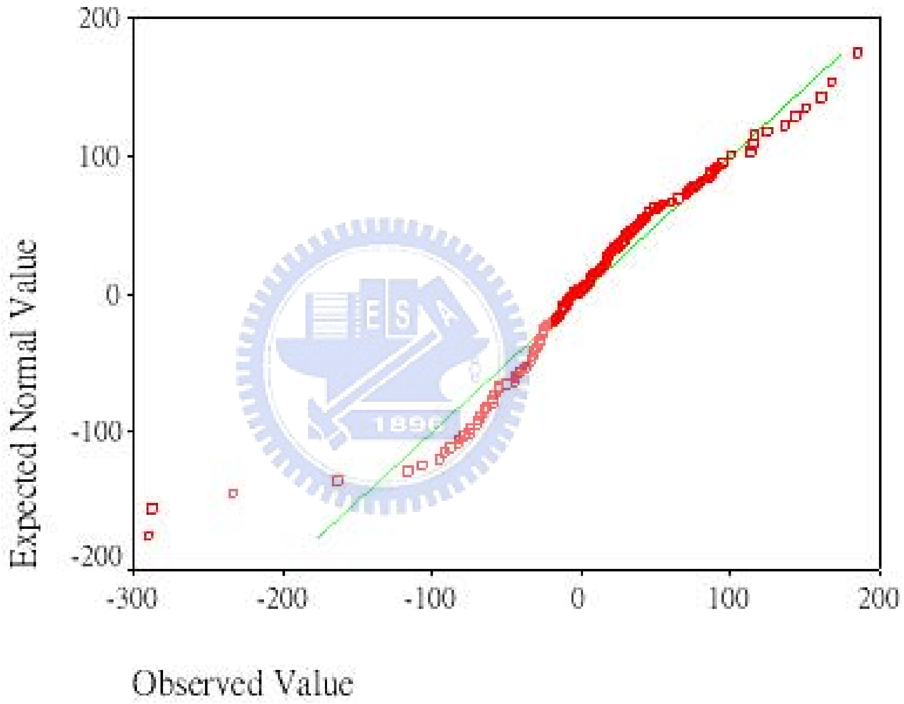


Figure 3-6. Normality test result by SPSS 11.0

The subjects of concern when using regression model are tested and none of them are violated. Therefore, it is unnecessary to transform the original data due to possible violations of assumptions of the models.

4. Simulation

As regression result reveals, overnight returns relate to three variables listed in Table 3-1. At time t, $Return_{t-1}^{oc}$ is already known. After opening period, overnight returns and $Return_{t}^{om}$ are also available. Hence, $Return_{t}^{oc}$ should be inferable according to formula (10). Since Overnight returns are significantly related to intraday returns, and should be considered while analyzing intraday trading activities.

After the importance of overnight effect is confirmed, this thesis would employ it to analyze intraday trading activities. Therefore, an artificial neural network is applied to simulate the relations between these returns. However, how to describe these given returns precisely and input them into neural network model is also an important topic in this paper.

4.1. Input Nodes of Neural Network

Although these returns, $Return_{t-1}^{oc}$, $Return_{t}^{oo}$ and $Return_{t}^{overnight}$ overlap in time-axis, briefly they can be observed from Figure 3-4 and described as these periods illustrated as Figure 4-1, respectively period between this Open price and previous Open price and period between ten minutes after opening and Close price in previous trading day. This thesis assumes that gap between previous Close and this Opening period shows most of the overnight information. Therefore, interval between Open price and Close price at day *t-1* and interval between Open and ten minutes after opening at day *t* should especially concerned to.

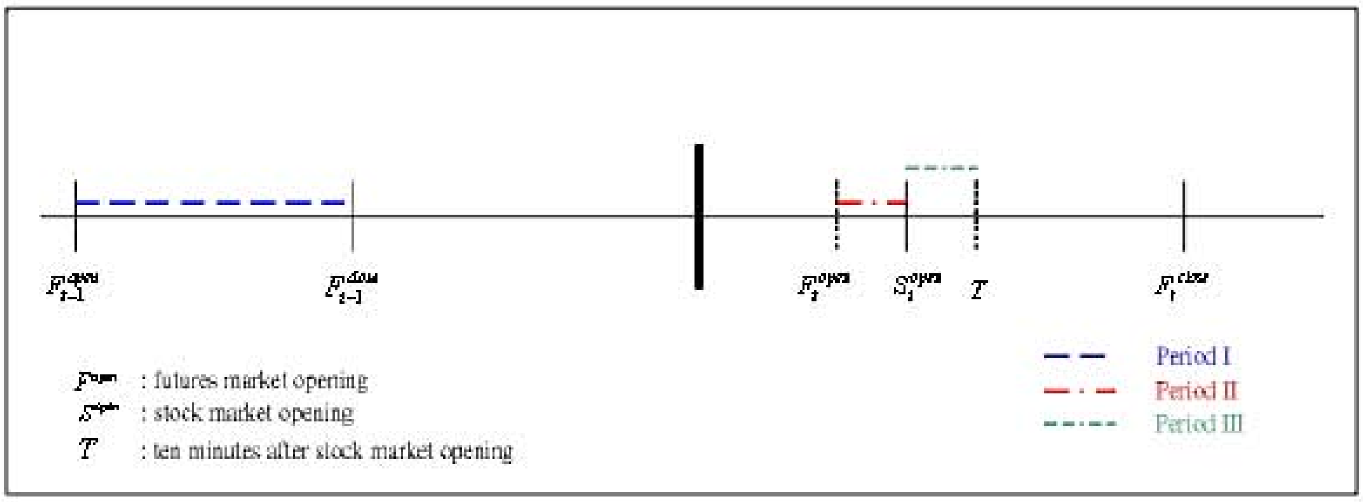


Figure. 4-1. An illustration of Period I, Period II and Period III.

