2015년 8월
석사학위 논문

Portfolio selection in Korean stock market : using Complex network method

조선대학교 대학원
경 영 학 과
안 석원

한국주식시장에서의 포트폴리오 선택 : 복잡계 네트워크 방법론을 이용하여

Portfolio selection in Korean stock market
: using Complex network method

2015년 8월 25일

조선대학교 대학원
경영학과
안 석원

Portfolio selection in Korean stock market : using Complex network method

지도교수 오 갑 진

이 논문을 경영학 석사학위신청 논문으로 제출함

2015년 4월

조선대학교 대학원
경영학과
안 석원

안석원의 석사학위 논문을 인준함

위원장 조선대학교 교수 김 규 영 (인)

위 원 조선대학교 교수 이 계 원 (인)

위 원 조선대학교 교수 오 갑 진 (인)

2015년 5월

조선대학교 대학원

Table of Contents

List of Tables

List of Figures

Abstract

한국 주식시장에서의 포트폴리오 선택 : 복잡계 네트워크 방법론을 이용하여

안 석원
지도교수: 오 갑 진
경영학과
조선대학교 대학원

본 연구는 Markowitz의 무작위 선택 포트폴리오와는 다르게 Winner-take-all이라는 네트워크 방법론을 이용해 두 가지 유형의 포트폴리오를 구성하고 그 성과를 분석해보았다. 네트워크 상 Degree[1])가 높은 주식들로 구성된 포트폴리오는 20~30여개의 구성 종목으로도 KOSPI와 높은 상관관계를 보였다. 기존 지수들은 시가 총액이 큰 주식에 의해 좌우된다는 단점이 있지만, 본 연구에서는 주식 간 연결 관계를 고려하였기에 주식 시장의 움직임을 보다 더 잘 설명할 수 있어 KOSPI의 보완지수로 이용될 수 있다고 본다. Degree가 낮은 주식들로 구성된 포트폴리오는 주가지수로서의 성과보단 포트폴리오 위험 분산에서 좋은 성과를 보였다. 30여개의 적은 주식으로도 Markowitz의 포트폴리오의 체계적 위험에 비해 훨씬 낮은 위험을 보임을 관찰하였다. 이는 systemic risk의 핵심인 interconnection 정보가 포트폴리오의 분산 효과를 높인다는 것을 시사한다. 마지막으로, Sharpe 비율을 통해 포트폴리오의 성과를 측정했고, 높은 분산 효과를 보이는 낮은 Degree의 포트폴리오가 Markowitz의 포트폴리오에 비해 높은 성과를 보이는 것을 확인할 수 있었다.

1) 상에서 다른 주식과의 연결 개수를 Degree라고 한다.

I Introduction

Should we ignore what we cannot see or measure but exist? Owing to the 'invisible hand' that maximizes individual benefit in financial market, we went through a severe financial distress called 'Sub-prime Mortgage Crisis' in 2008. Through the commodities based on sub-prime mortgage, tremendously large amount of investment money was concentrated to it, which will make intricate global financial crisis. Because what is considered is just the property of individual subject, not information on the interdependence among the financial objects that were a critical trigger of global systemic crisis, it looked innovative and safe financial commodity at that time. As the financial crisis has been erupted frequently than ever, the necessity of considering the interconnection information is growing larger. In recent times, there has been developed many distinctive methodologies to reflect this kind of interconnection factor, however, they are adopted by very limited range of financial studies.

This paper is motivated by this hypothesis: Whether the information on the interconnection among the risk assets is related to the portfolio selection strategy and how about the return or risk of the portfolio by the extent of interconnection among the assets included in the portfolio. In addition, we consider the total risk of individual company is induced by not only its own risk characterized by the both idiosyncratic risk and systematic risk, but also the systemic risk induced by interconnections with others. However, there are few studies that interconnection information is reflected to portfolio selection.

Usually, portfolio selection rule has a long history and forms an impregnable fortress for a long period. Their focus is how to allocate weights to assets and get an optimal performance. Since Harry Markowitz (1952)'s mean-variance portfolio optimization theory (MVT), many researchers conducted empirical studies, however, its boundary of studies are not deviated significantly from MVT. Even its defects in adopting the concept in real investment strategy, estimation error, and lots of studies have proposed various methodologies to overcome it. Because it causes portfolio weight to fluctuate extremely over time, it makes the out-of-sample return perform poorly. In spite of many efforts to reduce estimation error, it is still sensitive to estimation error and controversial among the researchers whether optimal rule can outperform the equal weight portfolio. As opposed to existing portfolio literature, it is the focus of this thesis that how to select the assets based on network properties and get an optimal performance.

Firstly, by what standard we regard it as interconnected state? For example, in case of the stock market, inter-connections among the individual stocks can be considered as additional information in portfolio selection, and they also should be determined by various financial relations. Here, we can consider this kind of inter-connectedness by using the correlation value between time-series return of individual stocks, and the pairs having high correlation value are considered as an interconnected one. In fact, significant network structure was investigated in financial market by many other network studies. According to the previous studies, the information on the interrelations among individual companies in financial market will be useful in portfolio selection in terms of the concept the diversification of portfolio selection.

Then, how can we reflect this interconnection information to portfolio selection? Two papers showed significant results to be referenced. In Garas et al.’s study (2008), it is observed that stock clusters are formed in accordance with their industry. They found that the interaction between individual companies shows a different structure in network according to their economic role. Furthermore, Chi K. Tse, et al (2010) proposed a distinctive approach called

'Winner-takes-all approach' to make a stock index fund portfolio, and emphasized the explanatory power of stocks to market that have lots of interconnection with many others (number of interconnections is called 'degree'). It was the first time that network approach was used to constructing the stock portfolio. In this paper, we developed the latter paper to portfolio selection approach making use of characteristic of correlation pairs proposed in former paper. Moreover, by varying the parameters, while it is conducted in limited setting in reference paper, optimal parameters for optimal performance of portfolio are investigated in this study[2].

The most noticeable result of this thesis is that the portfolios selected from outliers in the network (have a low degree with others) show substantially lower risk than other types of portfolio. As in Meir Statman (1987), Markowitz's risk of portfolio shows monotonic decreasing pattern, while the network based portfolio shows convex pattern with respect to the number of assets. Especially, it is diversified even more than systematic risk which is regarded as a non-diversifiable risk (systematic risk) that is compulsory risk factor to the traditional asset pricing model (capital asset pricing model, CAPM). It also leads to higher performance in terms of Sharpe ratio in the range of the number that shows convex figure.

This result can give an implication to systemic risk that is amplified by the intricate interconnection among investors. It is regarded that low interconnection can prevent the systemic risk from transmitting it to the others by 'too interconnected to fail'. Similarly, it is examined that the portfolios that is composed of stocks that have low interconnections with others show lower portfolio

---

2) There are two problems in Chi K. Tse, et al (2008)'s study in order to be used as a portfolio selection strategy. When the interconnectedness is defined how high value of correlation should be regarded as a significant value is the first. And the question that how many stocks should be included in portfolio is the second. By varying these two parameters, the performance of portfolio is measured.

risk than systematic risk of Markowitz's portfolio. That is to say, systemic risk should be diversified by the certain portfolio strategy using the interconnection information.

