The Dynamic Influence of Stock Future CNA 50 on Stock Index in 2015 Stock Market Crash: Based on VAR Model

Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

Mr. Shuyi Wang

Submitted in partial fulfilment of the requirement of the Masters in Financial Risk Management, School of Business, Trinity College, Dublin.

Supervised by Professor Taufiq Choudhry

July 2017

Trinity Business School

Trinity College Dublin

The University of Dublin

Signed Statement

MSc. Financial Risk Management, School of Business, Trinity College Dublin

I have read the University's code of practice on plagiarism. I hereby certify this material, which I now submit for assessment on the programme of study leading to the award of MSc. Financial Risk Management is entirely my own work and has not been taken from the work of others, save and to the extent that such work has been cited within the text of my work.

Student ID Number: 16300284

Name of Candidate: Shuyi Wang

Signature of Candidate:

Date: July 27th, 2017

The Dynamic Influence of Stock Future CNA 50 on Stock Index in 2015 Stock Market Crash: Based on VAR Model

Author: Shuyi Wang

Abstract

The well-known lead-lag relationship between future and spot market is often applied in price discovery and market efficiency. In addition, the leading power are normally considered generated by future market. Therefore, institutional investors use futures to hedge the risk of financial market or to arbitrage. However, this article proves that, in China mainland stock market during the most recent financial crash, the very leading effect of future is not significant. Moreover, two index futures from different market (namely CNA 50 is from foreign market Singapore Exchange, and CSI300 index future is from domestic market) are compared to see, during the financial crisis, if the foreign index future is the reason resulting in the crash of domestic stock market. Using a 5-minute high frequency data from both markets, this paper runs causality test and Impulse Response Function based on the VAR system to investigate influence of futures to their spot. In addition, the paper also establishes VECM to identify the causality because of the existence of co-integration. The paper finds out that the spot market also causes the foreign future to fall, and it is not safe to conclude that overseas capital can short the mainland market through CNA 50 index future. However, the domestic index future shows that the lead-lag relationship stably exists.

Executive Summary

This paper comes out to question the widely accepted idea that the future market is leading the spot market and to doubt the strategy that a foreign index future is able to short the domestic stock market. This paper chooses to investigate in the case in China mainland stock market during the most recent financial crash, involving the data of CNA 50 index future, CSI300 index future and SCI index which represents the performance of the whole stock market. From 15th , June to 26th , August in 2015, this paper uses 2548 5-minute high frequency data to build up VAR and VEC models. Moreover, the paper also identify the possible co-integration between the futures and the spot to see if there is a long-term relationship existing. To see the causality and how the leading impact last, the paper then applies Granger Causality Test and Impulse Response Function.

Based on these models and related tests, the paper proves that no long-term relationship exists between CNA 50 index future and the China mainland stock market, which proving that the leading effect is not significant during that specific period. Moreover, in the financial crash, CNA 50 index future is the granger cause of spot market, however, spot market exerts the same influence on the future and more importantly, showed by the IRF, the impact from spot market is stronger than that from the future. Compared to CNA 50 index future, CSI300 index future is co-integrated with the spot market, and the leading effect comes from the spot market and affects the future during that period.

Due to the results, this paper weakens the view about the leading effect of future market, especially in emerging market during a financial crisis. Also, manipulating a foreign future to short the spot market, say using CNA 50 in Singapore Exchange, is not reliable. At least, CNA 50 index future is much more powerless than CSI300 index future in doing so.

Acknowledgement and Thanks

I would like to express my gratitude to the experts who generously offered information, insights and opinions for this research project: [including but not limited to: Dr. Ranadeva Jayasekera, Associate Professor of Accounting and Finance, Trinity Business School; Dr. Shaen Corbet, School of Business, Trinity Business School; Prof. Michael Peardon, School of Mathematics, Trinity Business School; Dr. Darach Golden, support scientist at the Centre for High Performance Computing, Trinity College; Dr. Tapas Mishra, Associate Professor of Financial Econometrics at the Centre for Banking, Finance and Sustainable Development, School of Business, University of Southampton. Without the input and passionate participation of these individuals, the validation survey could have been successfully conducted.

Finally, I must express my profound gratitude to my supervisor, Professor Taufiq Choudhry, Professor of Finance, Department of Banking and Finance, Southampton Business School, University of Southampton, without whom accomplishment of this thesis would not have been possible. He was beyond generous with his time, valuable guidance and continuous encouragement.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

SSE	Shanghai Security Exchange
SGX	Singapore Exchange
OSE	Osaka Securities Exchange
SCI	Shanghai Securities Composite Index
CSI500	Shanghai-Shenzhen Smallcap 500 Index
CNA 50	Xinhua China A50 index
CSI300	Shanghai-Shenzhen 300 Index
QFII	Qualified Foreign Institutional Investor
QDII	Qualified Domestic Institutional Investor
DJIA	Dow Jones Industrial Average
VAR	Vector Auto Regression
VECM	Vector Error Correction Model
ECM	Vector Error Mechanism
EGARCH	Exponential Generalized Autoregressive Conditional Heteroscedasticity
VMA	Vector Moving Average
FTSE 100	Financial Times Stock Exchange 100 Index
S&P 500	Standard & Poor 500 Index
DAX30	Deutscher Aktienindex 30 Index
CFFE	China Financial Futures Exchange
CSRC	China Securities Regulatory Commission
CRDW	Co-integrating Regression Durbin-Watson
DGP	Data Generating Process
OLS	Ordinary Least Square
IRF	Impulse Response Function

List of Appendix

1. Introduction

Conception about lead-lag relationship is commonly considered in different financial markets, especially between futures and spots. Usually the future market possesses some earlier information that spot market is yet to obtain, thus resulting that price changes in future market are in advance. Moreover, it is also highly likely that the spots may follow those changes, and individual or institutional investors in the spot market predict prices through this mechanism. To explain why the mechanism exists, there are three main reasons. Take index spot and future markets of China as examples, first, index future markets start trading 15 minutes earlier than the opening of spot markets and close 15 minutes later than the closing of stock market. This results in different expectations of movement of spot price. Common things that investors prefer to inspect the price moving trend of futures before they start to trade in the stock market happen. Second, the spot market is always more frictional, that is to say the cost of transaction in the future market is lower than that of spot market. Since future markets can open an opposite position given the current position, investor tend to make their moves immediately in a future market, thus resulting in an advanced move of future market. Third, since an index is always a package of tens or hundreds of stocks, even huge fluctuation of single stock may hard to affect the index because of its diversification. However, the move of the index means big waves of its underlying stocks. The fact of diversification of index makes the futures more stable than the spots. Another question that is worth considering is that is this relationship significant between a future from foreign market and its underlying asset from domestic market. Also, one is reasonable to doubt the strength of the very connection from the remoted future may not be stronger enough to affect the spot. A study (Covrig et al. 2004) in researching this relationship between Singapore Exchange (SGX) and Osaka Securities Exchange (OSE) suggests that both markets have contributed in price behaviours, and it also reminds people of considering the contemporaneous future instruments which are traded in different countries when trading in the spot market. Although it has mentioned about the contribution of foreign markets to price movement of spot market, it does not take account of emerging markets, such as China mainland market. In addition, it is hard to conclude whether the connection becomes weaker or stronger in a financial crash without considering seriously. The situation may be distinct under the circumstances mentioned above.

This paper investigates this relationship in an extreme period which is the most recent financial crisis in China mainland stock market in 2015. It is common and obvious that investors use futures to hedge the risk of an overheated market. However, a vicious short-selling is unreasonable and results in economic depression quite often. From 15^{th} , June to 26^{th} , August in 2015 in China, Shanghai Securities Composite Index (SCI) slumps from 5178 points to 3373 points, seeing a 32.11% sharp drop in 17 days. CNA 50 future contract saw a similar drop in the same period. Media and public opinion consider this a vicious short-selling operated by foreign funds. Some studies have confirmed that such future contract as CSI500 does not affect the stock market much. The leading effect of this future is not the dominate power leading the market to fall. By contrast, the stock market was leading the future in that period. Besides, to compare with CNA 50, this paper also involves CSI300, which is the most popular future contract, to see if this domestic future can be more influential.

The view that CNA 50 future contracts exert influence on the stock market of another country based on several reasons. First, CNA 50 index takes stocks as its underlying assets from the China stock market, and those stocks cover about 33% value of the whole market. This provides the base that trading in another country is somewhat equal to trading in China mainland stock market. Moreover, due to the high barrier to enter into China mainland stock market for foreign investors, CNA 50 in Singapore Exchange (SGX) offers an available method to arbitrage, especially when there was no short positions in Chinese capital market. Second, Qualified Domestic Institutional Investors (QDII) allow domestic capital outflowing into international markets. Concept QDII means that institutional investors are approved to make oversea investment in stocks, bonds or any other securities controllably under the circumstances that Renminbi cannot be redeemed and capital market is not fully opened. This qualification authorizes some investors to build their positions in international markets. The last reason is that CNA 50 future contracts traded in SGX, a multi-sided market containing international investors, receive information earlier than some domestic investing instrument traded in some conservative market such as Taiwan Exchange. This has been proved by Roope and Zurbruegg (2002) in 2002. Therefore, it is reasonable to infer that China mainland stock market would response to price changes even more slowly due to its conservative.

This paper firstly tries to answer if the focused future contract CNA 50 is capable of impacting the spot stock market of China and to specify how the impacts last. This is aiming to identify the possibility that an institutional investor may short another country's spot market by shorting a foreign future during a specific period. In order to proceed, this paper investigates co-integration between different time series to see the possible equilibrium in their long run. Besides, based on the vector auto regreesion system, the research also need to determine the causality between those time series, identifying the order of happening of price changing. Apart from the foreign instrument, this paper will also apply the same mechanism in a domestic future, CSI300 index future, because it is meaningful to witness the very influence on the spot market may come more from a latercomer. In applying these methodologies, the paper decides to utilize 5-minute high frequency data from those time series mentioned. They can be collected from BLOOMBERG.