How to describe the two intervals more precisely? According to candlestick analysis, it is the belief that Open, High, Low and Close prices have higher information content than other prices to extract featured prices of intraday trading [7]. Therefore, these four prices in day *t-1* are selected as input nodes in neural network model. Based on fractal theory, pattern feature extraction is irrelevant to the length of period. The four prices found after the opening of future market and of stock market are also parts of the input nodes in neural network model. This is because that these prices are also Open, High, Low and Close prices in the overnight effect period. Consequently, four prices in the previous trading day and four prices in the ten minutes after stock market opening are considered as the factors to influence the intraday trading behaviors.

After input nodes determined, tuples of input nodes are computed as follows. Data of Taiwan Stock Index Future is collected and processed as follows.

1. According to candlestick analysis，every Open, High, Low and Close prices in the previous trading day are selected using daily minute ticks.
2. The four prices between futures market opening and stock market opening are found.
3. The four prices during ten minutes after stock market opening are also gathered.

The input tuples of neural network model will be obtained after completing the above steps. Then, the model will be trained with part of collected data. At last, there will be a performance validation with the rest of collected data. The preprocessing steps of modeling neural network are showed as Figure 4-2.

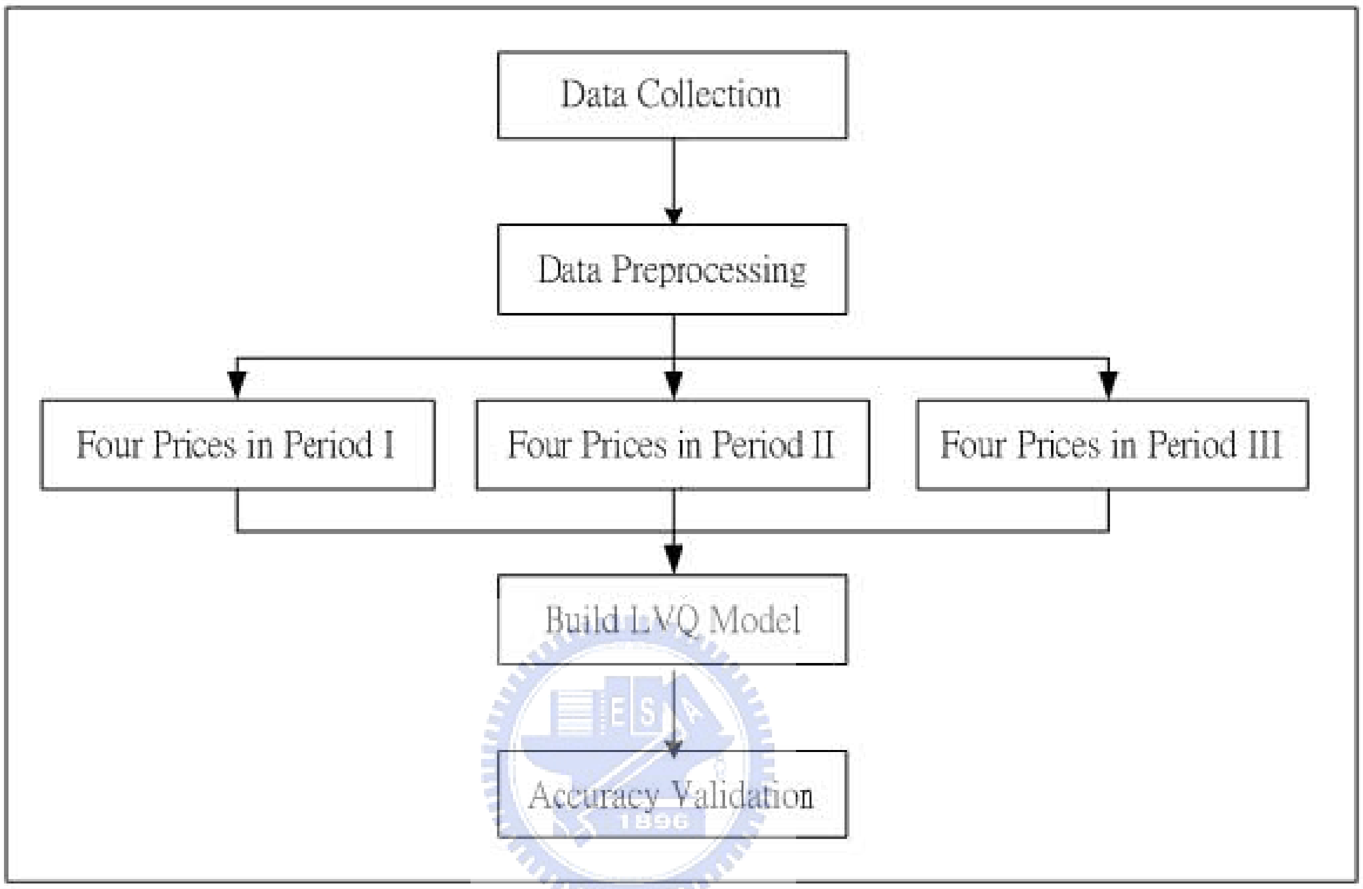
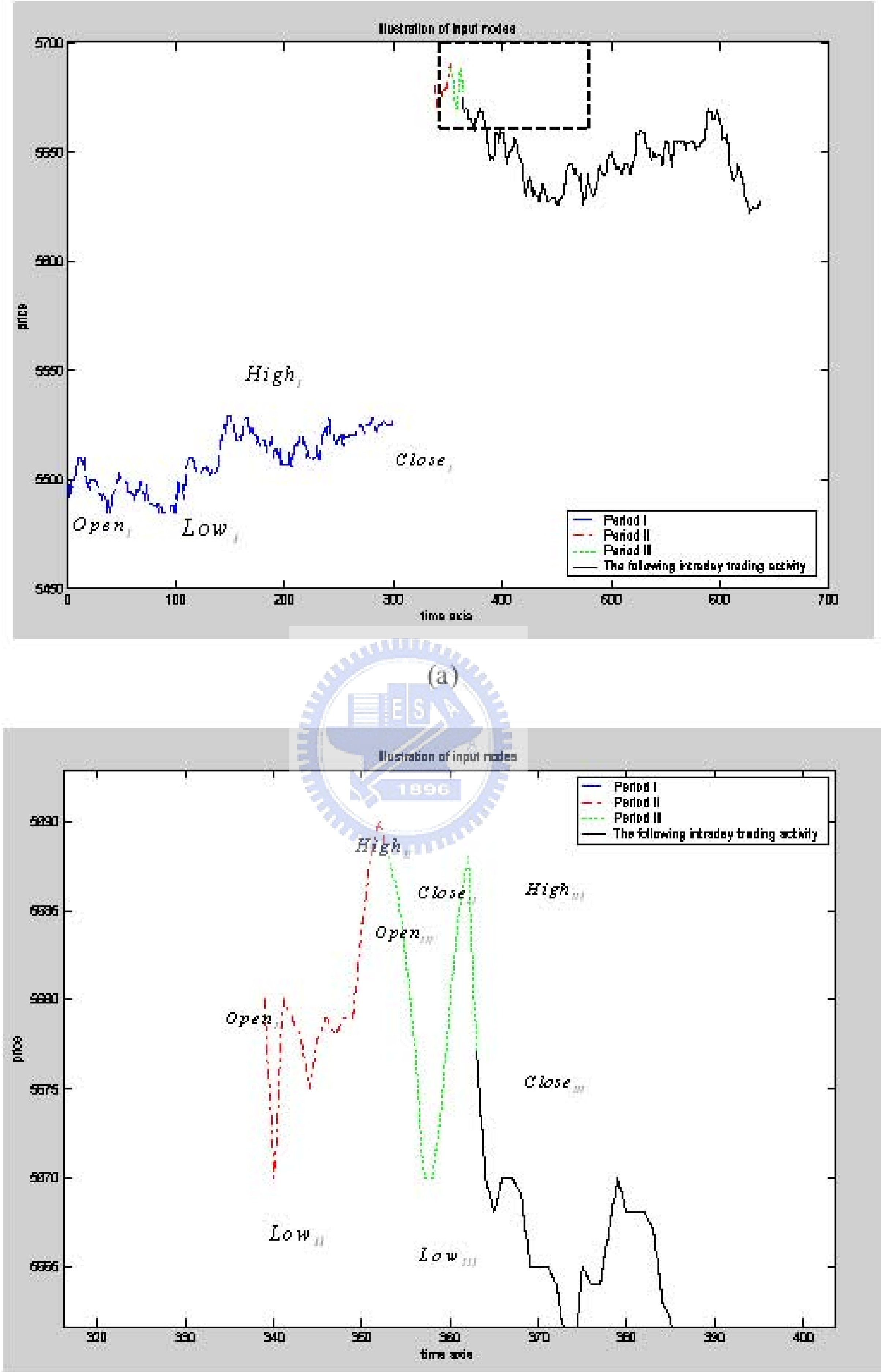