In addition, this result is very important in mathematics concept as well as in portfolio theory because the information on the interconnection among the financial subjects without taking into account the efficient market hypothesis can be considered as valuable information in the risk management area as well as the asset pricing theory.

Major contribution of this thesis is that interconnection information is applied to portfolio selection and diverse portfolio sets could be developed compared to previous studies. Furthermore, significantly outperformed results than other portfolio strategies could be obtained from this investigation, and it can be used as an innovative portfolio strategy. Lastly, it is observed that how interconnection information that is critical factor of systemic risk affects the portfolio's risk.

In Chapter 2, we reviewed earlier researches with respect to portfolio theories and network’ s methodologies, and the methodologies that we used in this study are described in Chapter 3. Chapter 4 illustrates the data we used and the empirical results and interpretation. At last, we conclude this study in Chapter 5.

Ⅱ. Literature review

A. Portfolio Optimization

There had been many portfolio theories that focused on the risk reduction through diversification. Since Harry Markowitz (1952) published 'Portfolio Selection' that proved superiority of diversification mathematically for the first time, and proposed optimal allocating weight concept, it has been the fundamental portfolio theory of many other findings.

In Tobin' s separation theorem (1958)[3], author argued that there is advantage of diversification of assets and investors are trying to do balanced investment in the portfolio based on MVT. William Sharpe (1964) proposed optimal portfolio combination model and Capital Asset Pricing Model (CAPM). He classified the risk of portfolio into idiosyncratic risk and systematic risk and considered the systematic risk as a market risk that should be counted in asset pricing model[4]. Evans and Archer (1968) observed that the benefit of risk reduction effect is blurred when a portfolio contains ten or so stocks. This had

3) Tobin, James proposed the advantage of asset diversification in his paper; Tobin, James. "Liquidity preference as behavior towards risk." The Review of Economic Studies 25.2 (1958): row 19, p.85.: "…Moreover, it has the empirical advantage of explaining diversification-the same individual holds both cash and " consols "-while the Keynesian theory implies that each investor will hold only one asset."

4) Sharpe, William F. "CAPITAL ASSET PRICES: A THEORY OF MARKET EQUILIBRIUM UNDER CONDITIONS OF RISK*." The journal of finance 19.3 (1964): row 15, p. 441-442.: "…Although the theory itself implies only that rates of return from efficient combinations will be perfectly correlated, we might expect that this would be due to their common dependence on the over-all level of economic activity. If so, diversification enables the investor to escape all but the risk resulting from swings in economic activity- this type of risk remains even in efficient combinations. And, since all other types can be avoided by diversification, only the responsiveness of an asset's rate of return to the level of economic activity is relevant in assessing its risk."

been quoted by many other researchers like Francis (1986). On the contrary, Meir Statman (1987) insisted that 30 or 40 stocks are needed to diversify enough by comparing the benefit of diversification with transaction cost of it. This result implied that the cost of diversification like transaction cost should be considered in measuring the portfolio performance.

After that, empirical works are followed. It was pointed out that implementing the optimal MVT rule in real investment. To make up of the problem of mean-variance optimal weighting, diverse literature have been published, which the Bayesian approach is the representative. Barry (1974), Bawa, Brown, and Jobson Klein (1979) implemented statistical approach based on diffuse-prior. Shrinkage estimators were also used by Jobson, Korkie, and Ratti (1979), and Jorion (1985). The approaches relying on an asset-pricing model for establishing a prior were proposed by Pastor and Stambaugh (2000). Best and Grauer (1992) proposed the method focused on reducing the estimation error of covariance matrix. Besides, moment restrictions given by factor structures of returns are exploited by MacKinlay and Pastor (2000) to reduce the estimation error. Although their efforts to improve the performance of portfolio, they are still vulnerable to estimation error. DeMiguel, Garlappi, and Uppal (2009) show empirical results that equal weight portfolio yields more better performance than MVT supplementary techniques. Kirby and Ostdiek (2012) refute the results of DeMiguel, Garlappi, and Uppal (2009) using volatility timing and reward-to-risk timing method, however, it is still unstable to estimation.[5)]

All of above studies basically assumes that the market is complete. It is supported by Eugene Fama (1970) who is representative of

5) All of the mentioned paper in this paragraph are empirical works for finding out the calibrating the and σ for allocating the weights. However, the results of papers are not consistent each other as the parameter is vulnerable to dataset and time-series. DeMiguel, Garlappi, and Uppal (2009) and Kirby and Ostdiek (2012) are representative

Efficient Market Hypothesis (EMH) and Nobel Prize Winner of 2013 for his contribution to portfolio theory and asset pricing. But, Robert Shiller (1980, 2000), who is joint Nobel Prize Winner of 2013, challenged EMH proposing that volatility of stock market was greater than could plausibly explained by existing rational view to market, and decision makings are driven by irrational behaviors of investor. His contributions to financial economics have motivated many other behavioral finance researches.

Recently, this behavioral concept is accommodated to portfolio management rule deviating from the established MVT. While only mean and variance are considered in MVT, Shefrin and Statman (2000) presented Behavioral Portfolio Theory (BPT) that considers expected wealth, desire for security and potential, aspiration levels, and probabilities of achieving aspiration levels. Das, Markowitz, Scheid, and Statman (2010) integrated appealing parts of MVT and Shefrin and Statman (2000)'s work into a Mental Accounting (MA) framework. These are distinguished from MVT in that irrational investors are assumed.

B. Network approach

Following the behavioral approaches to portfolio management, this thesis is motivated by that individual assets are not independent, but dependent. As finance has developed globally, the fluctuation of value of individual assets is affected by their intricate financial interconnections among them, even countries. Therefore, it is important to consider the network structure of financial market. Similar approach has been progressed in other fields. In biology, Anderson and May (1991) and Lloyd, Schreiber, Kopp, and Getz (2005) proposed a simple method that analyze influential power of "super spreaders" in the spread of infectious diseases with interconnection concept. Alsoand Kareiva and Levin (2003) investigated the importance of keystone species to whole ecosystem.

In this regard, network methodologies have been used in analyzing the stock market to identify inter-connections among them. This approach reflects that the market is inefficient and there is information asymmetry in stock market. Basically, the correlation value is calculated between two stocks and the value is transformed to distance value: high correlation pair is positioned closely in network, and vice versa. Mantegna and Rosario (1999) used Minimal Spanning Tree approach (MST) to investigate the financial market structure. The method has been used for filtering the edges, which makes the network much simpler to analyze. However, the essential information of the network structure is also lost, too. Although Planar Maximally Filtered Graph (PMFG) that is proposed by Tumminello, Michel, et al (2005) was applied in order to make up for shortcomings of MST, it also has similar defects. The two methods often filtered highly correlated pairs, but retained relatively un-correlated pairs because of their topological reduction criteria. In those studies, it is shown that the stocks in same industry are collected together significantly in network, because of short distance among them. However, these studies are limited to analyzing the structural properties of network.

