VSB – TECHNICAL UNIVERSITY OF OSTRAVA

FACULTY OF ECONOMICS

DEPARTMENT OF FINANCE

Informační efektivnost akciových trhů
Information Efficiency of Stock Markets

Student: Bc. Songyang Gao
Supervisor of diploma thesis: Ing. Petr Seďa, Ph.D.

Ostrava 2021

VŠB - Technical University of Ostrava
Faculty of Economics
Department of Finance

Diploma Thesis Assignment

Student: **Bc. Songyang Gao**

Study Programme: N0412A050005 Finance

Title: Information Efficiency of Stock Markets
Informační efektivnost akciových trhů

The thesis language: English

Description:

1. Introduction
2. Efficient Market Hypothesis
3. Models of Efficient Market and Methods for Testing Information Efficiency
4. Data Sample Analysis
5. Empirical Results
6. Conclusion

Bibliography
List of Abbreviations
Declaration of Utilization of Results from the Diploma Thesis
List of Annexes
Annexes

References:

CAMPBELL, J. Y., LO, A. W. and A. C. MACKINLAY. *The Econometrics of Financial Markets.* 1st ed. New York: Princeton University Press, 1997, ISBN 0-691-04301-9.
FRANKE, Jürgen and Christian M. HAFNER. *Statistics of Financial Markets: An Introduction.* 3rd ed. Berlin: Springer, 2011. ISBN 978-3-642-16520-7.
TSAY, Ruey S. *Analysis of Financial Time Series.* 3rd ed. New Jersey: John Wiley & Sons, Inc., 2010. ISBN 978-0-470-41435-4.

Extent and terms of a thesis are specified in directions for its elaboration that are opened to the public on the web sites of the faculty.

Supervisor: **Ing. Petr Seďa, Ph.D.**

Date of issue: 20.11.2020
Date of submission: 23.04.2021

VYSOKÁ ŠKOLA BÁŇSKÁ-TECHNICKÁ UNIVERZITA OSTRAVA
Ekonomická fakulta
1

Ing. Petr Gurný, Ph.D.
Head of Department

doc. Ing. Vojtěch Spáčil, CSc.
Dean

The declaration

I hereby declare that I have elaborated the entire thesis including annexes myself.

Ostrava dated ....

Student´s name and surname

Contents

1 Introduction

Until the early 1990s, the efficient market hypothesis was the predominant approach when explaining price changes in stock markets. Financial markets have undergone significant changes over the last 20-30 years. Traders use very sophisticated information systems that allow almost immediate execution of trades. This happens in seconds, or even faster. It is therefore very important for investors in financial markets to anticipate the future development of exchange rates of financial assets, which take the form of financial time series. If the investor manages to predict the future development of the asset price, one can achieve a profit which is higher than average. The question of whether it is possible to successfully predict the future development of financial asset prices has not yet been satisfactorily resolved.

After 2000, there was a significant increase in volatility in global financial markets. Financial markets have thus become the subject of intensive research not only by academics, but especially by institutional investors who invested large sums in research to understand the structure of financial markets in order to achieve a profit that will be higher than the market average. The financial market is in fact influenced by a large number of economic, political or psychological factors. There are probably both linear and non-linear relationships between these factors. Although a clear analytical model of financial asset behavior has not yet been found, investors have always sought to identify predictable patterns of behavior in asset price movements.

The issue of information efficiency is also topical in developed and emerging economies in Asia, including China. There are two major stock markets in mainland China (Shanghai and Shenzhen) that have been operating since the early 1990s. However, China also has a developed market in Hong Kong, which has a history of more than 100 years.

The intention of this thesis is not to find a final solution to the problem of testing a weak form of information efficiency, which is of course exactly very difficult in financial economics. The aim is also not to provide unambiguous statistical evidence of the efficiency or inefficiency of a selected stock market or index. The purpose is to show how it is possible to systematically apply selected linear and nonlinear testing methods with respect to different models of stock market behavior.

One of the characteristics of efficient market is that the stock prices move randomly and are not predictable. **The main goal of this thesis is an empirical verification of the properties of efficient market models on the returns of stock market indices in two selected Asian economies.** We select three stock indexes in two regions which are SSE 50 index and HSI index in China and Nikkei 225 in Japan. SSE 50 is the stock index of Shanghai Stock Exchange, representing the top 50 companies in the mainland of China. HSI index is also from China, but it is special administrative region Hong Kong. The most developed stock market in Asia is generally considered to be the Japanese market. The index Nikkei 225, which is the main performance measure of the Tokyo Stock Exchange, was chosen as a benchmark. In this thesis, we will apply selected linear and nonlinear testing methods using the daily data in the period from January 2010 to December 2020. The testing algorithm will be based on a combination of random walk models and various testing methods. If statistically significant deviations from the random walk models are identified in the return series, the returns could be conditionally predictable. This feature could be used in formulating a trading strategy that could lead to above-average returns, which runs counter to the weak form of the efficient market hypothesis.

The main goal of the thesis is also supported by three partial goals:

a) to compare the predictability of returns of indexes in mainland China (SSE50) and Hongkong (HSI),

b) to assess the dynamics of predictability of SSE50, HSI and NIKKEI225 returns,

c) to design a trading strategy that is based on the results of empirical tests.

In Chapter 2, we discuss the efficient market hypothesis, which is the theoretical basis of this thesis. First, we present the history and development of this hypothesis. Secondly, we will explain in detail what the term efficiency in finance means. Then the definition of the efficient market hypothesis will be given and the forms of the efficient market will also be defined. Finally, the assumptions and characteristics of the efficient market hypothesis will be presented.

The third chapter will describe the models of the efficient market, the properties of which we will verify in the practical part of the thesis. First, however, the properties of financial time series will be described with emphasis on their specifics. In this thesis, we will focus mainly on models of random walks. We will not empirically test the martingale model, which is the oldest model of financial asset prices. However, the connection between this model and random walk models will be shown as well. An important part of this chapter is also a description of linear and nonlinear statistical tests that will be applied in the verification of random walk models of the first and third types.

In Chapter 4, we will focus on the description of stock markets that will be the subject of empirical analysis. For the purposes of this thesis, we will choose stock markets in mainland China (Shanghai), Hong Kong, which is a developed market, but also a market in Japan, which is a kind of benchmark. These markets differ in their standards and trading rules, so differences in the predictability of returns can also be expected. The development of individual markets will be approximated by their main indices. We will work with the data on a daily basis.

Chapter five represents the most important part of the thesis and has empirical character. First, the general testing procedure will be described. Empirical testing will then be divided into two parts. First, a static testing approach will be applied by using five linear and nonlinear statistical tests. When applying the dynamic approach, the two best quality linear and nonlinear tests were chosen. This is specifically a variance ratio test and a BDS independence test. The basic one-year test window was moved by one month forward till the end of the test period. The result was a time series of empirical tests with a length of 120 observations, which describes the dynamics of the development of predictability of individual indices. The corresponding p-values were obtained by the bootstrap method. An important part of this chapter is also the design of a trading strategy that would be based on the results of empirical tests.

In the last chapter, there will be given the conclusion of the whole thesis.

2 Efficient Market Hypothesis

In chapter 2, we will introduce the history of the efficient market hypothesis (EMH) briefly. Then we will define the EMH. At last, we will describe the assumptions and characteristics of the EMH.

2.1 A historical review of the efficient market hypothesis

The efficient market hypothesis originates from the academic literature on random walk theory of stock price behavior in the late 1950s. In fact, the earliest studies on the distribution of securities prices date back to 1900. Bachelier's (1900) research on the prices of French goods found that the prices of these goods fluctuated randomly, and the current price of a commodity was an unbiased estimate of the future price of the commodity. No buyer or seller involved in commodity speculation can expect to profit from it because it is a fair game. Although Bachelier did not use the term random walk, his research results are indeed a major application of random walk theory.

Cowles (1933) research on market forecasting has found that the performance of professional investors and investment consultants is poor, and the average performance of the market as a whole is better than that of any forecaster. Even if the results of the random prediction are compared with a series of actual predictions, they still have a slight upper hand. Working's research on commodity prices found that although the trends and fluctuations of commodity prices show recognizable and repetitive patterns, and there is no random change pattern in commodity price levels, changes in commodity prices tend to be highly random, making it impossible to predict randomness. Commodity price fluctuations. Kendall (1953) analyzed the weekly average data of stocks in 19 industries such as finance and industry from 1928 to 1939 in the UK. After doing a lot of serial correlation analysis, he found that the shape of the stock price series is far more unsystematic than generally believed. The stock price is in fact in a roaming state, there is almost no possibility of predicting future stock prices.

A study by Roberts (1959) found that the sequence of numbers generated from the cumulative random number series has a shape similar to that of stock prices. The pattern of stock price series similar to the head and shoulders can also be found in the graph of the sequence of random numbers, and there is almost no difference between the two. To this end, Roberts suggests further research on whether the stock price series are random. Another important document in the same year caused widespread controversy. Osborne (1959), an outstanding physicist at the Naval Research Laboratory in Washington, studied the changes in stock price series out of interest in the stock market and tried to find out whether these stock price series are consistent with certain laws in physics. His research believes that the absolute number of the stock price itself does not have any meaning, but the change in the stock price is more meaningful. It is possible that investors make investment decisions based on the percentage of the stock price change rather than the absolute price. If the market is considered as a whole, the expected volatility of the stock price will be zero, and the possibility of the stock price rising and falling is the same. In addition, Osborne (1959) confirmed that the stock price's fluctuation range will increase with the increase of the square root of the time interval, which also confirms that the movement track of the stock price is consistent with Brownian motion.

Modern research on the efficient market hypothesis began with Samuelson (1965). He pointed out that because stocks do not have clear price reference standards, stocks have a lot of risks and are highly volatile. After emphasizing the importance of information, he pointed out that new information is the main reason for stock price fluctuations. Future events will cast their shadows before the event, but the emergence of new information does not have a certain predictable pattern. As a result, the reasonable expected price will fluctuate randomly, and because of the unpredictability of the stock market price, it proves that the market price is a reasonable prediction of the intrinsic value of the stock.

2.2 Meaning of efficiency in economics

Economics believes that efficiency is closely related to the concerns of the welfare of all members of the economy. It means the ability of economic activities to meet people's needs under certain resource constraints. It is assumed that individuals and enterprises are self-interested price receivers. Individuals care about maximizing their utility, and companies care about maximizing their profits. Through the adjustment of the invisible hand of market competition, the product price enables each product to achieve a competitive equilibrium in the market, thereby maximizing personal utility and maximizing corporate profits.

We distinguish efficiency into production efficiency and economic efficiency. When the input remains unchanged, without reducing other outputs, an increase in a part of output is called production efficiency, and the enterprise can realize the optimal production decision. Economic efficiency means that given the constraints, any change to Pareto's optimal state will not make people better. If production efficiency and economic efficiency are combined, the entire economy can achieve allocation efficiency.

There is also an efficiency called exchange efficiency. In a market economy, exchange is an important mechanism for the efficient allocation of resources. Through exchange, the division of labor and specialization of production is achieved. The configuration realized by exchanging at a certain price, in this configuration state, minimizes the cost paid by all exchange participants, and any other configuration inevitably causes participants to pay a higher price. And the balance between supply and demand. Therefore, when we refer to exchange efficiency, we mean that in this market, all participants have no incentive to create exchange arrangements that are not currently provided by the market. In this sense, exchange efficiency can also be called operational efficiency.

Economists divide market efficiency into three forms: operational efficiency, allocation efficiency and pricing efficiency:

a) Operational efficiency is measured by the relative costs and benefits of the financial process. A financial system with less friction is more efficient. Operational efficiency refers to whether the buyer and seller in the market can complete the transaction in the shortest time and at the lowest transaction cost.

b) Allocation efficiency refers to the effectiveness of directing savings funds to productive uses. The market for efficient production uses is the market for efficient allocation. At the same time, competition is the key factor to ensure fair, safe and effective capital allocation. When the market price is in a competitive equilibrium, the market can achieve operational efficiency and allocation efficiency.

c) Pricing efficiency refers to whether the price of securities in the market can fully, timely and accurately reflect all relevant information. If the market is priced effectively, the market price of the securities can fully, quickly and accurately reflect all relevant information and make corresponding adjustments accordingly. At this point, the market price represents the best estimate of the "intrinsic value" of the security. The pricing efficiency is what we usually call the information efficiency.

2.3 Definition of the efficient market hypothesis

When the market is assumed to have a large number of trading participants and they are rational. Then the market is efficient according to Fama (1970), if:

a) The prices of securities traded in the market fully reflect all available information.

b) The prices of these securities will respond quickly and in an almost unbiased manner. This means that the power of the market will push the price of securities back to where it should be, so that no one can consistently earn returns that exceed the market average.

The reason is that any predictions about the company, such as seasonal fluctuations in the company's business or the retirement of the company's key management personnel, will be immediately reflected in the company's securities prices as confirmed by the efficient market. In other words, the market's expectations of the impact of these events on the company are generally unbiased. The only reason for changes in security prices is related, unexpected information. Unexpected events happen randomly. In this way, if we look at the time series chart of the price of a particular security, we will find that it fluctuates randomly. This random series is also called "random walk".

In summary, if a stock market is efficient, it should have the following characteristics:

a) The stock price fully reflects all relevant information, so that the only reason that any stock price may change is unexpected new relevant information.

b) If relevant new information emerges, stock prices can respond to it quickly and accurately. The so-called rapid means that there should be no obvious delay between the stock market receiving new information and the corresponding response of the stock price, but to respond almost instantaneously. Accurate means that the stock price responds to new information in an unbiased manner, and this response can accurately reflect the impact of the information on the stock price at the beginning, without subsequent amendments and without excessive reaction or underreaction.

c) The expected return of the investor is composed of the risk-free interest rate obtained by transferring the right to use the funds and the risk premium obtained by bearing certain risks. The risk premium is affected by the average market risk or the degree of investor risk aversion. In this case, the total return of the investor includes the expected return and the random walk due to the relevant new information described in the second feature. Any systematic (or non-random) fluctuations in stock prices can only be related to risk-free returns and risk premiums, and once the effects of risk-free interest rates and risk premiums have been eliminated, other factors may cause investor returns or stock prices. Volatility must be random and unpredictable.

d) According to the above three-point analysis, in an effective market, no investor can earn a rate of return that exceeds a certain benchmark (determined by the risk-free interest rate and risk premium) through possible investment methods or investment strategies, that is, cannot be obtained excess income.

2.4 Forms of efficient market

Considering that people may raise the question of whether a transaction based on a clearly defined set of information can achieve excessive returns, Roberts (1967) classifies information as size and divides the effective market into three types, and this classification is also obtained by Fama (1970) recognition and promotion:

a) Weak form of efficiency means that the current stock price fully reflects all the information in the historical sequence of all stocks, such as trading volume and trading price. In this type of market, it is impossible for any investor to use historical information to formulate an investment strategy for securities trading and obtain excess returns.

b) Semi-strong form of efficiency means that the current stock price not only contains all historical information, but also fully reflects all public information, such as the company's profit announcement, stock split, dividend declaration, company financial performance announcement, competitor information, etc. In this type of market, no investor can use public information to formulate an investment strategy for securities trading and obtain excess returns.

c) Strong form of efficiency means that the current stock price fully reflects all relevant information. This information includes not only the information that has been published, but also various private and inside information. It is the highest form of an effective market. In this type of market, no investor can obtain excess returns from the trading of stocks, no matter what information he has, as is the company's internal staff.

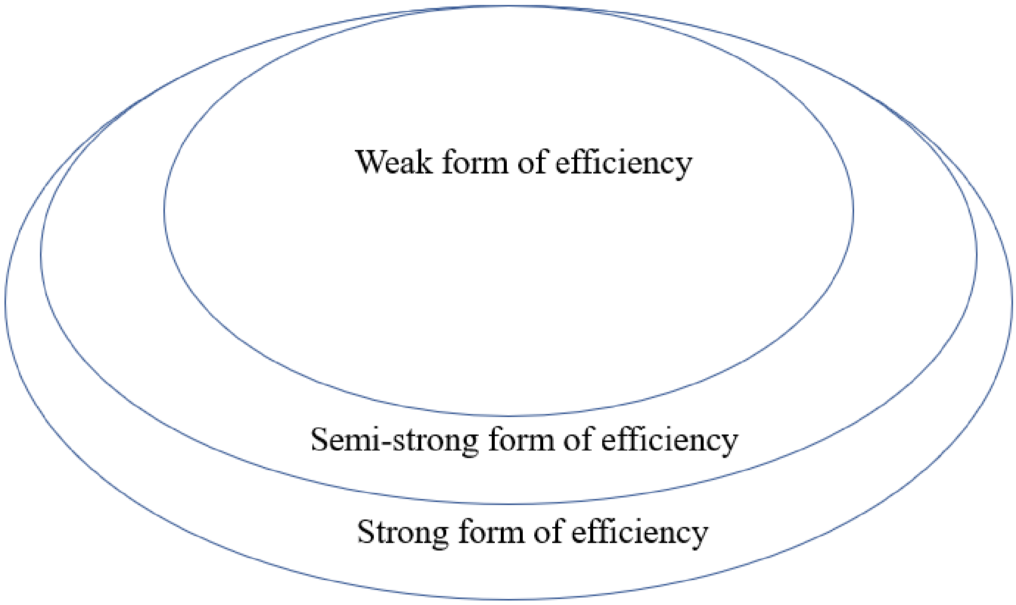
In fact, there is a certain level between these three forms of efficient markets and there is an embedded relationship, as shown in the Figure 2.1. We can also see from the figure that if the market is semi-strongly efficient, then the market must be weakly efficient, otherwise it is not necessarily. This is because the information set fully reflected by the semi-strong efficient market includes the information set reflected by the weak efficient market. Similarly, if the market is strong and efficient, the market must reach semi-strong and weakly efficient states.

With the more extensive study of EMH, Fama (1991) hanged the classification of effective markets in another retrospective literature and made the following adjustments:

a) Extend the weak efficient market hypothesis to a test of earnings predictability. The original weak test mainly studied the ability to predict the past returns, but now it also includes variables such as dividend yield and interest rate to predict returns. Because the EMH is inextricably linked to equilibrium pricing issues, discussion of predictability must also consider the cross-sectional predictive power of returns, that is, to examine asset pricing models and anomalies found in research (such as scale effects). In addition, research on seasonal benefits of returns (such as the January effect) and fluctuations in securities prices are also within the test of earnings predictability.

b) The scope of semi-strong-type efficient market and strong-type effective market inspection remains unchanged, but the name is changed accordingly: semi-strong inspection is changed to event research; strong inspection is changed to private information inspection

Figure 2.1 The relationship between three types efficient markets



Source: by author

2.5 Assumptions of the efficient market hypothesis

Fama (1991) believes that sufficient conditions for capital markets to be effective are:

a) No transaction costs for securities transactions;

b) All available information can be obtained by all market transaction participants at no cost;

c) All market trading participants have the same views on the meaning of the current price information of each security and the distribution of future prices. In such a market, it is clear that the current prices of securities will adequately reflect all available information and that the market is efficient.

However, Fama (1991) also realized that the real world cannot be a frictionless world, and it is difficult for investors to agree on the meaning of information. Fama (1991) pointed out that the above conditions are sufficient and not necessary conditions for the effective capital market. For example, as long as the transaction participants fully consider all available information, although the huge transaction costs hinder the transaction, this does not in itself mean that when the transaction occurs, the price does not fully reflect all available information. For another example, differences in the meaning of information among investors do not in themselves mean that the market is invalid, unless there are investors who can continuously obtain more information than the value implicit in the market price. Existing transaction costs, information acquisition costs, and inconsistencies among investors are just potential sources of market inefficiency. It is the central task of empirical research to measure and test the impact of these conditions on the formation of securities prices in the real world.

2.6 Characteristics of efficient market

The aim of this subchapter is to explain four characteristics of an efficient market. Reactions to new information. Random changes in quotations. Long-term returns on efficient markets, and business strategies in efficient markets, will be specially described. Many economists such as Chan (1988), Lo and Mackinlay (1988), have already offered comprehensive introduction of these characteristics of and efficient market.

a) Reactions to new information

An explanation for diversion from the EMH is that investors do not always react in proper way to new information. Sometimes, stocks have recent losses or gains, so investors can overreact by trading stocks. If investors are not rational, these overreactions will to make stock prices beyond their fair market prices. That means that investors only concentrate on recent information, and ignore the previous information. When investors understand actual meaning of an event, they will pull the stock prices in line eventually.

There are many effects that are troubling for the EMH. According to Chan (1988), people can get excess return from contrarian investment strategies, but this cannot prove that the EMH is not right. Moreover, Lo and Mackinlay (1988) demonstrated that at least half of the profits are not by overreaction, so there is not sufficient proof to conclude the failure of EMH.

b) Random changes in quotations

In the efficient market, there will be a lot of investors looking for clues at any time, and they use those evidences to make predictions about stock prices. Investors can use any useful information for analysis, trying to sell at a high price and buy at a low price, because they always want to make profits and not give up any chances. In fact, there are many investors taking part in this process at same time, so the changes in the stock price are fast. In other words, when investors use acquired information to predict the stock price, actually it is already too late. In the unpredictable market, the quotations will change without fixed scope.

The first test of the random walk hypothesis was developed by Cowles and Jones (1937). They compared the frequency of sequences and reversals in historical stock returns. Then Cootner (1964) and Fama (1965) described the tests of random walk hypothesis.

c) Long-term returns on efficient markets

Since the market is random, we can ignore the time. Under the framework of EMH, there will no relationship between current and past prices, or future and current prices. Moreover, regardless of the time variable, the price is also random. But in fact, the factor of time is actually very important. For example, the price changes may not be random, because of overreaction from irrational investors in the short-term. But in the long-term, rational investors will adjust the stock price. No investors can achieve above mean revenues at a given level of risk in the long period.

One of the first papers on long-term return anomalies was published by De Bondt and Thaler (1985). They find that when stocks are ranked on 3 to 5 year past returns, past winners tend to be losers in the future, and vice versa.

d) Business strategies in efficient markets

In efficient markets, business strategies are not correct. On the basis of the EMH, the market is introduced as a result of rational economic cycle, therefore there are no mistakes, or no irrational investors. Arbitrage is the main business strategy on financial markets. However, if the EMH is hold, it should be different. Arbitrageurs' motion cannot make financial markets deviate from their basic value. In other words, the market mechanism can always quickly adjust by mistakes.

2.7 Literature review

The efficient market hypothesis as proposed by professor Fama of the University of Chicago believes that in an efficient market, market prices will contain a certain degree of information. Price and value are always the same. The effective market hypothesis defines three types according to the various levels reflected in the stock price information. The form of the efficient market is weak, the semi-strong form of the efficient market and the strong form of the efficient market.

Fama (1965) used a random walk model to study the dynamic behavior of 30 stocks with an average price of 30 Dow Jones Industrial stocks from 1957 to 1962.

Song and Jin (1995) divided into two stages. In the first stage (1991-1992), Shanghai's stock market was still relatively small and in the early stages of development. In the second stage (1993-1994), the Shanghai stock market reached a certain scale, and the empirical analysis of the weekly return rate reflected in the second stage proved the random walk characteristics of the Shanghai stock market.

In the second stage, the Shanghai stock market was inefficient. The root is the conclusion that the Shanghai Stock Market is invalid (Yu, 1994). This means that choosing to include only the sample data of the first stage is invalid.

With the continuous deepening of financial system reforms, the development of China's stock market is impressive, and domestic scholars are shifting their focus to the effectiveness of semi-strong form markets. Most of them use the "event research method".

Investigate stock price trends before and after the incident and test the semi-strong stage.

The first study to describe the efficiency of the semi-strong form of the Chinese stock market was to select two holding events, namely the Baoyang event and the Shen event. The empirical results show that the stock market at that time did not reach a semi-strong form (Shen, 1996).

Contrary to the low interest rate environment of the past few years, the booming economy of developing countries like China has attracted the attention of many foreign investors. At the same time, the lack of investment opportunities3 due to investment restrictions has prompted Chinese investors to raise funds and invest overseas. Stock trading eliminates financial barriers between China and the rest of the world, and will undoubtedly expand the flow of information and increase the financial interaction between the two markets. Previous studies on multinational corporations in developing countries have cross-listed 4 in developed countries, providing some good theoretical understandings for the impact of linkages. In particular, it was found that cross-listing had no long-term liquidity impact. But regardless of the development of the listing location and the home country, it can improve the pricing efficiency of the home country (Liu and Wang, 2007). Other studies have shown that the strength of analyst coverage is negatively correlated with changes in company-specific returns for emerging market companies. In addition, the increasingly stringent disclosure requirements of developed countries may weaken the benefits of private information collection in emerging markets, thereby exaggerating the improvement of market information quality (Fernandes and Ferreira, 2008). Similarly, the study by Foucault and Gehrig (2008) found a compromise model between the benefits of strict disclosure rules and reducing the cost of private information discovery. This paper also finds that cross-listing will increase the sensitivity of investment decisions and stock prices, because stock prices tend to reflect the quality of management decisions. All the above studies point out the different effects of cross-listing and stock linkages on market information.