Figure. 4-2. An illustration of simulation steps

As candlestick analysis discussed in section 2.3 reveals, it is the belief that Open, High, Low and Close prices have higher information content than other prices to extract featured prices of intraday trading. These four prices are used as input nodes in neural network model. The four prices respectively between futures market opening and stock market opening (period II) and between stock market opening and its following ten minutes (period III) are also parts of the input nodes in neural network model. Based on fractal theory, it is known that pattern feature extraction is irrelevant to the length of period. The four prices in period II and period III are chosen to gather the featured prices during these two periods.

Consequently, four prices respectively in period I, period II and period III are considered as the factors to influence the following intraday trading activities and taken as input nodes of neural network. The input factors are shown as Figure 4-3.



(a)

(b)

Figure 4-3. An illustration of input factors of neural network, respectively Open, High, Low and Close prices of Period I, Period II and Period III.

Figure 4-3 (a) shows all input factors in neural network. Figure 4-3 (a) marks Open, High, Low and Close prices in Period I. Figure 4-3 (b) magnifies the rectangle area plotted by dash line. It also labeled the four prices in Period II and period III.

4.2. Output Nodes of Neural Network

As the regression model reveals, it is known that returns of these period are highly correlated to $Return_t^{CC}$. $Return_t^{CC}$ is the return between Close price at previous trading day and at this trading day. Since the previous Close price is already known. The output of neural network model is designed for Close price. It is intended for output nodes to analyze weather Close price would higher or lower than Open price at day t. In other words, it is aimed to classify the real body of this candlestick line at day t as black or white.

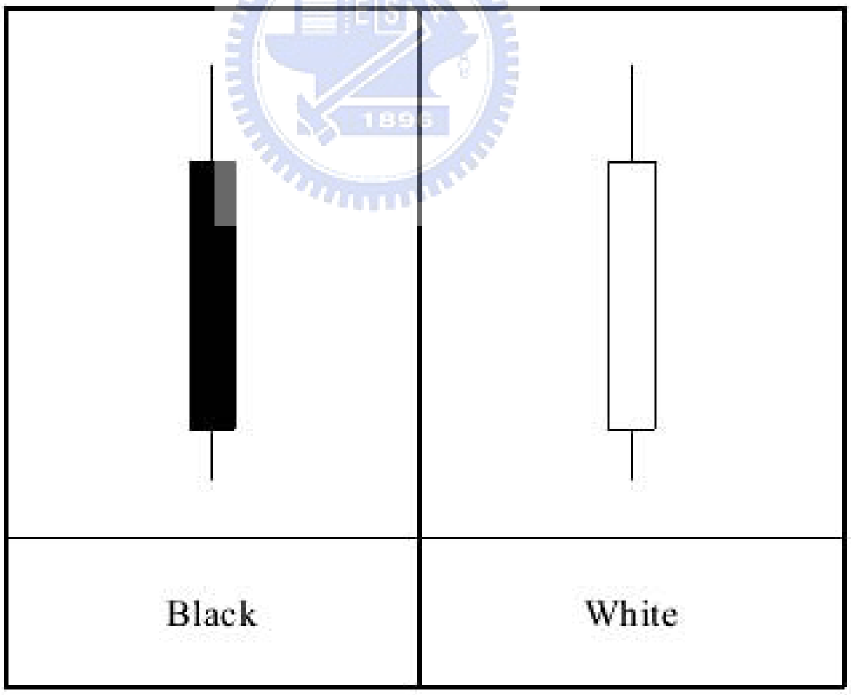


Figure. 4-4 Output node of neural network

Figure 4-4 illustrates the output nodes of neural network. If Close price were analyzed to be higher than Open price, it would be assigned into node "White". Oppositely, if Close price were diagnosed to be lower than Open price, it would be assigned into node "Black".