Especially, Garas, Antonios, Panos Argyrakis, and Shlomo Havlin (2008) analyzed the New York Stock Exchange (NYSE) in network perspectives. They found that weaker links (low correlation) contribute overall connectivity of the network, while strong links (high correlation) are clustered according to their industrial properties. Similar analysis is progressed by Chi K. Tse, et al (2010). They proposed a method that makes a stock index fund which explains stock market in terms of interconnection. The index fund is composed of the stocks that have strong links with many other stocks, which shows high correlation with existed stock index. This approach is different from previous network approach in that it makes binary decision on connecting two stocks according to the correlation value being larger than threshold value when edges of the network are established. Moreover, it was the first time that network approach is used in portfolio construction unlike previous network analysis trend. While the threshold values are set to certain value in Chi K. Tse, et al (2010)'s paper, diverse portfolio sets are made by varying it, and performances of them are investigated in this paper.

Ⅲ. Methodology

In many recent studies, the complex network in financial market is constructed by linear correlations among individual companies. In this thesis we establish a Pearson-correlation to create network using daily returns of individual stocks traded in the Korean Stock Market during the period from 2000 to 2012 year, 3217 business days, when 1081 companies had been listed. From the daily return we can a create correlation based network as proposed in Chi K. Tse, et al (2010).

First, we calculate the Pearson-correlation between individual companies used in this thesis and constructed the a lot of network over time with 1000 days of window, about 4 business years and moves the moves the window day by day. That's why there is 2217 number of window. However, there is the problem that the stocks can be delisted due to the default or other reasons, which could be included in the window. Therefore, we don't consider the stocks that have no return for 10 days consecutively in this procedure and it is regarded as a business-paused company.

Using daily return time series, we calculate the correlation value, denoted by as follows

$$C_{ij} = \frac{\sum_{t=1}^{N}[(x_i(t)-\overline{x_i})(x_j(t)-\overline{x_j})]}{\sqrt{\sum_{t=1}^{N}(x_i(t)-\overline{x_i})^2}\sqrt{\sum_{t=1}^{N}(x_j(t)-\overline{x_j})^2}}$$

where (t) and $x_j(t)$ is return time series at day t. The network can be viewed transforming the calculated correlation value to a distance one using the function described below:

$$_j = \quad 2(1-\rho_{ij}),\ 0 \le d_{ij} \le 2$$

where small distance value means that there are strong correlation between the stocks i and j, and vice versa.

Next, the distance values are sorted in descending order and the pairs which show high distance value than pre-set threshold value are filtered out. The other pairs are regarded that they are connected. The threshold value is the first parameter of this methodology. There will be isolated stock pairs or groups (=module) that are disconnected with other groups, because high distance pairs are filtered out. We consider only the stocks which belong to the largest module, because the other modules have a very limited number of nodes that cannot be used as a significant network structure as in Chi K. Tse, et al (2010).

From this network, the number of degrees of each node is counted with the correlation pairs the node is connected with. It is also limited by second parameter which decides the number of companies that will be included in portfolio. The degree of influence or explanatory power of one stock to other stocks is indicated by the degree of one stock. This is because the stocks that have many relationships with others would show high correlation value and low distance value, which makes the stock have higher degree than other stocks.

For the last procedure to select stocks from the network, the degree of stocks is also sorted in descending or ascending order. Stocks that is influential to other stocks are selected in descending ordered degree ordering, while comparably less correlated stocks will be selected in ascending ordered one. The former is named as a 'hub portfolio and the other is 'outlier portfolio'[6]. That is to say, high degree stocks are selected first in hub portfolio, and low degree

6) Mostly, the stocks that have lots of interconnections with others located in the hub of network. That's why we called the former as a 'hub portfolio'. And the stocks that have low degree position outside of the network, which is called 'outlier portfolio'.

stocks are included first in outlier portfolio. These two stock portfolios show different characteristics because different role does they have in network structure as in Garas, Antonios, Panos Argyrakis, and Shlomo Havlin (2008).

We assume that the portfolios are traded in daily and the all invested money including the money from capital return are invested again. To measure the trading return of portfolio, out-of-sample return (OSR) is used. This assumption is as follows.

$$_{+1} \quad M_t(1+R_{w+t}) \ (t=1,\cdots,n)$$

$$OSR_{t+1} = (M_{t+1} - M_t)/M_t$$

where w is window size and R is return. As the value t is increased the window is moving. M_t is the money invested at time t and by multiplying M_t by return $(1+R_{w+t})$ we can get M_{t+1}. And by arithmetic return of M_{t+1} and M_t, out-of-sample return is calculated. That is to say, portfolios are constructed from time series 1 to w and buy the stocks of the portfolios and hold them until next time. By equally weighting each stock return, the portfolio's return is decided. In this way, 2217 number of time series return of out-of-sample is calculated.

So far, we selected stocks from the network and calculate the return of portfolios. In summary, only the pairs that have high correlation value is selected and the degree of each stock is counted by the number of connections they have. With the portfolios mentioned above, we have to calculate the return of portfolios for various analysis.

Especially, the included stocks in these portfolios are varied not very much, as the first and second parameters are changed. But, we will investigate how the results including correlation with stock index, standard deviation, and Sharpe ratio of portfolio are changed by varying the parameters.

3-1. Correlation with stock index

Hub portfolio is composed of the stocks that have lots of connections with others. The thing that have many connections means that it has close correlation with the others. Therefore, it would be explainable to stock market with a few stocks. To measure how the network-based-portfolio is correlated with the stock index, we used Pearson-correlation with KOSPI return. As in Figure 1, hub portfolio shows very high correlation with the index even with a few stocks. The result of outlier portfolio shows high enough, but, it is not with a few stocks.

3-2. Standard Deviation of portfolio for risk

In portfolio management theory, the risk of portfolio is measured with standard deviation of portfolio return. Compared to randomly selected portfolio as in Markowitz (1952), higher risk is expected in hub portfolio, because it highly correlated with other stocks, which caused the lack of diversification among the stocks. However, the stocks in outlier portfolio would be superior to the other two portfolios from a diversification perspective, because it is not interconnected with other stocks.

From here, The comparison target portfolio to network-based portfolio is Markowitz's randomly selected and naive portfolio as in DeMiguel, Garlappi, and Uppal (2009)[7].

As in Figure 2, Risk of former one is considerably higher than that of random portfolio, while the risk of latter one is much lower than random even showing convex pattern. It will be described in detail in empirical result part.

3-3. Sharpe ratio of portfolio for performance

No matter how the portfolio’s risk is low, risk / return profile is also important concept in portfolio management, because it is connected to portfolio’s performance. To measure the performance, we use Sharpe ratio[8] described as below:

$$rtfolio\,Sharpe\,Ratio = \frac{Portfolio's\,Excess\,Return}{Portfolio's\,Risk}$$

In previous studies of portfolio theory or empirical investigation, Sharpe ratio has been used to compare their performance of portfolios to others.

As shown in Figure 3-B, the performance of outlier portfolio is much better than that of hub portfolio and random. This is because the former shows moderate return but the risk is much lower than others and shows convex figure that has minimum point. On the other hand, the latter’s

---

7) Demiguel, Garlappi, and Uppal (2009) compared the naive portfolios to others that the methodologies are different for estimation of μ and σ for portfolio weighting. This study observed that the naive portfolio is superior to the others in terms of Sharpe ratio. Although the other papers investigated the portfolios that outperform it, its difference was not critically different and it was vulnerable to the time-series of data.

8) For the portfolio's return, excess return, return of portfolio minus risk free return, should be used, however, the daily risk free return is so low that can be ignored. Furthermore, this study is for comparing the performance of each portfolio, it does not affect the results.

poor performance is due to the fact that the return is not so high compared to the risk, and its risk is comparably higher than outlier one.