In Japan, various studies related to the efficient market hypothesis have been conducted using parametric testing and non-parametric testing (such as operational testing, K-S testing, and serial correlation). Various researchers use daily, weekly or monthly data to test EMH empirically. EMH was rejected by Malafeyev et al. (2019).

3 Models of Efficient Market and Methods for Testing Information Efficiency

In this chapter, we will introduce in detail not only characteristics of financial time series. And There are three types of random walk models. Then there are verification and tests of models. We will describe relationship between random walk model and martingale models briefly.

3.1 Financial time series

In this subchapter, we introduce in detail the most important part: financial time series. And financial time series analysis is concerned with the theory and practice of asset valuation over time. Then we will introduce the components and classification of financial time series. At last, it's the methods of time series analysis and stylized statistical properties.

3.1.1 Description of financial time series

According to Ostrom (1990) time series refers to a series of numbers formed by arranging the observation values of certain statistical indicators in the order of observation time. There are a large number of time series data in the social, economic, scientific and other fields that need to be further processed by people. And some basic development laws of certain phenomena are thus found. So as much as possible to extract important information from the huge data according to their own needs, and use the extracted useful information with thousands of predictions, in order to master and control people's own future decisions.

Time series analysis is based on the time series data obtained by people's observation and statistics, through curve fitting and parameter estimation and other analytical methods to establish the theory of mathematical statistical model. Time series analysis is widely used in the macroeconomic regulation and control of the national economy, the operation and management of enterprises, and the research on the physiological and psychological aspects of weather forecast and biology. One of the main purposes of time series analysis is prediction. Prediction refers to people based on historical data of things and their interrelation and development, using their own scientific means to make a reasonable analysis, to infer the future development trend of things in advance. In short, prediction is based on historical and current data to scientifically estimate and infer the future development trend of things (Franke and Hafner, 2011).

As an important representative of time series, financial time series refers to a set of price data or income data obtained by arranging the prices of financial products in the financial market (including the stock market, foreign exchange market and futures market, etc.) in chronological order. As a special kind of time series, financial time series have very remarkable characteristics. Financial time series are relatively general time series data, and the data generation process is more complicated, including many random factors and the non-linearity between the data is strong.

The financial market is the core of the operation of the national economy. Under the leadership of the market economy, exploring the fluctuation laws of the financial market and conducting effective supervision and management to improve the investment efficiency of the financial market is one of the main goals of governments and social investment institutions. The financial industry is facing more and newer development opportunities and challenges. The general public is also paying more and more attention to the global financial market. People are investing in high-risk and high-return areas such as stocks, securities, and funds. Understanding and grasping the law of market fluctuations is directly related to the security, stability and efficiency of financial markets. Moreover, the financial market itself is a huge and complex system affected by various factors such as politics, economy, environment, and so on. As a comprehensive external manifestation of financial market fluctuations and changes, financial time series data must contain many of the internal financial systems. Objective law information with valuable value. Investors are eager to understand the operating rules of these financial data and more reliably predict the trend of these financial data in order to avoid high-risk investments. On the other hand, managers also hope to find out the inherent laws of the financial market, so that they can take effective measures to maintain the prosperity and stability of the financial industry. Therefore, it is undoubtedly of great significance for financial investment decision-making and risk management activities to find a certain pattern from these massive and complicated data in order to better understand and grasp the trend of financial markets.

3.1.2 The characteristics of asset returns

It is a highly empirical discipline, but like other scientific fields theory forms the foundation for making inference. There is, however, a key feature that distinguishes financial time series analysis from other time series analysis. Both financial theory and its empirical time series contain and element of uncertainty. For example, there are various definitions of asset volatility, and for a stock return series, the volatility is not directly observable. As a result of the added uncertainty, statistical theory and methods play an important role in financial time series analysis (Tsay, 2010).

Most financial studies involve returns, instead of prices, of assets. Campbell, Lo, and MacKinlay (1997) give two main reasons for using returns. First, for average investors, return of an asset is a complete and scale-free summary of the investment opportunity. Second, return series are easier to handle than price series because the former have more attractive statistical properties. There are, however, several definitions of an asset return.

Let P_t be the price of an asset at time index t. Assume for the moment that the asset pays no dividends. Several types of returns have been defined in financial theory:

a) One-period simple return

Holding the asset for one period from date $t - 1$ to date t would result in a simple gross return

$$1 + R_t = \frac{P_t}{P_{t-1}} \text{ or } P_t = P_{t-1}(1 + R_t). \tag{3.1}$$

The corresponding one-period simple net return or simple return is

$$R_t = \frac{P_t}{P_{t-1}} - 1 = \frac{P_t - P_{t-1}}{P_{t-1}}. \tag{3.2}$$

b) Multiperiod simple return

Holding the asset for k periods between dates $t - k$ and t gives a k-period simple gross return

$$1 + R_t|k| = \frac{P_t}{P_{t-1}} \times \frac{P_{t-1}}{P_{t-2}} \times \cdots \times \frac{P_{t-k+1}}{P_{t-k}} = (1 + R_t)(1 + R_{t-1}) \cdots (1 + R_{t-k+1}) = \prod_{j=0}^{k-1}(1 + R_{t-j}). \quad (3.3)$$

Thus, the k-period simple gross return is just the product of the k one-period simple gross returns involved. This is called a compound return. The k-period simple net return is $R_t|k| = (P_t - P_{t-k})/P_{t-k}$.

In practice, the actual time interval is important in discussing and comparing returns (e.g., monthly return or annual return). If the time interval is not given, then it is implicitly assumed to be one year. If the asset was held for k years, then the annualized (average) return is defined as

$$Annualized\{R_t|k|\} = \left[\prod_{j=0}^{k-1}(1 + R_{t-j})\right]^{1/k} - 1. \quad (3.4)$$

This is a geometric mean of the k one-period simple gross returns involved and can be computed by

$$Annualized\{R_t|k|\} = exp\left[\frac{1}{k}\sum_{j=0}^{k-1} \ln{(1 + R_{t-j})}\right] - 1, \quad (3.5)$$

where $exp(x)$ denotes the exponential function and $ln(x)$ is the natural logrithm of the positive number x. Because it is easier to compute arithmetic average than geometric mean and the one-period returns tend to be small, one can use a first-order Taylor expansion to approximate the annualized return and obtain

$$Annualized\{R_t|k|\} \approx \frac{1}{k}\sum_{j=0}^{k-1} R_{t-j}. \quad (3.6)$$

c) Continuously compounded return

The natural logarithm of the simple gross return of an asset is called the continuously compounded return or log return:

$$r_t = \ln(1 + R_t) = ln\frac{P_t}{P_{t-1}}, \quad (3.7)$$

Continuously compounded returns r_t enjoy some advantages over the simple net returns R_t. First, consider multiperiod returns. We have

$$r_t[k] = \ln(1 + R_t[k]) = \ln[(1 + R_t)(1 + R_{t-1}) \cdots (1 + R_{t-k+1})] = ln\,(1 + R_t) + \ln(1 + R_{t-1}) + \cdots \ln(1 + R_{t-k+1}) = r_t + r_{t-1} + \cdots + r_{t-k+1}. \tag{3.8}$$

Thus, the continuously compounded multiperiod return is simply the sum of continuously compounded one-period returns involved. Second, statistical properties of log returns are more tractable.

d) Portfolio return

The simple net return of a portfolio consisting of N assets is a weighted average of the simple net returns of the assets involved, where the weight on each asset is the percentage of the portfolio's value invested in that asset. Let p be a portfolio that places weight w_i on asset i. Then the simple return of p at time t is $R_{p,t} = \sum_{i=1}^{N} w_i R_{it}$, where R_{it} is the simple return of asset i.

The continuously compounded returns of a portfolio, however, do not have the above convenient property. If the simple returns R_{it} are all small in magnitude, then we have $r_{p,t} = \sum_{i=1}^{N} w_i r_{it}$, where $r_{p,t}$ is the continuously compounded return of the portfolio at time t. This approximation is often used to study portfolio returns.

e) Dividend payment

If an asset pays dividends periodically, we must modify the definitions of asset returns. Let D_t be the dividend payment of an asset between dates $t - 1$ and t and P_t be the price of the asset at the end of period t. Thus, dividend is not included in P_t . Then the simple net return and continuously compounded return at time t become

$$R_t = \frac{P_t + D_t}{P_{t-1}} - 1, r_t = \ln(P_t + D_t) - \ln(P_{t-1}). \tag{3.9}$$

f) Excess return

Excess return Z_t of an asset at time t is the difference between the asset's return and the return on some reference asset. The reference asset is often taken to be riskless such as a short-term U.S. Treasury bill return. The simple excess return and log excess return of an asset are then defined as

$$Z_t = R_t - R_{0t},\ z_t = r_t - r_{0t}, \tag{3.10}$$

where R_{0t} and r_{0t} are the simple and log returns of the reference asset, respectively. In the finance literature, the excess return is thought of as the payoff on an arbitrage portfolio that goes long in an asset and short in the reference asset with no net initial investment.

g) Summary of relationships

The relationships between simple return R_t and continuously compounded (or log) return r_t are

$$r_t = \ln(1 + R_t),\ R_t = e^{r_t} - 1. \tag{3.11}$$

If the returns R_t and r_t are in percentages, then

$$r_t = 100 \ln\left(1 + \frac{R_t}{100}\right),\ R_t = 100(e^{\frac{r_t}{100}} - 1). \tag{3.12}$$

Temporal aggregation of the returns produces

$$1 + R_t[k] = (1 + R_t)(1 + R_{t-1}) \cdots (1 + R_{t-k+1}), \tag{3.13}$$
$$r_t[k] = r_t + r_{t-1} + \cdots + r_{t-k+1}.$$

If the continuously compounded interest rate is r per annum, then the relationship between present C and future values A of an asset is

$$A = C \exp(r \times n),\ C = A \exp(-r \times n). \tag{3.14}$$

3.1.3 The components of financial time series

The change of the time series is affected by many factors, some of which play a long-term and decisive role, making them show a certain trend and certain regularity (Tsay, 2010). When analysing the changing laws of the time series, it is actually impossible to divide each influencing factor one by one and make an accurate analysis separately. The components of the series, and then analyse these types of components separately to reveal the regularity of the time series. The components that affect the time series can be summarized into the following four types.

a) Trend refers to the phenomenon that the phenomenon shows a continuous gradual rise and fall, a steady change or movement in a certain direction over time. This change is usually the result of many long-term factors.

b) Cycle, which means that the time series appears as a series of points that circulate above and below the trend line and continue to change regularly for a period of time. This factor has periodic changes, such as the period of moderate inflation immediately after the period of high-speed inflation, which will cause many time series to appear alternately above and below an overall increasing trend line.

c) Seasonal variation refers to the phenomenon that the phenomenon is affected by seasonality and changes in a periodical fluctuation according to a fixed cycle. Although the seasonal change in a time series is usually considered to be one year, the seasonal factor can also be used to represent a regular repeating pattern with a length of less than one year.

d) Irregular movement refers to the irregular fluctuations of phenomena caused by accidental factors. Such factors include the deviation between actual time series values and estimates that take into account trend, periodicity, and seasonal changes, and it is used to explain random changes in time series. Irregular factors are caused by short-term unpredicted and those factors that affect the time series that are not repeatedly discovered.

3.1.4 The classification of financial time series

According to different standards, time series have different classification methods. Commonly used standards and classification methods are described by Franke and Hafner (2011):

a) According to the number of objects studied, there is a unary time series and multivariate time series, such as a sales volume series of a certain commodity, which is a unary time series; Variables, such as temperature, pressure, and rainfall data sorted in order of year and month, and each time corresponds to multiple variables, then this sequence is a multivariate time series.

b) According to time continuity, time series can be divided into discrete time series and continuous time series. If the time parameter corresponding to each sequence value in a sequence is a discontinuity, the sequence is a discrete time series; if the time parameter corresponding to each sequence value in a sequence is a continuous function, the sequence It is a continuous time series.

c) According to the statistical characteristics of the series, it can be divided into stationary time series and non-stationary time series. The so-called stationarity of time series means that the statistical law of time series will not change with the passage of time. The timing chart of a stationary sequence should intuitively show that the sequence always fluctuates randomly around a constant value, and the fluctuation range is bounded, without obvious trends and without periodic characteristics. In contrast, the non-stationarity of time series means that the statistical law of time series changes with the passage of time.

d) According to the distribution of the sequence, there are two types of Gaussian and non-Gaussian time series.

3.1.5 Methods of time series analysis

Chatfield (2003) introduced time series analysis is a widely used data analysis method. It studies a series of interrelated digital series (dynamic data) that represent a phenomenon with time, and describe and explore the phenomenon. Regularity that develops and changes over time. The methods used for time series analysis can be analysed by intuitive and simple data graph method, index method, model method, etc. The model method is relatively more specific and deeper, and it can understand the internal structure and complex characteristics of the data more fundamentally to achieve the purpose of control and prediction. In general, time series analysis methods include the following two categories:

a) Deterministic time series analysis: refers to the method of temporarily filtering out random factors (such as seasonal factors and trend changes) for deterministic analysis. The basic idea is to fit a time series with a certain time function $y = f(t)$, different changes are described by different functional forms, and the superposition of different changes is described by different functional superpositions. It can be divided into trend prediction method (least square method), smooth prediction method, decomposition analysis method, etc.

b) Random time series analysis: The basic idea is to analyse the correlation between variables at different times to reveal their related structure, and use this correlation structure to establish a hybrid model of autoregression, moving average, and autoregressive moving average to time series make predictions.

3.1.6 Stylized statistical properties of asset returns

Let us start by stating a set of stylized statistical facts which are common to a wide set of financial assets (Cont, 2000).

a) Absence of autocorrelations: (linear) autocorrelations of asset returns are often insignificant, except for very petty intraday time scales (20 minutes) for which microstructure effects come into play.

b) Heavy tails: the (unconditional) distribution of returns seems to display a power-law or Pareto-like tail, with a tail index which is finite, higher than two and less than five for most data sets studied. In particular this excludes stable laws with infinite variance and the normal distribution. The precise form of the tails is hard to determine.

c) Gain/loss asymmetry: one observes huge drawdowns in stock prices and stock index values but not equally large upward movements.

d) Aggregational Gaussianity: as one increases the time scale Δt over which returns are calculated, their distribution looks more and more like a normal distribution. In particular, the shape of the distribution is different at different time scales.

e) Intermittency: returns display, at any time scale, a high degree of variability. This is quantified by the presence of irregular bursts in time series of a wide variety of volatility observations.

f) Volatility clustering: not same measures of volatility display a positive autocorrelation over several days, which quantifies the fact that high-volatility events tend to cluster in time.

g) Conditional heavy tails: even after correcting returns for volatility clustering (via GARCH-type models), the residual time series still exhibit heavy tails. However, the tails are less heavy than in the unconditional distribution of returns.

h) Slow decay of autocorrelation in absolute returns: the autocorrelation function of absolute returns decays slowly as a function of the time lag, roughly as a power law with an exponent $\beta \in [0.2,0.4]$. This is sometimes interpreted as a signal of long-range dependence.

i) Leverage effect: most measures of volatility of an asset are negatively correlated with the returns of that asset.

j) Volume/volatility correlation: trading volume is correlated with all measures of volatility.

k) Asymmetry in time scales: coarse-grained measures of volatility predict fine-scale volatility better than the other way round.

3.2 Random walk models

Bachelier (1900) for the first time proposed the assumption that financial asset prices follow a logarithmic normal distribution. He assumed that stock prices follow Brownian motion (an endless type of molecular particles in physics disordered movement), which became the common source of random walk theory and the efficient market hypothesis.

Fama (1965), a professor of economics at the University of Chicago, formally proposed the efficient market hypothesis: "*The premise of the hypothesis is that stock prices are always correct. Therefore, market trends are random, and no one can predict the future of the market. Direction. The premise of correct price is that the person who sets the price must be rational and have sufficient information*."

Campbell, Lo and MacKinlay (1997) described in The Econometrics of Financial Markets in detail about random walk models. Random walk theory means that the stock price changes are random and unpredictable. This randomness is considered to imply that the stock market is irrational. But on the contrary, random changes in stock prices indicate that the market is functioning normally or effectively.

All chart analysis schools are based on the assumption that they are stocks, foreign exchange, gold, bonds, etc. All investment will be affected by economic, political, and social factors, and these factors will continue to repeat itself like history. For example, if the economy recovers from the Great Depression, property prices, stock market, gold, etc. It will fall after rising, but it will rise even higher after falling. Even in the short term, the rules governing the value of all investments are inseparable from the aforementioned factors. As long as investors can predict which factors dominate prices, they can predict the future trend. This kind of mentality constitutes the reason, denying their income, age, understanding of the news, acceptance of digestion, and enthusiasm for confidence, all reflected in stock prices and transactions. According to the chart, we can predict the future stock price trend. But the random walk theory is opposed to this statement, and the opposite opinion.

Random walk theory points out that there are thousands of smart people in the stock market, not all of them are stupid people. Everyone knows the analysis, and the data flow into the market is all public. Everyone can know that there is no secret. The current price of the stock already reflects the relationship between supply and demand, or its value will not be too far. The so-called intrinsic value measurement method is determined by looking at basic factors such as asset value per share, price-earnings ratio, and dividend pay-out ratio. These factors are not big secrets, and everyone can find them by opening a newspaper or magazine. If a stock's asset value is ten yuan, it will never change to one hundred yuan or one yuan in the market. No one in the market will buy this stock for one hundred yuan or sell it for one yuan. The current market price of stocks has already represented the views of millions of eye-catching people and constitutes a reasonable price. The market price will fluctuate around the intrinsic value.

3.2.1 Causes of fluctuation

As mentioned earlier, the stock price fluctuates up and down around the market price, so why does the random trend produce stock price fluctuations, and the reasons for these fluctuations are proposed in theory by Tsay (2010):

a) New economic and political news is random and does not flow into the market in a fixed manner.

b) The news caused the fundamental analysts to re-estimate the value of the stock and make a buying and selling policy, resulting in new changes in the stock.

c) Because the news is nowhere to be found, it came suddenly, and no one can predict it beforehand. The stock trend speculates that this matter cannot be established. What the graphists said is just nonsense.

d) Since all stock prices in the market have reflected their basic value. This value is fair and is determined by both buyers and sellers. This value will not change again, unless there are news of good or bad news such as wars, acquisitions, mergers, interest rate cuts, oil wars, etc., it will fluctuate again. But the next news is that everyone will not know whether it is good or bad, so the stock currently has no memory system. The rise and fall of historical stock prices are not related to changes in current market value. Just like throwing a copper plate, this time the roll is positive, it does not mean that the next roll is the front, the next time the roll is the front, or the reverse takes 50% of the chance. No one will know that the next time will be the bottom or the reverse.

e) Since the stock price does not have a memory system, attempts to use stock price fluctuations to find a principle to overcome the market will eventually end in failure. The stock price has no direction at all, walks randomly, and rises and falls randomly. We cannot predict where the stock market is going. No one must be a winner, and no one will lose. As for the role of stock experts, it is actually not significant, and it can even be said to be meaningless. Because they are so specialized, they must use these theories to get rich themselves. Where will research be published to make others develop?

Simply put, the fluctuation of stock prices is determined by random market news, and such news cannot be predicted.

3.2.2 Types of random walk models

a) Random walk model I: *IID* Increments

Perhaps the simplest version of the random walk hypothesis is the independently and identically distributed (IID) increments case in which the dynamics of $\{P_t\}$ are given by the following equation:

$$P_t = P_{t-1} + \mu + \varepsilon_t, \quad \varepsilon_t \sim IID(0, \sigma^2), \tag{3.15}$$

where the P_t and P_{t-1} is the price for different turns. And μ is the expected price change or drift, and $IID(0, \sigma^2)$ denotes that ε_t is independently and identically distributed with mean 0 and variance σ^2. The independence of the increments $\{\varepsilon_t\}$ implies that the random walk is also a fair game, but in a much stronger sense than the martingale: Independence implies not only that increments are uncorrelated, but that any nonlinear functions of the increments are also irrelevant. We shall call this the random walk 1 model or RW1.

To develop some intuition for RW1, consider its conditional mean and variance at date t, conditional on some initial value P_0 at date 0:

$$E(P_t|P_0) = P_0 + \mu t, \tag{3.16}$$

$$var(P_t|P_0) = \sigma^2 t, \tag{3.17}$$

which follows from recursive substitution of lagged P_t in (3.15) and the *IID* increments assumption. From (3.16) and (3.17) it is apparent that the random walk is nonstationary and that its conditional mean and variance are both linear in time. These implications also hold for the two other forms of the random walk hypothesis (RW2 and RW3) described below.

b) Random walk model II: independent increments

Fama and Blume (1966) present a more detailed the restriction of identical distributions is clearly implausible, especially when applied to financial data that span several decades. However, testing for independence without assuming identical distributions is quite difficult, particularly for time series data. If we place no restrictions on how the marginal distributions of the data can vary through time, it becomes virtually impossible to conduct statistical inference since the sampling distributions of even the most elementary statistics cannot be derived.

Some of the non-parametric techniques described in this section, such as rank correlation, do not require test independence for the same distribution, but the number of individual peripheral distributions is usually limited and small. For example, the test of independence between IQ scores and academic performance involves two different peripheral distributions. One is for IQ scores and the other is for academic performance. To extract multiple observations from each peripheral distribution, various non-parametric tests can be designed to see whether the product of the peripheral distribution is equal to the simultaneous distribution of paired observations. Assuming that each observation of IQ and academic performance has a unique distribution, then this method is obviously unsuccessful.

Nevertheless, there are two lines of empirical research that can be viewed as a kind of "economic" test of RW2: filter rules, and technical analysis. Although neither of these approaches makes much use of formal statistical inference, both have captured the interest of the financial community for practical reasons. This is not to say that statistical inference cannot be applied to these modes of analysis, but rather that the standards of evidence in this literature have evolved along very unique paths. Therefore, we shall present only a cursory review of these techniques.

c) Random walk model III: uncorrelated increments

One of the most direct and intuitive tests of the random walk and martingale hypotheses for an individual time series is to check for serial correlation, correlation between two observations of the same series at different dates. Under the weakest method of the random walk, RW3, the increments or first differences of the level of the random walk are uncorrelated at all leads and lags. Therefore, we may test RW3 by testing the null hypothesis that the autocorrelation coefficients of the first-differences at various lags are all zero.

This seemingly simple approach is the basis for a surprisingly large variety of tests of the random walk, and we shall develop these tests in this chapter. For example, tests of the random walk may be based on the autocorrelation coefficients themselves. More powerful tests may be constructed from the sum of squared autocorrelations. Linear combinations of the autocorrelations may also have certain advantages in detecting particular departures from the random walk. Therefore, we shall devote considerable attention to the properties of autocorrelation coefficients in the coming sections.

3.2.3 Verification of random walk models

After the random walk theory is well-known to investors, many economists have verified the random walk theory. The following are some of the more famous cases by MacKinlay (1997):

a) There was a study that used the Standard & Poor's stocks for long-term research, and found that stocks rose or plummeted, rose four or five times, or fell 99%, the proportion was only a small number. Most stocks range from 10% to 30%. There is a phenomenon of normal distribution in statistics. That is, the greater the percentage of increase or decrease, the smaller the proportion. There is no single trend in stock prices. Buying stocks depends on whether you are lucky or not. You have equal opportunities to buy a rising stock or a falling stock.

b) In another experiment, a US senator used darts to throw a financial newspaper and picked 20 stocks as a portfolio. As a result, this messy portfolio was similar to the overall performance of the stock market, and it was not inferior to experts. Portfolios we recommend even perform better than some experts suggest.

c) Some people have studied the performance of unit funds and found that this year's performance is good, and it may perform the worst next year. Some disappointing funds in the past may stand out this year and become the top of the increase. So nowhere to be found, buying funds also depends on your luck, investment skills are not practical, because the stock market has no memory, everyone just bets on probability.

In addition to the above-mentioned more famous cases, there is also a personal case, which is quite interesting. Let me talk to you here. During the bull market in 2015 in China, a financial channel once launched alpaca stock picking. The stock symbol was selected by alpaca grazing and positions were opened and lightened. As a result, the results of investment by the alpaca stock pool outperformed most analysts. Portfolio, which also justifies the random walk theory.

We put aside the psychological factors of investors, and it is a joke to predict the current stock price through historical stock prices. If the probability of success is high, it can only show that this analysis forms a reasonable market, not how scientific this analysis is.

The composition of the stock price is two parts, the actual value of the enterprise and the enterprise's expectations. If the stock price only reflects the actual value of the enterprise, it should be equal to the net worth. Looking at the stock market, there are very few situations where the stock price and net worth are the same, so the different part is the company's expectations. Here to explain the problem of breaking the net, which means that the market expects a loss, so the stock price will be lower than the net asset. Suppose that the actual value part is a fixed value at a certain stage, and the expectation is a variable. This variable is random and no trace is found. The fluctuation of the stock price is generated by the expectation. It cannot be concluded that the performance of the company at this stage will increase at the next stage, and the probability of increase and decrease is 50%. This can return to the theory of coin tossing.

In the long term, as the economy shows an overall upward trend, the prediction of stock prices will be relatively accurate. But in fact, this is also a problem of probability.