After the input factors and output nodes are determined, the different simulation on intraday trading activities is launched by Learning Vector Quantization algorithm, a classification algorithm discussed in section 2.5.

5. Empirical Results and Discussion

The minute ticks of Taiwan Stock Index Future between 2001/04/13 and 2002/04/12 are used in this paper. Every minute price in one trading day is characterized as a record, totaling 204 records. 80% of records are taken as the training samples to build the neural network model. The remaining 20% records are tested for the accuracy of this model.

Before this simulation, some indicator must be defined first to see the accuracy of neural network model. Accuracy indicator, ratio of success, is the percentage of correct identification. The formula is listed as follows:

$$S = \frac{T}{T + F} \tag{14}$$

S ：ratio of success

T ：the correct identification number

F ：the incorrect identification number

For a record of data, after it is inputted into the neural network model, if it is classified as white (black) candlestick line while actually it is, the record is defined as correct identification. If it is classified wrong, it is defined as incorrect identification. After all testing samples are finished, the correct identification number and incorrect identification number could be counted. Then Ratio of success is computed according to formula (14).

5.1. Accuracy

The first step is to process the raw data according to Figure. 4-2. According to the regression formula, a sliding window size is chosen as two days. At day t, the minute ticks data of day *t-1* are computed to get four prices in period I, while the minute ticks data of day *t* are used for period II and period III. However, it is necessary for these prices to be normalized. Therefore, these prices are all transformed into the concept of ratio. They are all divided by the Close price at day *t-1*. This transformation is aimed to adapt to the scope of Taiwan Stock Index Futures. To avoid the different index level, the normalization such as ratio is considered as essentially.

These three groups of four prices are inputted as ratio into input vector while black or white candlestick line is outputted as classification results. After data processing, Learning Vector Quantization algorithm is used to build a Neural Network model.

The sequence of the input vectors is randomly selected. There are two different ways to get initial codebooks; one is random selection while the other is exploiting SOM (Self-Organizing Maps) to initialize. Kohonen said that it might be a difficult problem while initializing codebook. Then he suggested that it might be a better strategy to first use SOM for initialization.

The following table reports representative experiment results. 5-1 (a) records results of several experiment initialed randomly; the 5-2 (b) records results of several experiment runs initialed by SOM. The iteration is all 100 for both experiments initializing with random selection and with SOM initialization.

Table 5-1 (a) Statistics of the experiment (Initializing randomly)

experiment run	initial codebooks	learning rate	Accuracy indicator
1	(W_1 , W_2)	0.04	72.5 %
2	(W_1 , W_2)	0.07	70 %
3	(W_1 , W_2)	0.08	70 %
4	(W_1 , W_2)	0.1	67.5%

Table 5-1 (a) Statistics of the experiment (Initializing by SOM)

experiment run	initial codebooks	learning rate	Accuracy indicator
1	(W_1 , W_2)	0.04	75 %
2	(W_1 , W_2)	0.07	72.5 %
3	(W_1 , W_2)	0.08	70 %
4	(W_1 , W_2)	0.1	72.5 %

The initial learning rates in different runs are all smaller than 0.1. It is recommended by Kohonen that learning rate should be initially smaller than 0.1. (W_1 , W_2) are initial codebooks assigned differently in different runs of experiment. $W_1 = \{w_{1i}, i = 1,2,3,...,12\}$, $W_2 = \{w_{2i}, i = 1,2,3,...,12\}$. w_{1i} is the connection between output node "Black" and all input nodes; w_{2i} is the connection between output node "White" and all input nodes. (W_1 , W_2) are expressed in Figure 5-1:

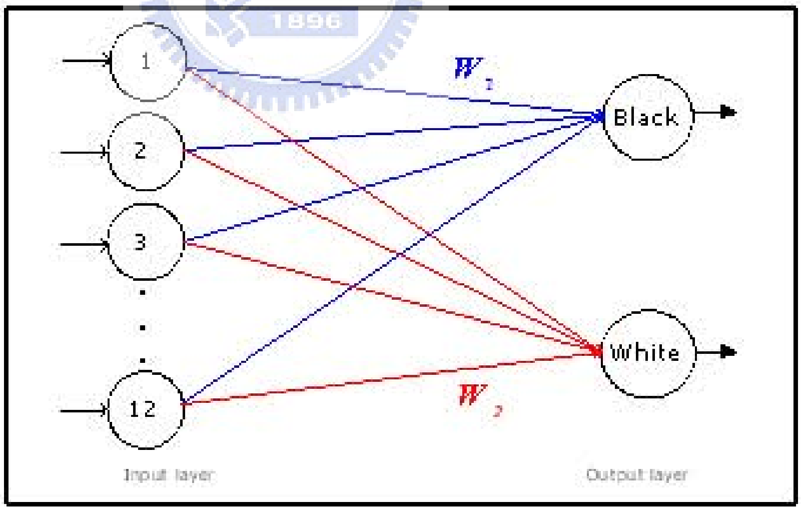


Figure 5-1. The structure of neural network model

Table 5-1 reports the experiment results. 5-1 (a) lists several experiment runs initialized randomly; 5-1 (b) lists experiment runs initialized by SOM. However, the results are not as good as expected. The best ratio of success is only 75%. The input vectors may be represented improperly. Hence, the other way to express the input factor is adopted. The differences between Open and High, Open and Close, Close and Low, High and Low are emphasized and inputted as input vectors in the next experiment. Of course, these differences are formatted as concept of ratio. These inputs are illustrated as Figure 5-2.