2. Literature Review

Since future market is able to affect the behaviours in price changing of spot, this phenomenon or effect is called leading effect. The leading effect of index futures on spot market has attracted much attention during the past twenty years. Studies use this effect to examine the market efficiency and price discovery between different financial markets. An early study using 1-minute high frequency irregularly spaced data observes that the futures on S&P index lead its spot at least by ten minutes (de Jong & Nijman 1997), inferring a leading role the future market plays. This article applies a covariance estimator to measure the added information that cannot be omitted based on the discrete 1-minute data. This research not only shows that new information is captured more quickly by future market than spot index, but also indicates that the future has led its spot ten-minutes. A ten-minute leading in the daily trading, if true, can be a good indicator for fund managers or even a government in financial crisis to response properly. Moreover, indicated by volatility of each market, Tse (1999) proves that the Dow Jones Industrial Average (DJIA) future contributes more in price discovery than its spot does, especially when considering impact of bad news. After the confirmation of co-integration between the future and the spot, he quantifies this issue by taking both vector error correction model (VECM) and Hasbrouck common-trend model into consideration, and the latter indicates that DJIA futures share 88.3% of market information. The article also researches in volatility spillover effect by an EGARCH (1,1)-*t* model in a bivariate VAR environment. Although the effect comes from every direction of different products, the manipulation of the methods gives conclusion in a more stronger influence from future market to other markets, an evidence which reinforces a leading role of future market. Information arrival and aggregation can be mainly used by future market and be reflected in the mispricing of spot market. Alphonse (2000) employing error correction mechanism (ECM) and vector moving average (VMA) shows that futures represents 95% in price discovery, given the example from CAC 40 future and its spot. This paper also implies that the existence of co-integration relationship affects the choose of model. Another study from Brooks, Rew and Ritson (2001) also apply this theory to build a trading strategy for FTSE 100. Given the good forecasting ability the VAR model possesses, the authors apply a 10-minutes leading period into predicting the price moving trend. However, under the real-world circumstances (which means taking transaction costs into consideration), the paper nevertheless reveals that the strategy is unable to outperform a passive benchmark even if given the ten-minute forecasting. The academic results are not always applicable in commercial analysis in this case. To investigate this relationship between different markets in different countries, Roope and ZurbRuegg (2002) compare Taiwan exchange to Singapore exchange. Using the Hasbrouck and Gonzalo–Granger methodologies that can extract information from different markets, the authors find that the information from each exchange has influence on future products of each other. They also remind both exchanges and investors to consider relative markets that have a similar future product or instrument. Most early studies mainly agree with the theory that the spillover effect exerts influence on spot market, and they do ignore somewhat that the very impact may do so from spot market to future market. Continuously, doubt about the lead-lag relationship rises due to a research, using a genetic programming approach, on the Nikkei spot index and future price under an extreme period but not normal periods. This study highlights "major changes" from spot market affect future prices more than the changes on future prices affect spot markets (Lien et al. 2003). Therefore, it is highly likely that, during a financial crash period, the transmission of negative information is against the lead-lag direction. Further study in Mexico (Zhong et al. 2004) and Greece (Floros, Christos; Vougas 2007) also reinforce the lead-lag theory that future market is more informationally efficient. In the emerge market of Mexico, the authors report that a two lag of spot index is found, indicting a two-day leading of the future instruments. This conclusion was also based on the idea of error correction and, more innovatively, was extended by utilizing EGARCH model. However, it should be noted that the Greek article still prove it true under an extreme circumstance—the crisis period 1999-2001. Many published papers so far have proved the lead-lag relationship true in the domestic spot market and future market. Futures play a role in implying and serving the changes in spot market. However, this relationship remains uncertain between different markets in different countries. Article in 2002 (Roope and Zurbruegg) cited that Singapore exchange may react more quickly than Taiwan exchange, but it also mentioned that regulatory force can be a factor resulting in a later response of Taiwan and the multi-trading environment of Singapore exchange may lead to a faster response. This analysis can be applied in many emerge markets, such as China mainland stock market. Using a more complicated but more accurate method, Li (2007) applies Markov-switching vector error correction model to examine the dynamic relationship between futures (S&P 500, FTSE 100 and DAX 30) in mature markets and spot market in emerging markets (Brazil and Hungary) under both low and high variance regimes. It proves that the price discovery is not as informationally efficient in future market as in spot market, or, the lead-lag relationship cannot be proved significant in the different types of markets. Those later studies running more complicated models to question the leading effect of future market.

It is also reasonable to consider the interaction between home exchange and satellite exchanges, which can be translated into the reaction of one future to another issued in overseas markets. An empirical research (Covrig et al. 2004) investigates in the information linkage of Nikkei 225 index and its domestic future in home exchange, a foreign future in SGX, representing the satellite exchange. This research concludes that the foreign futures contribute 42% of the price discovery, however, the domestic one contributes 33% of the discovery. An Indian research (Sehgal & Dutt 2016) comparing Indian NSE Nifty index to its future products traded on three international exchanges, namely, SGX, OSE and CME, suggests that its home exchange dominate the trading. Given the two researches, the information linkage is blurred, and none of foreign exchanges can control or dominate other exchanges. However, those studies confirm that even futures from different markets have influence on domestic spot or future market.

So far, studies mentioned above have firmed a logic in how the futures may lead the spots, or say, the changes resulted by bad news in future markets affect the pricing issue of spot markets. Apart from that, many recent studies using different kinds of data and methodologies try to reveal some unusual phenomena. Those studies give a positive attitude that the very influence can be reversed in some extreme circumstances, such as a financial crash.

On the other hand, the situation is not the same in China when considering the domestic index future lead the spot future market. This was investigated in 2010 after the issuing of the first index future CSI300. A study (Yang et al. 2012) weakens this lead-lag relationship in the Chinese market and argues that cash market plays a more essential role in functioning the leading effect, it also attributes this unusual statistical result to the high barrier to entry and the infancy stage of an emerging market. This finding partly agrees with the Li's theory that future products in emerging market do not serve the price discovery well. Evidence from Thailand (Judge & Reancharoen 2014) market also shows that there was a leading power from spot market directing future market, showing a reverse connection against empirical conclusion. The article also applies a conventional error correction model to find out the lead-lag effect between SET50 index future and its spot. Furthermore, scholars also try to define whether it is because of the issuing time of CSI300IF affect its underlying. Before and after introduction are taken into consideration, He, Wang and Du (2014), employing multifractal detrended fluctuation analysis, indicate that the factor resulting in multifractal transits from long-range correlation to fat-tail probability distribution in Chinese domestic market. Although the article does not mention which market is in the leading state, it does conclude that the introduction of CSI300IF improve the market efficiency in the aspect of reducing risks. Meanwhile, another article (Cao et al. 2014), employing the same methodology but running a high frequency data, points out a bidirectional relationship is casual between spot and future market, but the index future possesses a stronger impact. Combined the two articles, china mainland spot and future market do not necessarily follow the conventional theory about the leading role of future market. Many studies using high frequency data (one-minute or five-minute data) strengthen the relationship between spot and future market. Based on the thermal optimal path (TOP), the most recent study (Gong et al. 2016) modeling the relationship between CSI300 and HSI, S&P 500 states that the local future CSI300IF leads the spot two days, but the futures from Hong Kong and the US only lead one day. Another valuable perspective of this study is that its method (TOP) gives a dynamic evolution in measuring the lead-lag relationship. Moreover, the author states that there is no need to consider any other factors in the very methods.

This paper tends to find out whether a future CNA50 traded in Singapore market will considerably affect Chinese stock market, or the underlying market during a financial crash period in China. Meanwhile, compared to the product from different market, a domestic future product CSI300IF is considered as well in finding out arbitrage opportunities. This paper is also seeking to support or undermine the leading effect of future, or to qualify the effectiveness of the effect in different market during a specific period, based on VAR or VEC environment and high frequency data.

3. Data and Institution Difference

Shanghai Security Exchange (SSE) is the most important component of China's capital market on mainland. SSE was initially named Securities Brokers' Association and was the earliest stock exchange of China in 1860s took place in Shanghai. In the meantime, both national and international investors can trade stocks, bonds, government bonds and futures through brokers in the very association. The modern well-known exchange was reopened in 1990. With a 25-year development, the exchange has facilitated its service with world class trading system, the biggest security trading database and fully functional website. The 1382 listed companies in Renminbi ordinary shares (A-shares) and special Renminbi denominated shares (B-shares) constitute the Shanghai Security Composite Index (SCI). By the first half year of 2017, the exchange has achieved $4434.64 billion for market capitalization. This index aims to reflect the performance of stock market on Chinese mainland and is able to catch any information may affect the market, therefore the index is applied to represent Chinese stock market.

In April 2010, The Shanghai and Shenzhen 300 index future (CSI300IF) was successfully launched by China Financial Futures Exchange (CFFE), an institution authorized by China Securities Regulatory Commission (CSRC), and was aiming to offer a proper hedging tool for investors and institution. This event is not only a signal that China has its index future but also a beginning that the Chinese stock market is no longer a "one-side market". This future is based on its underlying index CSI300, consisting of 300 stocks contributing about 70% of the whole market value and about 59% of the circulation market value. Moreover, the index is correlated with SCI in a very strong sense. The index has gained its recognition since it was established and so has its underlying stocks. The high coverage of market value and separated weight of constituent stocks make this index hard to be manipulated and the most appropriate one to be an underlying of futures.