Ⅳ. Empirical results

4-1. Data Description

To calculate the Pearson-correlation values, daily return for individual companies listed on KOSPI is used. During the period from January 04, 2000 to December 31, 2012, 1073 stocks had been listed on KOSPI. Two types of daily return are used, daily adjusted return and daily return. The daily return of stock on day t, denoted by $r(t)$, is defined as

$$r_i(t) = \frac{p_i(t) - p_i(t-1)}{p_i(t-1)},\ p_i(t) = stockprice$$

where $p(t)$ is stock price at day t. Because the two data types don't affect the result significantly, the results below are based on daily adjusted return, which the dividend factor is reflected. The statistical description of used data sets is reported in Table 1. In Table I, the linear statistics of the return time series in KOSPI stock market shows different features compared to the random walk process of efficient market hypothesis.

Table 1. Statistical data description of KOSPI stock data

	Mean	Standard Deviation	Skewness	Kurtosis
2000	-0.0026	0.0287	-0.142	3.8452
2001	0.0014	0.0213	-0.5851	7.3223
2002	0.0013	0.0203	-0.0712	3.7052
2003	0.0089	0.0165	-0.0422	3.6998
2004	0.0055	0.0149	-0.4299	4.4689
2005	0.0018	0.0105	-0.2129	3.3342
2006	0.0023	0.0115	-0.4183	3.6778
2007	0.0013	0.0145	-0.6011	6.2128
2008	-0.0019	0.0246	-0.0634	7.4217
2009	0.0017	0.0155	-0.3037	4.7663
2010	0.0084	0.0095	-0.5189	3.6103
2011	-0.0032	0.0165	-0.2859	4.4537
2012	0.0039	0.0097	-0.0319	4.0429

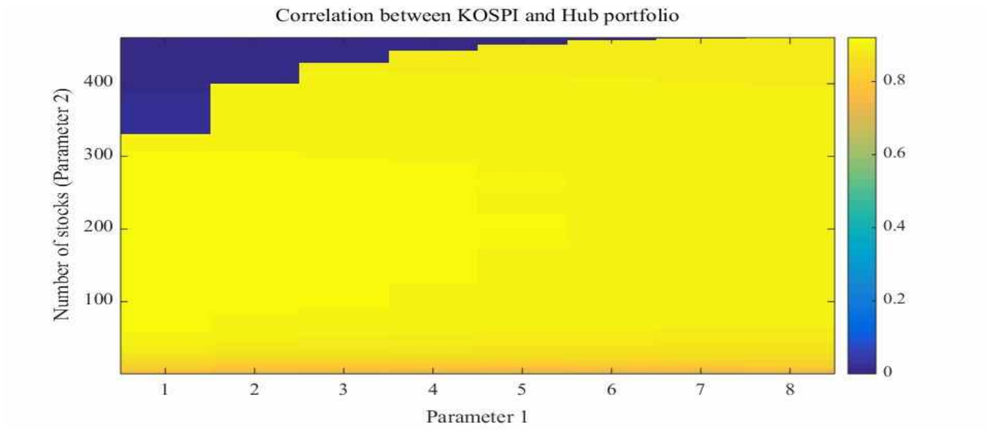
4-2. Empirical Results

4-2-1. Correlation between KOSPI and portfolios

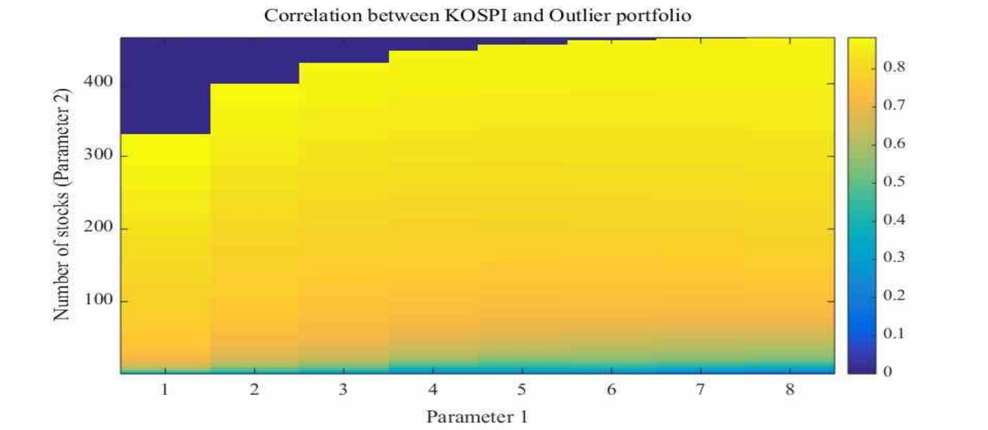
As mentioned above, by varying the first and second control parameters, diverse portfolio sets with complicated correlation structure are constructed. However, the sets are too much to explain their results in tables, summary results are shown in the tables as follows, and the whole outcomes are described in the figures with colors map.