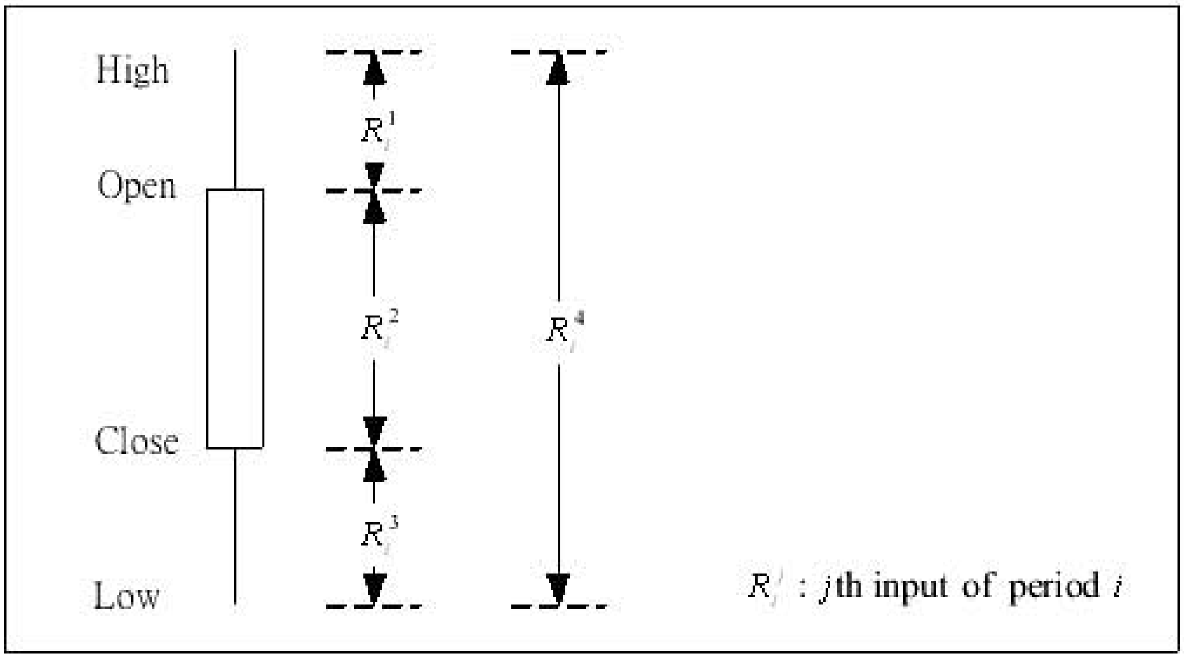


Figure. 5-2. A different formation to describe input factors of neural network

R_i^j is the jth ratio of period i inputted into neural network and is computed in terms of the following formula.

$$R_i^1 = \frac{High_i - Open_i}{Close_i} \tag{14}$$

$$R_i^2 = \frac{Close_i - Open_i}{Close_i} \tag{15}$$

$$R_i^3 = \frac{Close_i - Low_i}{Close_i} \tag{16}$$

$$R_i^4 = \frac{High_i - Low_i}{Close_i} \tag{17}$$

In these formulas, $i = \{I, II, III\}$, which respectively refers to period I, period II and period III. These four ratios describes difference between High and Open, Close and Open, Close and Low and High and Low during period I, Period II and Period III. Note the denominators of these formulas are all $Close_i$. It is intended for this transformation to describe candlestick chart more completely. The experiment results are listed at the following table.

Table 5-2 (a) Statistics of the experiment (Initializing randomly)

experiment run	initial codebooks	learning rate	Accuracy indicator
1	(W_1, W_2)	0.04	82.5 %
2	(W_1, W_2)	0.07	72.5 %
3	(W_1, W_2)	0.08	72.5 %
4	(W_1, W_2)	0.1	70 %

Table 5-2 (b) Statistics of the experiment (Initializing by SOM)

experiment run	initial codebooks	learning rate	Accuracy indicator
1	(W_1, W_2)	0.04	75 %
2	(W_1, W_2)	0.07	70 %
3	(W_1, W_2)	0.08	72.5 %
4	(W_1, W_2)	0.1	72.5 %

It is obviously that the experiment results listed in Table 5-2 outperform these listed in Table 5-1. The ratio of success could achieve 82.5% while the initial learning rate and iteration are all the same. Therefore, Model II is also the neural network adapted in this thesis. Although the best accuracy indicator 82.5% is obtained by initializing randomly, initializing with SOM seems to be more stable. This distinction between Table 5-1 and Table 5-2 could also show the different formats to represent input factor may result in different outcome. However, it is not the focus in this thesis.

5.2. Profitability

The accuracy has been verified, however, the profitability is still the concern. If investors trade according to the output of neural network model, how many returns they could earn? For a record of data, it could be classified correctly or incorrectly. If it is assigned right into black candlestick line, investors could have a short position, and offset it at closing period.

Figure 5-3 shows the prices change of intraday trading. If there is any price higher than the dotted line illustrated in Figure 5-3, investors could have a short position and offset at closing period as Figure 5-3 expressed.