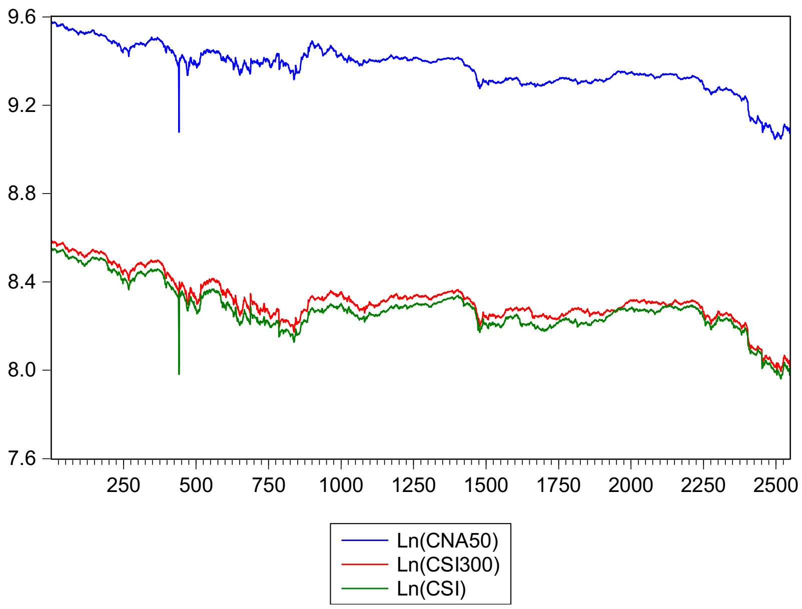
Before CSI300, Xinhua China A50 (CNA 50) index, on the other hand, was launched by FTSE Xinhua Index Ltd., a corporation found by Xinhua Financial Network and Financial Times and London Stock Exchange (FTSE), and its future was first traded in Singapore Exchange (SGX) in 2006 aiming to reflect the performance of Chinese stock market. The index consists of the top 50 market value stocks of China, representing 33.2% of the whole value of market. Designed for investment needs, the future is available for both Chinese investors and Qualified Foreign Institutional Investors (QFII). International investors tend to trade on this index futures contract due to a high barrier to entry the Chinese mainland market. Some more specific differences of both futures are listed below:

Table 1 Comparison of CNA 50 and CSI300 Future Contract

Product	Xinhua FTSE China A50 Future Contract	Shanghai and Shenzhen 300 Index Future Contract
Code	CN	IF
Underlying Asset	FTSE Xinhua China A50	Shanghai and Shenzhen 300 Index
Contract Size	$1*China A50	¥300*CSI300
Settlement Month	Mar, Jun, Sep, Dec	Current month, Following month, and next two seasonal months
Trading Hours	9:15am-11:35am, 1:00pm-3:05pm	9:15am-11:30am, 1:00pm-3:00pm
Margin	6%-10% of contract value	10% of contract value
Final Settlement Price	Last price of the spot	The arithmetic means of price of the last two hours before closing
Settlement Date	The last second day before the month	The final settlement date of the month
Transaction Mode	Electronic Transaction	Electronic Transaction

Based on the different trading period mentioned in the table above, this paper chooses a contemporaneous period that both markets are open to investors. These periods are from 9:25 am to 4:25 am, and from 1:00 pm to 2:55 pm. Besides, this paper utilizes 5-minute high frequency data and three time series are present in a logarithm form. The paper aims to analyse the issue of financial crisis from 15th June 2015 to 26th August 2015. Removing the data that are mismatching the same trading period, this paper gets an amount of 2548 for each time series. Besides, to present properly, three series are in log arithmetic form and can be seen in the figure 1 below. It can be seen that these time series follow a quite similar decline trend after 0 period. They suffer a sudden plummet at around 400 periods. Moreover, the decline of China A50 index is seemingly a bit earlier than the same changes of SCI and CSI300, and this could be a warning signal of financial crash.

Figure 1 Index Movement



4. Methodology

Based on a high frequency data analysis, the main goal of this study is to identify the lead and lag relationship between SCI, which represents the performance of the whole stock market of mainland China, and two futures may affect stock market during a financial crash. The main methodologies were built in a general vector auto regression (VAR) system. The paper will also take account causality and impulse response into overall analysis.

4.1. Stationarity

Stationarity can be one of the most important issues in econometrics. It requires a time series to be applicable and the regressive results to be meaningful. The main idea of the method is to ensure the existence of a unit root. If there is a unit root, the series is non-stationary; If there is no unit root in the series, the series is stationary. The very first step of time series analysis is to identify their stationarity, since non-stationary series cause spurious regression in application of ordinary least square (OLS). Fountis and Dickey (1979) investigated in unit root testing and named it DF testing. DF test is based on an autoregressive process, which can be viewed below:

$$y_t = \rho y_{t-1} + \varepsilon_t \tag{4.1}$$

where, ε_t is a white noise process based on the assumptions:

$$E(\varepsilon_t) = \mu$$
$$Var(\varepsilon_t) = \delta$$
$$Cov(\varepsilon_i \varepsilon_j) = 0, i \neq j$$

where, μ and δ are constants that do not change throughout the time. In the equation *(3.1)*, to determine the stationarity of series y_t, if $|\rho|$<1, then the series is stationary. However, if $|\rho|$>1, then y_t is explosive and is meaningless. In practice, equation *(3.1)* can be transformed into:

$$\Delta y_t = (\rho - 1)y_{t-1} + \varepsilon_t \tag{4.2}$$

In the hypothesis testing:

$$H_0: \rho - 1 = 0; H_1: \rho - 1 < 0$$

where, rejecting null hypothesis is to say that the series is stationary, and rejecting alternative hypothesis is to say the series is non-stationary. Moreover, stated in their theory, their article shows a data generating process (DGP) under three circumstances. However, DF test has its weakness in indicating possible auto regression in the error process. Therefore, their further study (Fountis & Dickey 1989) in multivariable autoregressive time series involves metrics into the regression, which is also mentioned as vector auto regression. Moreover, Sims, Stock and Watson (1990) model a linear time series, with some or all of variables have a unit root, to conclude that those variables are integrated or co-integrated. This provides evidence that such time series can be characterized by Granger causality test. In confirming the stationarity of time series, the autocorrelation of residuals makes DGP far too complicated in the first order auto regression. To avoid this complexity and to enhance the power of testing unit roots, the augmented Dickey-Fuller (ADF) test tries to increase the order of lagged term in eliminating the autocorrelation, which can be called augmented term.

$$\Delta y_t = \delta * y_{t-1} + \sum_{i=1}^{k} \delta_i * \Delta y_{t-i} + \varepsilon_t \tag{4.3}$$

The unit root of function can be examined through the equation in calculating student-t ratio for δ. Similarly, to accept null hypothesis is to confirm the existence of unit roots and is implying the series will cause spurious regression.

Based on the same idea as ADF test does, Phillips and Pierre (1986) offer another unit root test, namely Phillips-Pierre test (PP test). The test creates a new statistic t_p that is *t*-distributed to measure $\rho - 1$. It is presented below:

$$t_p = t_{\widehat{\rho-1}} (\frac{\gamma_0}{f_0})^{\frac{1}{2}} - \frac{T(f_0 - \gamma_0) s_{\widehat{\rho-1}}}{2 f_0^{\frac{1}{2}} \hat{\sigma}} \tag{4.4}$$

In the equation *(3.4)*, $t_{\widehat{\rho-1}}$ is t-statistic of $\rho - 1$, $s_{\widehat{\rho-1}}$ is the standard deviation of $\rho - 1$. To accurate the estimation, this paper will manipulate this test to determine the existence of unit root.

4.2. Co-integration and Causality

Non-stationary time series cannot be regressed properly. However, time series, such as stock price and index price, are subject to some kind trend. Both stochastic and deterministic trend can lead a time series to be non-stationary. In this case, it is also meaningful for investors to find out a relative stable relationship that are between two non-stationary time series and can combine the two series into one stationary series, making the series more predictable. This relationship is called co-integration. The testing of co-integration should be carried out after the conformation of stationarity. It should be noted that the original time series are always non-stationary, they nevertheless can be determined as non-stationarity data. A more common sense is that if the first or the second difference of series of time series is stationary, then the data can be reported as weak stationary and can be used into analysis. In the case, in a two dimension of time series, co-integration (Engle & Granger 1987) test aims to determine whether there is a long-term equilibrium relationship between variables.

Considering the co-integration in a two-dimensional environment, Engle and Granger (1987) provide an efficient two-step procedure to identify as well as to estimate the relationship. This two-step approach is based on an assumption that the two series, namely x_t and y_t, are integrated of order d, noted as $x_t \sim I(d)$ and $y_t \sim I(d)$. Given a simple linear regression in estimating the equation:

$$y_t = \beta_1 * x_t + \varepsilon_t \tag{4.5}$$

the main idea is to identify whether the error term $\varepsilon_t = y_t - \beta_1 * x_t$ is a stationary series at the 0 order of integration, denoted as $\varepsilon_t \sim I(0)$. If ε_t is tested stationary, it can be concluded that y_t and x_t are co-integrated of order zero, and that β_1 is called the co-integrating parameter. In the process of test the stationarity of ε_t, accepting null hypothesis means a non-stationary error term, implying no co-integration exists. In this paper, there will be three possible co-integration relationships among SCI, CSI300 and CNA 50.

In the process of unit test of residual series, ADF test is normally manipulated and accepting null hypothesis means no co-integration. The formula of ADF test is presented as equation *(3.3)*. Another testing method, Co-integrating Regression Durbin-Watson (CRDW), reported by Sargan and Bhargava (1983) can also be used to determine if the residual series is stationary showing co-integration between those time series. However, another article (Engle & Yoo 1987) states that DW-statistic is not robust for the lack of limiting distribution. Therefore, the CRDW is valid only for preliminary testing. This paper will mainly depend on the conclusion based on ADF testing. The co-integration test is available not only in finding the specific linear relationship between variables but also in avoiding the spurious regression by introducing the error correction mechanism (ECM). Reported by Johansen (1988), a paper efficiently describes the relationship, both static and dynamic, between variables. This mechanism, aiming to correct possible error in the short run, can be more accurate and useful if the co-integration has been confirmed. A simple form of this mechanism is presented below:

$$ECM = y_t - y_t^* = y_t - \beta_1 * x_t = u_t \tag{4.6}$$

where u_t represents the deviation of y_t in the long run.