Figure 1. Correlation between KOSPI return and network-based portfolio return

A. Hub portfolio



B. Outlier portfolio



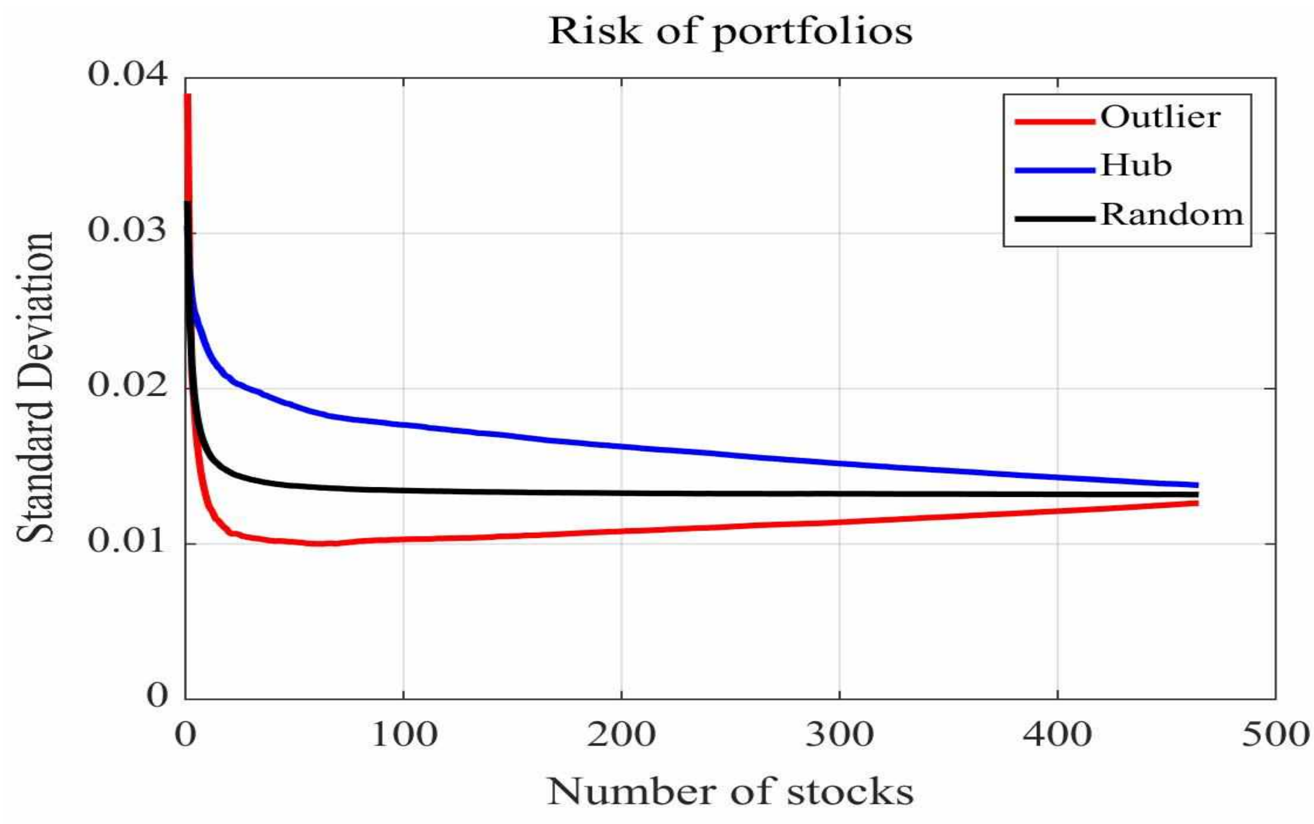
The previous paper made a hub stock portfolio and calculated the correlation value between hub portfolio and stock index. Based on the methodology proposed by Chi K. Tse, et al (2010), investigation has conducted using KOSPI data and almost same result can be obtained. Because this study is for making stock index fund portfolio, out-of-sample return is used. Figure 1-A and 1-B illustrate the correlation values between KOSPI return and two types of portfolio return. X-axis means parameter1 (portion of highest correlation) and Y-axis means parameter2 (the number of companies in portfolio). There is an index bar indicating the correlation value at the right side of each graph. The correlation value can be known by the color. As in Figure 1-A that almost all of the each parameter's color are yellow, correlation values are entirely high (0.8 ~ 0.9). This result is consistent with the result of Chi K. Tse, et al (2010): The stocks located in hub of network are highly correlated with more stocks than others, which makes the portfolio constructed based on the stock network property explain the market portfolio well. However, the values in Figure 1-B are comparably lower than 1-A. This is because the portfolios are constructed with the stocks that have low degree with other stocks and have less explanatory power to stock market. But, it is still correlated with KOSPI or so (0.7 ~ 0.8) Here, one meaningful suggestion can be made. KOSPI and others including KOSPI 200, and KOSPI 100 have a defect that they are critically driven by the total market value of stocks, which is not representative of individual stocks. However, when the network information, interconnection, is used to prove what kinds of stocks can explain the other stocks, detailed stock market can be explained with the network-based index that is composed of fewer stocks than previous one.

In sum, we can prove the explanatory power of hub portfolio over the stock market, which can be used as an alternative stock index that explains the stock market in terms of interconnection even with smaller number of stocks. Furthermore, moderate correlation value of each portfolio suggests that it can be a useful index fund portfolio management methodology.

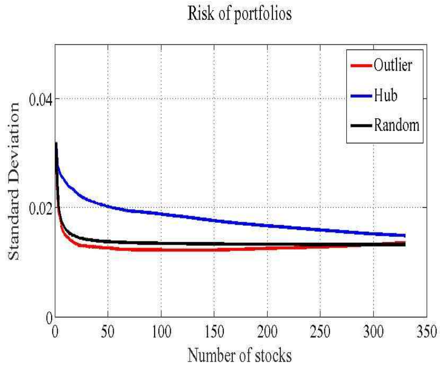
4-2-2. Risk of portfolios

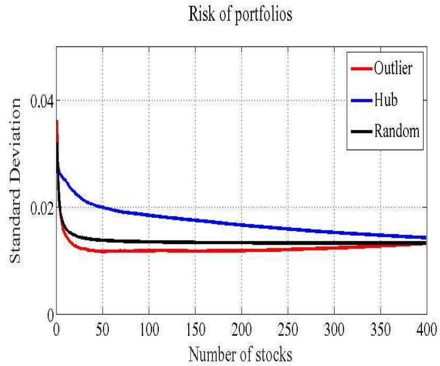
Figure 2. Standard deviation of each portfolio set

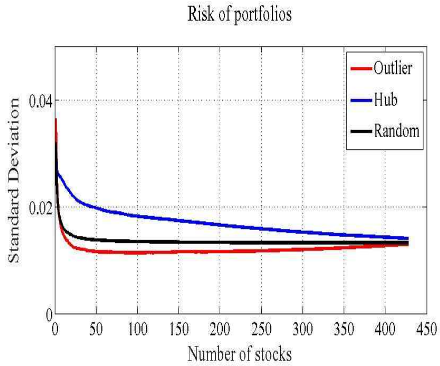
A. 80% of correlation pairs are remained - Parameter 1

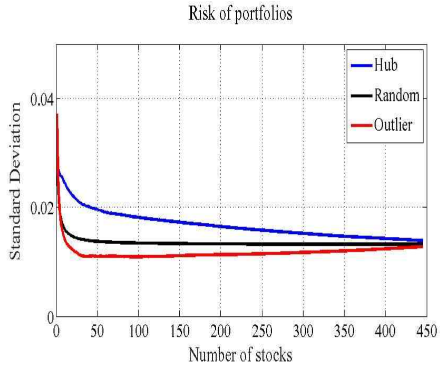


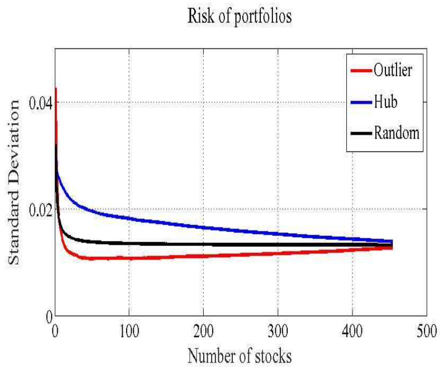
B. 10~70% of correlation pairs are remained each - Parameter 1

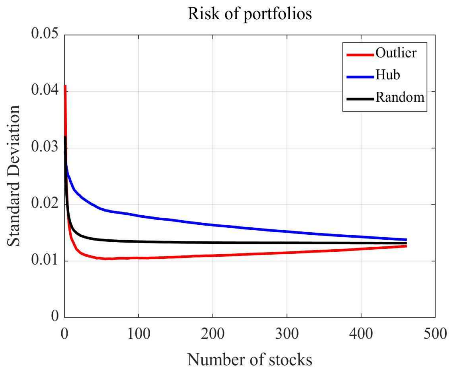












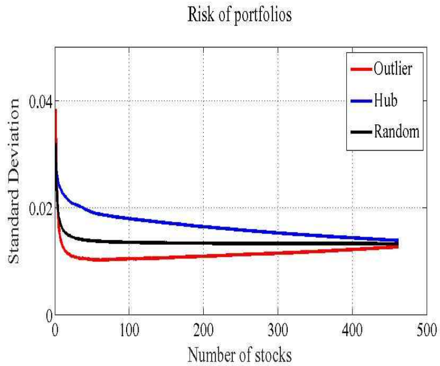


Figure 2-A and 2-B show the risk of network-based portfolio and Markowitz's randomly selected portfolio. Figure 2-A is the graph that chosen parameter is setting 80%. This means that 80% of highest correlation pairs are considered in constructing stock network. This is why almost all of the stocks listed on stock market at certain period are included in portfolio. Y-axis means the standard deviation. A blue, red, and black line indicates the risk of hub, outlier, and randomly selected portfolio. As the number of companies in a portfolio are increased (the number of companies), the risk of portfolio shows decreasing pattern, because the risk of each portfolio is diversified.