If the national economy goes into decline, the stock market will definitely go into decline. This variable is also random, but we make this analysis based on experience and national economic trends by MacKinlay (1997).

Since investing is betting on the expected probability, if you believe in this theory, you don't need to invest, nor is it. Now that we know the variables of stock prices, we can analyse the variables, but not by historical stock prices, but by the real situation of the market to predict. This does not contradict the theory. The stock price reflects the recognition of the stock price in the real-time market. When the real market changes, the stock price will inevitably change, but this change is not known to us. The analysis we have made is Pre-judgment can only represent a probability, and our analysis of the message is to increase this probability.

A theory, from generation to verification, has been observed and practiced for a long time, and this theory can be widely circulated because its principle is desirable. There is an old saying in China, it is better to have a book than a book without faith. Everyone must have their own cognition when learning this theory, so as to get what you want from the theory. Survival in the market, because many classic theories themselves are opposed in principle, seeking common ground while reserving differences is the way to grow.

3.3 Martingale Models

Perhaps the earliest model of financial asset prices was the martingale model, whose origin lies in the history of games of chance and the birth of prob ability theory. The prominent Italian mathematician Cardano proposed an elementary theory of gambling, in which he wrote:' The most fundamental principle of all in gambling is simply equal conditions, e.g., of opponents, of bystanders, of money, of situation, of the dice box, and of the die itself. To the extent to which you depart from that equality, if it is in your opponent's favour, you are a fool, and if in your own, you are unjust.

This subchapter clearly contains the notion of a fairgame, a game which is neither in your favor nor your opponent's, and this is the essence of a martingale, a stochastic process $\{P_t\}$ which satisfies the following condition:

$$E(P_{t+1}|P_t, P_{t-1}, \dots) = P_t, \tag{3.18}$$

or equivalently,

$$E(P_{t+1} - P_t|P_t, P_{t-1}, \dots) = 0. \tag{3.19}$$

If P_t represents one's cumulative winnings or wealth at date t from playing some game of chance each period, then a fair game is one for which the expected wealth next period is simply equal to this period's wealth (see equation (3.18)), conditioned on the history of the game. Alternatively, a game is fair if the expected incremental winnings at any stage is zero when conditioned on the history of the game (see equation (3.19)).

If P_t is taken to be an asset's price at date t, then the martingale hypothesis states that tomorrow's price is expected to be equal to today's price, given the asset's entire price history. Alternatively, the asset's expected price change is zero when conditioned on the asset's price history; hence its price is just as likely to rise as it is to fall. From a forecasting perspective, the martingale hypothesis implies that the "best" forecast of tomorrow's price is simply today's price, where "best" means minimal mean-squared error.

Another aspect of the martingale hypothesis is that nonoverlapping price changes are uncorrelated at all leads and lags, which implies the ineffectiveness of all linear forecasting rules for future price changes based on historical prices alone. The fact that so sweeping an implication could come from as simple a model as (3.18) foreshadows the important role that the martingale will play in the modelling of asset price dynamics.

In fact, the martingale was long considered to be a necessary condition asset market, one in which the information contained in past prices is instantly, fully, and perpetually reflected in the asset's current price. If the market is efficient, then it should not be possible to profit by trading on the information contained in the asset's price history; hence the conditional expectation of future price changes, conditional on the price history, cannot be either positive or negative (if shortsales are feasible) and therefore must be 0. This notion of efficiency has a wonderfully counterintuitive and seemingly contradictory flavor to it: The more efficient the market, the more random is the sequence of price changes generated by the market, and the most efficient market of all is one in which price changes are completely random and unpredictable.

However, one of the central tenets of modern financial economics is the necessity of some trade-off between risk and expected return,and, although the martingale hypothesis places a restriction on expected returns, it does not account for risk in any way. In detail, if an asset's expected price change is positive, it may be the reward necessary to attract investors to hold the asset and bear its associated risks. Therefore, despite the intuitive appeal that the fair-game interpretation might have, it has been shown that the martingale property is neither a necessary nor a sufficient condition for rationally determined asset prices.

In this thesis we will focus on verifying the properties of random walk models, because the martingale hypothesis is not in itself well testable. This is, for example, due to the need to define the relationship to risk. Table 3.1 shows the relationship between random walk and martingale models.

Table 3.1 Classification of random walk and martingale hypotheses

$\boldsymbol{cov[f(r_t, g(r_{t+k})] = 0}$	$\boldsymbol{g(r_{t+k}),\ \forall g(\cdot) Linear}$	$\boldsymbol{g(r_{t+k}), \forall g(\cdot)}$
$\boldsymbol{f(r_t)\ , \forall g(\cdot) Linear}$	Uncorrelated Increments, Random Walk 3: $proj(r_{t+k}\|r_t) = \mu$	–
$\boldsymbol{f(r_t),\ \forall g(\cdot)}$	Martingale/Fair Game: $E(r_{t+k}\|r_t) = \mu$	Independent Increments, Random Walks 1 and 2: $pdf(r_{t+k}\|r_t) = pdf(r_{t+k})$

Source: Campbell, Lo, and MacKinlay (1997)

3.4 Methods for testing the information efficiency

One of the earliest and most enduring questions in financial metrology is whether the prices of financial assets are predictable. Perhaps due to the obvious similarities between financial investment and luck-dependent games, the mathematical models of asset prices have a very rich history before all other aspects of economic analysis. The fact that many outstanding mathematicians and scientists use a lot of techniques to predict the price of financial securities proves the appeal and challenge of this problem. Indeed, modern financial economics is firmly rooted in early attempts to "beat the market", which is still attracting attention and is discussed in journal articles, conferences and cocktail parties.

In this chapter, we will consider the problem of predicting future price fluctuations only by forecasting using past price fluctuations. Limiting forecasts based on past price fluctuations seems too strict to arouse people's interest, but after all, investors are always getting a lot of information. Nonetheless, this is simple because it can provide surprisingly rich insights into asset price behavior. Facts have proved that the two most important ideas in random theory and financial economics, namely mar and random walk, all come from this relatively elementary exercise.

3.4.1 Test of random walk model I

In this subchapter, we will briefly introduce the four tests which are test of sequences, autocorrelation test, BDS test and variance ratio test for RW1.

a) Test of sequences and reversal

This implies that for any pair of consecutive returns, a sequence and a reversal are equally probable; the Cowles-Jones ratio $\widehat{CJ} = N_s/N_r$, should be approximately equal to one. N_s is the number of sequence and N_r is the number of reversals. More formally, this ratio may be interpreted as a consistent estimator of the ratio CJ of the probability π_s, of a sequence to the probability of a reversal $1 - \pi$, since:

$$\widehat{CJ} = \frac{N_s}{N_r} = \frac{N_s/n}{N_r/n} = \frac{\widehat{\pi_s}}{1-\widehat{\pi_s}} \to \frac{\pi_s}{1-\pi_s} = CJ = \frac{1/2}{1/2} = 1. \tag{3.20}$$

However, the assumption of a zero drift is critical in determining the value of *CJ*. In particular, *CJ* will exceed one for an *IID* random walk with drift, since a drift--either positive or negative--clearly makes sequences more likely than reversals. (Cowles and Jones, 1937). To see this, suppose that log prices follow a normal random walk with drift:

$$p_t = \mu + p_{t-1} + \epsilon_t, \epsilon_t \sim N(0, \sigma^2). \qquad (3.21)$$

where $p_t = lnP_t$. Then the indicator variable I_t is no longer a fair coin-toss but is biased in the direction of the drift, i.e.,

$$I_t = \begin{cases} 1 \; with \; probility \; \pi \\ 0 \quad with \; probility \; 1 - \pi' \end{cases} \qquad (3.22)$$

where

$$\pi = \delta(\frac{\mu}{\sigma}). \qquad (3.23)$$

If the drift π is positive then $\pi > 1/2$, and if it is negative then $\pi < 1/2$. Under this more general specification, the ratio of π, to $1 - \pi$, is given by

$$CJ = \frac{\pi^2 + (1-\pi)^2}{2\pi(1-\pi)} \geq 1. \qquad (3.24)$$

As long as the drift is nonzero, it will always be the case that sequences are more likely than reversals, simply because a nonzero drift induces a trend in the process. It is only for the "fair-game" case of $\pi = 4$ that CJ achieves its lower bound of 1.

Testing statistics t_{CJ} has the form:

$$t_{CJ} = \frac{\widehat{CJ} - \pi_S/(1-\pi_S)}{(var\widehat{CJ})^{1/2}} \sim N(0,1), \qquad (3.25)$$

$$var\widehat{CJ} = \frac{\pi_S \cdot (1-\pi_S) + 2 \cdot [\pi^3 + (1-\pi)^3 - \pi_S^2]}{n \cdot (1-\pi_S)^4}, \qquad (3.26)$$

where the probability of the sequence $\pi_s = \pi^2 + (1-\pi)^2$ and the probability of positive return $\pi = \delta(\frac{\mu}{\sigma})$. Then null hypothesis is rejected at α significance level if: $|t_{Cj}| > Z_{1-\alpha/2}$, where $Z_{1-\alpha/2}$ is the $1 - \alpha/2$ quantile of $N(0,1)$ distribution.

b) Autocorrelation test

According to any of the three versions of the random walk hypothesis, there are a number of returns. Uncorrelated and in the case of the RW1 model it is also stationary. As a result, the corresponding autocorrelation function for all delays has the following form:

$$\rho(k) = \frac{cov[r_t, r_{t+k}]}{var(r_t)} = \frac{\gamma(k)}{\gamma(0)} = 0, \tag{3.27}$$

where the second and third equality are the result of stationarity in covariance r_t. Tests have been proposed in the literature to analyze the non-correlation condition required for all three types of random walk models. The test statistics of these tests are based on an estimate according to equation (3.27) such that for a sample of size n:

$$\hat{\rho}(k) = \frac{\hat{\gamma}(k)}{\hat{\gamma}(0)}, \tag{3.28}$$

$$\hat{\gamma}(k) = \frac{1}{n}\sum_{t=1}^{n-k}(r_t - \bar{r})(r_{t+k} - \bar{r}), 0 \leq k < n. \tag{3.29}$$

Anderson (1942) shows that the distribution of $\rho\,(k) = 0$ is approximately normal. Several path-correlation tests can be found in the literature. These are the Bartlet test (Bartlet, 1946), the Box and Tenkins defined test (Box and Jenkins, 1970) or the Box – Pierce test (Box and Pierce, 1970).

Ljung and Box (1978) defined the Q statistic as an alternative to different autocorrelation hypotheses with different time delays. This test eliminates the shortcomings of the above autocorrelation tests. The tested hypotheses have the form:

$$\begin{cases} H_0: \rho(1) = \rho(2) = \cdots = \rho(k) = 0 \\ H_1: At\ least\ one\ \rho(j) \neq 0, j \leq k \end{cases}. \tag{3.30}$$

Ljung and Box (1978) proposed to test hypotheses (3.30) based on the correction of the test statistics of the Box – Pierce test for final selections. The Ljung-Box Q statistic is a linear combination of autocorrelation squares, each of which has the same weight and is defined as:

$$Q(k) = n(n+2)\sum_{j=1}^{k}\frac{\widehat{\rho^2}(j)}{n-j}. \tag{3.31}$$

which follows $\chi^2\ (k)$ by distribution. This test is much stronger than, for example, the Box – Pierce autocorrelation test.

c) BDS test

This series of views runs independent BDS tests as described in Brock et al. (1996). BDS testing is a set of time-based dependency tests. It can be used to test various possible deviations of independence, including linearity, nonlinearity or chaos.

We can apply this test to a set of estimated residuals to see if the residuals are independent and identical in distribution (IID). For example, we can test the residuals in the ARMA model to see if the series has nonlinear dependence after fitting a linear ARMA model.

The idea behind the test is very simple. To run the test, first select the distance. Next, consider two points. For any pair of points, if the observations in the series are indeed, the probability that the distance between these points is less than or equal to epsilon is constant. We can also consider a pair of points. One way to select a set of pairs is to cycle through continuous observations of the sample. In other words, given observations t and Series X observations, we can create a set of pairs of the form:

$$\{\{x_\delta, x_\iota\}, \{x_{\delta+1}, x_{\iota+1}\}, \{x_{\delta+2}, x_{\iota+2}\} \cdots \{x_{\delta+m-1}, x_{\iota+m-1}\}, \tag{3.32}$$

where m is the number of consecutive points used in the set or embedding dimensions. It represents the simultaneous probability of all point pairs in the set that satisfy the $c_m(\epsilon)$ condition.

The BDS test continues under the assumption of independence and indicates that the probability is the product of the probabilities of each pair of individuals. That is, if the observations are independent,

$$c_m(\epsilon) = c_1^m(\epsilon). \tag{3.33}$$

When processing sample data, or when not directly observing. It can only be estimated based on samples. As a result, we cannot expect this relationship to be properly maintained, but we will only get an error. The larger the error, the smaller the error caused by random sample changes. The BDS test provides a formal basis for determining the size of this error.

To estimate the probability of a particular dimension, look at all possible sets of that length that can be extracted from the sample and count the number of sets that meet the criteria. The ratio of the number of sets that meet the criteria divided by the total number of sets provides an estimate of the probability. Given a sample of γ series observations, this condition can be expressed in mathematical notation, and the n observation is to write the correlation integral for m dimensions:

$$c_{m,n}(\epsilon) = \frac{2}{(n-m+1)(n-m)} \sum_{\pi=1}^{n-m+1} \sum_{\iota=\pi+1}^{n-m+1} \prod_{j=0}^{m-1} I_\epsilon(\gamma_{\pi+j}, \gamma_{\iota+j}). \tag{3.34}$$

where I_ϵ is the indicator:

$$I_\epsilon(x,y) = \begin{cases} 1 \; if |x-y| \leq \epsilon \\ 0 \; otherwise. \end{cases} \tag{3.35}$$

Note that the statistics $c_{m,n}$ are often referred to as correlation integrals.

Brock et al. (1996) defined testing statistics:

$$W_{m,n}(\epsilon) = \frac{\sqrt{n-m+1}\left[C_{m,n}(\varepsilon) - \left(C_{1,n-m+1}(\varepsilon)\right)^m\right]}{\hat{\sigma}_{m,n}(\varepsilon)} \sim N(0,1), \tag{3.36}$$

where $\hat{\sigma}_{m,n}(\varepsilon)$ is an estimation of asymptotic standard deviation $C_{m,n}(\varepsilon) - \left(C_{1,n-m+1}(\varepsilon)\right)^m$.

d) Variance ratio test

Under conditions (3.15), the null hypothesis of RW1 is

$$H_0: r_t = \mu + \epsilon_t, \epsilon_t \sim IID(0, \sigma^2). \tag{3.37}$$

The normality assumption is considered for convenience, the Lo's and MacKinley's (1988) results apply more generally to process with *IID* increments with finite forth moments.

Assume stationary time series of $2n+1$ log prices $p_0, p_1, \dots, p_{2n}$. The estimators of σ^2 have forms:

$$\hat{c} = \frac{1}{2n}\sum_{k=1}^{2n}(p_k - p_{k-1}) = \frac{1}{2n}(p_{2n} - p_0), \tag{3.38}$$

$$\widehat{\sigma_a^2} = \frac{1}{2n}\sum_{k=1}^{2n}(p_k - p_{k-1} - \hat{c})^2, \tag{3.39}$$

$$\widehat{\sigma_b^2} = \frac{1}{2n}\sum_{k=1}^{n}(p_{2k} - p_{2k-2} - 2\hat{c})^2. \tag{3.40}$$

Estimators (3.38) and (3.39) are maximum-likelihood estimators of c and σ^2. Estimator (3.40) exploits the random walk property, i.e. the fact that the mean and variance of increments are linear in the increment interval. The above mentioned estimators are consistent and

$$\sqrt{2n}\left(\widehat{\sigma_a^2} - \sigma^2\right) \sim_a N(0,2\sigma^4), \tag{3.41}$$

$$\sqrt{2n}(\widehat{\sigma_b^2} - \sigma^2) \sim_a N(0,4\sigma^4), \tag{3.42}$$

where "$\sim_a$" means asymptotic distribution. Lo and MacKinley (1988) proved that

$$\sqrt{2n}\left(\widehat{\mathrm{VR}}(2) - 1\right) \sim_a N(0,2), \tag{3.43}$$

where

$$\widehat{\mathrm{VR}}(2) = \frac{\widehat{\sigma_b^2}}{\widehat{\sigma_a^2}} \tag{3.44}$$

is the two-period variance ratio statistic. The null hypothesis can be tested by the standardized statistic $\sqrt{2n}\left(\widehat{\mathrm{VR}}(2) - 1\right)/\sqrt{2}$ which has asymptotically standard normal distribution. It means that if the value of standardized statistic lies outside interval $[-1.96, 1.96]$, RW1 can be rejected at the 5% significance level.

Lo and Mackinlay (1988) proposed a normalisation of (3.44) in this form:

$$\psi(k) = \sqrt{n\,k}\left(\widehat{VR}(k) - 1\right)\left(\frac{2\,(2k-1)\,(k-1)}{3\cdot k}\right)^{-\frac{1}{2}} \sim N(0,1). \tag{3.45}$$

The null hypothesis is rejected at α significance level if $|\psi(k)| > Z_{1-\alpha/2}$, where $Z_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of $N(0,1)$ distribution.

Lo and Mackinlay (1988) derived such sampling distributions. We develop some intuition for the population values of the variance ratio statistic under various scenarios. Consider again the ratio of the variance of a two-period continuously compounded return $r_t(2) = r_t + r_{t-1}$ to twice the variance of a one-period return r_t, and for the moment let us assume nothing about the time series of returns other than stationarity. Then this variance ratio, which we write as $VR(2)$, reduces to:

$$VR(2) = 1 + 2\rho(1), \tag{3.46}$$

where $\rho(1)$ is the first-order autocorrelation coefficient of returns $\{r_t\}$. For any stationary time series, the population value of the variance ratio statistic $VR(2)$ is simply one plus the first-order autocorrelation coefficient. In particular, under RW1 all the autocorrelations are zero, hence $VR(2) = 1$ in this case, as expected.

In the presence of positive first-order autocorrelation, $VR(2)$ will exceed one. If returns are positively autocorrelated, the variance of the sum of twoone-period returns will be larger than the sum of the one-period return's variances; hence variances will grow faster than linearly. Alternatively, in the presence of negative first-order autocorrelation, the variance of the sum of two one-period returns will be smaller than the sum of the one-period return's variances; hence variances will grow slower than linearly.

For comparisons beyond one- and two-period returns, higher-order autocorrelations come into play. In particular, a similar calculation shows that the general q-period variance ratio statistic $VR(q)$ satisfies the relation:

$$VR(q) = 1 + 2\sum_{k=1}^{q-1}\left(1-\frac{k}{q}\right)\rho(k), \tag{3.47}$$

where $r_t(k) = r_t + r_{t-1} + \ldots + r_{t-k+1}$ and $\rho(k)$ is the kth order autocorrelation coefficient of $\{r_t\}$. This shows that $VR(q)$ is a particular linear combination of the first $k-1$ autocorrelation coefficients of $\{r_t\}$, with linearly declining weights.

Under RW1, (3.42) shows that for all q, $VR(q) = 1$ since in this case $\rho(k) = 0$ for all $k > 1$. Moreover, even under RW2 and RW3, $VR(q)$ must still equal one as long as the variances of r_t, are finite and the "average variance" $\sum_{t=1}^{T} Var[r_t]/T$, converges to a finite positive number. But (3.47) is even more informative for alternatives to the random walk because it relates the behavior of $VR(q)$ to the autocorrelation coefficients of $\{r_t\}$ under such alternatives. For example, under an $AR(1)$ alternative, $r_t = \emptyset r_{t-1} + \epsilon_t$, (3.42) implies that

$$VR(q) = 1 + 2\sum_{k=1}^{q-1}(1 - \frac{k}{q})\emptyset^k = 1 + \frac{2}{1-\emptyset}\left[\emptyset - \frac{\emptyset^q}{q} - \frac{\emptyset - \emptyset^q}{q(1-\emptyset)}\right]. \quad (3.48)$$

Relations such as this are critical for constructing alternative hypotheses for which the variance ratio test has high and low power.

3.4.2 Test of random walk model II

In this subchapter, we will describe the two tests which are filter rules and technical Analysis for RW2.

a) Filter Rules

To test RW2, Alexander (1964) applied a filtering rule that buys assets when the asset price rises $x\%$ and sells (short-term) assets when the price drops $x\%$. Such rules are called $x\%$ filter, proposed by Alexander (1961), and the reason is as follows: suppose you temporarily assume that there are stock market price trends, but think that they are hidden by market fluctuations. It can exclude all movements smaller than the specified size and check the remaining movements. The total return of this dynamic portfolio strategy is considered a measure of the predictability of asset returns. By comparing total returns with the returns of Dow Jones and Standard & Poor's industry average buy-and-hold strategies, Alexander concluded: "*There is a trend in stock market prices*."

Fama (1965) and Fama and Blume (1966) conducted a more detailed empirical analysis of filter rules, revised dividends and transaction costs, and concluded that such rules are not as effective as buy and hold strategies. In the absence of transaction costs, small filters (1% for Alexander (1964) and 0.5% to 1.5% for Fama and Blume (1966)) give good returns, but small filters are quite reasonable.

b) Technical Analysis

As a measure of predictability, filtering rules have practical advantages. This is a specific and easy-to-implement trading strategy. The indicator of its success is total return. Filter rules are just examples of a large class of trading rules derived from technical analysis or charts. Technical analysis is an investment management method based on the belief that historical price series, trading volume and other market statistics show regularity-usually (but not always) double bottoms, head and shoulders, etc. The form of geometric patterns and support and resistance levels can be used to estimate future price fluctuations (for example, Murphy (1986)). Technical analysis is a science that records the actual trading history of a particular stock or "average" (price fluctuation, trading volume, etc.), usually in a graphical format. Then, from the drawn history, we infer possible future trends.

Historically, technical analysis has always been a "failure" in academic finance. Many scholars believe that this is a pursuit between astrology and voodoo, and technical analysis has never been as widely accepted as basic analysis. Although the distinction between technical analysis and fundamental analysis is becoming blurred, this situation continues to this day.

Perhaps some of the prejudice against technical analysis may be due to semantics. Fundamental analysis is a natural bridge to academic literature because it is based on quantities familiar to most financial economists (such as income, dividends, other balance sheet and income statement items). In contrast, the vocabulary of technical analysts is completely unfamiliar to scholars, but often mysterious to the general public. For example, consider the following problems that may be found in recent academic journals: the first 12 autocorrelation amplitudes and attenuation patterns and the statistical significance of the Box-Pierce Q statistics have been restocked.

Compare this to the following statement. Clearly determined support and resistance levels and the one-third retracement parameters when the price is between the two indicate that there are strong trading opportunities.

These two statements have the same meaning. Past prices can be used to predict future prices in the short term. However, because these two sentences contain many technical terms, the type of response they cause is highly dependent on the individual reading them.

Although the technical terminology is different, recent empirical evidence suggests that technical analysis and more traditional financial analysis may have a lot in common. The study of Treynor and Ferguson (1985) show that financial scientists are increasingly interested in technical analysis, which may lead to more active research fields in the near future.

3.4.3 Test of random walk model III

In this subchapter, we will only introduce only one way of test for RW3 which is variance ratio test.

a) Variance ratio test

Lo and Mackinlay (1988) formulated an alternative version of the variance ration test, which allows the random walk hypothesis that admit posiible heteroscedasticity in the returns. The RW3 can be tested using testing statitic that is based on (3.45):

$$\psi(k) = \frac{\sqrt{n \cdot k} \cdot (\widehat{VR}(k) - 1)}{\sqrt{\hat{\theta}(k)}} \sim N(0,1), \tag{3.49}$$

where

$$\hat{\theta}(k) = 4 \cdot \sum_{i=1}^{k-1} (1 - \frac{i}{k})^2 \cdot \hat{\delta}_i, \tag{3.50}$$

$$\hat{\delta}_i = \frac{n \cdot k \cdot \sum_{j=i+1}^{n \cdot k} (p_j - p_{j-1} - \hat{\mu})^2 \cdot (p_{j-i} - p_{j-i-1} - \hat{\mu})^2}{[\sum_{j=i+1}^{n \cdot k} (p_j - p_{j-1} - \hat{\mu})^2]^2}. \tag{3.51}$$

3.5 Summary

In this chapter, we described the models of efficient market in detail. In particular, our attention was focus on the random walk and martingale models considering their statistical properties. Next, the procedures for detecting the selected type of random walk process were described. Specially, the test of sequences and reversals, autocorrelation test, BDS test and variance ratio tests were introduced. We will use such methods to analyse the data we selected in chapter 5. The test of sequences and reversals, Ljung-Box test and variance ratio test in its homoscedastic version are proper linear tests that can be applied when investigating the random walk model I, while the BDS test is considered very powerful nonlinear test. For the random walk model II, the technical analysis and filter rules can be applied. Heteroscedastic version of the variance ratio test can be utilized for detecting the random walk model III.

4 Data sample analysis

In the whole thesis, we will find out whether there is predictability between China markets and Japan market. We choose two indexes of Chinese markets: The SSE 50 which is the stock index of Shanghai Stock Exchange, including the top 50 companies by "float-adjusted" capitalization and other criteria; HSI index which is calculated from the market value of 52 constituent stocks.