It can be difficult to define a cause and outcome, neither from econometrical or statistical perspective. Earlier studies mainly consider covariance a path to track the cause. However, it is obviously that $cov(x, y) = cov\ (y, x)$ cannot give a clear illustration about their causality. Granger (1987) suggests that, in one of a bivariate VAR system (listed below), this basic idea is to estimate the parameters of the function and the existence of parameters indicates a causal relationship between dependent and independent variables.

$$y_t = \alpha_0 + \sum_{i=1}^{k} \alpha_i y_{t-i} + \sum_{j=1}^{k} \beta_j x_{t-j} + \varepsilon_t \tag{4.7}$$

Based on this idea, it can be easier to conclude a cause for an outcome, especially after the building of VAR. The null hypothesis that $\beta_1 = \beta_2 = \cdots = \beta_j = 0$, if cannot be rejected significantly, means *x* does not granger cause *y*.

4.3. Vector Auto Regression and Vector Error Correction Model

Vector auto regression was first introduced by Sims (1980), concerning in foreseeing the interaction of mutually connected time series and in analysing the dynamic impacts of stochastic disturbance. The original idea of this method is about to treat each endogenous variable as lagged term, which means, on the right side of the equation, there will be only endogenous variables, thus avoiding the structural requirements. As a reduced form, VAR takes advantages of the noneconomic restrictions. The regression results under OLS are more consistent and efficient. A general form of VAR can be represented as:

$$y_t = G_0 + G_1 y_{t-1} + G_2 y_{t-2} + \cdots + G_p y_{t-p} + \varepsilon_t \qquad (4.8)$$

where G_0 is a $(n\times1)$ vector of constants, G_p is a $(n\times n)$ vector of coefficients and ε_t is $(n\times1)$ vector of white noise innovations. An important issue of utilizing VAR is about how to choose lagged order wisely. To determine a proper lagged order, Schwarz offered a criterion, namely Schwarz Information Criterion (SIC), to decide the optimal lag length to ensure a more efficient order that can be used in co-integration. It should be noted that there are many standards can be employed to get the optimal lag, such as mean square error (MSE) and Akaike (1974) Information Criterion (AIC). However, stated by Reimers (1992) in his paper, SIC performs better in identifying the optimal lag length. Therefore, this paper uses SIC to get the most suitable lag length.

Questioning a lead-lag relationship between spot and future market, this paper focuses on determining the leading effects that may be produced from each market. Based on bivariate VAR system, this paper sets variables in the equations below:

$$R_{s,t} = \alpha_0 + \sum_{i=1}^{n} \alpha_i R_{s,t-i} + \sum_{j=1}^{n} \beta_0 R_{f,t-j} + \varepsilon_{s,t} \qquad (4.9a)$$

$$R_{f,t} = \alpha_0 + \sum_{i=1}^{n} \alpha_i R_{f,t-i} + \sum_{j=1}^{n} \beta_j R_{s,t-j} + \varepsilon_{f,t} \qquad (4.9b)$$

Where $R_{s,t}$ represents the changes of spot index and $R_{f,t}$ represents the changes of index that a future underlies. Each equation treats influence that affects the left-hand side of the equation as lagged term of both spot and future market. Besides, given another circumstance that two time series are co-integrated, to build a more accurate model, vector error correction model (VECM) should be considered. This is based on the 'Granger Representation Theorem' which gives certainty in an error correction mechanism. The mechanism takes the existence of co-integration as a necessary condition and was described by Hylleberg and Mizon (1989) as an approach that combines both long run equilibrium and short run disequilibrium. The general VEC model can be seen as a co-integrating restricted form of VAR, and VEC aims to correct an unusual shock happening at time *t-1* in the following period *t*. Like a bivariate VAR mentioned, VEC model can be presented:

$$\Delta R_{s,t} = \delta_0 + \sum_{i=1}^{m} \delta_i \Delta R_{s,t-i} + \sum_{j=1}^{m} \delta_j \Delta R_{t-j} - \lambda(R_{s,t} - R_{s,t-1}) + \nu_t \qquad (4.10a)$$

$$\Delta R_{f,t} = \delta_0 + \sum_{i=1}^{m} \delta_i \Delta R_{f,t-i} + \sum_{j=1}^{m} \delta_j \Delta R_{s,t-j} - \lambda(R_{f,t} - R_{f,t-1}) + \nu_t \qquad (4.10b)$$

Due to this logic, this paper will build VAR model for those series that is not co-integrated and build VEC model for those do co-integration.

4.4. Impulse Response Function

Impulse Response Function (IRF) aims to measure the dependent variables' changes triggered by a shock of one unit standard deviation added to the innovations in a VAR system. This method was first introduced by Pesaran and Shin (1993) and is able to reflect the responsiveness of endogenous variables at the left-hand side of an equation in a VAR system. The paper utilizing IRF tends to see how will the dependent variables response to the shock. Considering a VAR (p):

$$y_t = A_0 + B_1 y_{t-1} + B_2 y_{t-2} + \cdots + B_p y_{t-p} + \varepsilon_t \qquad (4.11)$$

This equation can be transformed into:

$$y_t = C + \sum_{s=0}^{\infty} (\Psi_s P)(P^{-1}\,\varepsilon_{t-s}) = C + \sum_{s=0}^{\infty} (\Psi_s P)\omega_{t-s} \qquad (4.12)$$

where, Ψ_s is coefficient matrix, P is non-singular matrix that satisfies $PP^{-1} = \Omega$; ω_t is a white noise process.

5. Empirical Analysis

5.1. Descriptive Statistics

As mentioned, the 5-minute high frequency time series among three indices were chosen to be traded contemporaneous from 15th June 2015 to 26th August 2015. Due to a deterministic trend that three series possess, one is reasonable to consider the stationarity of the data. Combined with descriptive information given below, the conclusion that the series are not stationary can be further proved positive. For instance, the skewness is not zero indicting a deviation of mean, and kurtosis is greater than 3 implying the observations are not normally distributed. Jarque-Bera statistics of series are different from zero also means that the residual does not obey normal distribution. In addition, earlier studies have proved that financial time series always possess 'fat tail' in their distributions, thus a stationarity test will be required. To identify the co-integration in the series, one should ensure the stationarity of series and a stationarity test is necessary.

Table 2 Descriptive Statistics

	Ln(CNA50)	*Ln(CSI300)*	*Ln(SCI)*
Mean	9.368885	8.315455	8.276860
Median	9.385508	8.306588	8.269386
Maximum	9.578654	8.585601	8.551482
Minimum	9.045154	7.994885	7.961394
Std. Dev.	0.099487	0.109541	0.108968
Skewness	-0.732004	0.106326	0.246758
Kurtosis	4.226787	3.909642	3.845959
Jarque-Bera	387.3309	92.64835	101.8354
Probability	0.000000	0.000000	0.000000
Sum	23871.92	21187.78	21089.44
Sum Sq. Dev.	25.20952	30.56209	30.24322
Observations	2548	2548	2548

5.2. Stationarity Test

Since the non-stationary data results in spurious regression and makes the whole research meaningless, the paper carries unit root test to identify this property of data. Moreover, co-integration requires the series to be non-stationary in the first place, otherwise there is no need to apply the theory. The original data was put into both ADF and PP testing formula. As expected, the original data is not stationary. This conclusion comes from the chart given below. The original series, marked as *ln*(*SCI), ln(CSI300)* and *ln(CNA50)*, have a *t*-statistic that is significantly less than 2 and p-value greater than 5%, meaning the null hypothesis is failed to reject. Therefore, unit roots exist in the series and those series are non-stationary. However, this cannot be the final conclusion for the econometric defines the data weak stationary under a condition that the first or second difference of the series can be tested stationary. In this case, the first difference form of data was retested and those time series are defined stationary. It can be seen in the same chart that marks *D(SCI), D(CSI300)* and *D(CNA50)* as the first difference form of the series. The *t* ratios are greater than two, and p-values are less than 5%. This indicates that the null hypothesis that the series has a unit root is rejected, and the alternative hypothesis is accepted, which proves those time series are stationary in their first difference form.

Table 3 Unit Root Test

TEST	*Augmented Dickey-Fuller*		*Phillips-Pierre*	
	t-statistic	p-value	*t*-statistic	p-value
Ln(SCI)	-0.573513	0.8739	-0.769893	0.8268
Ln(CSI300)	-1.114421	0.7123	-1.315644	0.6244
Ln(CNA50)	-1.128637	0.7066	-1.380509	0.5934
D(ln(SCI))	-38.43337	0.0000	-87.50389	0.0001
D(ln(CSI300))	-39.62322	0.0000	-84.77885	0.0001
D(ln(CNA50))	-40.22021	0.0000	-84.30560	0.0001

The conclusion that the original time series are not stationary or can be defined as weak stationary after first difference can be drawn. The fact that those series are at the first order of integration is the base of co-integration test.

5.3. Co-integration Test

Since the co-integration requires the series to be non-stationary in the first place and tries to prove an equilibrium relationship between two series in the long run, the order of integration of series must be the same. Different order of integration of time series cannot be proved co-integration. Given by the unit root test that the three series are stationary at their first difference, which means they have the same order of integration, thus there might be co-integration relationship between them. As mentioned, in a two-dimension series, a two-step procedure given by Engle and Granger (1987) is most effective. Using Engle-Granger two-step method, the paper mainly concerns the error term of each equation below:

$$\ln(CSI) = \beta_1 \times \ln(CSI300) + \varepsilon_{t_{csi300}} \tag{5.1}$$

$$\ln(CSI) = \beta_2 \times \ln(CNA50) + \varepsilon_{t_{cna50}} \tag{5.2}$$

In the process of testing residuals of each function, this paper chooses both ADF test and Phillips-Pierre test to ensure the existence of a unit root in the residual series. The testing results of residual series are listed below:

Table 4 Unit Root Test of Residual Series

Test	*Augmented Dickey-Fuller*		*Phillips-Pierre*	
	t-statistic	p-value	t-statistic	p-value
Residual1	-3.960993	0.0017	-4.443961	0.0002
Residual2	-1.679452	0.4417	-1.883256	0.3405

In the table above, *residual1* represents the error term series $\varepsilon_{t_{csi300}}$, equally after subtracting the co-integration term from dependent variable. Rejecting the null hypothesis, both ADF and PP test indicate that the series does not contain a unit root, implying that the series is at order zero of integration, namely $\varepsilon_{t_{csi300}} \sim I(0)$. This results in a co-integration relationship between *ln(SCI)* and *ln(CSI300)*. On the other hand, similarly, *residual2* reports that $\varepsilon_{t_{cna50}}$ is not a stationary series, resulting in that the *ln(SCI)* index and *ln(CNA50)* index are not co-integrated. The results of regression that takes *ln(SCI)* as dependent variable and its two futures as independent variables are listed below:

Table 5 Co-integration Test

Independent Variables	*Coefficient*	*Std. Error*	*t-statistic*	*Prob.*
ln(CSI300)	0.987667	0.002352	419.9817	0.00000
ln(CNA50)	0.985984	0.009453	104.3034	0.00000

Even if the coefficient of *ln(CNA50)* is statistically significant based on *t*-statistic and p-value, the residual series producing from the regression is not stationary at the order zero of integration. Therefore, the only on co-integration relationship can be presented in the equation below:

$$\ln(CSI) = 0.063959 + 0.987667 \times \ln\,(CSI300) + \varepsilon_{t_{csi300}} \quad (5.3)$$

Citing in the definition, the error correction mechanism can be only operated in the circumstance that equilibrium has been recognized in the long run, or equally the existence of co-integration. Therefore, it is more suitable to build a vector error correction model subsequently. On the other hand, since co-integration cannot be proved true between *ln(CNA50)* and *ln(SCI)*, vector auto regression can be built up.