Besides, the table 3 indicate the number of stocks included in portfolios that have minimum standard deviation of portfolio. From the table, we can know the detailed number of stocks and the value of risk. The noticeable point of the table is that the number of stocks that have minimum variance is very stable by the first parameter.

The black line is the same as the risk of portfolio that is described in Markowitz (1952). The risk of portfolio shows monotonic decreasing pattern, and almost all of the standard deviation is diversified at about 30~40 stocks in portfolio as in Statman, Meir (1987). Furthermore, minimum risk is observed at the end of the x-axis, which is considered as a systematic risk.

Comparably high risk is shown in the blue line, which is due to the lack of diversification. In figure 1, high correlation with KOSPI is observed and this is due to the high correlation with many other stocks. However, this property is not effective in terms of risk diversification. In other words, the stocks included this portfolio should make a relatively large influence into the market. This leads to lower performance of portfolio compared to randomly selected portfolio, as shown in Figure 3-A, due to the fact that the return of portfolio is not high enough with higher risk.

However, we find that the meaningful result is outlier portfolio's performance. Outlier portfolio, literally, is composed of the stocks that are positioned outside of the network. These stocks are correlated with fewer stocks than others, which makes them not correlated or even anti-correlated among the stocks in portfolio. Interestingly, the risk of outlier portfolio shows convex pattern. It is less clear in the smaller percent of correlation pairs. This is because the loss of network information is notable when we use limited number of correlation pairs. As in Figure 2-A, its risk is much lower than random portfolio, which is opposed to the existing portfolio theory insisting that the benefit of diversification is exhausted when a portfolio contains 30 or so stocks as in Statman, Meir (1987).

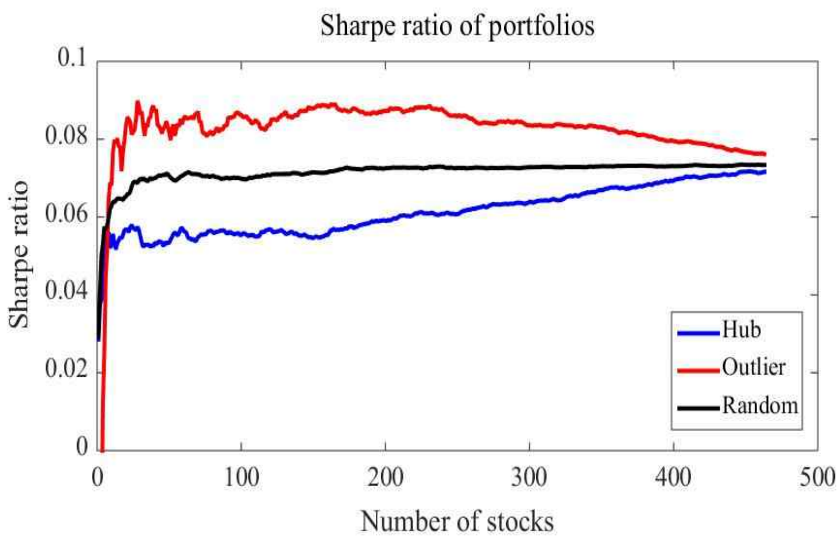
Besides, there are two more noticeable meanings. First, the risk of outlier portfolio shows convex pattern. Generally, to fully diversify the idiosyncratic risk of portfolio, all of the stocks have to be included, and only systematic risk remained. However, the risk can be much more diversified than Markowitz's randomly selected portfolio with fewer stocks, and even the systematic risk considered that is to be taken. We can say that this result is from the network property that is characterized by interconnection information. This is related to systemic risk which is controversial that it is idiosyncratic risk or systematic risk. In the view that systemic risk can be explained by the concept of interconnection, this result can give a significant implication: systemic risk is related to idiosyncratic risk and the risk of portfolio can be diversified even more than the risk of randomly selected portfolio that is previously considered as a systematic risk when the interconnection information is reflected to portfolio selection.

4-2-3. Sharpe ratio of portfolios

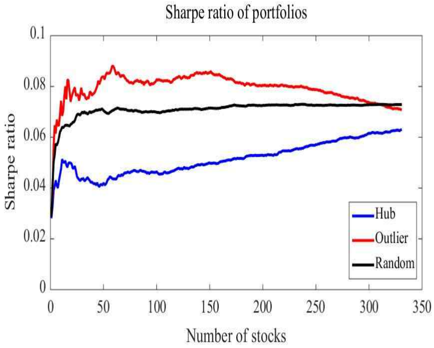
Second, convex pattern in the relationship between portfolio risk and the number of companies means that there is minimum point of risk. Low risk does not always guarantee sound performance, however, there is high probability that there can be an optimal point of portfolio in terms of traction cost, because minimum point should have the lowest risk as well as the proper transaction cost. Figure 3 shows the Sharpe ratio of each portfolio. Overall Sharpe ratio of outlier portfolio is superior to random one. Somewhat unstable Sharpe ratio, not concave, is shown in 10% ~ 40% control parameter 1. But, its performance of minimum point of risk is getting better as the value of parameter 1 is increased. Furthermore, when the transaction cost is considered here, the relative performance of portfolios that have a few stocks in their portfolios would be much better than others that have more stocks. These results imply that minimum point of portfolio risk makes the performance optimal in terms of Sharpe ratio.

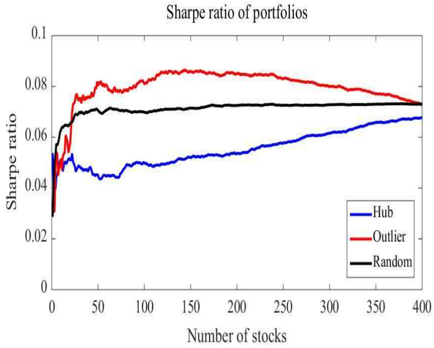
Figure 3. Sharpe ratio of each portfolio set

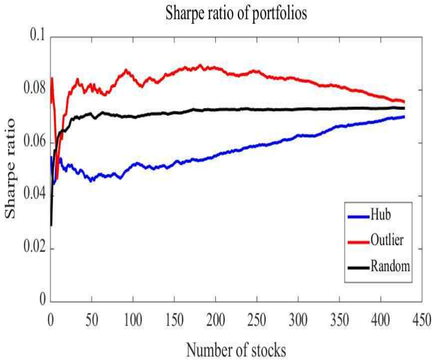
A. 80% of correlation pairs are remained - Parameter 1