Nikkei 225 which is a stock market index for the Tokyo Stock Exchange. In the first sub-chapter, we will learn the characteristics of the selected stock market. Secondly, we will learn the stock indexes in detail. At last, we will compute the descriptive statistics of stock indexes. We will compare two Chinese stock indexes with the Japanese stock index which is the most standard in the Asia.

4.1 Characteristics of stock markets

For testing the predictability hypothesis between stock markets, we have chosen three stock markets from Asia: Shanghai stock market which is from China mainland; Hong Kong stock market which is also from China; Tokyo stock market which belongs to Japan. There are also other reasons to choose such markets. China is a developing country, and the stock market is still not consummate. But Japan is one of the most powerful countries in Asia. There has fairly good economic environment and financial markets.

4.1.1 Chinese mainland stock market

Form the Ft Chinese news, the China stock market is an emerging market with an investor structure dominated by retail investors, and short-term trading is the mainstay. The market trading is quite active. The biggest feature of China’s stock market is that state-owned shares and legal person shares are promised not to be circulated when they are listed. Therefore, only tradable shares are traded in the market according to the stock price. It was restored during the transition from a planned economy to a market economy. The deep government intervention during its establishment and initial development made the stock market clearly different from other countries.

a) Trading rules

T+1 delivery: Both parties to the transaction complete the transaction-related securities and payment receipts and payments on the next day of the transaction, that is, the buyer receives the securities and the seller receives the payment. Shanghai stock exchanges both implement T+1 settlement for A shares.

Limits on fluctuations: In order to curb excessive speculation and prevent excessive surges and falls in the market, the stock exchange stipulates the range of fluctuations in the trading price of the securities of the day on the basis of the closing price of the previous trading day. Today, the Shanghai and Shenzhen Stock Exchanges impose a 10% price limit. (ST shares and S shares that have not completed the share reform are limited to 5%.)

b) The peculiarities of China's stock market

The Chinese stock market was formed during the transition period from the planned economy financial system to the market economy financial system. It was born when the financial situation of finance, banks and state-owned enterprises became increasingly difficult. The function of the Chinese stock market was set up for the special needs of economic transformation. The reform of the economic system has caused major changes in the national income distribution pattern. The central government's funds are relatively small, and the funds at the disposal of enterprises and individuals have relatively increased, forming a large amount of social precipitation funds. Therefore, the country began to experiment with innovation in the investment and financing system. On the one hand, it raises national construction funds from the society through the issuance of various bonds; on the other hand, it encourages and supports the pilot joint-stock system. Therefore, using the raising function of the securities market to raise funds for economic construction is the foothold that the securities market can generate in China. As China's securities market is under the pressure of state-owned enterprises to finance, it can only be an institutional arrangement of the financing market, while the securities market of a mature market economy country is an institutional arrangement of the investment market. Existing research has shown that the initial institutional arrangements of the financing market have brought many inherent defects to the Chinese securities market.

c) Insufficient functions of the Chinese stock market

The Chinese stock market did not undergo a natural process of gestation like the Western developed countries, but was largely catalysed and promoted by the government. Therefore, a certain degree of functional dislocation and alienation occurred.

The stock market has been given the historical responsibility of raising construction funds, invigorating state-owned enterprises, and smoothly switching from a planned economy to a market economy by the government, which determines that the functional positioning of my country's stock market must focus on raising large amounts of funds and supporting the reform of state-owned enterprises in a certain period of time. However, the basic financing function of the stock market is still weak. Due to the implementation of listing quota management and the asymmetry of market information, large-scale enterprises have difficulties in issuing, non-state-owned enterprises are difficult to obtain issuance quota, and many outstanding enterprises cannot go public to raise funds. What can be seen is that the financing and resource allocation functions of the stock market have been extremely weakened.

In addition, the function of institutional innovation is also weak. Due to the absence of state-owned enterprise owners and insider control, the establishment of a modern enterprise system by listed companies is imperfect, and the transformation of operating mechanisms is not complete; as state shares cannot be circulated, state shares cannot enjoy the benefits of securities appreciation, and there are many speculative acquisitions in the secondary market Regarding operational acquisitions, state-owned enterprises determined by government actions have not been successful after reorganization, which has affected the maintenance and appreciation of state-owned assets, and so on.

Information asymmetry in the Chinese stock market is widespread. Due to the short development time of my country's stock market and the low degree of market regulation, the problem of asymmetric information is more serious. Listed companies have an information advantage, while regulators are at an information disadvantage. Due to the particularity of China's listing selection mechanism and the fundamental defects of the corporate principal-agent mechanism in which the personality of state-owned capital owners absented, the adverse selection risk of information asymmetry in the listing process is exacerbated, and the cost of fraud is far less than the revenue of fraud. The information disclosure of listed companies is false, forming a source of false information. In the process of information transmission and processing, due to the internal control of state-owned enterprises and the dominant shareholding structure of state-owned equity, the lack of supervision of shareholders of state-owned listed companies and the inability to protect shareholders' interests have caused small and medium shareholders to "free ride" in the process of collecting information. "In the process of information transmission and price discovery, problems such as "herding phenomenon", strong speculation, vicious speculation using inside information, market manipulation, and the formation of bubbles have appeared.

4.1.2 Hong Kong stock market

The status of the Hong Kong stock market as the international financial core of Hong Kong depends on its extremely active stock market. In Hong Kong, investors can buy and sell financial products on exchanges, including securities market (also known as spot market) products and derivatives market products. Among them, the securities market includes the Main Board and the Growth Enterprise Market. The Main Board accepts companies that meet certain profit and financial standards and have a certain scale; the Growth Enterprise Market provides a financing and trading platform for companies with growth potential, and there is no requirement for past performance records. Financial products in the main board market are divided into: stocks, warrants, bonds, unit trusts and mutual funds, CBBCs and linked notes. The only financial products in the ChiNext market are stocks and warrants.

a) Trading rules

Trading Hours: The trading hours of the Hong Kong Stock Exchange are divided into two periods: morning and noon. The morning time is from 10:00 am to 12:30 noon, and the lunch time is from 2:30 pm to 4:00 in the currency year. whole. There is no afternoon trading on Christmas Eve, New Year’s Eve or Lunar New Year’s Eve.

Price: Every security traded on an exchange is traded at a designated "price" (minimum price change unit), which represents the smallest range that the price can increase or decrease, and is related to the price range of the security. At present, the exchange's price list stipulates the price range of stocks from the price range of 0.01 to 0.25 Hong Kong dollars (the price is 0.001 Hong Kong dollars) to the price range of 1,000 to 9995 Hong Kong dollars (the price is 2.50 Hong Kong dollars). When the price of a stock rises or falls to another price range, its price will also change. The complete minimum price change unit comparison is shown in Table 1.10 below. This is different from the rules of the mainland market. The mainland market does not distinguish between high and low stock prices, and the minimum price change unit is RMB 0.01.

Minimum trading unit: Unlike the minimum trading unit in the mainland market which is 100 shares, in Hong Kong, the trading unit of each listed security is determined by each listed company and issuer, so it is different. Some stocks require a lot size of 500 shares, and some stocks require a lot size of 2,000 shares.

Opening quotation: The exchange stipulates that the "opening quotation" should be carried out in accordance with procedures to ensure the continuity of prices between two adjacent trading days and prevent drastic market fluctuations when the market opens. The first input on each trading day. The buying or selling orders of the trading system are regulated by the opening quotation rules. The price of the first order cannot exceed 4 price points above and below the closing price of the previous day.

Trading and settlement: The Hong Kong stock exchange adopts the T+0 system for stock trading, that is, the stock can be sold on the same day that the stock is bought. The settlement and delivery of eligible securities traded on the Main Board and

Growth Enterprise Market of the Hong Kong Stock Exchange is the responsibility of the Hong Kong Clearing House. The Hong Kong market adopts the T+2 days settlement system, that is, the trading day plus two trading days for settlement. Exchange participants (i.e. securities firms) matching or reporting transactions through the automatic order matching system must be on each trading day (T Settlement will be completed with the Central Clearing System before 3:45 pm on the second trading day after the following day. In addition, the Hong Kong Stock Exchange has other trading rules that are different from those of the mainland exchanges. For example, the Hong Kong Stock Exchange does not have price limit restrictions but the mainland market has them. The Hong Kong market allows short-selling transactions but the mainland market does not.

b) The industry structure of listed companies is still relatively single, and the market value is highly concentrated

Although the Stock Exchange introduced domestic state-owned enterprises to go public in 1993, it has greatly improved the market structure of the tertiary industry in the Hong Kong stock market listed companies in the past, but in general, the industry structure of Hong Kong listed companies is unbalanced and the market value is too concentrated.

The phenomenon of minority stock companies is still relatively serious. The industry distribution of Hong Kong listed companies, especially the 33 constituent stocks, is basically concentrated in real estate, banking and the information industry represented by China Telecom, while the market value is highly concentrated in the 33 constituent stocks of the Hang Seng Index. Especially in the top 7 stocks by market capitalization. For example, in 2001, the total market value of seven listed companies, including HSBC Holdings, China Mobile, Hutchison Whampoa, Hang Seng Bank, Cheung Kong Holdings, Sun Hung Kai Properties, and China Unicom, accounted for 59.06% of the total market value of the entire listed company market. Historically, the market value of the four stocks of China Mobile, HSBC, Hutchison Whampoa, and Cheung Kong have accounted for about 60% of the total market value of Hong Kong. This structure of the Hong Kong stock market makes the Hong Kong stock market lack of strong support and is prone to stock market crashes under external influence. This is because the real estate industry is the industry that has caused the most serious bubble economy in Hong Kong, and it is an industry that is undergoing adjustment. Although the banking industry has developed relatively steadily, it is greatly affected by the overall economic environment.

c) Increased market volatility, frequent ups and downs

As far as the quality of the entire market is concerned, the Hong Kong stock market undoubtedly has a higher investment value with its average price-earnings ratio of about 10 times. However, the speculative nature of the Hong Kong stock market is also incomparable to other international stock markets. A Japanese scholar once called the Hong Kong stock market "the largest financial casino in Asia." It is precisely because of this high degree of speculation that the Hong Kong stock market fluctuates sharply, and the gap between high and low points within a year is often more than several times. For example, in 1973, the Hang Seng Index had a maximum of 1774.9 points and a minimum of 400.01 points, a difference of more than 4 times. In recent years, the Hong Kong stock market has fluctuated frequently and gradually increased. Looking at the trend of the Hong Kong stock market over the past four years, the most obvious feature is the abnormal volatility, with great ups and downs. In addition, the Hong Kong stock market is too sensitive to the trend of the U.S. stock market, leading to a lack of independence in many cases. The U.S. stock market is the world's largest and most powerful market. The trend of U.S. stocks is definitely an important and even dominant force in the trend of stock markets around the world. With the Hong Kong dollar pegged to the U.S. dollar, Hong Kong stocks have basically been regarded as a dollar asset zone. Therefore, it is normal for Hong Kong stocks to be affected and affected by it. But the problem is that the Hong Kong stock market has almost become the "satellite" of Wall Street, which is obviously abnormal. Judging from the shocks of stock markets around the world caused by the recent fluctuations in U.S. stocks, on January 4, 2000, the Dow Jones Index fell, causing global stock markets to fall. However, the Hong Kong stock market fell the most, with a decline of 11% for two consecutive days. It greatly exceeds the decline in the stock markets of international financial centres such as Tokyo and London, and even far exceeds the decline in Singapore, Seoul, Bangkok, Manila, Jakarta, and Sydney.

d) The degree of openness is higher, and the proportion of overseas investors has increased

In recent years, the Hong Kong securities market has become more attractive to foreign investors. The Hong Kong Stock Exchange's research report shows that as of September 2006, the proportion of overseas investors whose main investment group is institutional investors has reached the past ten years. The highest level since 2005, during the one-year period from October 2005 to September 2006, the proportion of foreign investors' transactions reached 42%. In short, among the different investor categories, overseas institutional investors have the largest proportion of transactions.

4.1.3 Japanese stock market

According to Reuters authority, in the Japanese stock market, new share issuance can be divided into two types according to the purpose of issuance, one is issuance to raise funds (ordinary new share issuance), and the other is issuance for other purposes (special new share issuance).

a) Trading rules

T+3: Both parties to the transaction complete the transaction-related securities and payment receipts and payments on the fourth day of the transaction, that is, the buyer receives the securities and the seller receives the payment.

The trading methods on the exchange can be divided into three types according to the delivery date. The first is ordinary transactions, accounting for 99% of the total transaction volume, and is divided into two forms: spot transactions and credit transactions; the second is settlement transactions on a specified date, but the longest cannot exceed 15 days; the third is the issue date Settlement transactions, that is, actual delivery after the issuance of new stocks, are generally adopted when the company allocates shares to original shareholders or implements stock splits. The store market is mainly used for the trading of newly issued stocks and stocks that do not meet the listing conditions of the exchange. Since the three major stock exchanges opened the second part of the market, the development of transactions in the store market was slow. All traders in the company are not allowed to open or exit positions by placing large orders to create momentum, otherwise they will be punished by stopping trading; they are also not allowed to place meaningless orders. The orders placed must be at the price you want to trade. Scan as many orders as possible.

Up-limit and down-limit: Each stock has a daily price change range. When the order price exceeds this range, the order will be rejected. For example, if the closing price of a stock was 500 the previous day, the allowable trading price fluctuation range for that day is 400-600. When the stock price rises to 600, the stock has a daily limit, and when it drops to 400, it is a limit. It is understood (I haven't personally experienced it) that shipments and positions cannot be closed during the daily and low limit. If the goods are not stopped before the market closes, they can only be counted as overnight, and the goods will be sold after the market opens the next day. Generally, the probability of occurrence is small.

b) Method of issuing shares

According to whether the stock issuance is required to pay the share capital, the issuance of new shares (capital increase) can be divided into two categories: paid issuance (paid capital increase) and free issuance (free capital increase). In the past, there was also a mixed issuance for compensation, but at the end of 1973, due to the invalidation of relevant legal provisions, this method could no longer be used.

- Paid increase in capital:

a) The paid capital increase refers to the stock issuance methods in which the subscribers of new stocks receive payment, which are divided into the following categories:

b) Allocate new shares to original shareholders. That is, the right to subscribe for new shares is only allocated to the original shareholders, the issue price of the new shares is the face value, and the difference between the market price and the face value is obtained by the shareholders as a profit.

c) Third party allotment. That is, the new stock options are handed over to a third party that has a specific relationship with the company, such as the company's customer or the related bank. This approach is conducive to strengthening the relationship with specific customers

d) Public offering. That is, extensively solicit capital from the society, generally using the current stock price in the circulation market as the issue price, so it is also called current issue, but the actual issue price is usually lower than the current price.

e) Mid-price issue. That is to say, the distribution of new shares to the original shareholders is based on the median value of the current price and the face value of the new shares. The difference between current price and face value is shared by the company and shareholders. This method is rarely used in Japan.

- Free capital increase

a) Refers to when the company’s internal retention or profits are incorporated into capital, corresponding new shares must be issued and distributed to shareholders. According to the different sources of capital included in the capital, it can be divided into the following two forms:

b) Free delivery. This method of capital increase is used when capital reserves or profit reserves are included in capital.

c) Share dividends. That is, the dividend is not paid in cash, but the dividend is first incorporated into the capital, and then the corresponding shares are issued and delivered to the shareholders as dividends. Share dividends must be made within the range of possible profits.

d) In the past, Japan has always paid for capital increase through the denomination issuance method of allotment to original shareholders. However, after 1969, this type of issuance has decreased, while public offerings have gradually increased. Although there have been repeated changes since then, by the 1980s, public offerings accounted for 70-80% of corporate stock financing. It can be seen that public offerings have become the main method of corporate financing.

In the Table 4.1, we classify the different trading hours of three markets.

Table 4.1 Trading hours among two markets

Market	Local trading hours (Mon-Fri)
Shang Hai	9:30-11:30 and 13:00-15:00
Hong Kong	9:30-12:00 and 13:00-16:00
Japan	09:00-11:30 and 12:30-15:00

Source: by author

4.2 Description of stock indexes

Having briefly learnt the fundamental knowledge of these three stock markets, for the aim of this thesis, we will select the main index from each market as analysed purpose. The stock index is to measure the performance of a plenty of securities in order to represent a market. In case of this, in this subchapter, we will choose three representative indexes from China and Tokyo. They are SSE 50, HSI and Nikkei 225. And about the time periods, we select basic testing period from the beginning of 2010 to the end of 2020. Since this period is after the 2008 financial crisis, the overall economy is recovering and developing steadily. The data source is from: https://finance.yahoo.com/.

4.2.1 Mainland of China stock index SSE 50

The SSE 50 Index is a stock index of constituent stocks compiled and published by the Shanghai Stock Exchange. It is officially defined as a sample index of the overall situation of a group of leading companies with the most market influence. Based on the principle of combining sample stability and dynamic tracking, the SSE 50 Index adjusts its constituent stocks every six months, and the adjustment time is consistent with the SSE 180 Index. In special circumstances, the sample may also be temporarily adjusted. The proportion of each adjustment is generally not more than 10%. Sample adjustments set up a buffer zone. New samples ranked before 40 are given priority to enter, and old samples ranked before 60 are given priority to retain. The benchmark date of the index is December 31, 2003, with a benchmark number of 1,000 points and 50 constituent stocks. Based on the closing price on September 27, 2019, the tradable market value of the SSE 50 Index constituent stocks is approximately RMB 14.35 trillion.

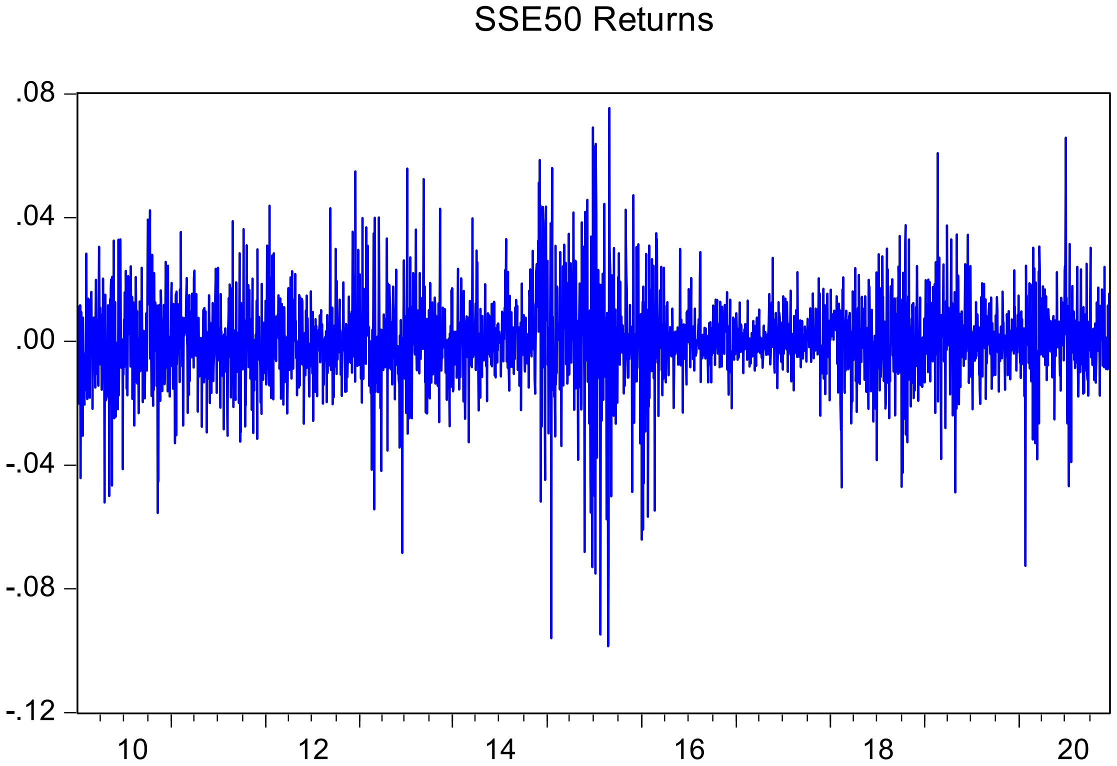
In Figure 4.1, We can see that the overall index is showing an upward trend. On the March 20 of 2014, the line of prices came to its lowest point at approximately 1406 points. The growth rate has fallen, investment has fallen, and the local currency has depreciated. Among them, real estate, infrastructure, and manufacturing investment all showed a downward trend. Although since 2014, the government has gradually begun to loosen policy controls in the areas of currency, finance, and real estate, there has been a time lag from policy implementation to effect. On the December 1 of 2020, index stands at the peak about 3554 points. That was the first time the SSE50 index reached such a high value.

Figure 4.1 Development of SSE50 index from 04/01/2010 to 19/12/2020

Source: own calculation in EViews

As for the aim of thesis, we also produce the charts of the returns of three index. As shown in Figure 4.2, we can learn the volatility of returns directly. During 2014 and 2015, the returns of index volatile largely. During that period, the index price changed greatly and reached its lowest point.

Figure 4.2 Daily returns of SSE50 index from 04/01/2010 to 19/12/2020

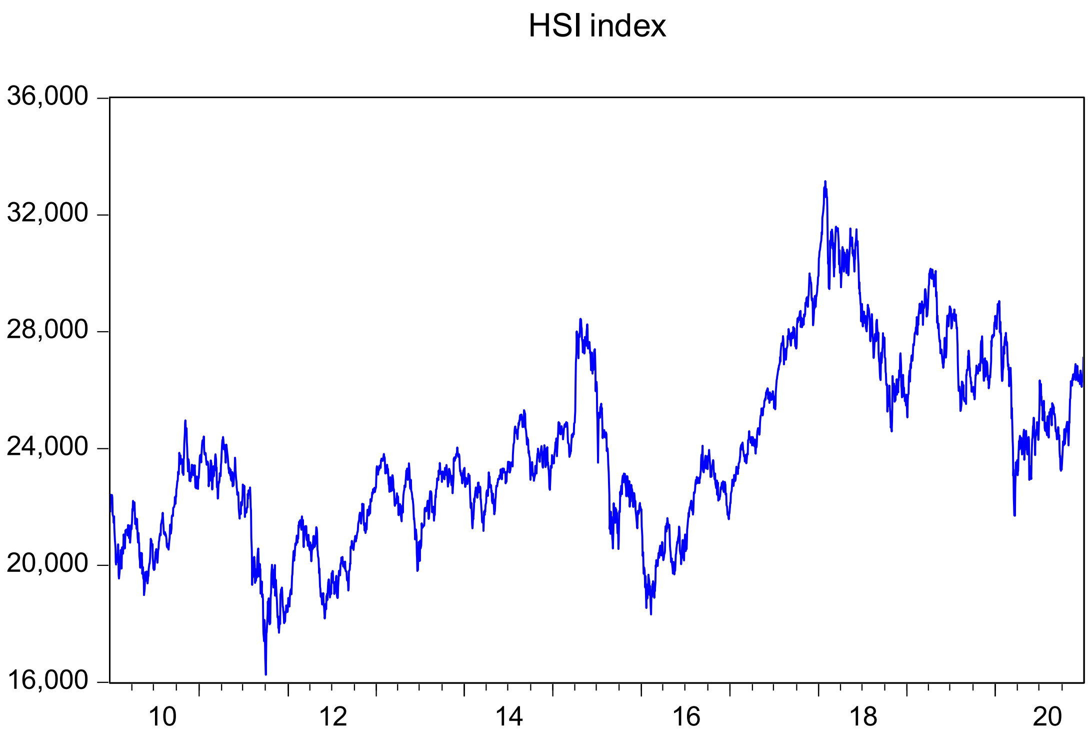


Source: own calculation in EViews

4.2.2 Hong Kong stock index HSI

The Hang Seng Index is an important indicator that reflects the Hong Kong stock market. The index is calculated from the market value of 52 Hang Seng Index constituent stocks and represents 63% of the 12-month average market value coverage rate of all listed companies on the Hong Kong Stock Exchange. The constituent stocks of the Hang Seng Index are Hong Kong's blue-chip stocks. The Hang Seng Index is calculated and reviewed on a quarterly basis by the Hang Seng Index Co., Ltd., and adjustments to constituent stocks are announced. The Hang Seng Index was created in 1964 by Guan Shiguang, the head of Hang Seng Bank's research department. The Hang Seng Index was officially released on November 24, 1969. Over the years, the Hang Seng Index has been cited as an important indicator of the performance of the Hong Kong stock market. The index includes companies listed on the main board of the Hong Kong Stock Exchange with the largest market capitalization and the most active transactions. In the early days of its launch, the HSI was only quoted once a day; however, driven by computerization, the HSI has already moved close to real-time quotes every ten seconds, and in April 2013 it has accelerated to every two-second quotes.