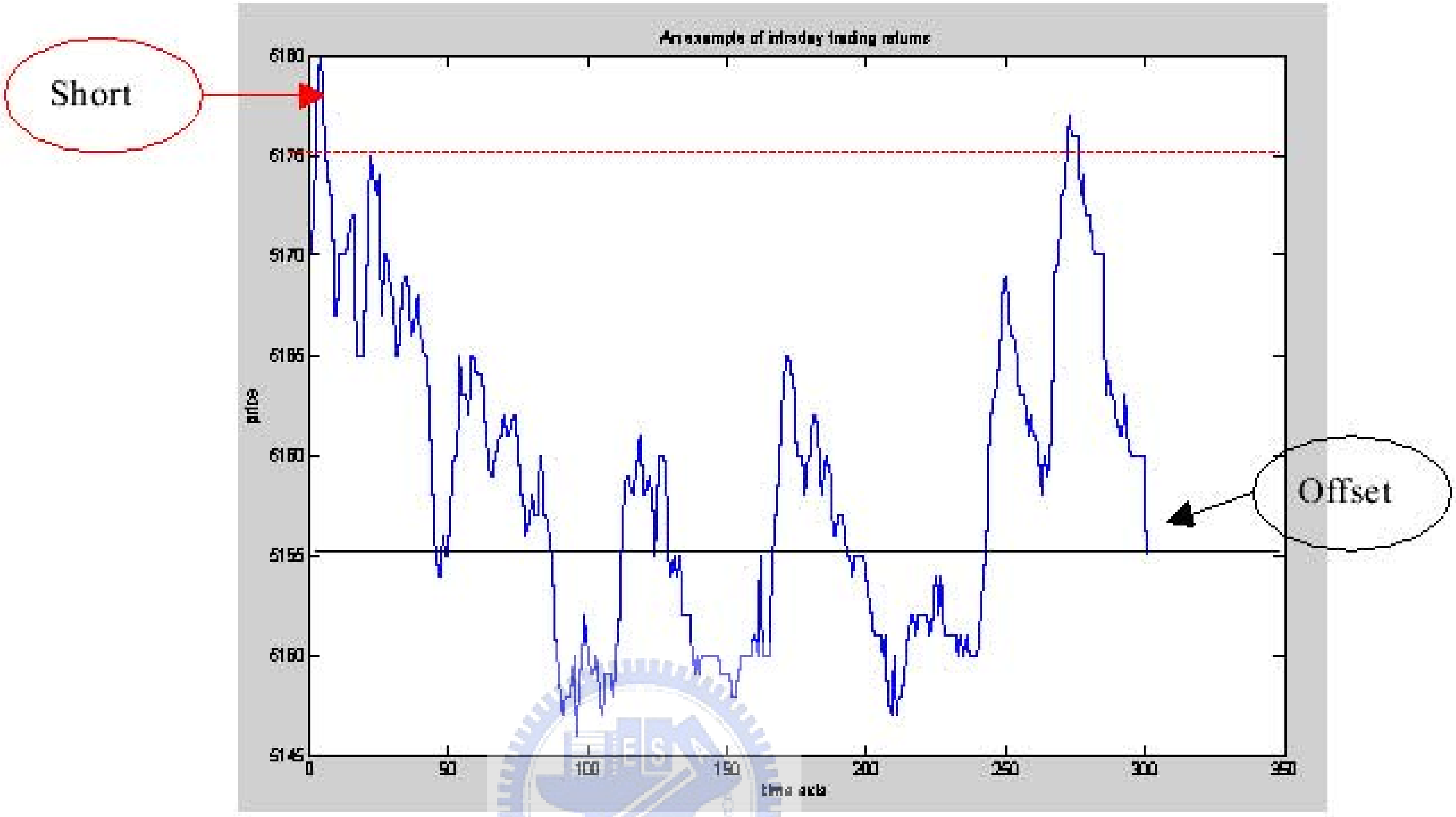


Figure 5-3. An illustration of simple trading rules.

In Figure5-3, the dotted line is a threshold defined by investors. Because Open price and Close price might be close to each other even though the trading day is assigned into “white” or “black” correctly. If investors buy at Open price and sell at Close price, there is no guarantee for revenues. However, by this threshold, there is a fixed range of returns if this trading day is classified correctly. On the contrary, if this trading day is assigned into wrong classification, there is a loss. For example, if Figure 5-3 is classified as white candlestick line, investors would have a loss between the price they buy and Close price.

By this simple trading rule, the following trading returns could obtain while learning rate of neural network model is 0.04 and iteration is 100. 40 tuples of data set are chosen randomly to test the profitability. Threshold is 0.005% higher or lower than Open price.

Table 5-4 Returns of a simple trading rule (Initializing randomly)

experiment run	initial codebooks	learning rate	Accuracy indicator	Profits	Trading times
1	(W_1 , W_2)	0.04	70%	417	19
2	(W_1 , W_2)	0.04	70 %	942	24
3	(W_1 , W_2)	0.04	72.5%	769	29

As Table 5-4 reports, even though learning rate and iteration are all the same, the returns and trading times are different. Trading times is the number of trading. Because even though one trading day is assigned into "white", it is uncertain for prices during trading hours to be 0.005% higher than Open price. If so, there is no trading in the trading day. This are several possible reasons, One is different trading day is selected. Different trading days may have different scope of returns. Other is different accuracy. Definitely, lower accuracy makes fewer returns, however, but it is not necessary for higher accuracy to have more returns. Trading rules are not focus in this thesis. Table 5-4 is just a contrast for the accuracy.

6. Conclusion and Future Work

6.1. Conclusion

This study investigates how information accumulated during non-cash period influences the next trading activities. This phenomenon, overnight effect, originates form restrictions of trading hours. All impact on stock prices is derived by time, but trading hours are discontinuous sections in time-axis. The information accumulated overnight could not be responded directly because there are no trading activities during non-cash period. Overnight information could be reflected only after opening. The phenomenon, overnight effect, is first discussed in this thesis. Discovering the relations of overnight effect and the following trading day is also the intention. A regression model is utilized to test and, then confirm significance of overnight effect. The result of regression model supports the assumption that the gap between two trading days shows most of the overnight information and also has influences on the next trading activities.

After overnight effect has been verified, overnight information should be studied for intraday trading activities. This thesis employs Learning Vector Quantization, one of neural network models, to classify intraday trading activities as white or black candlestick line. Learning Vector Quantization is an important algorithm to identify different classifications. Open, High, Low and Close prices from candlestick analysis are employed to describe intraday activities in terms of higher information contents. Intraday trading activities concerning about overnight effect are simulated by artificial intelligence techniques. Consequently, the importance of overnight effect would be shown by the simulation results. The phenomenon between two following trading hours has been concluded to be an important factor as valuable in this thesis.