Given two futures that are tightly connected with its spot, it is surprised to find that no co-integration exists between the spot and one of the futures, but another combination does have such a relationship. This results report that, even though CNA 50 captures more than one third of value of stock market, no long-term correlation exist between the future and its underlying during a financial crash. However, CSI300 index future represents co-integration in the same case, and leading effect is thus significant. Although CSI300 was launched only five years before the crash, long-term relationship is more likely to exist than CNA 50. The research attributes this to a four-year emulation trade from 2006 on CSI300 index future before the launch of the index. Investors were then quite familiar with trading mechanism and system. This fact leads to the CSI300 index future capture the properties which a mature future product possesses. Therefore, the later initiation does not affect the importance of CSI300 index future in the long run.

5.4. Vector Auto Regression and Vector Error Correction Model

VECM is designed to measure or to correct disequilibrium in short run so that this disharmony will not impact the equilibrium in the long run. Therefore, a VECM needs the variables being measured to be co-integrated. In this case, to be specific, only *ln(SCI)* and *ln(CSI300)* can be built with a VECM. In addition, those series are taken in the form of first difference. This is to eliminate the probability that non-stationary series cause spurious regression in OLS. On the other hand, since no co-integration can be proved positive between *ln(SCI)* and *ln(CNA50)*, they are able to be built within VAR.

5.4.1. Vector Auto Regression Model

5.4.1.1. Vector Auto Regression

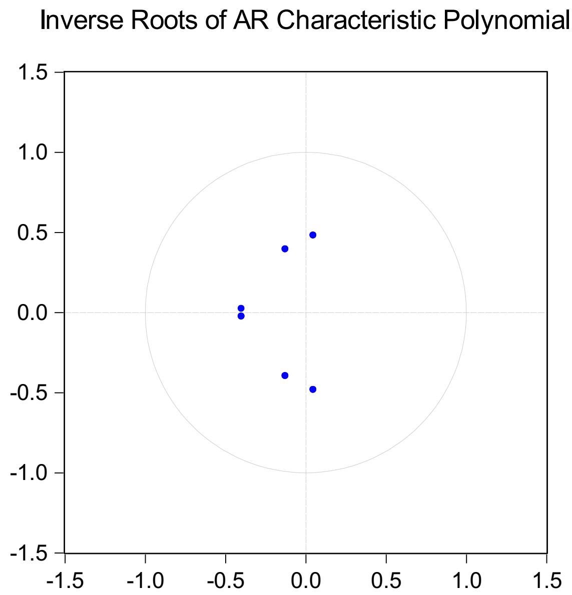
Based on the conclusion drawn from last section, firstly a VAR model will be built between the *ln(SCI)* and *ln(CNA50)*. The lagged length was chosen automatically by Eviews 9.0 and was two that exist in the relationship. Based on the Schwarz Information Criterion (SIC), the final equation chooses three as the lag length. Rerunning the VAR to determine that a three-lagged term is not problematic. The three-lagged term in the equation means that there are three lags in the independent variables affecting the changes of dependent variable. In the real-world environment of this case, there is a 15-minute lag of both market. However, the exact market that affects the other remain unseen before the causality test and this test will be carried out in the next part. From the table below, it is clearly that *t*-statistic of each independent variable is greater than 2, thus the coefficients of those explanatory variables are significantly different from 0.

Table 6 Vector Auto Regression

	D(ln(SCI))	*D(ln(CNA50))*
D(ln(SCI) (-1))	-0.309278	0.184453
Std. Error	(0.04712)	(0.04196)
t-statistic	[-6.56406]	[4.39562]
D(ln(SCI) (-2)	-0.110085	0.128835
Std. Error	(0.05070)	(0.04515)
t-statistic	[-2.17135]	[2.85329]
D(ln(SCI) (-3))	-0.039953	0.035355
Std. Error	(0.04659)	(0.04150)
t-statistic	[-0.85751]	[0.85203]
D(ln(CNA50) (-1))	-0.207741	-0.658900
Std. Error	(0.05301)	(0.04721)
t-statistic	[-3.91875]	[-13.9559]
D(ln(CNA50) (-2))	-0.183810	-0.326114
Std. Error	(0.05754)	(0.05152)
t-statistic	[-3.19429]	[-6.36339]
D(ln(CNA50) (-3))	-0.066470	-0.104730
Std. Error	(0.05243)	(0.04669)
t-statistic	[-1.26783]	[-2.24295]
C	-0.000412	-0.000330
Std. Error	(0.00021)	(0.00018)
t-statistic	[-1.99578]	[-1.79286]

One should consider the stability of the stationary VAR system that currently obtained from table above. This is different from estimating the coefficients of each endogenous variable. Cited in definition of VAR that the model treats all the exogenous variables as a lagged form of endogenous variables put in the right side of the equation, another test should be carried out to see if all the endogenous are identified. This test is called auto regression (AR) test. To a VAR system possessing $n*k$ (where n means the number of endogenous variables and k represents the maximum lag length) roots, the inverse roots of characteristic polynomial should be within in the circle with one as its radius. The representation of AR test can be showed in a circle graph. If the points all fall into the circle, then it is reliable to conclude that the unit roots are wiped out under the current lagged length and the VAR system is stable; If the inverse roots plot outside the circle, the VAR is not stable and unstable VAR system is not able to operate impulse response function (IRF) subsequently. Therefore, the result of AR test is showed in the figure below and it is clearly that all the points are in the circle, proving the stability of VAR system.

Figure 2 AR Test



5.4.1.2. Granger Causality test

The causality test is taken after the confirmation of stability of VAR system. Even if choosing an optimal lag length and involving enough endogenous variables, the current VAR is yet to confirm the causes and outcomes of the system. Therefore, causality test is designed to answer if it is independent variables cause the dependent variable to change. The table showed below demonstrates the result from statistical perspective. The chi-square statistic is to compare the variance of a population and that of a sample. To be specific, the table illustrates that, through *Prob*. less than 5%, all the coefficients in the VAR equations are significantly different from zero. For both *D(ln(CNA50))* and *D(ln(SCI))*, they can affect each other. To be exact, *D(ln(CNA50))* is the granger cause of *D(ln(SCI), and D(ln(SCI)* is granger cause of *D(ln(CNA50).* Therefore, the existence of price changes of CNA 50 to change, and vice versa. Even though there is no long-term relationship between CNA 50 and SCI, it is reasonable to report that a huge price change of CNA 50 is capable of affecting the changes in SCI, and so is SCI. To see this impact more specifically, impulse response function (IRF) will be applied next.

Table 7 Granger Causality Test

Dependent Variable: D(ln(SCI))			
Excluded	*Chi-sq*	*df*	*Prob.*
D(ln(CNA50))	20.63815	3	0.0001
All	20.63815	3	0.0001

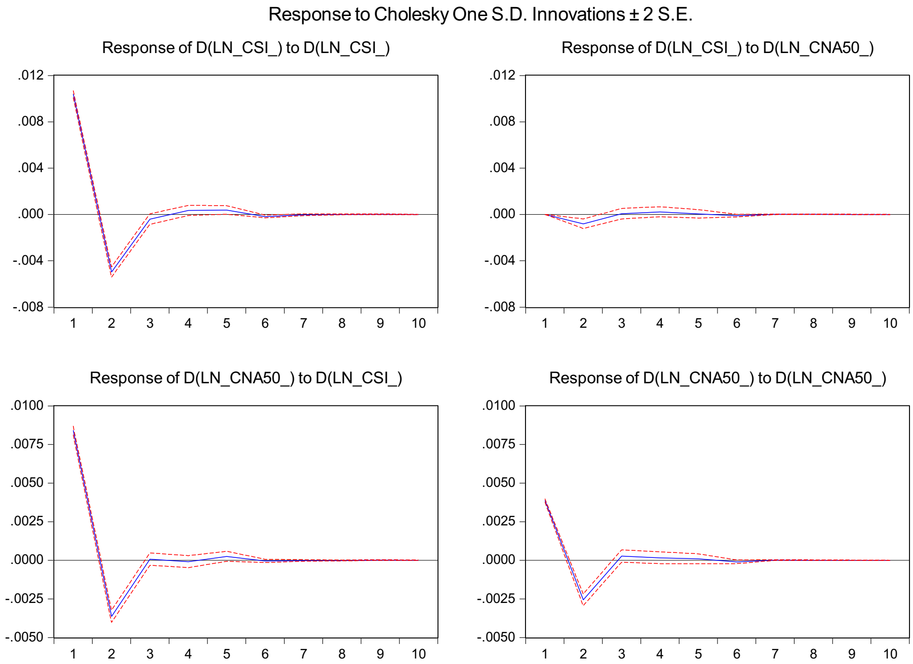
Dependent Variable: D(ln(CNA50))			
Excluded	*Chi-sq*	*df*	*Prob.*
D(ln(SCI))	18.10243	3	0.0004
All	18.10243	3	0.0004

5.4.1.3. Impulse Response Function

Although the statistics give people many method to determine causality, it should be mentioned that the exact reason causing something happen is extremely hard to define. Even granger causality test cannot tell which one is the real cause for it only tells the order that which issue happens firstly. The test only offers a way to see the causality statistically, it is nevertheless unable to show a positive or negative reaction of dependent variable to the changes of independent variables. Besides, the test is not capable of detailing how long the impacts generated by the right-hand side of equation will last. One of the innovation accounting named Impulse response function can settle this problem. After adding a standard deviation of change to the innovation, the method ought to present the influence of endogenous variables and how long the influence lasts. The results is showed in the figure below:

The graph clearly shows the reaction that a shock added to different variables. SCI reacts positively to its own disturbance in the first period. This trend turns to negative in the next period and tends to be stable after the third period, and eventually it fades away to zero in the sixth period. On the other hand, this innovation change from SCI negatively affects the CNA 50 in the first period and is far less stronger than the impact to SCI itself. The response from CNA 50 to SCI is quite similar to that from SCI. Besides, the responses CNA 50 are positive to both SCI and itself and die away after the sixth period.