Figure 4.3 Development of HSI index from 04/01/2010 to 19/12/2020

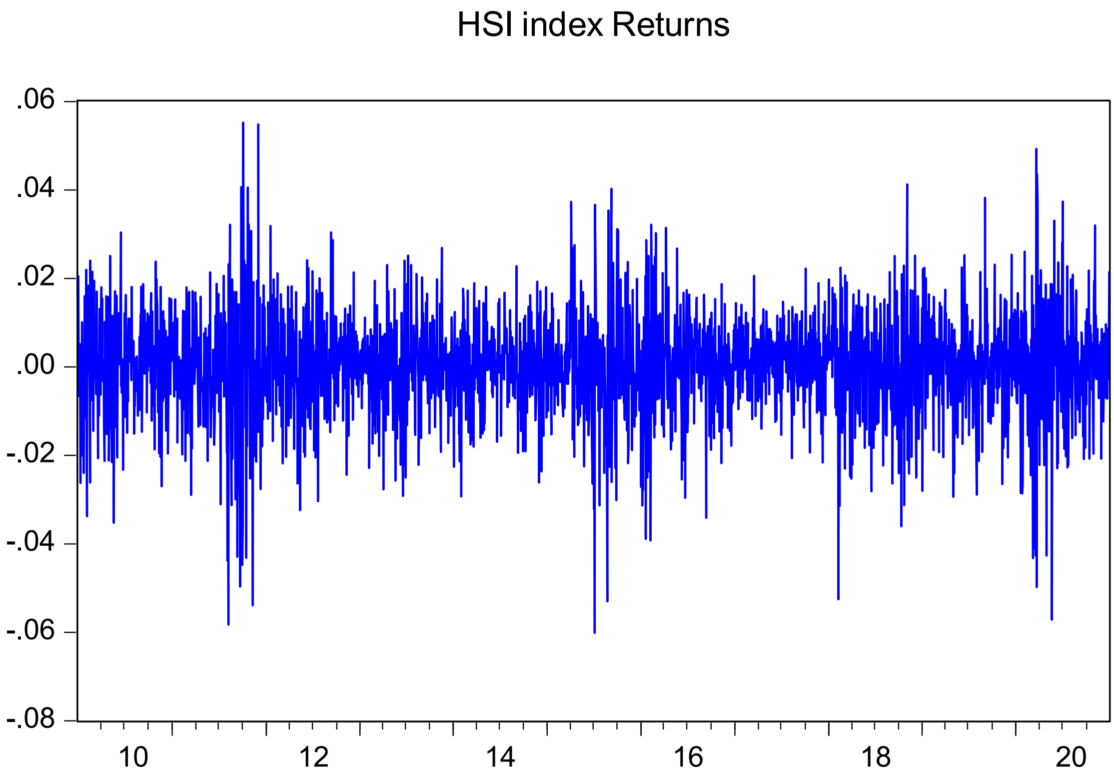


Source: own calculation in EViews

As shown in Figure 4.3, nn 2011, the international economic situation was turbulent and Europe and the United States staged debt crises one after another. The index dropped to 16,170.35 points in October 2011. In the following four years, the economy gradually recovered. The influence of mainland China's economy on the Hong Kong stock market took an important position. Rising, the opening of Shanghai-Hong Kong Stock Connect on November 17, 2014, following the big bull market of A shares in 15 years, rose all the way to a maximum of 28588.52 points, and fell by 18278.80 points in February 2016. From February 2016 to the present, the Hang Seng Index has been in an upward state, and the form of the slow bull is basically determined. As of now, the Hang Seng

Index has regained 26,000 points. The prosperity or depression of China, either due to economic changes in Europe and America, or due to the development of China, are all triggered by sudden political or economic events to form a bull-bear turn.

Figure 4.4 Daily returns of HSI index from 04/01/2010 to 19/12/2020



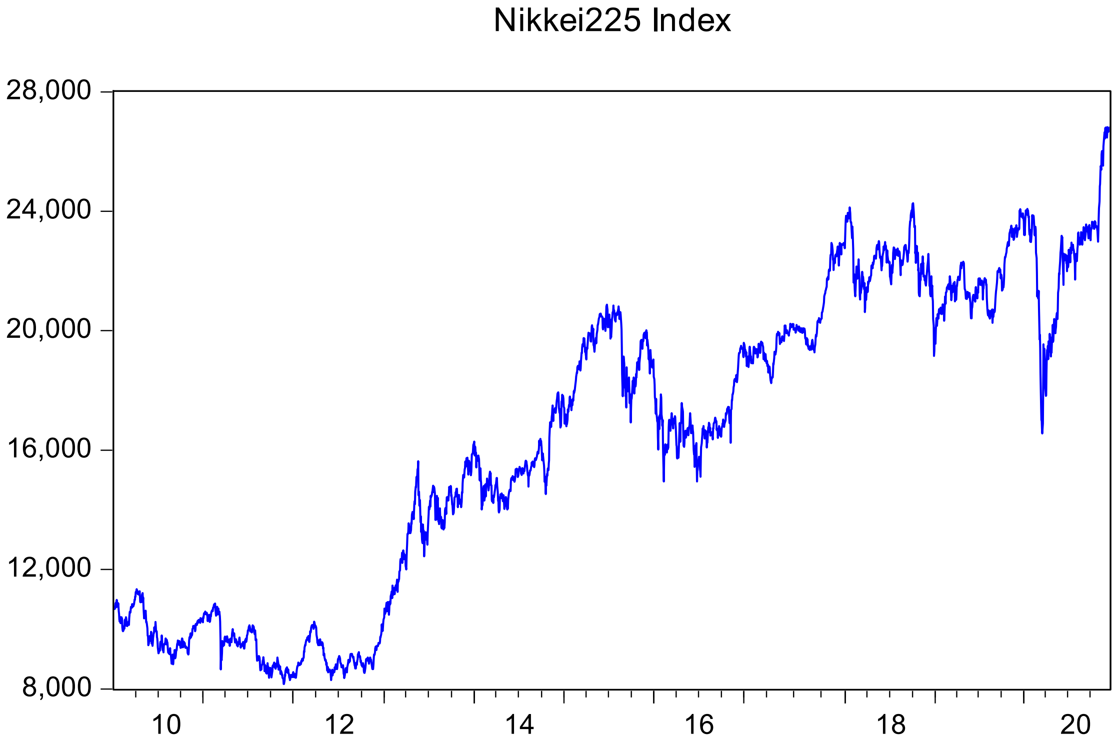
Source: own calculation in EViews

In Figure 4.4, Hong Kong stock market has four periods of bigger volatility: 2011, 2015, 2018 and 2020 respectively. The reason is same as the trend as the price we descried.

4.2.3 Japan stock index Nikkei 225

The Nikkei 225 Index is the 225 stock prices index of the Tokyo Stock Exchange launched by the Nikkei. Therefore, this index lasts for a long time and has good comparability. It has become the most commonly used and most reliable indicator to examine the long-term evolution and the latest changes of the Japanese stock market stock prices. The Nikkei Index cited by the media refers to this index. The compilation of the Nikkei 225 Index began in 1949. It is composed of the prices of 225 stocks listed on the Tokyo Stock Exchange. This index, calculated and managed by Nippon Keizai Shimbun Co., Ltd. (NKS), is disseminated through major international price reporting media and is widely used by various countries as a reference for representing the Japanese stock market. In September 1986, the Singapore International Financial Exchange (SIMEX) launched the Nikkei 225 stock index futures, which became a major historical development milestone. Since then, the trading of Nikkei 225 stock index futures and options has also become an integral part of the investment strategies of many Japanese securities firms.

Figure 4.5 Development of Nikkei225 index from 04/01/2010 to 19/12/2020

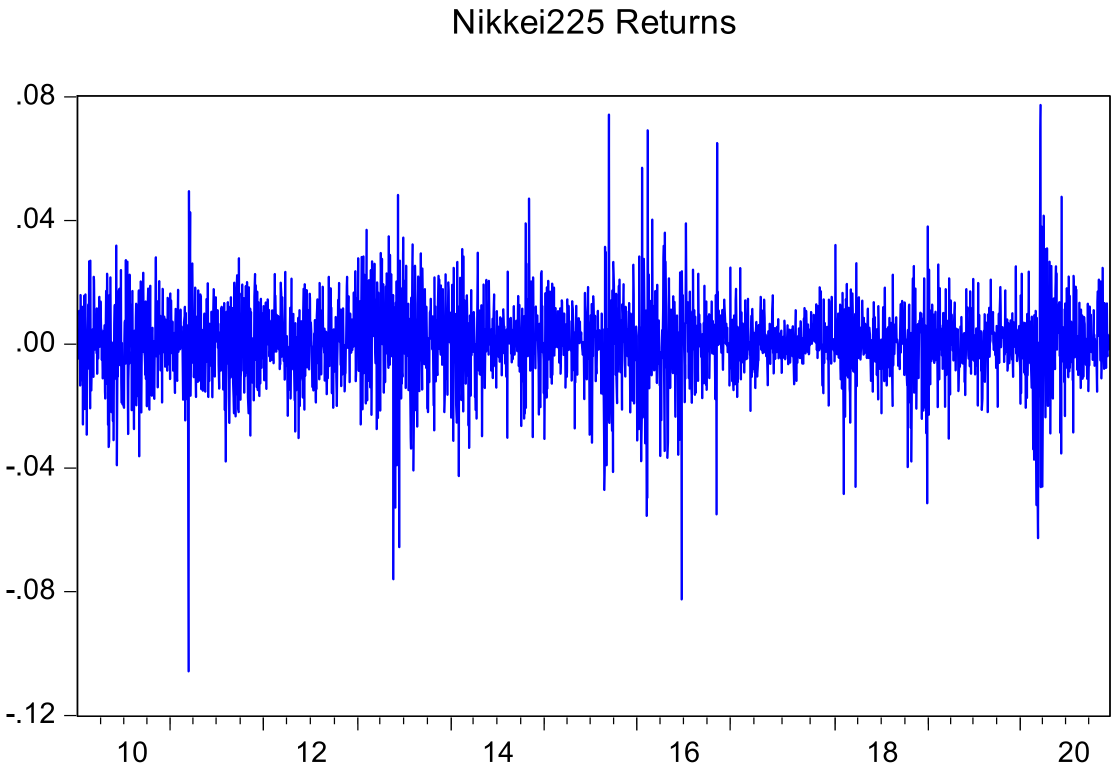


Source: own calculation in EViews

As shown in Figure 4.5, Compared with the SSE 50, the overall trend of the Nikkei 225 is soft. In the past ten years, the Nikkei 225 has shown an overall upward trend and it has more investment value. After the 2008 global financial crisis, Japan took about 5 years to recover its economy. This can be seen from the index image. And it has been rising, and finally reached a peak of more than 27,000 points recently.

In Figure 4.6, we can obviously know the biggest of volatility after 2008. That is in the 2011, severely affected by the financial crisis. Moreover, in 2015, the Nikkei255 index welcome its second big volatility.

Figure 4.6 Daily returns of Nikkei225 index from 04/01/2010 to 19/12/2020



Source: own calculation in EViews

4.3 Descriptive statistics of stock indexes

In this thesis, for the purpose of testing, we choose the basic testing period from the January of 2010 to the end of 2020. This testing period also called as post-crisis periods. With the announcement of NBER authority, the global financial crisis came to its ending on June 30, 2009. So that the economy of all markets slowly recovered during the ten years after crisis. We can see the general trend in Figures 4.1, 4.3 and 4.5.

In Table 4.2, we can see the results of mean are all positive which means each stock market were running normal and even great averagely during this period. As for the standard deviation, SSE50 has a little higher than Nikkei225 and HSI which means Chinese mainland stock market were with higher risk. As for the skewness, the objective markets reached negative values which means the distribution of returns were right skewed than normal distribution. The kurtosis of any univariate normal distribution is 3. The SSE50 and Nikkei225 are above 8 which means the return of two markets were not normal distribution. The HIS index is not normal distribution, either. The values of Jarque-

Bera (*J-B*) test of Nikkei index is higher than SSE and HSI. The larger *J-B* value indicates that returns are not normally distributed. The probabilities of *J-B* test statistics are approaching to zero. Detailed results of descriptive statistics are given in Annex 1.

Table 4.2 Descriptive statistics of daily returns during 04/01/2010 to 19/12/2020

	SSE 50	**HSI**	**Nikkei 225**
Mean	0.000125	0.000061	0.000340
Maximum	0.075470	0.055187	0.077314
Minimum	-0.098523	-0.060183	-0.105758
Std. Deviation	0.014790	0.011727	0.013290
Skewness	-0.405210	-0.336200	-0.436341
Kurtosis	8.280691	5.427889	8.268587
***J-B* test**	3169.404	711.8953	3220.336
Probability	0.000000	0.000000	0.000000

Source: own calculation in EViews

5 Empirical results

The focus of the fifth chapter is purely empirical. The fifth chapter of this thesis is devoted to empirical testing of the properties of efficient market models on the main indexes of the Chinese mainland, Hong Kong and Japanese stock markets, which are in line with the main goal of this thesis declared in the introduction. The introduction of this chapter focuses on the formulation of the testing procedure. This procedure will be based on the knowledge contained in the previous chapters.

The following subchapter is devoted to empirical testing from a static point of view. The static view will be based on a comparison of the results of linear and nonlinear statistical tests of first and third type random walk models. The test results are compared between individual indexes.

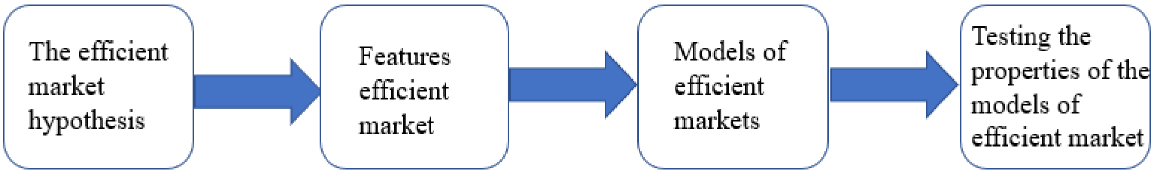
Then, the dynamic of the development of statistical test results is assessed. The non-linear BDS independence test was chosen as the most suitable test for the RW1 model, while in the case of the RW3 model the variance ratio test was chosen. Then the proposal of a trading strategy based on nonlinear relationships in returns will described in subchapter 4. We would learn the trading strategy base on the approaches we applied.

All empirical calculations will be showed in the EViews software, with the exception of the test of sequences and reversals that will be carried out in Microsoft Excel.

5.1 Procedure for testing efficient market models

The procedure of empirical testing is based on the findings presented in the previous chapters. The efficient market hypothesis was defined in subchapter 2.3. Subchapter 2.6 was focused on selected features of the efficient market. In this thesis, the randomness of change in asset prices is verified on real data. It is therefore a matter of verifying that these changes are independence order to test this feature, it is firstly necessary to define an efficient market model. In subchapter 3.2 and 3.3, selected mathematical models of the efficient market were defined. For the purposes of empirical testing, random walk models of the first and third types were chosen and will be utilized to test on real data from financial markets. The procedure for testing the weak form of efficiency is shown in Figure 5.1.

Figure 5.1 The procedure of testing the efficient market models



Source: by author

In the case of the application of statistical tests, it is not a question of testing a weak form of efficiency, but of verifying the predictability of returns. The notion of revenue predictability should not be confused with the notion of market inefficiency. While it is possible to predict future returns to some extent, it can still be an information-efficient market if the returns from this prediction are lower than transaction costs. We cannot rule out a degree of predictability in time series of returns, but this rate may not be economically significant. The difference between economic and statistical significance may be due, for example, to non-compliance with model conditions. In practice, the existence of transaction costs that must be considered when interpreting the results.

To provide statistical evidence of the (in) effectiveness of the stock market, it is necessary to identify a strategy that provides a risk-adjusted return higher than the risk-free rate after deducting transaction costs, and to test this strategy empirically on the data. But that is not the aim of this thesis.

The results of statistical tests can be evaluated from several perspectives. In this thesis, the results will be evaluated from a static point of view, but also from a point of view that includes market dynamics. The test results will be presented in the form of clear tables. More detailed results for selected tests will then be given in the annexes.

However, a comparatively static view has only a limited explanatory power. If deviations from the random walk model are found in the selected test period, it is advisable to verify the stability of the results over time. For this reason, testing of the properties of the first and third type random walk model will be performed using the most suitable tests. The period from January 2010 to December 2010 was chosen as the initial testing period. This is a period of stability and continuous growth of all indices. The dynamics of development is then monitored on monthly changes in test results in the period from January 2011 to December 2020. The test results will be shifted with a monthly frequency, then presented in graphical form and also in the form of a clear table.

5.2 Empirical results - static approach

Empirical tests will first be applied to the returns of the SSE50, HIS and Nikkei225 indices. The testing procedure will be based on subchapters 3.4.1 and 3.4.3, in which the relevant statistical tests of models RWI and RW3 have been defined. The corresponding p-values for the statistical tests were obtained in this subchapter in an asymptotic manner. The symbols *, **, *** indicate that the test criterion is statistically significant at the 10%, 5%, and 1% significance levels, respectively. The null hypothesis is always a random walk model of the first or third type, depending on the test used. This is a verification of the predictability of revenues, not their effectiveness.

The first step is to verify the random walk models of the first type. Linear and nonlinear tests are used for this purpose. From the linear tests, the test of sequence and reversal and the Ljung-Box autocorrelation test will be applied. In addition to linear tests, the RWI will be tested with the strongest nonlinear BDS independence test. The next step is to test the RW3 models, which is less restrictive. In addition, there will be applied two versions of the variance ratio test. The first one is stricter and does not allow heteroskedasticity in returns, and is used to test a random walk model of the first type, while the second option is adapted to the presence of heteroskedasticity.

In order to capture the main aim of the thesis we will prove the following hypothesis:

H_0: The stock index is not predictable (follows the random walk model)

H_1: The stock index is predictable (doesn´t follow the random walk model)

If the stock index follows random walk model, we fail to reject H_0, so the stock market is not predictable.

5.2.1 Test of sequences and reversals

The first linear method applied to verify the RWI model is the test of sequences and reversals. The results of this test for each sub-test period and the indices tested are shown in Table 5.1.

In the first step, it was necessary to identify the number of sequences and reversals in the returns of the indices and then also calculate the value of the Cowls-Jones (*CJ*) ratio. Since there is a positive drift in the returns, it is also necessary to calculate the value of the probability of a positive return for each index. For both indices, this value is greater than 50%. Probabilities, of course, may vary between indices and periods because the sample means and sample standard deviations that affect the calculation of this ability may differ. Finally, the values of the test criteria are calculated based on the equation (3.25) and the corresponding *p*-values are calculated.

Table 5.1 Results of test of sequences and reversals

	SSE 50	**HSI**	**Nikkei 225**
$\boldsymbol{\pi}$	0.5034	0.5030	0.5102
$\boldsymbol{\pi_s}$	0.5000	0.5000	0.5002
$\boldsymbol{N_s}$	1301	1347	1326
$\boldsymbol{N_r}$	1364	1366	1387
$\boldsymbol{CJ}$	0.9538	0.9861	0.9560
$\boldsymbol{varCJ}$	0.0015	0.0015	0.0015
$\boldsymbol{t_{CJ}}$	-1.1944	-0.3640	-1.1654
$\boldsymbol{p-value}$	0.2323	0.7159	0.2439

Source: own calculation in Excel

Table 5.1 shows that for none of the indices is it possible to reject the null hypothesis that returns are *IID* in any test period. The value of the *CJ* ratio is less than one for all indexes. Thus, a higher number of reversals than sequences was recorded. The values of the *CJ* ratio do not differ statistically significantly from the value of 1. However, it is necessary to address the stability of such a phenomenon over time. If the value of *CJ* differed statistically significantly from 1, it would be possible to design a profitable trading strategy. For example, if the number of sequences were greater than the number of reversals, it could be inferred that a positive yield could be obtained by using "a buy strategy" if the previous return was positive and "a sale strategy" if the previous return was negative. The question is whether the return after deducting transaction and information costs would be higher than the risk-free rate.

5.2.2 Autocorrelation test

The Ljung-Box test is used to test the linear dependence in logarithmic returns. To verify the correlation dependence, the values of autocorrelation functions and partial autocorrelation functions will be estimated. In addition, the values of the Ljung — Box *Q* statistic will also be estimated.

Table 5.2 Results of autocorrelation test

	SSE 50		**HSI**		**Nikkei 225**	
Lag	*Q*-Stat	Prob.	*Q*-Stat	Prob.	*Q*-Stat	Prob.
1	0.457	0.499	0.103	0.748	1.377	0.241
2	2.092	0.351	0.162	0.922	4.235	0.120
3	3.520	0.318	1.403	0.705	4.437	0.218
4	5.083	0.279	4.056	0.399	7.430	0.115
5	5.375	0.372	4.478	0.483	7.448	0.189
6	21.070***	0.002	8.795	0.185	7.804	0.253
7	24.923***	0.001	11.192	0.130	8.862	0.263
8	26.652***	0.001	11.812	0.160	9.011	0.341
9	30.807***	<0.001	13.085	0.159	15.369	0.081
10	30.892***	<0.001	14.298	0.160	15.449	0.117

Source: own calculation in EViews

In the table 5.2, we can see the probabilities of HSI index and Nikkei 225 index are all more than 5% significance level which means it cannot reject the null hypothesis. For the 1 to 5 lags of SSE 50 index, it is the same. But after lags 5, there are probabilities less than 5% which means a rejection of the null hypothesis.

Table 5.2 shows the values of the autocorrelation function for the first delay, the

Ljung — Box Q-statistics for lags from 1 to 10 and the corresponding p-values for the individual tested indices assuming homoskedasticity, which is in accordance with model RWI. Detailed results of the Ljung — Box autocorrelation test including graphical outputs of autocorrelation functions and partial autocorrelation functions for individual indices are given in Annex 2.

The results in Table 5.2 can be viewed from several perspectives. For the purpose of estimating autocorrelation functions, the effect of asynchronous trading was neglected as it is expected to be minimal for all indices. The same is true for the effect of transaction costs, as their quantification is a very complicated matter for indices and their amount may not be stable over time. Part of the autocorrelation caused by non-trading and transaction costs may affect conclusions about the validity of the random walk hypothesis, but may not be economically viable for achieving above-average returns.

The results of autocorrelation tests can theoretically be used to create a trading strategy. In the case of negative values of autocorrelation coefficients of the 1st order, it is possible to predict that the return rate of a given index will develop in the opposite direction than at time t-1. For positive values of autocorrelation coefficients, the opposite is true. Due to the systematic presence of linear dependencies in the case of the SSE index, it cannot be ruled out that there exists an investment strategy that could be used to take advantage of short-term trends and thus achieve above-average turns.

5.2.3 BDS test

This test is a portmanteau test for time-based dependence in a series. It can be used for testing against a variety of deviations from independence including linear dependence and nonlinear dependence of chaos. To test the hypothesis of a random walk of the first type, the strongest BDS test of independence will also be applied. Using this test, it is also possible to verify the dependence of a non-linear character. The results of this test for each index are shown in Table 5.3.

In small samples or in series that have unusual distributions, the distribution of the BDS test statistic can be quite different from the asymptotic normal distribution. To compensate for this, EViews offers us the option of calculating bootstrapped p-values for the test statistic. When bootstrapped p-values are requested, EViews first calculates the test statistic for the data in the order in which it appears in the sample. EViews then carries out a set of repetitions where for each repetition a set of observations is randomly drawn with replacement from the original data. Also note that the set of observations will be of the same size as the original data. For each repetition, EViews recalculates the BDS test statistic for the randomly drawn data, then compares the statistic to that obtained from the original data. When all the repetitions are complete, EViews forms the final estimate of the bootstrapped p-value by dividing the lesser of the number of repetitions above or below the original statistic by the total number of repetitions, then multiplying by two (to account for the two tails).

Table 5.3 Results of BDS test

	SSE 50			**HSI**			**Nikkei 225**		
Dimens.	*z*-stat.	Asym. Prob.	Boots. Prob.	*z*-stat.	Asym. Prob.	Boots. Prob.	*z*-stat.	Asym. Prob.	Boots. Prob.
2	5.777 ***	<0.001	<0.001	3.147 ***	0.002	0.002	8.417 ***	<0.001	<0.001
3	7.756 ***	<0.001	<0.001	4.731 ***	<0.001	<0.001	12.844 ***	<0.001	<0.001
4	8.773 ***	<0.001	<0.001	6.348 ***	<0.001	<0.001	15.438 ***	<0.001	<0.001
5	10.206 ***	<0.001	<0.001	7.828 ***	<0.001	<0.001	17.168 ***	<0.001	<0.001
6	11.589 ***	<0.001	<0.001	9.169 ***	<0.001	<0.001	18.669 ***	<0.001	<0.001

Source: own calculation in EViews

To perform this test, the nesting dimensions m of order 2-6 were chosen. The selected distance ϵ was set to 0.7, which is a very good starting point if we test the lower nesting dimensions. If the null hypothesis is not rejected, it can be stated that the time series of logarithmic daily returns are independent and have the same probability distribution.

As shown in Table 5.3, the probability of *z*-statistic of these three indexes are all lower than 0.01. The results of the nonlinear BDS independence test are different from the linear tests. For all indices, the null hypothesis is rejected at the 1% level of significance. Thus, time series of returns are not independent and do not have the identical probability distribution. The stability of BDS test results will be verified in subchapter 5.3. Detailed results of BDS independence tests are given in Annex 3.