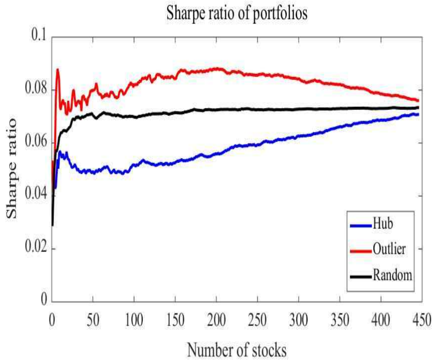


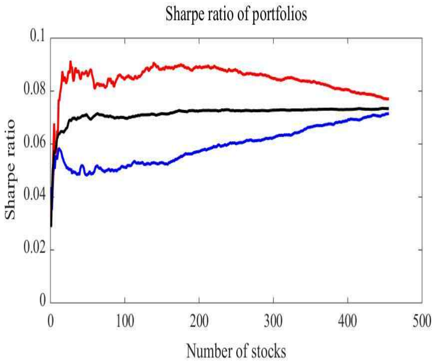
B. 10 ~ 70% of correlation pairs are remained each - Parameter 1

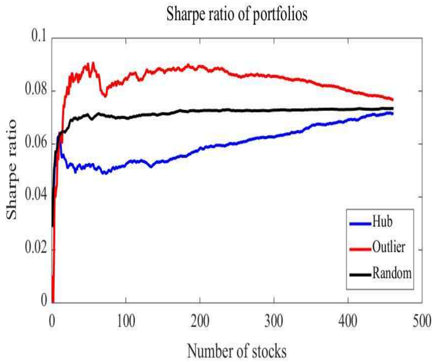


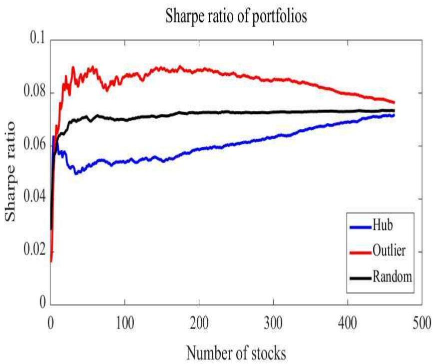












4-2-4. Industry sector of stocks included in each portfolio

We observed that each network-based portfolio shows different performance. To investigate why they show unlike results, we check the industry sector of the stocks where they are included. For simplification, the tables are based on the result that the number of stocks are 50 and 100[9], which shows minimum point of risk and sound performance in outlier portfolio. There are 16 number of industry sector[10], and the manufacturing industry makes up of almost half of the whole stocks by Korea's industry structure[11]. Because of this biased distribution to manufacturing, the weight of stocks of portfolio will also be skewed. That is why we focus on the difference of the weight between hub and outlier portfolio in table 2. For the weight, the number of each industry sector are counted, and they are divided by 50 and 100. It can be known that which industry sector is influential in portfolio from the weight.

The row line of table 2 means the industry sector and column one is first parameter. It is observed that the weight of manufacturing sector is dominant and finance and insurance sector is also high. On the other hand, the finance and insurance sector is lowest and the percentage of hub one become higher in outlier portfolio. This means that the stocks of financial sector are located in hub of the network and whether the stocks of financial sector are included in the portfolio affects the

9) The table of portfolio composed of 100 stocks is in Appendix 1.

10) A : Agriculture, forestry, and fishing / B : Mining / C : Manufacturing / D : Electricity and water supply / F : Construction / G : Whole, retail trade / H : Transfortation / I : Accomodation and food / J : Publishing, motion picture, broadcasting, telecommunications, and information service / K : Finance and insurance / L : Real estate and renting/ M : Professional, scientific, and technical service / N : Business facilities management and support / P : Education / R : Art, sports, and leisure / S : Membership organization, maintenance, and personal service

11) The number of stocks belong to each industry sector are as follows; A : 4 / B : 2 / C : 628 / D : 12/ F : 58 / G : 86 / H : 22 / I : 2 / J : 35 / K : 130 / L : 3 / M : 71 / N : 5 / P : 1 / R : 5 / S : 1, and industry sector of 5 stocks are not reported. We excluded the calculating the weight of each industry and ranking including A, B, D, I, L, N, P, R, and S, because it distort the weight by its small increase.

Table 2. Weight of industry sector in each portfolio[12)]

A. Hub portfolio composed of 50 stocks.

	C	F	G	H	J	K	M
10%	0.388	0.064	0.051	0.030	0.012	0.383	0.072
20%	0.403	0.060	0.057	0.029	0.013	0.361	0.075
30%	0.414	0.061	0.060	0.031	0.012	0.344	0.076
40%	0.422	0.058	0.061	0.028	0.014	0.339	0.076
50%	0.423	0.058	0.061	0.030	0.013	0.334	0.078
60%	0.422	0.061	0.062	0.030	0.014	0.328	0.078
70%	0.416	0.066	0.060	0.030	0.017	0.327	0.079
80%	0.419	0.062	0.065	0.034	0.018	0.317	0.078

B. Outlier portfolio composed of 50 stocks.

	C	F	G	H	J	K	M
10%	0.695	0.041	0.060	0.022	0.054	0.036	0.055
20%	0.652	0.039	0.067	0.027	0.046	0.031	0.063
30%	0.675	0.031	0.067	0.031	0.038	0.020	0.062
40%	0.680	0.022	0.083	0.037	0.034	0.016	0.055
50%	0.672	0.019	0.096	0.042	0.036	0.014	0.048
60%	0.672	0.017	0.105	0.040	0.039	0.013	0.042
70%	0.669	0.014	0.108	0.038	0.049	0.012	0.042
80%	0.667	0.014	0.111	0.035	0.052	0.012	0.042

risk and the performance of portfolio. To investigate how the financial sector affects the portfolio, more detailed investigation is needed.

12) C : Manufacturing / F : Construction / G : Whole, retail trade / H : Transfortation / J : Publishing, motion picture, broadcasting, telecommunications, and information service / K : Finance and insurance / M : Professional, scientific, and technical service

However, when we consider that there is large difference in risk and performance between hub and outlier portfolio, it should be regarded that we cannot expect high diversification effect with the stocks of financial sector.

4-2-5. Comparison before and after financial crisis

There had been economic stable state from 2004 to 2007, and various types of financial commodities had been developed and lots of investment was concentrated to them ignoring its potential financial risk as a systemic risk. However, after the sub-prime crisis in 2008, financial uncertainty was expanded and global economic recession made matters worse. It was an stimulant for investor to care more about the risk diversification than ever. Following this concept, to observe how the portfolio risk and performance is changed before and after the financial crisis, the time-series data is divided in two period: from 2004 to 2007 and 2008 and 2012.

As in Figure 4-A, it is observed that the overall risk of former period is comparably lower than the latter, however, risk diversification effect of outlier portfolio is more outstanding at latter (It is diversified more with smaller stocks). This can be regarded that the investors are more sensitive to risk diversification, which makes the assets less correlated than ever, and causes better condition to diversify. However, because the economic state was very unstable and had lots of uncertainty, the overall risk of former was higher than latter as in risk of randomly selected portfolio.

Moreover, There are significantly large gap of Sharpe ratio between former and latter period. Although, the risk of each period is not significantly different, the Sharpe ratio of former shows about 5 times larger than the latter. This is due to the fact that there was economic recession in latter period and lots of investments were out of stock market. Therefore, no matter how the risk of both of the periods are not different each other, it is assumed that the return of latter time was much lower than the former.