Figure 3 Impulse Response Function



The results of IRF figure above clearly show that the impact from different index last until the third period. More meaningfully, the impact from CNA 50 to SCI is much more slight than other impacts. Based on this fact, it is reasonable to conclude that price changes from SCI influence CNA 50 more than the other way around.

5.4.2. Vector Error Correction Model

Considering the co-integration between *ln(SCI)* and *ln(CSI300)*, VECM is capable of catching the properties of co-integration that has already been tested and of being more accurate. The results are presented in the table below:

Table 8 Vector Error Correction Model

Co-integration Eq:	*CoinEq1*
D(ln(CSI300) (-1))	1.000000
D(ln(SCI) (-1))	-79.93286
Std. Error	1.96466
t-statistic	[-40.6854]
C	-0.017796

Error Correction	*D(ln(CSI300),2)*	*D(ln(SCI),2)*

CoinEq1	0.022341	0.23289
Std. Error	(0.0006)	(0.00058)
t-statistic	[37.2611]	[40.0402]
D(ln(CSI300) (-1),2)	-0.919149	-0.097809
Std. Error	(0.0621)	(0.06025)
t-statistic	[-14.8000]	[-1.62347]
D(ln(CSI300) (-2),2)	-0.506229	-0.10937
Std. Error	(0.06213)	(0.06028)
t-statistic	[-9.14739]	[-1.81452]
D(ln(SCI) (-1),2)	1.241617	0.461608
Std. Error	(0.06907)	(0.06701)
t-statistic	[17.9756]	[6.88905]
D(ln(SCI) (-2),2)	0.595155	0.212596
Std. Error	(0.0633)	(0.06141)
t-statistic	[9.40195]	[3.46205]
C	7.36E-06	7.86E-06
Std. Error	(0.00021)	(0.00021)
t-statistic	[0.03449]	[0.03799]

The chart shows that only one co-integration relationship in the system, which is incorporated with the conclusion that *ln(SCI)* and *ln(CSI300)* are co-integrated. Reported as *CoinEq1*, the error correction term in the table cites that this long run equilibrium is between one-lag SCI and one-lag CSI300. The coefficient of the term is significant since the *t*-statistic is greater than 2 and represents λ in the *(3.10a)* and *(3.10b).*

5.4.2.1. Granger Causality Test

As does in the VAR, it is necessary to see the cause and outcome in the VEC system. The results are showed in the table below. Depend on the *Chi-square* statistic and *Prob.,* SCI is the granger cause of CSI300. However, CSI300 is not the granger cause of SCI for the reason that *Prob.* is greater than 5% and null hypothesis cannot be rejected. The parameters of SCI is not significantly different from zero. The exogeneity has not been eliminated. Resulting from these statistics, the paper believe that the CSI300 index future is not capable of leading the spot stock market, and other reasons or factors should be considered. On the other hand, SCI representing the spot market is leading the its index future because changes of SCI are the granger cause of changes of CSI300 index. This result is incorporated with former studies concluding that the leading effect was exerted to future market from spot market in emerging market.

Table 9 Granger Causality Test

Dependent Variable: D(ln(CSI300),2)			
Excluded	*Chi-sq*	*df*	*Prob.*
D(ln(SCI),2)	326.1673	2	0.0000
All	326.1673	2	0.0000

Dependent Variable: D(ln(SCI))			
Excluded	*Chi-sq*	*df*	*Prob.*
D(ln(CSI300),2)	3.765798	2	0.1521
All	3.765798	2	0.1521

6. Empirical Results

6.1. Conclusion

This paper applies vector auto regression model (VAR) and vector error correction mechanism (VECM) to investigate whether a foreign future can affect the domestic spot market during an extreme period. In order to accomplish this, the paper utilizes five-minute high-frequency time series during the financial crash. To be more accurate, the paper sorts out the data and only focuses on the contemporary trading period.

In the analysis part, the paper identifies the possible co-integration relationship between the variables and reports that the very relationship only exists between CSI300 index future and SCI spot market. However, this long-term relationship does not happen between CNA 50 index future and SCI spot market. The paper reaches the co-integration results through the two-step procedure proved the most effective by Engle and Granger. Since CNA 50 index future was established and traded in 1999, it is surprising to find that the firstly traded index future does not affect its underlying in the long run. On the other hand, SCI300 index future, even if it was traded from 2010 and was 11 years later than CNA 50 index future did, plays a more important role in leading and affecting the price changing of spot market. The paper attributes this results to some perspectives listed:

1. The four-year emulation trades and pre-education for investors provide the trading of CSI300 index future with mature environment. Most intuitional investors are only allowed to invest in domestic market and they have practiced before the formal launch of the future.

2. The trading volume of CNA 50 index future is much less than that of CSI300 index future. For example, about $2.2 billion was trading on CNA 50 index future monthly, however, more than $53 billion was trading on CSI300 index future monthly. It is reasonable that such a small trading volume of CNA 50 index future leads to the fact that CNA 50 index is not co-integrated with SCI in the long run. In the case relating to a financial crisis, CNA 50 is not capable of affecting SCI in the long run.
3. CSI300 index covers more than 60% of the market value, however, CNA 50 index only covers about 30%. These number can be other reason that the leading effect of CNA 50 index future is not significant.

Based on the result of co-integration, the paper builds VAR and VECM to CNA 50 and CSI300 respectively. This is to investigate the causality between those variables. Since the VAR system has treated exogenous variables as lagged endogenous variables, the Granger causality is able to conclude the causes and outcomes by the order of time. This paper reports that, in the VAR model (which is built between spot SCI and CNA 50 index future), the spot and the future is the granger cause to each other. Therefore, this paper has proved the statement that the price fluctuation of CNA 50 causes or leads the spot market to wave is not a safe conclusion because the spot market also affects the future price much. Furthermore, in the impulse response function (IRF), the result is more specific. The response of SCI to innovation changes of CNA 50 is much slighter than the that of CNA 50 to innovation changes of SCI. The *figure 5* clearly presents this trend. Combining those figures and causality test results together, the paper concludes that during the financial crash, CNA 50 index future and SCI spot are not co-integrated, implying no long-term relationship exists. Besides, CNA 50 index future is the granger cause of changing of SCI spot, however, that influence is very slight; SCI spot is also the granger cause of CNA 50 index future, and the influence is very strong. Therefore, during this specific, shorting CNA 50 index future cannot be exact reason resulting the spot market falling.

This conclusion can be proved further from the fact of another index future CSI300. Domestic CSI300 index future is co-integrated with the spot market SCI. Based on this long-term relationship, this paper builds up VEC model to take account the co-integration. In addition, the paper applies granger causality test, and results show that CSI300 index future is not the granger cause of the spot market; however, changes of SCI is the granger cause of changes of CSI300 index future. To be specific, CSI300 index future was affected or led by the spot market, or equally the leading effect is from spot market to future market, not vice versa. Therefore, even if there was a stable relationship between CSI300 index future and SCI spot in the long run, predicting price changes of SCI through the index future during the extreme period was not reliable.

6.2. Suggestion

Based on the results of the whole research, the paper offer some instructions about trading in both foreign market and domestic market, especially in extreme period:

1. CNA 50 index future is not a proper instrument for hedging the spot market during the financial crisis. Investors should also know that shorting CNA 50 index future is not enough to short the spot market in the long run. VAR instructs that there was a 15 minutes lag between the markets, however, the spot can affect the future more significant.
2. CSI300 index future is capable of hedging the down-trend risk of spot market in the long run, but it should be noticed that the spot is more likely to affect the future. Investors should feel more comfortable consulting CSI300 index future when they are trading in the stock market.
3. Since CNA 50 index future is traded in Singapore Exchange, compared to CSI300 index future, it is more suitable for international capital to hedge or to cross-arbitrage. Furthermore, CNA 50 index future is not bounded to regulation of China mainland market and is more easy to entry. Overseas capital is more likely to influence the mainland market through Singapore Exchange.

7. Reference

Akaike, H., 1974. A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control*, 19(6), pp.716–723.

Alphonse, P., 2000. Efficient Price Discovery in Stock Index Cash and Futures Markets. , (60).

Brooks, C., Rew, A.G. & Ritson, S., 2001. A trading strategy based on the lead-lag relationship between the spot index and futures contract for the FTSE 100. *International Journal of Forecasting*, 17(1), pp.31–44.

Cabras, S. & Morales, J., 2007. Extreme Value Analysis within a Parametric Outlier Detection Framework. *Applied Stochastic Models in Business and Industry*, 23(January), pp.157–164.