5.2.4 Variance ratio test of random walk model 1

The last linear test applied to test the properties of the random walk model is the variance ratio test as described in subchapter 3.4.1. This test will be used in two different versions. The first one verifies the null hypothesis in the form of a random walk model of the first type, while the second variant is not so restrictive and allows heteroskedasticity in returns. It therefore verifies predictability in the form of a random walk model of the third type. If the null hypothesis is not rejected, the magnitude of the variance ratio must be around one. This means that the logarithmic returns will be linearly independent. The variance ratio test will be applied for time horizons of 2, 4, 8 and 16 business days. Table 5.4 shows that the results of the variance ratio test of RW1 from 2010 to 2020.

Table 5.4 Results of variance ratio test of RW1

	SSE 50			**HSI**			**Nikkei 225**		
Lags	Var. Ratio	*z*-stat.	Prob.	Var. Ratio	*z*-stat.	Prob.	Var. Ratio	*z*-stat.	Prob.
2	1.020	1.045	0.296	1.012	0.598	0.550	0.981	-0.951	0.342
4	0.999	-0.018	0.985	1.040	1.114	0.265	1.011	0.321	0.748
8	0.988	-0.201	0.841	1.033	0.580	0.562	0.996	-0.077	0.938
16	1.000	0.012	0.991	0.990	-0.115	0.909	0.987	-0.153	0.878

Source: own calculation in EViews

In Table 5.4, we can see that the SSE 50, HSI and Nikkei 225 probabilities of *z*-statistics are greater than 0.05. It means that we fail to reject the null hypothesis. The results show that, assuming homoskedasticity, the results of the variance ratio test indicate no presence of linear dependences in returns. However, in this test it is appropriate to formulate a conclusion on the basis of the heteroskedastic version of this test. Detailed results of variance ratio tests (RW1) are given in Annex 4.

5.2.5 Variance ratio test of random walk model 3

Table 5.5 shows the values of *z*-statistics and the corresponding p-values. In this case, the null hypothesis is a random walk of the third type, which allows heteroskedasticity in returns (Lo and Mackinlay (1988)).

It can be seen that the hypothesis of a random walk of the third type is not rejected for any index. This fact is supported by the results of partial tests for the time horizon of 2, 4, 8 and 16 business days. In sum, a statistically significant value of the test criterion was not identified. The results of both tests of the variance ratio are thus in a harmony with each other. Detailed results of variance ratio tests (RW3) are given in Annex 5.

Table 5.5 Results of variance ratio test of RW3

	SSE 50			**HSI**			**Nikkei 225**		
Lags	Var. Ratio	*z*-stat.	Prob.	Var. Ratio	*z*-stat.	Prob.	Var. Ratio	*z*-stat.	Prob.
2	1.020	0.660	0.510	1.012	0.525	0.600	0.981	-0.520	0.603
4	0.999	0.026	0.979	1.040	0.936	0.349	1.011	0.229	0.818
8	0.988	-0.064	0.949	1.033	0.546	0.585	0.996	0.009	0.993
16	1.000	0.086	0.931	0.990	0.013	0.989	0.987	-0.015	0.987

Source: own calculation in EViews

5.2.6 Summary

In above five subchapters, we used five methods to test the predictability of the indexes. For the test of sequences and reversals, we got three indexes p-values which are more than 0.05. It means we fail to reject the null hypothesis in the 5% significance level. As for the autocorrelation test, the firstly five lags of SSE 50 index got the probabilities which are more than 5%. The other are lower than 0.01 which means we reject the null hypothesis by 1% significance level. The HSI and Nikkei 225 all got the probabilities bigger than 0.05. Then the BDS test, the all probabilities are lower than 1% which lead to a rejection of the null hypothesis.

At last, the two versions of variance ratio test were applied. For the SSE 50 and Nikkei 225 indexes, they both got the probabilities more than 5% which we fail to reject the null hypothesis. As for the HIS index, the probabilities of two tests are both lower than 1% which we reject the null hypothesis. It means that it can be predicted.

5.3 Empirical results - dynamic approach

The previous subchapter gave an overview of the results of statistical tests of random walk models of the first and third types for the whole data sample. The results of these tests indicate whether or not returns are predictable. However, it is not clear whether these results are stable. In this subchapter, the predictability of revenues from the dynamic point of view will be verified.

To verify the predictability of returns over time, the most appropriate statistical tests for testing the first and third type random walk models were selected. The linear variance ratio in its heteroskedastic version was chosen as the most suitable test of the random walk model of the third type. To verify the random walk model of the first type, a nonlinear BDS independence test was chosen.

In order to describe the development of predictability of returns over time, a time window with a fixed length was chosen. This time window will be slid one month forward (Urquhart and McGroarty, 2014). For the purposes of this thesis, a first period of 12 months from January 2010 to December 2010 was chosen. The test statistics were therefore calculated on the interval from the first business day of January 2010, when the base test period begins, to the last business day of December 2010. The next step was a shift from February 2010 to January 2011. The whole process of the shift by one month was repeated until December 2020, when the basic testing period ends. The test window contains a sufficient number of observations to generate reliable results. Moreover, we get sufficient results to analyse how the level of predictability has evolved over time.

The first applied test is the nonlinear BDS independence test. The values of BDS statistics were calculated for nesting dimensions 2 and 3, which are described in the graphs by the symbols BDS2 and BDS3. The second applied test was variance ratio test in its heteroskedastic version. In fact, the values of particular z-statistics for the delays of 2 and 4 business days will be calculated.

One of the problems of the BDS and variance ratio tests is the fact that the test statistics are based on asymptotic estimates, which can lead to erroneous conclusions for smaller samples (Richardson to avoid this problem, the bootstrap method will be applied, which leads to improved results of the test statistics for smaller samples (Kim, 2006). This method involves a calculation of individual and combined statistical tests on T-observation samples generated by weighing the original data with average random variables with a mean of 0 and a variance of 1. These results will be used to generate bootstrapped distributions of test statistics. Bootstrapped *p*-values are calculated directly from a fraction of replications that fall outside the limits defined by the estimated statistics. The number of replications was set at 2500 for each test.

The obtained results show that for each index were identified periods when returns are statistically significantly predictable, but also periods when returns are predictable. These changing levels of predictability depend on the tests used, as the tests vary in strength and verify different aspects of predictability. Therefore, direct market comparisons can only be made if the testing procedure is consistent. In order to be able to compare both indices, it is advisable to examine the results also relatively.

Since the graphic outputs are not completely clear, it is advisable to create an overview of test results also in the form of a clear table. Tables 5.6 shows the percentage of test statistics that do not reject the null independence hypothesis for the BSD test with nesting dimensions 2 and 3. Table 5.7 and 5.8 demonstrates the same results for the variance ratio test in its heteroskedastic version. The results are confronted with a 5% level of significance (Smith, 2012).

5.3.1 BDS test

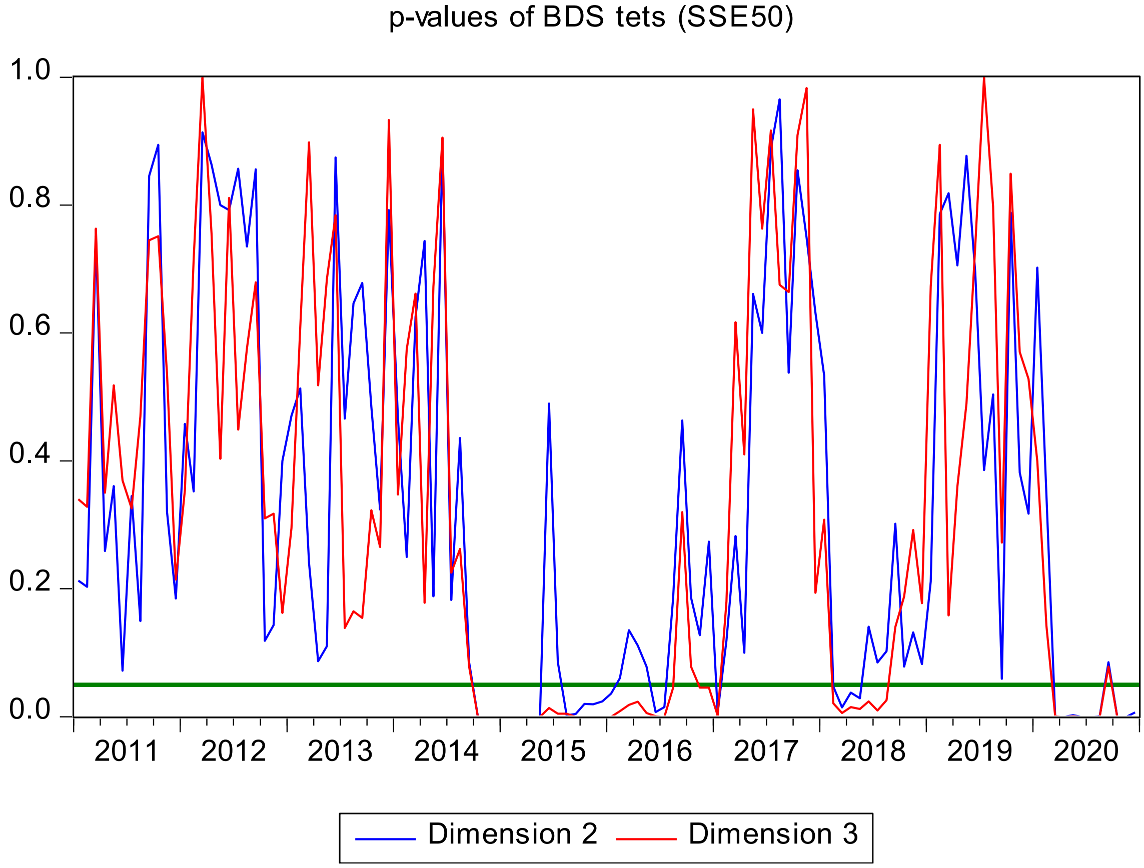
One of the problems of the BDS test is the fact that the test statistic is based on asymptotic estimates, which can get erroneous conclusions especially for smaller samples (Richardson and Stock, 1989). To avoid this situation, the method of bootstrap was applied. This method involves the calculation of individual and combined statistical tests using T-observation samples generated by weighing the original data with average random variables with a mean of 0 and a variance of 1. These results were used to generate bootstrapped distributions of test statistics. Bootstrapped p-values are calculated directly from a fraction of replications that fall outside the limits defined by the estimated statistic. The number of replications was set at 2500.

Figure 5.2 shows the trend of the p-values of the non-linear BDS independence test statistics for the return of the SSE 50 index. The values of BDS statistics were calculated for nesting dimensions 2 and 3, which are described in the graph by the blue line and red line. The statistics significance of BDS tests will again be evaluated on the 5% significance level, which is indicated in the graphs by the green line.

The p-values were obtained based on a time series lasting 12 months. Thus, for example, the p-values as of June 2015 corresponds to the data sample for the period from May 2014 to May 2015. The obtained p-values are variable over time and differ between indexes and nesting dimensions.

In the case of the SSE 50 index, the values of BDS statistics are statistically insignificant from January 2010 to October 2013 for both dimensions 2 and 3. Then there is a period when the values of BDS statistics are significant lasts until June 2014 (for nesting dimension 2) and October 2015 (for nesting dimension 3). Until October 2015, for the nesting dimension 2, there are July 2014, August 2014, March 2015, April 2015, May 2015, June 2015 and September 2015 which are insignificant periods. Then February 2016 is significant for nesting dimension 2 and 3. Then the significant period is from March 2016 to June 2016(for nesting dimension 2) and from March 2017 to September 2017(for nesting dimension 3), respectively. Then it is a period when the values of BDS statistics are insignificant. The last period from March 2020 to December 2020 is significant when the values are all smaller than 0.05.

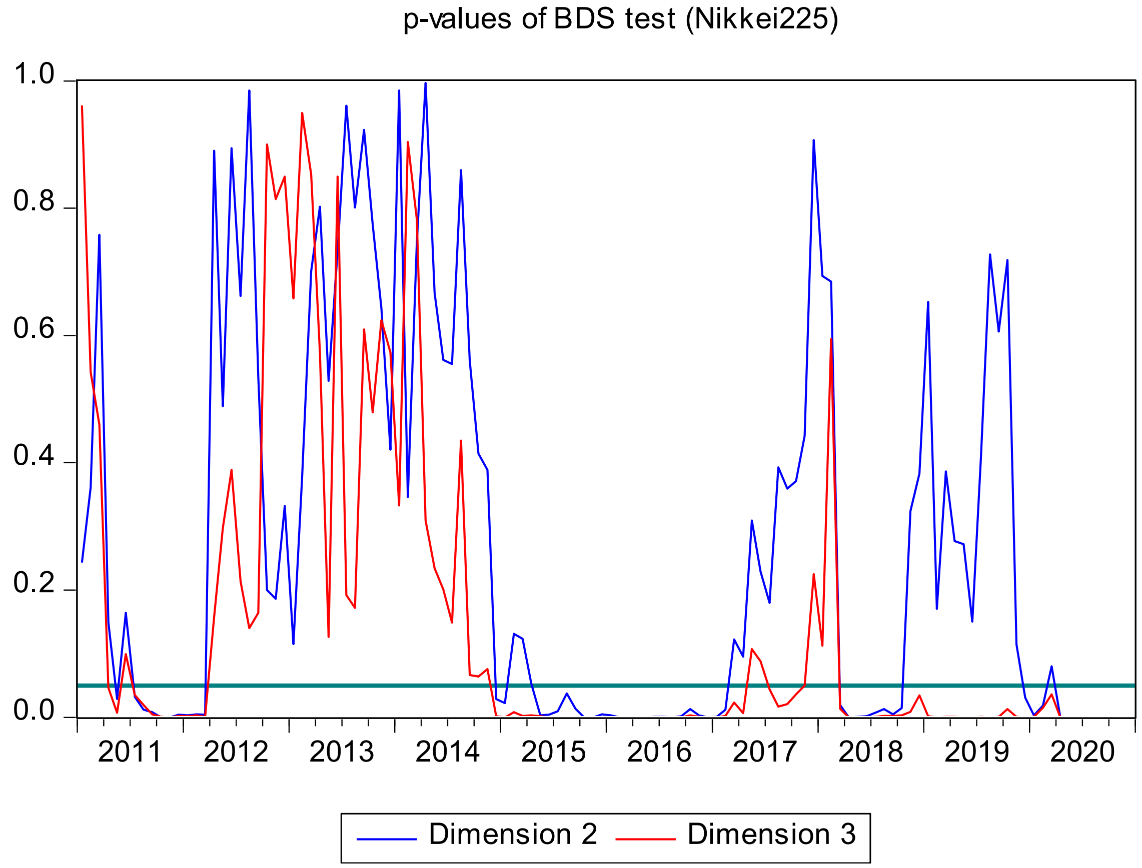
Figure 5.2 Bootstrapped p-values of the BDS independence test with nesting dimensions 2 and 3 for SSE 50 index returns



Source: own calculation in EViews

Figure 5.3 shows the trend of the p-values of the non-linear BDS independence test statistics for the return of the Nikkei 225 index. The values of BDS statistics were calculated for nesting dimensions 2 and 3, which are described in the graph by the blue line and red line. The statistics significance of BDS tests will again be evaluated on the 5% significance level, which is indicated in the graphs by the green line. The p-values were obtained based on a time series lasting 12 months as same as SSE 50 index.

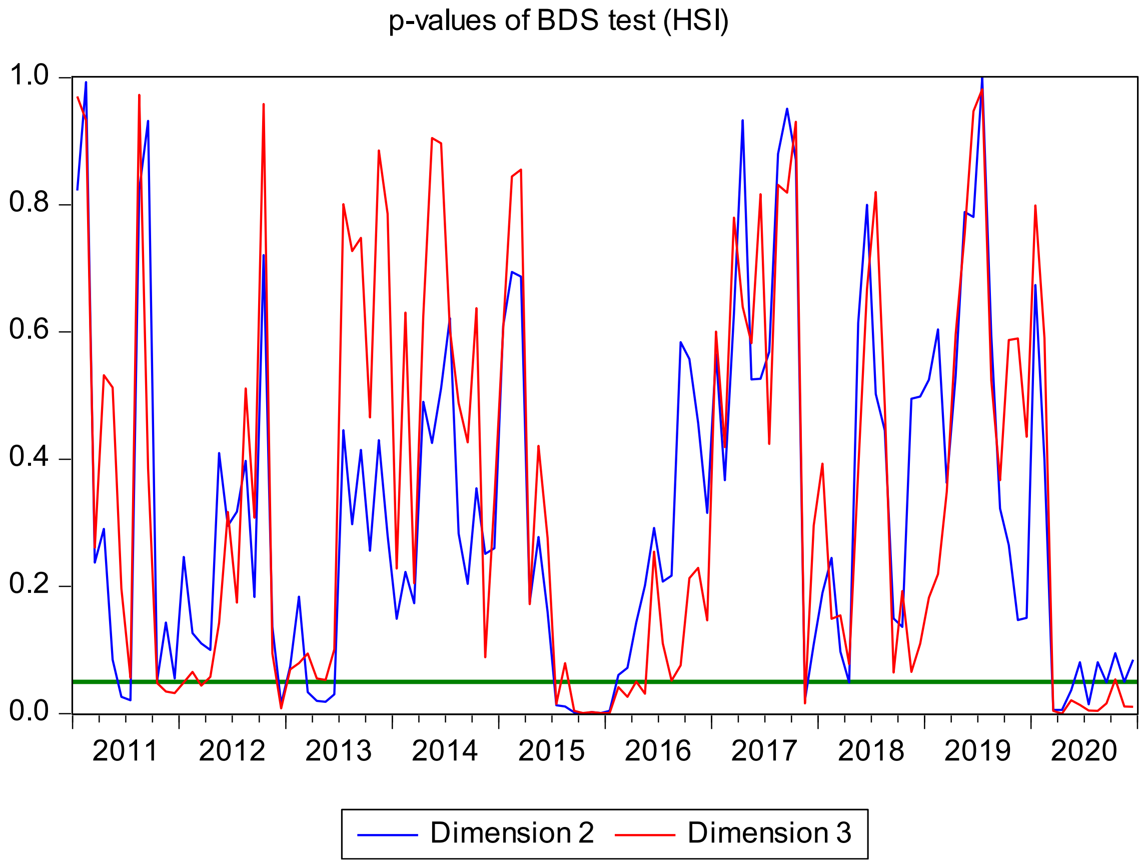
Figure 5.3 Bootstrapped p-values of the BDS independence test with nesting dimensions 2 and 3 for Nikkei 225 index returns



Source: own calculation in EViews

As for the Nikkei 225 index, the results are significantly different with the SSE 50 index. The BDS statistics for the nesting dimension 2 and 3 are not statistically significant in the period from the beginning to June 2010. Then there is a period from July 2010 to March 2011 when the values are statistically significant for the nesting dimension 2 and 3. Then there is a long period when the values of BDS statistics are insignificant. The BDS statistics are statistically significant from December 2014 to April 2016 for nesting dimension 3. In the period, for the nesting dimension 2, the p-values of February 2015, March 2015, April 2015, March 2016 and April are not statistically significant. However, for the nesting dimension 3 till the end, the BDS statistics are significant except the p-values of May 2016, June 2016, July 2016, December 2016, January 2017 and February 2017. For the nesting dimension 3, till the December 2020, the values of BDS statistics are insignificant in this period: May 2016 to December 2016, January 2017, December 2017, November 2017, December 2017 and January 2018 to November 2018.

Figure 5.4 Bootstrapped p-values of the BDS independence test with nesting dimensions 2 and 3 for HSI index returns



Source: own calculation in EViews

Figure 5.4 shows the trend of the p-values of the non-linear BDS independence test statistics for the return of the HSI index. The values of BDS statistics were calculated for nesting dimensions 2 and 3, which are described in the graph by the blue line and red line. The statistics significance of BDS tests will again be evaluated on the 5% significance level, which is indicated in the graphs by the green line. For the dimension 2, there are 22 months that the values are lower than 5% which below the significance level. The June and July of 2010 got not statistically significant p-values. Then it's a significant period till the end of 2011. The p-values of December 2011, March, April, May, June of 2012 are not statistically significant. Then the next insignificant period is from July 2015 to January 2016. November 2017 and April 2018 are also not statistically significant. Then the other statistically insignificant period is the year of 2020 which the months are March, April, June, August, December and November, respectively. Compare to the dimension 2, there are 25 insignificant months of dimension 3. The first period is from October 2010 to March 2011 except February 2011 which p-values are not statistically significant. In the next three years, only the December 2011 are not statistically significant. Then it is a significant period. Then from July 2015 to May 2016 except the August 2015 and April 2016 are statistically insignificant. Then the next is November 2017. It is same as the dimension 2, the last insignificant period is in 2020. It is from March 2020 to the end of 2020. HSI index get the last statistically insignificant period.

The results which we obtained show that for three indexes, there were identified periods when returns are statistically significantly predictable, but also some periods are not statistically significantly predictable. These changing levels of predictability depend on the tests used, as the tests vary in their strength. Therefore, a direct comparison can only be used if the procedure is consistent.

Table 5.6 Relative predictability based on a comparison of the percentage of non-rejection of the null hypothesis in the case of the BDS test against the 5% level of significance

Index	**Dimension 2**	**Dimension 3**
SSE 50	75.00%	65.00%
HSI	81.67%	79.17%
Nikkei 225	55.00%	34.17%

Source: own calculation in Excel

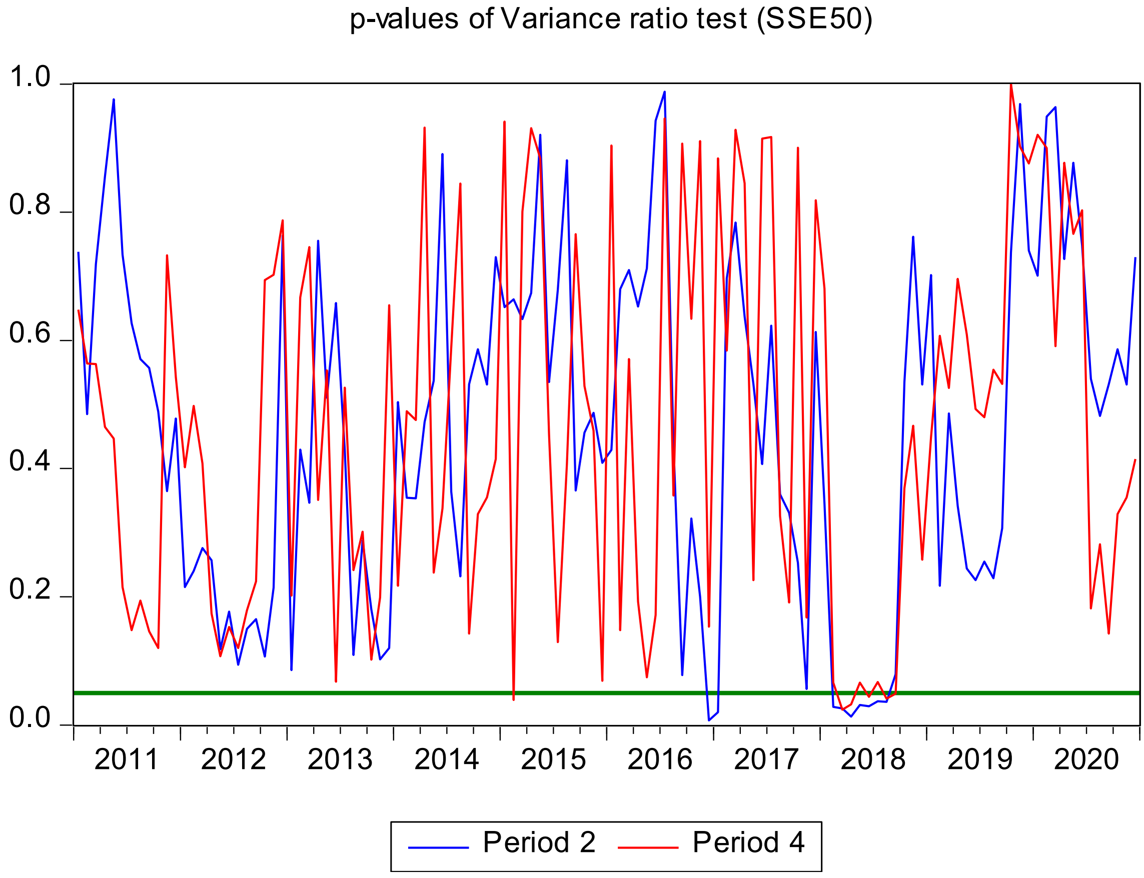
Table 5.6 shows the percentage of test statistics that fail to reject the null hypothesis for the BDS test with nesting dimensions 2 and 3. The results are confronted with a 5% level of significance. It shows that the SSE 50 index can be considered less predictable than Nikkei 225 index. However, it can be stated that for all indexes the periods when returns are predictable alternate with those when they are not predictable.

5.3.2 Variance ratio test

Lo and MacKinlay (1988) recommends the use of overlap because it may improve the ability of variance-ratio test, but overlap data can be used to analyse the exact distribution of variance-ratio test statistics. It also becomes difficult to do. However, little is really known about the exact distribution of variance-ratio test statistics using duplicate data, even at that moment.