6.2. Recommendatory Investment Strategy

For the strategy of arbitrage and speculation would be an extend issue. This conclusion could also be applied to the strategy of arbitrage and speculation. As arbitrage between Taiwan Stock Index Futures (TX) and MSCI Taiwan Stock Index Futures (STW), arbitragers should have a short position on TX and have a long position on STW or have a opposite position on STW and TX. If the analysis on TX is correct, it is possible for investors to lower the trading cost of one. For example, arbitrager has a long position on TX and a short position on STW. If Open price of TX is high, but the analysis shows that there will be a black candlestick line in this trading hours. Arbitragers can dthe long position of TX, and repurchase the position back before closing in order to avoid overnight risks. For arbitrager, total profits by lowering the cost at the loss position should be able to increase, and the fluctuation of prices between the two trading times and make fixed profits should be capable to be avoided. For speculators, if Open price of TX is low but the analysis shows that there will be a white candlestick line in this trading day. Having a long position during trading hours if there is any price lower than Open price to some extent is necessary to be concerned. According to this speculation strategy, purchasing futures at a lower price and offsetting holding positions at a higher price, the profit is earned due to a proper "view" about the intraday trading activities.

Furthermore, overnight effect could also be used in option activities. As we know, options are derivative instruments with high financial leverage. If intraday trading activities were analyzed correctly by using overnight effect, options could be reasonable tools to avoid risk or speculate, even better than investment in futures activities.

6.3. Future Works

Although the importance of overnight effect is verified, how to employ it to analyze intraday trading activities more precisely is always the concern. Further researches should focus on the ability of neural network model to incorporate overnight information into intraday trading analysis. Characterizing the intervals that are respectively between Open price and Close price at day *t-1*, between futures Open and stocks open and between Open and ten minutes after opening at day *t* is an important topic worthy of discussion. The purpose is to approve the accuracy of analysis on intraday trading activities and have a better view to the following trend.

Investment strategy is another future work. Although Close price could be analyzed by neural network model in this thesis, however, no explicit trading rules are proposed. Future works could focus on trading rules because how to hedge and speculate well is always the most interesting issue for investors.

Reference

1. Abhyankar, A.; Copeland, L.S.; & Wong, W., "Uncovering nonlinear structure in real-time stock-market indexes: The S&P 500, the DAX, the Nikkei 225, and the FTSE-100," Journal of Business & Economic Statistics, Vol. 15, 1-14, 1997.

2. Anderson, Torben G. & Bollerslev Tim, "Intraday periodicity and volatility persistence in financial markets," Journal of Empirical Finance, Vol. 4, 115-158, 1997. Anderson, Torben G.; Bollerslev, Tim & Cai, Jun, "Intraday and interday volatility in the Japanese stock market," Journal of international Financial Markets, Institutions and Money, Vol. 10, 107-130, 2000.

4. Balvers, R. J.; Cosimano, T. F. & McDonald, B., "Predicting stock returns in an efficient market," Journal of finance, Vol. 55, 1109-1128, 1990.

5. Berry Thomas D. & Howe Keith M., "Public Information Arrival," Journal of Finance, Vol. 49, 4, 1331-1346.

6. Campbell, JY & Shiller, RJ. , "Stock prices, earnings, and expected dividends," Journal of Finance, Vol. 43, 661-676, 1988.

7. Chan, Kalok; Chockalingam, Mark & Lai Kent W.L., "Overnight information and intraday trading behavior: evidence from NYSE cross-listed stocks and their local market information," Journal of Multinational Financial Management, Vol. 10, 495-509, 2000.

8. Chan, Leo & Lien, Donald, "Using high, open, and closing prices to estimate the effects of cash settlement on futures prices," International Review of Financial Analysis, Vol. 12, 34-47, 2003

9. Chen, N. F.; Roll, R., & Ross, S. A., "Economic forces and stock market," Journal of business, 59, 383-403, 1986.

10. Curcio, R.& Goodhart, C. D., "When support resistance levels are broken, can profit be made? Evidence from the foreign exchange market," Discussion Paper Series, Discussion Paper No.142. LSE Financial Market Group. July 1992.

11. Darrat, Ali F.; Rahman, Shafiqur & Zhong, Maosen, "Intraday trading volume and return volatility of the DJIA stocks: A note," Journal of Banking & Finance, Vol. 27, 2035-2043, 2003.

12. DeGrauwe, P. & Decupere, D., "Psychological barriers in the foreign exchange market," Discussion Paper No. 621. Centre for Economic Policy Research, London, 1992.

13. Feeny, M., "Charting the foreign exchange markets. In: Dunis, C., Feeny, M. Eds. , Exchange Rates Forecasting," New York, 99-163. 1989

14. Fiessa, Norbert M. & MacDonald, Ronald, "Towards the fundamentals of technicalanalysis: analysing the information content of High, Low and Close prices," Economic Modelling, Vol. 19, 353-374, 2002.

15. Gençay, Ramazan "Non-linear prediction of security returns with moving average rules," Journal of Forecasting, Vol. 15,165-174, 1996

16. Gerety, Manson S. & Mulherin, Harold J., "Trading Halts and Market activity: An Analysis of Volume at the Open and the Close, " Journal of Finance, 47, 5, 1765-1784.

17. Kanas, Angelos & Yannopoulos Andreas, "Comparing linear and nonlinear forecasts for stock returns," International Review of Economics and Finance, Vol. 10, 383-398, 2001.

18. Kangas, Jari & Kohonen, Teuvo, "Developments and applications of the self-organizing map and related algorithms," Mathematics and Computers in Simulation, Vol. 41, 3-12. 1996.

19. Kohonen, Teuvo, "Learning vector quantization," Helsinki University of Technology,

Laboratory of Computer and Information Science Report, 1986TKK-F-A-601.

20. Kohonen, Teuvo & Somervuo, Panu, "Statistical Pattern Recognition with Neural Networks: Benchmarking Studies," Neural Networks, 61-68, IEEE International Conference on San Diego, 24-27 July 1988,1, 1988.

21. Kohonen, Teuvo, "The self-organizing map," Proc. IEEE 781464-1480, 1990

22. Kohonen, Teuvo, "Self-organizing maps, Springer Series in Information Sciences," Vol. 30, 1995

23. Kohonen, Teuvo, "Improved visions of learning vector quantization," Neural Networks, 1990. IEEE International Conference, 17-21 June 1990.