Cao, G. et al., 2014. Multifractal detrended cross-correlations between the CSI 300 index futures and the spot markets based on high-frequency data. *Physica A: Statistical Mechanics and its Applications*, 414, pp.308–320. Available at: http://dx.doi.org/10.1016/j.physa.2014.07.065.

Covrig, V., Ding, D.K. & Low, B.S., 2004. The contribution of a satellite market to price discovery: Evidence from the Singapore exchange. *Journal of Futures Markets*, 24(10), pp.981–1004.

Dickey, D.A. & Fuller, W.A., 1979. Distribution of the Estimators for Autoregressive Time Series With a Unit Root. *Journal of the American Statistical Association*, 74(366), p.427. Available at: http://www.jstor.org/stable/2286348?origin=crossref.

Engle, R.F. & Granger, C.W.J., 1987. Co-Integration and Error Correction: Representation, Estimation, and Testing. *Econometrica*, 55(2), p.251. Available at: http://www.jstor.org/stable/1913236?origin=crossref.

Engle, R.F. & Yoo, B.S., 1987. Forecasting and testing in co-integrated systems. *Journal of Econometrics*, 35(1), pp.143–159.

Floros, Christos; Vougas, D. V., 2007. Lead-Lag Relationship between Futures and Spot Markets in Greece : 1999 - 2001. *International Journal of Finance and Economics*, 7(7), pp.168–174.

Fountis, N.G. & Dickey, D.A., 1989. Testing for a Unit Root Nonstationarity in Multivariate Autoregressive Time Series. *Annals of Statistics*, 17(1), pp.419–428.

Gong, C.C. et al., 2016. The lead-lag relationship between stock index and stock index futures: A thermal optimal path method. *Physica A: Statistical Mechanics and its Applications*, 444, pp.63–72. Available at: http://dx.doi.org/10.1016/j.physa.2015.10.028.

He, X., Wang, H. & Du, Z., 2014. The complexity and fractal structures of CSI300 before and after the introduction of CSI300IF. *Physica A: Statistical Mechanics* *and its Applications*, 414, pp.76–85. Available at: http://dx.doi.org/10.1016/j.physa.2014.07.040.

Johansen, S., 1988. Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, 12(2–3), pp.231–254. Available at: http://linkinghub.elsevier.com/retrieve/pii/0165188988900413 [Accessed June 15, 2017].

de Jong, F. & Nijman, T., 1997. High frequency analysis of lead-lag relationships between financial markets. *Journal of Empirical Finance*, 4(2–3), pp.259–277. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0927539897000091.

Judge, A. & Reancharoen, T., 2014. An empirical examination of the lead-lag relationship between spot and futures markets: Evidence from Thailand. *Pacific Basin Finance Journal*, 29, pp.335–358. Available at: http://dx.doi.org/10.1016/j.pacfin.2014.05.003.

Lien, D., Tse, Y.K. & Zhang, X., 2003. Structural change and lead-lag relationship between the Nikkei spot index and futures price: a genetic programming approach. *Quantitative Finance*, 3(2), pp.136–144. Available at: http://www.tandfonline.com/doi/abs/10.1088/1469-7688/3/2/307.

Pesaran, M.H. et al., 1993. *Cointergration and speed of convergence to equilibrium.*, University of Cambridge, Department of Applied Economics.

Phillips, P.C.B. et al., 1986. *Testing for a unit root in time series regression*, Montréal: Dép. de Science économique, Université de Montréal.

Reimers, H.-E., 1992. Comparisons of tests for multivariate cointegration. *Statistical Papers*, 33(1), pp.335–359. Available at: http://dx.doi.org/10.1007/BF02925336.

Roope, M. & Zurbruegg, R., 2002. The Intra-day Price Discovery Process between the Singapore Exchange and Taiwan Futures Exchange. *Journal of Futures Markets*, 22(3), pp.219–240.

Sargan, J.D. & Bhargava, A., 1983. Testing Residuals from Least Squares Regression for Being Generated by the Gaussian Random Walk. *Econometrica*, 51(1), pp.153–174.

Sehgal, S. & Dutt, M., 2016. Domestic and international information linkages between NSE Nifty spot and futures markets : an empirical study for India. *Decision*, 43(3), pp.239–258.

Sims, C.A., 1980. COMPARISON OF INTERWAR AND POSTWAR CYCLES : in Economic Fluctuations . Any opinions expressed are those of the Thomas Doan executed the computations . Funds to support this Monetarism Reconsidered in the interwar period . Furthermore , the money stock eme. , (1430).

SIMS, C.A., H., J.S. & WATSON, M.W., 1990. INFERENCE IN LINEAR TIME SERIES MODELS WITH SOME UNIT ROOTS. *Econometric*, 58(1), pp.113–

144.
Society, R.E. & Journal, T.E., 1989. Cointegration and error correction mechanisms. *The Economic Journal*, 99(395), pp.113–125.

Tse, Y., 1999. P rice D iscovery and V olatility S pillovers in the DJIA I ndex and F utures M arkets. , 19(8), pp.911–930.

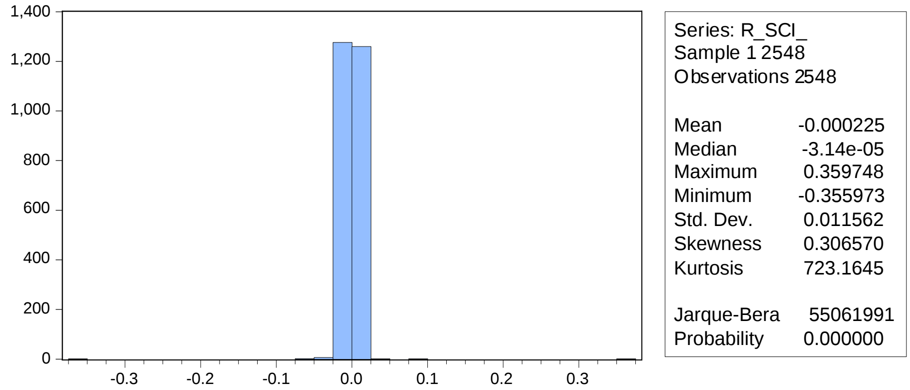
Yang, J., Yang, Z. & Zhou, Y., 2012. Effects of environmental temperature on the onset and the duration of oviposition period of Caiman latirostris. *The Journal of Futures Markets*, 32(2), pp.99–121.

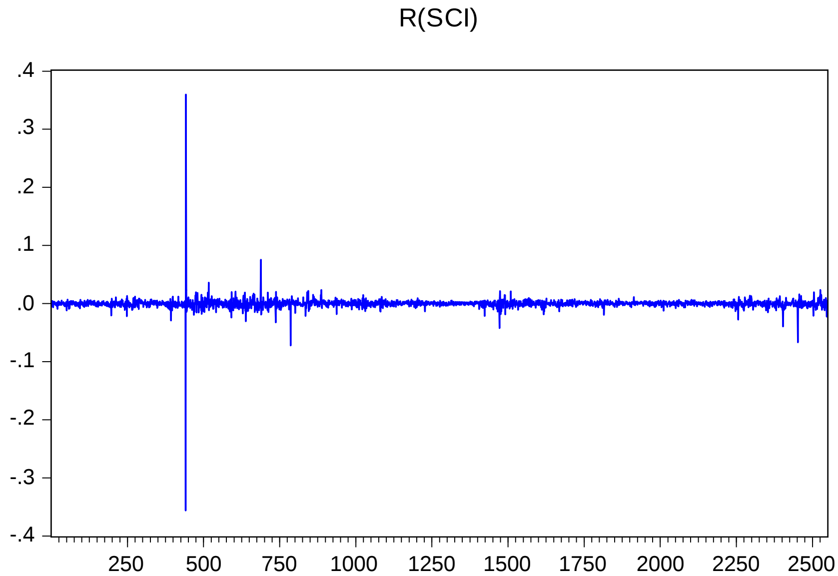
Zhong, M., Darrat, A.F. & Otero, R., 2004. Price discovery and volatility spillovers in index futures markets: Some evidence from Mexico. *Journal of Banking and Finance*, 28(12), pp.3037–3054.