The time window with a fixed length was selected as same as the BDS independence test. The first period of 12 months was defined from January 2010 to December 2010. This time window will be always moved one month forward until December 2020, when the testing period ends.

Figure 5.5 Bootstrapped p-values of the variance ratio test with period 2 and 4 for SSE 50 index prices

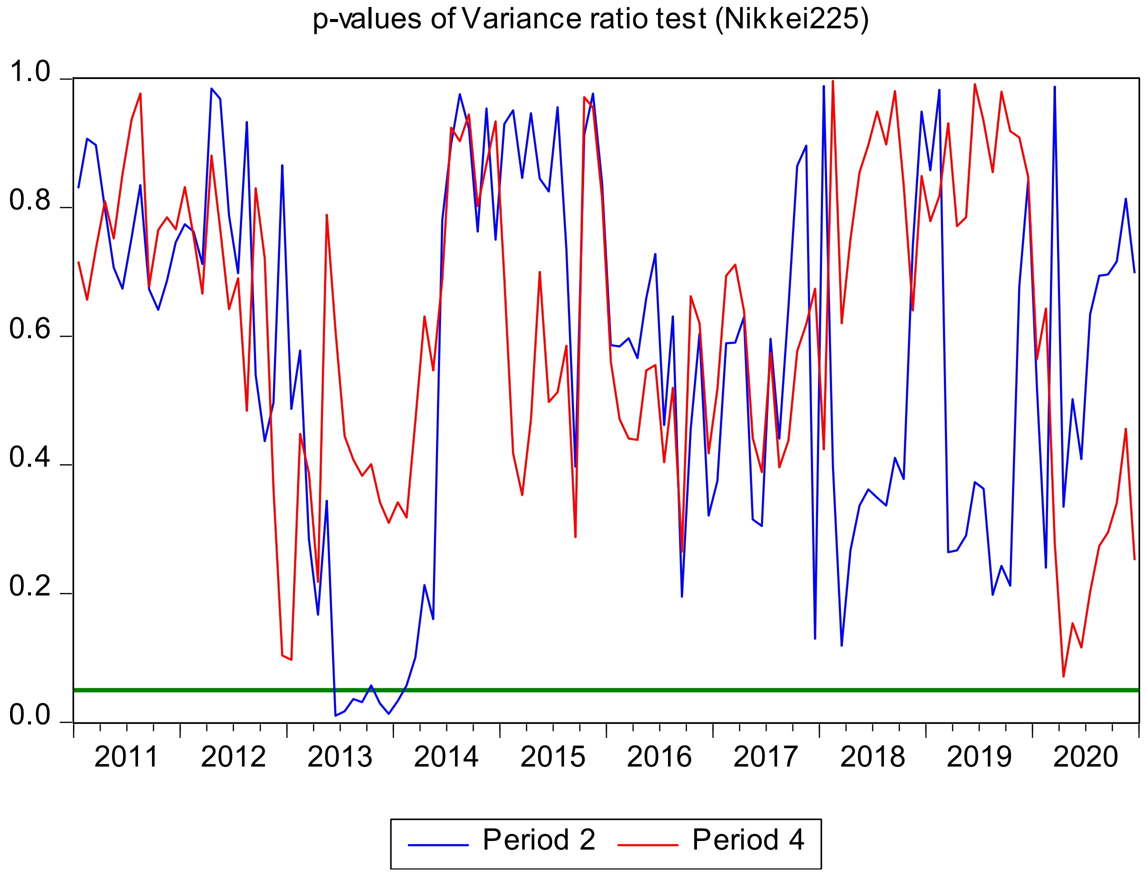


Source: own calculation in EViews

Figures 5.5, 5.6 and 5.7 show the evolution of the p-values of the variance ratio test that allows the heteroscedasticity for the prices of the SSE 50, HSI and Nikkei 225 indexes. The values of variance ratio test were calculated for period 2 and 4, which are described in the graphs by the different lines. Blue line is for period 2 and red line is for period 4. The statistical significance of variance ratio test will again be evaluated 5% significance level, which is indicated in the graphs by a green horizontal line.

The p-values were obtained based on a time series lasting 1 year. Thus, for example, the p-values as of February 2011 corresponds to the data for the period from February 2010 to January 2011. The p-values we got which were variable over time and differ between both indexes and periods.

Figure 5.6 Bootstrapped p-values of the variance ratio test with period 2 and 4 for Nikkei 225 index prices

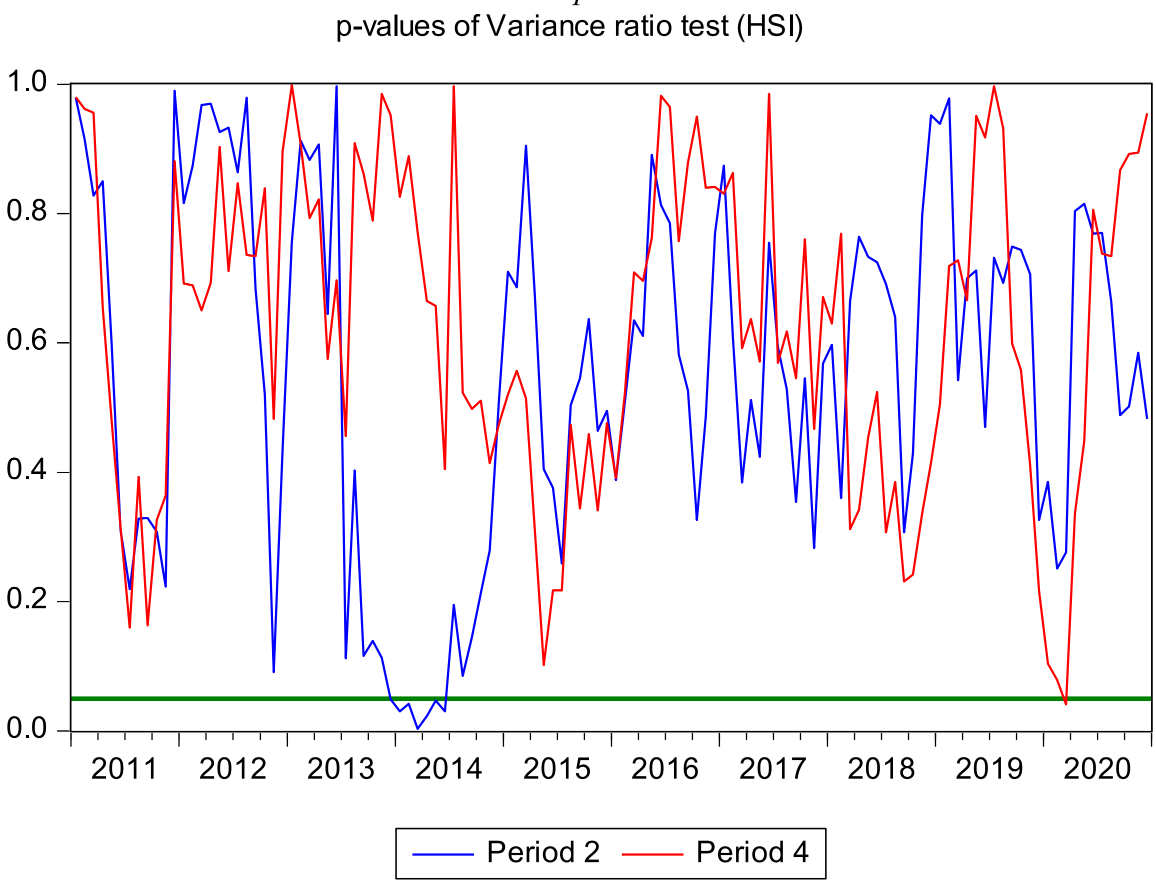


Source: own calculation in EViews

In the case of the SSE 50 index, the values of variance ratio statistics are statistically insignificant for a lone period which is from January 2010 to November 2016 for period 2 and 4. Then the values of two months for period 2 which are December 2016 and January 2017 are statistically significant. Then the next period is from February 2018 to August 2018 for period 2. It's statistically significant. For the period 4, there are six values which are February 2015, March 2018, April 2018, June 2018, August 2018 and September 2018. These are the statistically significant values for period 4. The other values for period 2 and 4 are insignificant.

As for the Nikkei 225 index, the results are easy to identify compare with the SSE 50 index. Firstly, for period 4, the all values of variance ratio statistics are statistically insignificant. The variance ratio test statistics for the period 2 are not significant in the period from January 2010 to April 2012. Then from May 2012 to December 2012 except September are statistically significant for the period 2. Then there is a long period when the values are insignificant till the testing period ends.

Figure 5.7 Bootstrapped p-values of the variance ratio test with period 2 and 4 for HSI index prices



Source: own calculation in EViews

For the HSI index, there are only few insignificant p-values. For the period 2, it is a period from December 2013 to June 2014. For the period 4, only one insignificant p-values which is March 2020.

The results we got shot that for each index, there were identified periods when returns are statistically significantly predictable except Nikkei 225 index for the period 4, but also some periods when returns are not statistically predictable. These changing of predictability depend the test vary in their strength. Therefore, it is wise to create an overview of the results of a clear table.

Table 5.7 Relative predictability based on a comparison of the percentage of non-rejection of the H_0 in the case of the variance ratio test against the 5% significance level

Index	Period 2	Period 4
SSE 50	92.50%	95.00%
HSI	94.17%	99.17%
Nikkei 225	94.17%	100.00%

Source: own calculation in Excel

Table 5.7 shows the percentage of test statistics that do not reject the null independence hypothesis for the variance ratio test with period 2 and 4. The results are confronted with a 5% level of significance. And the results of HSI index are similar with Nikkei 225 index. It shows that the Nikkei 225 index can be considered less predictable than the SSE 50 index. Overall, for the period 4 of Nikkei 225, it can be considered follow-up random walk model totally.

5.4 Proposal of a trading strategy based on nonlinear relationships in returns

The last logical step should be the design of a trading strategy, which will be based on the results of statistical tests applied in subchapters 5.2 and 5.3. Given the results, it was found that returns are rather predictable based on the results of the nonlinear BDS test. It is rational to design a trading strategy on the basis of nonlinear relationships in returns. The non-linear method of nearest neighbor (NN), seems to be a suitable method for generating signals for buying or selling. This method is based on the selection of geometric segments of historical values of the time series, which are similar to the last segment that precedes the observation that must be predicted (Fernández-Rodríguez et al., 1999). This approach is philosophically different from the Box-Jenkinson methodology. The NN method selects the relevant previous observations on the basis of their levels and geometric trajectories, not their location in time.

Prediction using the NN method can be described in several steps. The time series x_t, where $t = 1, \dots, n$, must be transformed into a series of segments of the same length, which take the form of vectors $x_t^{m,\tau}$ and contain a sample of m observations of the original time series on intervals $\tau \in N$:

$$x_t^{m,\tau} = \left(x_t, x_{t-\tau}, \dots, x_{t-(m-1).\tau}\right), \tag{5.1}$$

where $t = 1 + \tau\,(m-1), \dots, n$, m represents the nesting dimension, while τ is the delay parameter. These m-dimensional vectors are referred to in the literature as m-histories, while the m-dimensional space R^m is the phase space of the time series. The proximity of two m-histories in the phase space of the time series R^m can be marked as the nearest neighbours in the dynamic behaviour of two segments in the time series x_t. The prediction of the time series x_{n+1}^f using the NN method consists of the application of linear autoregression, the coefficients of which are estimated by the least-squares method. This is a total regression the *k* *m*-histories $x_{i_1}^m, x_{i_2}^m, x_{i_3}^m, \dots, x_{i_k}^m$.

It is important that the prediction by the NN method depends on the value of the nesting dimension *m* and also on the number of closest *k* points in the phase space of the time series R^m. For example, heuristic methods can be applied to determine these two parameters. However, the use of genetic algorithms that allow the simultaneous determination of optimal m and k values seems to be the most appropriate. Genetic algorithms represent a class of adaptive search and optimization techniques that have the advantage of being able to evaluate loss functions associated with predictor parameters without assuming continuity or the differentiability of the loss function (Holland, 1975). In addition, the application of genetic algorithms is also advantageous because it eliminates the problem of data snooping. The sample is divided into a training and test set.

However, before calculating predictors using the NN method, it is appropriate to verify the existence of nonlinear dependence in the time series of returns. Evidence of nonlinearity would confirm the correctness of our thinking about the application of the

NN method. The most suitable test seems to be the BDS independence test, which has similar logic as the NN method. It is based on the calculation of the correlation integral as a measure of the spatial correlation between two points in m-dimensional space (see equation 3.34).

The proposed trading strategy is based on a simple market timing, which consists of investing in the stock market (or index or asset) or risk-free asset. The forecasts obtained on the basis of the NN method are applied for the classification of the next trading day into two variants. The first of these means the presence of the investor on the market (brings market return), the second means the absence on the market and investment in a risk-free asset (brings risk-free return). The trading strategy specifies the position for the next trading day, considering the current state (presence or absence on the market) and generating signals for buying or selling using the NN method. If under market presence conditions, prices are expected to fall based on a forecast using the NN method, the asset is sold and the funds are invested in a risk-free asset. Conversely, under the absence in a market, price increases are expected in the near future based on the NN method, a signal to purchase the asset is generated. The risk-free asset is sold and the funds are invested in the market. With the remaining two variants, the current state is maintained.

When applying the trading rule for the entire trading period from time 1 to time T, the return r can be written as follows:

$$r = \sum_{t=1}^{T} r_m(t)\, I_b(t) + \sum_{t=1}^{T} r_f(t)\, I_s(t) + n \log \frac{1-c}{1+c}, \qquad (5.2)$$

where $r_m(t)$ is the market return, $I_b(t)$ and $I_s(t)$ are the indicator variables equal to one if the NN method generates a buy or sell signal, or zero in other cases. The condition $I_b(t) \cdot I_s(t) = 0, \forall t \in (1, T)$ must be met. The value of n is the number of transactions and the value of c means transaction costs, which can be expressed, for example, as a share of the price.

When using this strategy empirically, it is appropriate to modify a simple buy or sell rule with a filter that reduces the number of spurious buy or sell signals if the prediction of time t is close to the closing price at time t-1. The filter can be interpreted as a risk that the investor is willing to transfer. The filter rule will generate buy (sell) signals at time t if the expected value obtained by the NN method is greater (less) than the closing price at time t-1 adjusted by the percentage δ of the standard deviation σ of the time series of price differences in the interval from 1 to t -1. If $\widehat{P_t}$ is the predicted value of the price P_t, then if $\widehat{P_t} > P_{t-1} + \delta \cdot \sigma$ and at the same time the investor is not present in the market, a buy signal is generated. Assuming a market presence, we should not leave the market. Conversely, if $\widehat{P_t} \leq P_{t-1} + \delta \cdot \sigma$ holds and at the same time the investor is present in the market, a signal for sale is generated. Assuming absence from the market, we should continue to hold a risk-free asset.

In case of application of this strategy based on the NN method, it would be appropriate to compare the results with other strategies. For example, a prediction strategy based on the ARIMA model or a risk-adjusted buy-and-hold strategy is offered (Allen and Karjalainen, 1999).

5.5 Summary of results

In the previous paragraphs, the performance of the main indices of the Chinese, Hong Kong and Japanese markets was measured against the model of random walks of the first and third types. If we did not identify statistically significant linear or non-linear dependencies when applying statistical tests, the price movement would be random and the given market could be efficient in its weak form. It would be a kind of test of absolute efficiency.

Table 5.8 is derived from results of statistical test obtained by own calculation. In this table, letter R means the null hypothesis is rejected, and letter N means the null hypothesis in not rejected.

It is appropriate to follow up on the above partial conclusions of individual tests by verifying the relative efficiency of both indexes. Table 5.8 provides a summary of the results of applied tests at 5% significance level. The results shown in Table 5.8 show that the selected indices of the Chinese, Hong Kong and Japanese markets behave partially heterogeneously in terms of the characteristics of the random walk models.

The test of sequences and reversals examines just static approaches in the data sample. According to the results, there are no statistically significant so that these three markets are random during ten years.

Then the autocorrelation test delivered different results. We found that there are statistically significant in SSE 50 index after lag 5. But the results for Nikkei and HSI are always insignificant which means the they are not predictable.

The BDS test gets very miraculous values. The values are all zero which means they are all significant. And for the three indexes, they are predictable. As shown in Table 5.8, we can consider the most of markets are efficient when applying static approach. However, the main difference between static methods and dynamic approaches is whether to consider the time factor. Dynamic approaches describe the changing process of a state over a period of time, comparing before and after, while static methods are an analysis of an equilibrium state. That's why we chosen 12 months as a time window and always slid one month forward.

Overall, it can be concluded that there are deviations from the random walk model and the performance of some indices in several periods is therefore predictable. However, it must be emphasized that according to the EMH the market is not efficient in its weak form means if exists business strategy, which would bring the investor above-average return in the long run. Thus, also efficient market may not exclude a certain degree of predictability, for example, autocorrelation in time series, unless this degree is economically significant. The difference between economic and statistical significance may be due to non-compliance with the strict conditions of the model. In the practical analysis, it is mainly the existence of transaction costs and the need to take them into account when interpreting the results.

As shown in Table 5.8, the HSI index and Nikkei 225 index has the similar predictability. Compare with these two indexes, the SSE 50 index is more predictable. Then we can learn that there is strong non-linear dependency in all indexes according to the BDS test, because the all BDS tests reject null hypothesis. As for the variance ratio tests. The null hypothesis was rejected for all indexes and both types of the variance ratio test. However, these tests can investigate just linear dependency. To sum up, although linear tests did not reveal a linear dependence in the data, with the exception of the autocorrelation test for the SSE 50 index, a non-linear dependence was demonstrated for all indexes using a strong non-linear BDS test.

Table 5.8 Comparison of results

Static approach					
Index	Test of sequences and reversals	Autocorrelation test	BDS test	Variance ratio test (RW1)	Variance ratio test (RW3)
SSE 50	N	R	R	N	N
HSI	N	N	R	N	N
Nikkei 225	N	N	R	N	N

Source: own calculation

If we want to evaluate the relative predictability of returns statistically, it would be possible to compare the results for individual stocks, which are included in the indexes SSE50, HSI and Nikkei225, and compare the relative frequencies of results using, for example, one-sided Fisher factorial test of relative frequency agreement.

In terms of dynamic results, the most suitable tests for the random walk model I and III were applied. It is a BDS test and a variance ratio test that allows heteroskedasticity in the data. The results were presented in the form of overview tables. Based on the nonlinear BDS test, it can be stated that the SSE 50 and HIS indexes can be considered less predictable than Nikkei 225 index. As for the linear variance ratio test, the Nikkei 225 index can be considered least predictable of all indexes.

6 Conclusion

The efficient market hypothesis is one of the possible explanations for the development of asset prices in the stock markets. According to this hypothesis, prices develop completely randomly. However, critical opinions about the efficient market hypothesis gradually began to emerge, both from academics and stock market traders. In the last two decades, empirical evidence that is in direct conflict with the efficient market hypothesis has been identified. Thus, it may seem that stock markets are not really information efficient or their effectiveness is not as high. The issue of information efficiency is also relevant in stock markets in Asia, including China.

This thesis consisted of six chapters including the introduction as the first chapter. In chapter 2, we learned the fundamental knowledge about the EMH in detail. The third chapter included the models of efficient markets and related statistical tests. Then, in the fourth chapter, we delivered the description of the Chinese and Japanese stock markets. We selected the most representative index from Japan as the benchmark to compare its predictability with the two indexes from China. Chapter five, the most important part of this thesis, contains the results of statistical tests of random walk models from a static and dynamic point of view, trading strategy proposal and gave the final summary of achieved results.

The main goal of the thesis was an empirical verification of the properties of efficient market models on the returns of stock market indices in selected economies in Asia. The object of research of this thesis were the main stock market indices of mainland China, Hongkong and Japan, such as the SSE50, HSI and NIKKEI225 indexes. Selected linear and nonlinear testing methods were applied using daily data in the period from January 2010 to December 2020. The testing algorithm was based on a combination of random walk models and statistical tests. The null hypothesis was not efficiency but market unpredictability. The results of empirical tests are given in chapter 5. Overall, it can be stated that the results of testing vary from market to market. It does not make much sense to compare the results presented in this thesis with the results of other studies presented in chapter 1. The reason is the difference in the data sample and also the inconsistency in applied methods. Very important is also the fact that authors do not distinguish between different types of random walk models according to their statistical properties.

The static view of the results was based on a comparison of the results of linear and nonlinear statistical tests of random walk models of the first and third types. The results of statistical tests were compared between individual indices. The most frequently statistically significant deviations from the random walk model were observed for the SSE50 index. As for the HSI and NIKKEI225 indexes, the returns are conditionally predictable based on the nonlinear BDS test only. The null hypothesis was not rejected by any linear test. Taking the overall view on the test results, it can be concluded that the null hypothesis was always rejected by the strongest nonlinear BDS independence test, which was applied to verify the RW1. However, the assumption of homoscedasticity is far from the real behaviour of returns in stock markets.

The main goal of the thesis was also supported by three partial goals. The first partial goal was to compare the predictability of returns of indexes in mainland China (SSE50) and Hongkong (HIS) because both stock markets differ by trading rules and regulatory standards in general. It can be concluded that conditional predictability was higher for the SSE 50 index than for the HSI index, which showed behaviour similar to the Japanese NIKKEI225 index. For the SSE index, even a linear dependence was identified during the basic testing period of 2010-2020 years.

The second sub-goal was to assess the dynamics of the development of predictability of returns. Unlike papers published in journals, which were focused on Asian stock markets, this thesis also monitors the dynamics of the development of results since we have a relatively long time series. It is therefore appropriate to also verify the stability of the test results over time. The dynamic view is much more convincing. For this purpose, the most suitable tests for both RW models were selected. In the case of the RW1 model, a nonlinear BDS independence test was chosen, while the properties of the RW3 model were then verified using a heteroskedastic version of the variance ratio test.

The BDS independence test was applied for nesting dimensions 2 and 3. As for the variance ratio test, the lags by 2 and 4 periods were chosen.

A period of a fixed length of 12 months from January 2010 to December 2010 was defined. This time frame was then slid forward by one month. The whole process was repeated until December 2020. In the end, we got a total of 120 results for each test. The test window contained a sufficient number of observations to generate reliable results to analyse how the level of predictability of the indices evolved over time. However, to get stable estimates, the respective *p*-values for both tests were obtained by the method of bootstrap.

The test results show that for each index, periods were identified periods when returns were predictable, but also periods when returns were not predictable. Of course, the variable level of predictability depends on the tests applied. This is consistent with the adaptive market hypothesis, not the efficient market hypothesis. The returns of the SSE50 index are relatively more predictable than returns of the HSI index. In the case of the NIKKEI225 which could be considered a benchmark of the developed stock market in Asia, the results are relatively surprising. Based on the BDS test, the null hypothesis was rejected more often than for both Chinese indexes.

The last, third sub-goal was to discuss the design of an investment strategy, which is based on the results of empirical tests. The aim was not to test this strategy on real data and provide statistical evidence of the efficiency or inefficiency of the market, but only to describe the principle of this strategy. The relationship between the results of statistical tests and a possible investment strategy was also briefly discussed in relation to all applied tests. In addition, subchapter 5.4.4 describes the investment strategy, which is directly linked to the results of statistical tests. Due to the detection of predictability of returns based on the BDS test, a nonlinear NN method was proposed to generate signals for purchase or sale. This method is based on a similar principle as a nonlinear BDS independence test.

The presented thesis also indicated the direction that further research in the field of information efficiency would focus on. When using classical statistical methods, we encounter barriers to statistical paradigms, which very often prevent or limit their use. Moreover, in the case of application in the field of financial time series, it is not possible to suffice with the assumption of normality and linear models. It is definitely necessary to enter the space of nonlinear models and tests. In addition to the analytical apparatus of mathematical statistics, there are methods of so-called artificial intelligence. The issue of the EMH implied the development of analytical tools towards the development of methods based on data entropy or uncertainty.

In this thesis, time series of daily frequency were used in order to ensure a sufficiently long series of observations. However, we work just with closing values. That is why, it is impossible to gain an understanding of the dynamics of processes, such as the formation of market trends. Thus, dynamics can only be detected when switching to data of a higher frequency. The hypothesis that the dynamics hidden in the analysis of daily data can be revealed in the analysis of intraday data has its own logic. Investors are not homogeneous and therefore operate in different time horizons.

In this text, the main indices of selected Asian stock markets were at the core of the empirical analysis. More detailed results could be obtained if a similar analysis were performed on individual shares. Better results could also be achieved if returns are adjusted for the effect of dividends.

Bibliography

Professional book

BOX, G. and G. M. JENKINS. *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day, 1970. 712 p. ISBN 978-1-118-67502-1.

CAMPBELL, J. Y., A. W. LO and A. C. MACKINLAY. *The Econometrics of Financial Markets*. 1st ed. New York: Princeton University Press, 1997. 632 p. ISBN 0-691-04301-9.

CHATFIELD, CH. *The Analysis of Time Series: An Introduction*. New York: Chapman and Hall, 2003. 352 p. ISBN 978-1584883173.

COOTNER, P. H. *The Random Character of Stock Market Prices*. Cambridge: MIT Press, 1964. 510 p. ISBN 978-0262530040.