24. Lam, K.P., "Predictability of intraday stock index," Neural Networks, Proceedings of the 2002 International Joint Conference on 12-17 May 2002, 3, 2156-2161

25. Lee C.I. & Mathur L., "The influence of information arrival on market microstructure: evidence from three related markets," Financial Review, Vol. 34, 1-26, 1999.

26. Lee, K.H. & Jo, G.S., "Expert system for predicting stock market timing using a candlestick chart," Expert Systems with Applications, Vol. 16, 357-364, 1999.

27. Lee, Tian Shyue & Chen, Nen Jing, "Investigating the information content of non-cash-trading index futures using neural networks," Expert Systems with Applications, Vol. 22, 225-234, 2002.

28. Leshno, Moshe & Spector, Yishey, "Neural Network Prediction Analysis: The Bankruptcy Case," Neurocomputing, Vol. 10, 125-147, 1996.

29. Lo, Andrew W.; Mamaysky, Harry & Wang, Jiang, "Foundations of Technical Analysis: Computational Algorithms, Statistical Inference, and Empirical Implementation," The

Journal of Finance, Vol. 4, 1705-1765, 2000.

30. Malliaris, Mary & Salchenberger, Linda. “Using Neural Networks to Forecast the S&P 100 Implied Volatility,” Neurocomputing, Vol. 10, 183-195, 1996.

31. Mandelbrot, Benoit B. "The Variation of Certain Speculative Prices," Journal of Business, 36, 394-419, 1963.

32. McCulloch, W. S. & W. Pitts., “A Logical Calculus of the Ideas Immanent in Nervous Activity,” Bulletin of Mathematical Biophysics, Vol. 5, 115-133, 1943.

33. Menkhoff, L., “The noise trader approach _ questionnaire evidence from the foreign exchange,” J. Int. Money Finance, Vol. 3, 547-564, 1998.

34. Mok, Debby M.Y.; LAM K. & LI W., “Using daily high/low time to test for intraday random walk in two index futures markets,” Review of Quantitative Finance and Accounting, Vol. 14, 381-397, 2000.

35. Stoll, Hans R. & Whaley Robert E., “The Dynamics of Stock Index and Stock Index Futures Returns,” Journal of Financial and Quantitative Analysis, Vol. 25, 4, 441-468, 1990.

36. Qi, Min & Maddala, G.S., “Economic Factors and the stock Market: A New Perspective,” Journal of Forecasting, Vol.18, 151-166, 1999.

37. Yao, Jingtao & Tan Chew L., “A case study on using neural networks to perform technical forecasting of forex,” Neurocomputing, Vol. 34, 79-98, 2000.

Appendix A (SPSS results)

A-1 Correlations between $Return_t^{overnight}$, $Return_t^{cc}$, $Return_t^{oo}$ and $Return_{t-1}^{oc}$

		$Return_t^{overnight}$	$Return_t^{cc}$	$Return_t^{oo}$	$Return_{t-1}^{oc}$
$Return_t^{overnight}$	Pearson Correlation	1	.541**	.312**	-.071
	Sig. (2-tailed)	.	.000	.000	.321
	N	216	200	200	200
$Return_t^{cc}$	Pearson Correlation	.541**	1	.079	-.123
	Sig. (2-tailed)	.000	.	.269	.082
	N	200	200	200	200
$Return_t^{oo}$	Pearson Correlation	.312**	.079	1	.895**
	Sig. (2-tailed)	.000	.269	.	.000
	N	200	200	200	200
$Return_{t-1}^{oc}$	Pearson Correlation	-.071	-.123	.895**	1
	Sig. (2-tailed)	.321	.082	.000	.
	N	200	200	200	200

** Correlation is significant at the 0.01 level (2-tailed).

A-2. Stepwise regression results

Variables Entered/Removed

Model	Variables Entered	Variables Removed	Method
1	$Return_t^{cc}$	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
2	$Return_t^{oo}$	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
3	$Return_{t-1}^{oc}$	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).

a Dependent Variable: $Return_t^{overnight}$

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.541	.293	.290	45.96067
2	.605	.366	.360	43.63629
3	.865	.748	.744	27.59655

a Predictors: (Constant), $Return_t^{cc}$

b Predictors: (Constant), $Return_t^{cc}$, $Return_t^{oo}$

c Predictors: (Constant), $Return_t^{cc}$, $Return_t^{oo}$, $Return_{t-1}^{oc}$

ANOVA

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	173512.959	1	173512.959	82.141	.000
	Residual	418251.916	198	2112.383		
	Total	591764.875	199			
2	Regression	216652.023	2	108326.012	56.890	.000
	Residual	375112.852	197	1904.126		
	Total	591764.875	199			
3	Regression	442497.209	3	147499.070	193.678	.000
	Residual	149267.666	196	761.570		
	Total	591764.875	199			

a Predictors: (Constant), $Return_t^{cc}$

b Predictors: (Constant), $Return_t^{cc}$, $Return_t^{oo}$

c Predictors: (Constant), $Return_t^{cc}$, $Return_t^{oo}$, $Return_{t-1}^{oc}$

d Dependent Variable: $Return_t^{overnight}$

Coefficients

		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
Model		B	Std. Error	Beta		
1	(Constant)	-.312	3.250		-.096	.924
	C2C	.280	.031	.541	9.063	.000
2	(Constant)	-.543	3.086		-.176	.861
	C2C	.269	.029	.520	9.143	.000
	O2O	.134	.028	.271	4.760	.000
3	(Constant)	1.405	1.955		.719	.473
	C2C	.114	.021	.222	5.549	.000
	O2O	.824	.044	1.667	18.793	.000
	O2C	-.861	.050	-1.534	-17.221	.000

a Dependent Variable: *Return*

Excluded Variables

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics Tolerance
1	$Return_t^{oo}$	.271	4.760	.000	.321	.994
	$Return_t^{cc}$	-.004	-.064	.949	-.005	.985
2	$Return_t^{cc}$	-1.534	-17.221	.000	-.776	.162

a Predictors in the Model: (Constant), $Return_t^{cc}$

b Predictors in the Model: (Constant), $Return_t^{cc}$, $Return_t^{oo}$

c Dependent Variable: $Return_t^{overnight}$