8. Appendices

Appendix A

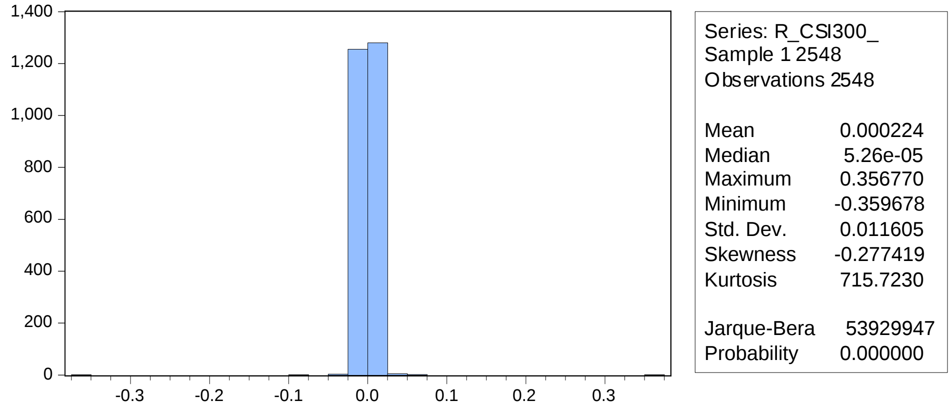
The Return of SCI

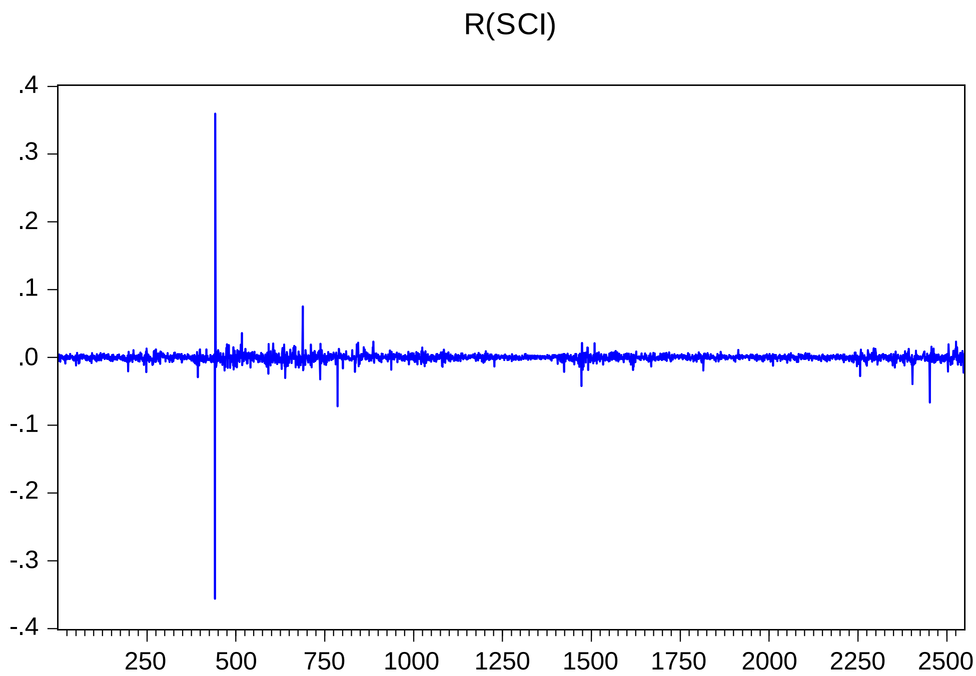




Appendix B

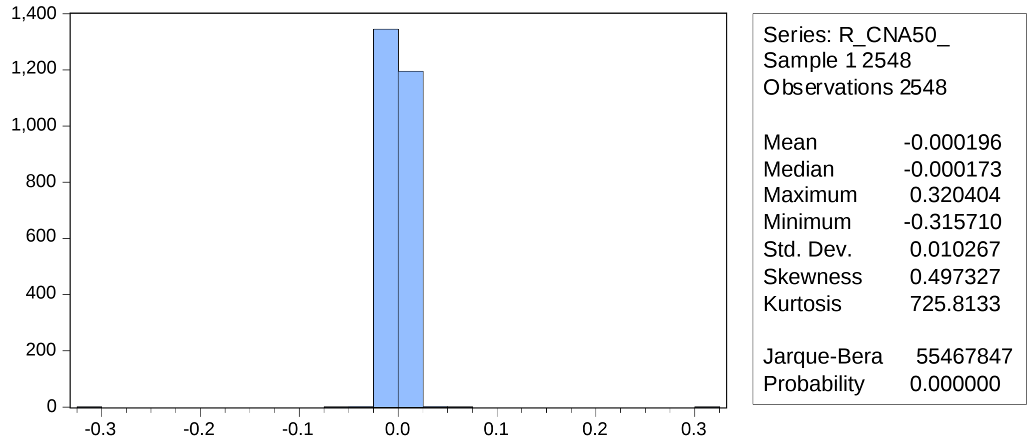
The Return of CSI300

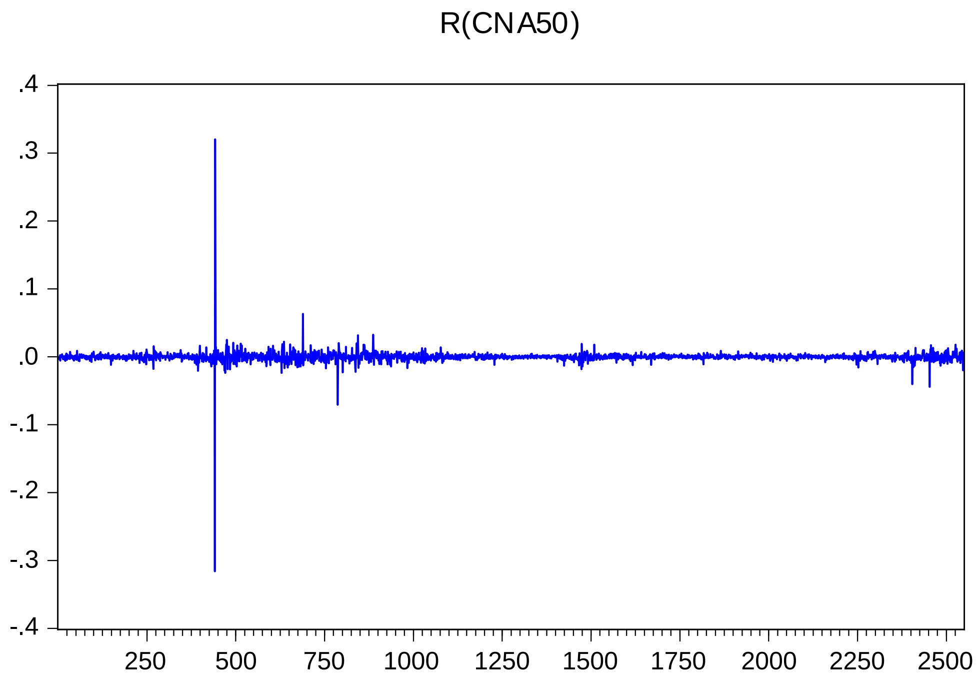




Appendix C

The Return of CNA 50





Appendix D

Granger Causality Test:

$$H_0: \beta_{yx1} = \beta_{yx2} = \cdots = \beta_{yxp} = 0 \quad \text{or} \quad H_0: \beta_{xy1} = \beta_{xy2} = \cdots = \beta_{xyp} = 0$$

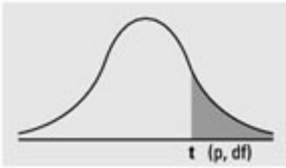
The outcomes of Granger Causality Test:

	Fail to reject: $\beta_{yx1} = \beta_{yx2} = \ldots = \beta_{yxs} = 0$	Reject: $\beta_{yx1} = \beta_{yx2} = \ldots = \beta_{yxs} = 0$
Fail to reject: $\beta_{xy1} = \beta_{xy2} = \ldots = \beta_{xys} = 0$	$y \not\Rightarrow x$ $x \not\Rightarrow y$ (no Granger causality)	$y \not\Rightarrow x$ $x \Rightarrow y$ (x Granger causes y)
Reject: $\beta_{xy1} = \beta_{xy2} = \ldots = \beta_{xys} = 0$	$y \Rightarrow x$ $x \not\Rightarrow y$ (y Granger causes x)	$y \Rightarrow x$ $x \Rightarrow y$ (bi-directional Granger cau-sality, or feedback)

Appendix E

Student *t*-statistic:(*Source: Statistical Analysis with R for Dummies*)

Numbers in each row of the table are values on a *t*-distribution with (*df*) degrees of freedom for selected right-tail (greater-than) probabilities (*p*).



df/p	0.40	0.25	0.10	0.05	0.025	0.01	0.005	0.0005
1	0.324920	1.000000	3.077684	6.313752	12.70620	31.82052	63.65674	636.6192
2	0.288675	0.816497	1.885618	2.919986	4.30265	6.96456	9.92484	31.5991
3	0.276671	0.764892	1.637744	2.353363	3.18245	4.54070	5.84091	12.9240
4	0.270722	0.740697	1.533206	2.131847	2.77645	3.74695	4.60409	8.6103
5	0.267181	0.726687	1.475884	2.015048	2.57058	3.36493	4.03214	6.8688
6	0.264835	0.717558	1.439756	1.943180	2.44691	3.14267	3.70743	5.9588
7	0.263167	0.711142	1.414924	1.894579	2.36462	2.99795	3.49948	5.4079
8	0.261921	0.706387	1.396815	1.859548	2.30600	2.89646	3.35539	5.0413
9	0.260955	0.702722	1.383029	1.833113	2.26216	2.82144	3.24984	4.7809
10	0.260185	0.699812	1.372184	1.812461	2.22814	2.76377	3.16927	4.5869
11	0.259556	0.697445	1.363430	1.795885	2.20099	2.71808	3.10581	4.4370
12	0.259033	0.695483	1.356217	1.782288	2.17881	2.68100	3.05454	43178
13	0.258591	0.693829	1.350171	1.770933	2.16037	2.65031	3.01228	4.2208
14	0.258213	0.692417	1.345030	1.761310	2.14479	2.62449	2.97684	4.1405
15	0.257885	0.691197	1.340606	1.753050	2.13145	2.60248	2.94671	4.0728
16	0.257599	0.690132	1.336757	1.745884	2.11991	2.58349	2.92078	4.0150
17	0.257347	0.689195	1.333379	1.739607	2.10982	2.56693	2.89823	3.9651
18	0.257123	0.688364	1.330391	1.734064	2.10092	2.55238	2.87844	3.9216
19	0.256923	0.687621	1.327728	1.729133	2.09302	2.53948	2.86093	3.8834
20	0.256743	0.686954	1.325341	1.724718	2.08596	2.52798	2.84534	3.8495
21	0.256580	0.686352	1.323188	1.720743	2.07961	2.51765	2.83136	3.8193
22	0.256432	0.685805	1.321237	1.717144	2.07387	2.50832	2.81876	3.7921
23	0.256297	0.685306	1.319460	1.713872	2.06866	2.49987	2.80734	3.7676
24	0.256173	0.684850	1.317836	1.710882	2.06390	2.49216	2.79694	3.7454
25	0.256060	0.684430	1.316345	1.708141	2.05954	2.48511	2.78744	3.7251
26	0.255955	0.684043	1.314972	1.705618	2.05553	2.47863	2.77871	3.7066
27	0.255858	0.683685	1.313703	1.703288	2.05183	2.47266	2.77068	3.6896
28	0.255768	0.683353	1.312527	1.701131	2.04841	2.46714	2.76326	3.6739
29	0.255684	0.683044	1.311434	1.699127	2.04523	2.46202	2.75639	3.6594
30	0.255605	0.682756	1.310415	1.697261	2.04227	2.45726	2.75000	3.6460
z	0.253347	0.674490	1.281552	1.644854	1.95996	2.32635	2.57583	3.2905
CI	———	———	80%	90%	95%	98%	99%	99.9%