FRANKE, J. and CH. M. HAFNER. *Statistics of Financial Markets: An Introduction*. 3rd ed. Berlin: Springer, 2011. 599 p. ISBN 978-3-642-16520-7.

HOLLAND, J. H. *Adaptation in Natural and Artificial Systems*. Ann Arbor: University of Michigan Press, 1975. 232 p. ISBN 978-0262581110.

MURPHY, J. *Technical Analysis of the Futures Market*. Cambridge: Prentice Hall Press, 1986. 576 p. ISBN 978-0735200661.

OSTROM, CH. W. *Time Series Analysis: Regression Techniques (Quantitative Applications in the Social Sciences)*. Thousand Oaks: SAGE Publications, 1990. 96 p. ISBN 978-0803931350.

TSAY, R.S. *Analysis of Financial Time Series*. 3rd ed. New Jersey: John Wiley & Sons, Inc., 2010. 720 p. ISBN 978-0-470-41435-4.

An article in a journal or proceedings

ALEXANDER, S. Price Movements in Speculative Markets: Trends or Random Walks, *Industrial Management Review*. 1964, 5(2), 25–46. ISSN 0884-8211.

ALLEN, F. and R. KARJALAINEN. Using genetic algorithms to find technical trading rules. *Journal of Financial Economics*. 1999, 51(2), 245–271. ISSN 0304-405X.

ANDERSON, R. L. Distribution of the serial correlation coefficient. *The Annals of Mathematical Statistics*. 1942, 13(1), 1–13. ISSN 0003-4851.

BACHELIER, L. Theorie de la Speculation. *Annales Scientifiques de l'École Normale Supérieure*. 1900, 3(17), 21–86. ISSN 0012-9593.

BOX, G. and D. PIERCE. Distribution for residua autocorrelations in autoregressive average time series models. *Journal of American Statistical Association*. 1970, 65(332), 1509–1526. ISSN 1804-2112.

BROCK, W., D. DECHERT, J. SCHEINKMAN and B. LEBARON. A test for independence based on the correlation dimension. *Econometric Reviews*. 1996, 15(3), 197–235. ISSN 0747-4938.

CHAN, K. C. On the contrarian investment strategy. *Journal of Business*. 1988, 61(2), 147-164. ISSN 1075-6124.

CONT, R. Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*. 2001, 1(2), 223–236. ISSN 1469-7688.

COWLES, A. Can stock market forecasters forecast? *Econometrica*. 1933, 1(3), 309–324. ISSN 1468-0262.

COWLES, A. and H. E. JONES. Some a posteriori probabilities in stock market action. *Econometrica.* 1937, 5(3), 280–294. ISSN 1468-0262.

DE BONDT, W. F. M. and R. H. THALER. Does the stock market overreact? *Journal of Finance*.1985, 40(3), 793-808. ISSN 1540-6261.

FAMA, E. Efficient capital markets: A review of theory and empirical work. *Journal of Finance*. 1970, 2(25), 383–417. ISSN 1540-6261.

FAMA, E. Efficient capital markets II. *Journal of Finance*. 1991, 45(6), 1575–1610. ISSN 1540-6261.

FAMA, E. The behavior of stock-market prices. *Journal of Business*. 1965, 38(1), 34–105. ISSN 0021-9398.

FAMA, E. and M. E. BLUME. Filter rules and stock market trading. *The Journal of Business.* 1966, 39(1), 226-241. ISSN 0021-9398.

FERNANDES, N. and M. FERREIRA. Does international cross-listing improve the information environment. *Journal of Financial Economics*. 2008, 88(2), 216-244. ISSN 0304-405X.

FERNANDEZ-RODRIGUEZ, F., S. SOSVILLA-RIVERO and J. ANDRADA-FELIX. Exchange-rate forecasts with simultaneous nearest-neighbour methods: Evidence from the EMS. *International Journal of Forecasting*. 1999, 15(4), 383–392. ISSN 0169-2070.

FOUCAULT, T. and T. GEHRIG. Stock Price Informativeness, Cross-Listings and Investment Decisions. *Journal of Financial Economics*. 2008, 88(1), 146-168. ISSN 0304-405X.

KENDALL, M. The analysis of economic time series, Part I: Prices. *Journal of Royal Statistical Society*. 1953, 96, 85–99.

KIM, J. Wild bootstrapping variance ratio tests. *Economics Letters*. 2006, 92(1), 38–43. ISSN 0165-1765.

LIU, W. and Y. WANG. Test on efficiency of Shanghai stock market. *Journal of Management Science & Statistical Decision*. 2007.

LJUNG, J. M. and G. BOX. On a measure of lack of fit in time series models. *Biometrika*. 1978, 65(2), 297–303. ISSN 0006-3444.

LO, A. W. and A. C. MacKINLAY. Stock market prices do not follow random walks: Evidence from a simple specification test. *Review of Financial Studies*. 1988, 1(1), 41–66. ISSN 1351-847X.

MacKINLAY, C. Event Studies in Economics and Finance. *Journal of Economic Literature*, 1997, 35(1), 13-39. ISSN 0022-0515.

MALAFEYEV, O., A. AWASTHI, K. S. KAMBEKAR, and A. KUPINSKAYA. Random Walks and Market Efficiency in Chinese and Indian Equity Markets. *Statistics, Optimization & Information Computing*. 2019, 7(1), p. 1-25. ISSN 2311-004X.

OSBORNE, M. F. Brownian motion in the stock market. *Operations Research*. 1959, 7(2), 145–173. ISSN 0030-364X.

RICHARDSON, M. and J. STOCK. Drawing inferences from statistics based on multiyear asset returns. *Journal of Financial Economics*, 1989, 25(2), 323–348. ISSN 0304-405X.

ROBERTS, H. *Statistical versus Clinical Prediction of the Stock Market*. Chicago: University of Chicago. 1967. Unpublished manuscript.

ROBERTS, H. Stock market patterns and financial analysis: methodological suggestion. *Journal of Finance*. 1959, 14(1), 7–16. ISSN 1540-6261.

SAMUELSON, P. Proof that properly anticipated prices fluctuate randomly. *Industrial Management Review*. 1965, 6(2), 41–49. ISSN 2336-8225.

SHEN Y. Accounting information disclosure and semi-strong form efficiency of Chinese stock market. *Accounting Research*. 1996, 1, 14-17.

SONG S. and W. JIN. The study on the effectiveness of Shanghai stock market. *Economist*. 1995, 4, 107-128.

SMITH, G. The changing and relative efficiency of European emerging stock markets. *The European Journal of Finance*. 2012, 18(8), 689–708. ISSN 1351847X.

TREYNOR, J. and R. FERGUSON. In defence of technical analysis. *Journal of Finance.* 1985, 40(3), 757–773. ISSN 1540-6261.

URQUHART, A. and F. MCGROARTY. Calendar effects, market conditions and the adaptive market hypothesis: Evidence from long-run U.S. data. *International Review of Financial Analysis*. 2014, 35(C), 154–166. ISSN 1057-5219.

YU Q. Effective market, periodic abnormal and the fluctuation of stock price. *Economic Research Journal*. 1994, 4, 43-50.

List of Abbreviations

ACF – Autocorrelation Function

ADF – Augmented Dickey-Fuller

ARMA – Autoregressive Moving Average

BDS – Brock, Dechert, Scheinkman

CJ – Cowles-Jones

EMH – Efficient Market Hypothesis

HIS – Hang Seng Index

IID – Independently and Identically Distributed

JB – Jarque-Bera

NBER – National Bureau of Economic Research

NN – Nearest Neighbor

RW1 – Random Walk Model I

RW2 – Random Walk Model II

RW3 – Random Walk Model III

SSE – Shanghai Stock Exchange

VR – Variance Ratio

Declaration of Utilisation of Results from the Diploma Thesis

Herewith I declare that

- I was acquainted with the fact that my diploma thesis is fully covered by Act No. 121/2000 Coll. - Copyright Act, in particular § 35 - use of the work in the framework of civil and religious ceremonies, in school performances and use of school work and § 60 - school work;
- I acknowledge that by submitting my diploma (bachelor) thesis, I agree to the publication of my work under Act No. 111/1998 Coll. on universities and on amendments to other acts (the Higher Education Act), as amended, regardless of the outcome of the defence;
- I acknowledge that the VSB – Technical University of Ostrava (hereinafter VSB-TUO) has a non-profit right to use the diploma (bachelor) thesis for its own internal needs (Section 35 (3));
- I agree that the thesis (bachelor's thesis) will be archived in electronic form in the Central Library of VSB-TUO. I agree that bibliographic data on the thesis will be published in the VSB-TUO information system;
- it was agreed that I would conclude a license agreement with VSB-TUO, if it was interested in it, with the right to use the work within the scope of Section 12 (4) of the Copyright Act;
- it was agreed that I can use my work, diploma thesis (bachelor's thesis), or grant a license for its use only with the consent of VSB-TUO, which is entitled to request a reasonable contribution from me because of costs incurred by VSB-TUO for the creation of the work (up to their actual amount).

Ostrava dated

....

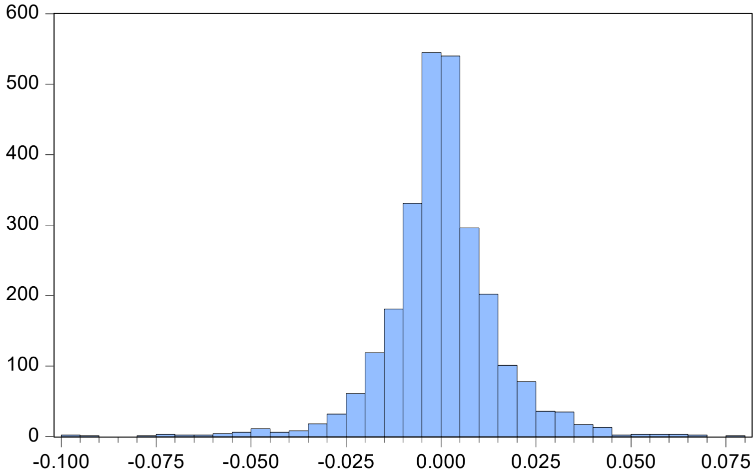
....

Student’s name and surname

List of Annexes

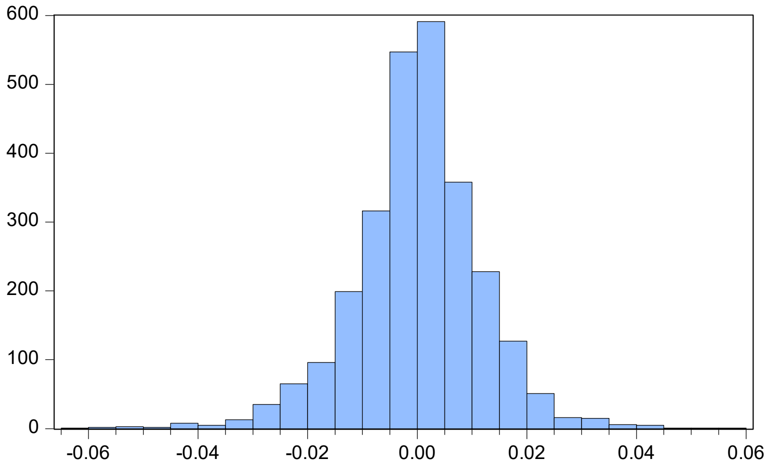
Annex 1

SSE 50



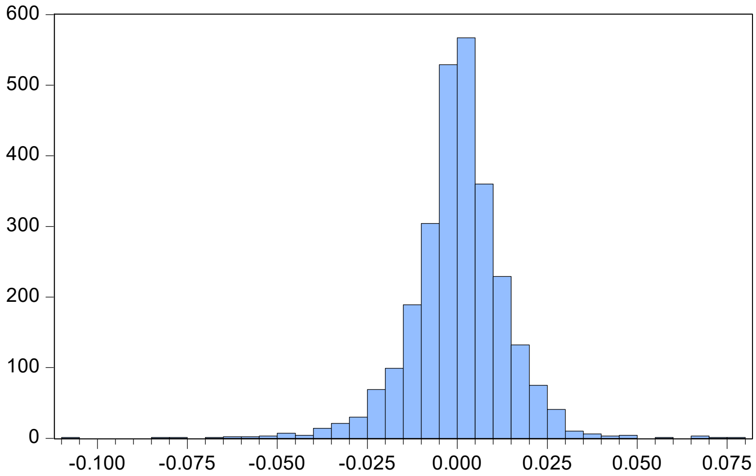
Series: RETURN	
Sample 1/04/2010 12/18/2020	
Observations 2665	
Mean	0.000125
Median	-1.69e-05
Maximum	0.075470
Minimum	-0.098523
Std. Dev.	0.014790
Skewness	-0.405210
Kurtosis	8.280691
Jarque-Bera	3169.404
Probability	0.000000

HSI



Series: RETURN	
Sample 1/04/2010 12/30/2020	
Observations 2692	
Mean	6.01e-05
Median	0.000535
Maximum	0.055187
Minimum	-0.060183
Std. Dev.	0.011727
Skewness	-0.336200
Kurtosis	5.427889
Jarque-Bera	711.8953
Probability	0.000000

Nikkei 225



Series: RETURN	
Sample 1/04/2010 12/18/2020	
Observations 2710	
Mean	0.000340
Median	0.000608
Maximum	0.077314
Minimum	-0.105758
Std. Dev.	0.013290
Skewness	-0.436341
Kurtosis	8.268587
Jarque-Bera	3220.336
Probability	0.000000

Annex 2

SSE 50

Date: 03/01/21 Time: 20:42
Sample: 1/04/2010 12/18/2020
Included observations: 2665

Autocorrel	Partial Correlation		AC	PAC	Q-Stat	Prob
| |	| |	1	0.013	0.013	0.4573	0.499
| |	| |	2	-0.025	-0.025	2.0923	0.351
| |	| |	3	0.023	0.024	3.5208	0.318
| |	| |	4	0.024	0.023	5.0838	0.279
| |	| |	5	-0.01	-0.01	5.375	0.372
*| |	*| |	6	-0.077	-0.076	21.07	0.002
| |	| |	7	0.038	0.039	24.923	0.001
| |	| |	8	0.025	0.021	26.652	0.001
| |	| |	9	0.039	0.045	30.807	0
| |	| |	10	0.006	0.007	30.892	0.001

HSI

Date: 03/01/21 Time: 20:27
Sample: 1/04/2010 12/30/2020
Included observations: 2692

Autocorrel	Partial Correlation		AC	PAC	Q-Stat	Prob
| |	| |	1	0.006	0.006	0.1034	0.748
| |	| |	2	0.005	0.005	0.1619	0.922
| |	| |	3	0.021	0.021	1.4036	0.705
| |	| |	4	-0.031	-0.032	4.0556	0.399
| |	| |	5	0.013	0.013	4.4784	0.483
| |	| |	6	-0.04	-0.04	8.7948	0.185
| |	| |	7	0.03	0.032	11.192	0.13
| |	| |	8	-0.015	-0.017	11.812	0.16
| |	| |	9	0.022	0.024	13.085	0.159
| |	| |	10	-0.021	-0.026	14.298	0.16

Nikkei 225

Date: 03/01/21 Time: 20:44
Sample: 1/04/2010 12/18/2020
Included observations: 2710

Autocorrel	Partial Correlation		AC	PAC	Q-Stat	Prob
| |	| |	1	-0.023	-0.023	1.377	0.241
| |	| |	2	0.032	0.032	4.2349	0.12
| |	| |	3	-0.009	-0.007	4.4368	0.218
| |	| |	4	-0.033	-0.035	7.4298	0.115
| |	| |	5	-0.003	-0.004	7.4478	0.189
| |	| |	6	-0.011	-0.009	7.804	0.253
| |	| |	7	-0.02	-0.021	8.8619	0.263
| |	| |	8	-0.007	-0.009	9.0111	0.341
| |	| |	9	0.048	0.049	15.365	0.081
| |	| |	10	0.006	0.007	15.449	0.117

Annex 3

SSE 50

BDS Test for RETURN					
Date: 03/01/21 Time: 20:43					
Sample: 1/04/2010 12/18/2020					
Included observations: 2666					
Dimension	BDS Statist	Std. Error	z-Statistic	Normal Pr	Bootstrap Prob.
2	0.010989	0.001902	5.777329	0	0
3	0.023409	0.003018	7.755776	0	0
4	0.031489	0.003589	8.773097	0	0
5	0.038134	0.003736	10.20622	0	0
6	0.041711	0.003599	11.58947	0	0
Raw epsilon		0.01843			
Pairs within epsilon		4989477	V-Statistic	0.702523	
Triples within epsilon		1.03E+10	V-Statistic	0.542628	
Dimension	C(m,n)	c(m,n)	C(1,n-(m-1	c(1,n-(m-1	c(1,n-(m-1))^k
2	1788791	0.504294	2491341	0.702357	0.493305
3	1311413	0.36999	2489723	0.702428	0.346582
4	973299	0.274804	2487512	0.702332	0.243315
5	739284	0.208889	2485254	0.702222	0.170754
6	571488	0.161598	2483317	0.702202	0.119887

HSI

BDS Test for RETURN					
Date: 03/01/21 Time: 20:32					
Sample: 1/04/2010 12/30/2020					
Included observations: 2715					
Dimension	BDS Statist	Std. Error	z-Statistic	Prob.	
2	0.005089	0.001617	3.146588	0.0017	
3	0.012155	0.002569	4.731194	0	
4	0.01941	0.003058	6.348119	0	
5	0.024935	0.003185	7.828397	0	
6	0.028148	0.00307	9.168712	0	
Raw epsilon		0.015942			
Pairs within epsilon		5103850	V-Statistic	0.704284	
Triples within epsilon		1.05E+10	V-Statistic	0.537968	
Dimension	C(m,n)	c(m,n)	C(1,n-(m-1	c(1,n-(m-1	c(1,n-(m-1))^k
2	1814769	0.501401	2549842	0.704494	0.496312
3	1308381	0.36176	2547842	0.704465	0.349605
4	959721	0.265555	2545586	0.704365	0.246145
5	715678	0.198176	2543297	0.704255	0.173241
6	541667	0.150103	2541228	0.704206	0.121955

Nikkei 225

BDS Test for RETURN					
Date: 03/01/21 Time: 20:44					
Sample: 1/04/2010 12/18/2020					
Included observations: 2711					
Dimension	BDS Statist	Std. Error	z-Statistic	Normal Pr	Bootstrap Prob.
2	0.014234	0.001691	8.416917	0	0
3	0.034411	0.002679	12.8443	0	0
4	0.0491	0.00318	15.43849	0	0
5	0.056735	0.003305	17.16814	0	0
6	0.059315	0.003177	18.66911	0	0
Raw epsilon		0.017227			
Pairs within epsilon		5156326	V-Statistic	0.702105	
Triples within epsilon		1.07E+10	V-Statistic	0.536961	
Dimension	C(m,n)	c(m,n)	C(1,n-(m-1	c(1,n-(m-1	c(1,n-(m-1))^k
2	1859257	0.506888	2574533	0.701893	0.492654
3	1392953	0.38004	2572233	0.701784	0.345629
4	1067676	0.29151	2569943	0.701677	0.24241
5	829708	0.226705	2567671	0.701575	0.16997
6	652637	0.178455	2565390	0.701471	0.11914

Annex 4

SSE 50

Included observations: 2665 (after adjustments)				
Standard error estimates assume no heteroskedasticity				
Use biased variance estimates				
User-specified lags: 2 4 8 16				
Joint Tests		Value	df	Probability
Max |z| (at period 2)*		1.044868	2665	0.7545
Wald (Chi-Square)		3.36737	4	0.4983
Individual Tests				
Period	Var. Ratio	Std. Error	z-Statistic	Probability
2	1.02024	0.019371	1.044868	0.2961
4	0.999334	0.03624	-0.01837	0.9853
8	0.98849	0.0573	-0.20088	0.8408
16	1.000993	0.085265	0.011646	0.9907
*Probability approximation using studentized maximum modulus with parameter value 4 and infinite degrees of freedom				
Test Details (Mean = 0.372727954972)				
Period	Variance	Var. Ratio	Obs.	
1	1223.34	--	2665	
2	1248.1	1.02024	2664	
4	1222.52	0.99933	2662	
8	1209.26	0.98849	2658	
16	1224.55	1.00099	2650	

HSI

Included observations: 2692 (after adjustments)				
Standard error estimates assume no heteroskedasticity				
Use biased variance estimates				
User-specified lags: 2 4 8 16				
Joint Tests		Value	df	Probability
Max |z| (at period 4)*		1.114101	2692	0.7085
Wald (Chi-Square)		3.049303	4	0.5496
Individual Tests				
Period	Var. Ratio	Std. Error	z-Statistic	Probability
2	1.011526	0.019274	0.598025	0.5498
4	1.040172	0.036058	1.114101	0.2652
8	1.033088	0.057012	0.580368	0.5617
16	0.99025	0.084837	-0.11493	0.9085
*Probability approximation using studentized maximum modulus with parameter value 4 and infinite degrees of freedom				
Test Details (Mean = 1.50505621434)				
Period	Variance	Var. Ratio	Obs.	
1	75105.7	--	2692	
2	75971.4	1.01153	2691	
4	78122.8	1.04017	2691	
8	77590.8	1.03309	2685	
16	74373.4	0.99025	2677	

Nikkei 225

Included observations: 2710 (after adjustments)				
Standard error estimates assume no heteroskedasticity				
Use biased variance estimates				
User-specified lags: 2 4 8 16				
Joint Tests		Value	df	Probability
Max |z| (at period 2)*		0.950622	2710	0.8123
Wald (Chi-Square)		6.22055	4	0.1833
Individual Tests				
Period	Var. Ratio	Std. Error	z-Statistic	Probability
2	0.981739	0.019209	-0.95062	0.3418
4	1.011537	0.035938	0.321024	0.7482
8	0.995614	0.056822	-0.07719	0.9385
16	0.987032	0.084554	-0.15336	0.8781
*Probability approximation using studentized maximum modulus with parameter value 4 and infinite degrees of freedom				
Test Details (Mean = 5.94413284133)				
Period	Variance	Var. Ratio	Obs.	
1	46420.7	--	2710	
2	45573	0.98174	2709	
4	46956.2	1.01154	2707	
8	46217	0.99561	2703	
16	45818.7	0.98703	2695	

Annex 5

SSE 50

Sample: 1/04/2010 12/18/2020
Included observations: 2665 (after adjustments)
Heteroskedasticity robust standard error estimates
User-specified lags: 2 4 8 16

Joint Tests	Value	df	Probability
Max |z| (at period 2)*	0.659566	2665	0.9421

Individual Tests

Period	Var. Ratio	Std. Error	z-Statistic	Probability
2	1.021006	0.031849	0.659566	0.5095
4	1.001589	0.061727	0.025735	0.9795
8	0.993704	0.097996	-0.06425	0.9488
16	1.012359	0.142966	0.086449	0.9311

*Probability approximation using studentized maximum modulus with parameter value 4 and infinite degrees of freedom

Test Details (Mean = 0.372727954972)

Period	Variance	Var. Ratio	Obs.
1	1223.8	--	2665
2	1249.5	1.02101	2664
4	1225.74	1.00159	2662
8	1216.09	0.9937	2658
16	1238.92	1.01236	2650

HSI

Sample: 1/04/2010 12/30/2020
Included observations: 2692 (after adjustments)
Heteroskedasticity robust standard error estimates
User-specified lags: 2 4 8 16

Joint Tests	Value	df	Probability
Max |z| (at period 4)*	0.935937	2692	0.8207

Individual Tests

Period	Var. Ratio	Std. Error	z-Statistic	Probability
2	1.012278	0.023401	0.524679	0.5998
4	1.04172	0.044575	0.935937	0.3493
8	1.038483	0.0705	0.545851	0.5852
16	1.001381	0.103565	0.013331	0.9894

*Probability approximation using studentized maximum modulus with parameter value 4 and infinite degrees of freedom

Test Details (Mean = 1.50505621434)

Period	Variance	Var. Ratio	Obs.
1	75133.6	--	2692
2	76056.1	1.01228	2691
4	78268.2	1.04172	2691
8	78025	1.03848	2685
16	75237.4	1.00138	2677

Nikkei 225

Sample: 1/04/2010 12/18/2020				
Included observations: 2710 (after adjustments)				
Heteroskedasticity robust standard error estimates				
User-specified lags: 2 4 8 16				
Joint Tests		Value	df	Probability
Max |z| (at period 2)*		0.520123	2710	0.9752
Individual Tests				
Period	Var. Ratio	Std. Error	z-Statistic	Probability
2	0.982464	0.033715	-0.52012	0.603
4	1.013781	0.060163	0.229052	0.8188
8	1.000778	0.089315	0.008712	0.993
16	0.998052	0.128166	-0.0152	0.9879
*Probability approximation using studentized maximum modulus with parameter value 4 and infinite degrees of freedom				
Test Details (Mean = 5.94413284133)				
Period	Variance	Var. Ratio	Obs.	
1	46437.8	--	2710	
2	45623.5	0.98246	2709	
4	47077.7	1.01378	2707	
8	46473.9	1.00078	2703	
16	46347.4	0.99805	2695	