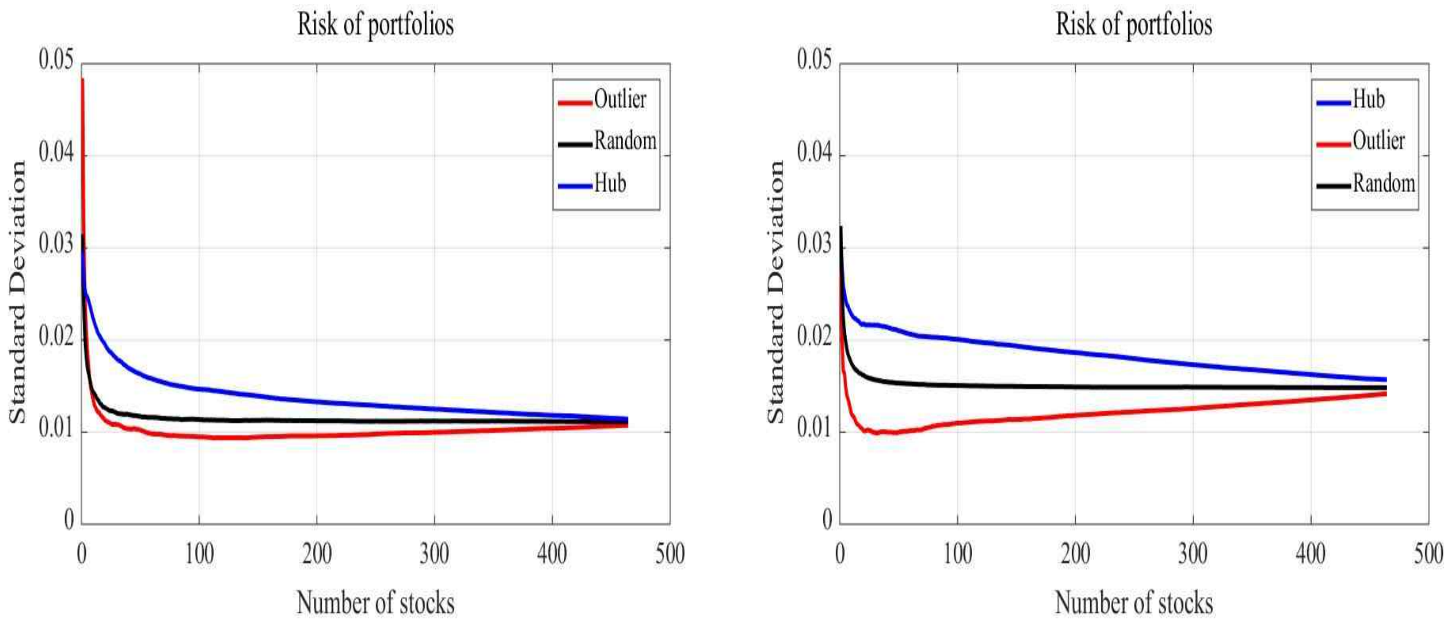
Although, the risk diversification effect is more significant at latter, the overall portfolio risk is higher than former over the first parameter as in Figure 4-B in Appendix. Furthermore, when we consider that there was stock market boom at former and recession at latter, lower return can be expected at latter. This is why it shows considerably higher Sharpe ratio at former.

Figure 4. Comparison the portfolio risk before and after financial crisis[13)]

A. 80% of correlation pairs are remained - Parameter 1

1. Before crisis (2000 ~ 2007)

2. After crisis (2008 ~ 2012)

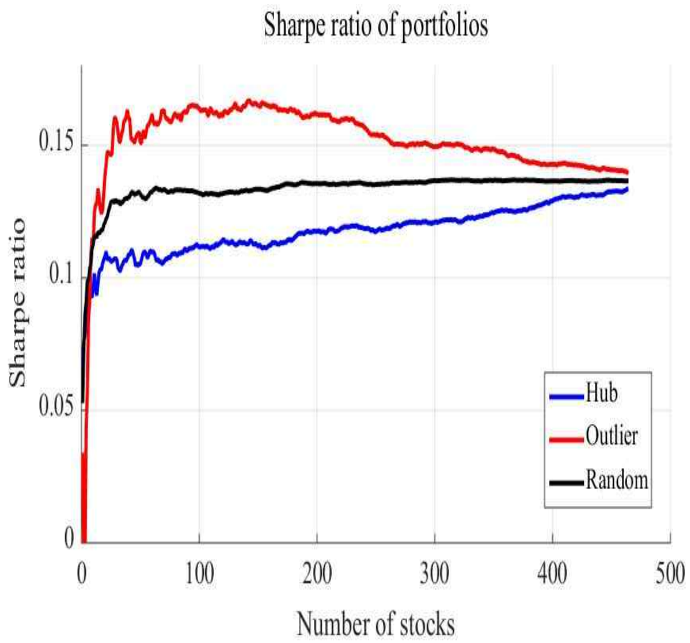


13) 10% ~ 70% of first parameter is in Appendix.

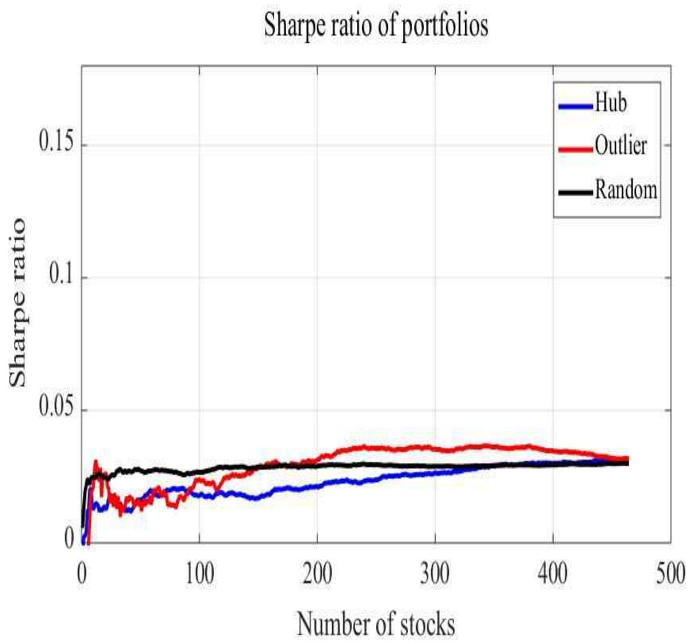
Figure 5. Comparison the portfolio performance before and after financial crisis[14)]

A. 80% of correlation pairs are remained - Parameter 1

1. Before crisis (2000 ~ 2007)



2. After crisis (2008 ~ 2012)



14) 10% ~ 70% of first parameter is in Appendix.

Ⅴ. Conclusions

We have used complex network analysis methodologies to construct portfolio in KOSPI equity market activity up to 232 companies from 2000 to 2014. Our main argument is that the information of interconnection among equity activity-customarily used to access systemic risk-are informative in portfolio selection process. We could know that how Interconnection information affect the portfolio risk and the performance. Diverse portfolio sets are made by the first and second parameters and it is observed that the results are consistent over them. The summary empirical results are as follows.

To the best of our knowledge, ours is the first attempt to construct the portfolio sets of KOSPI equity market. Moreover, we do not focus on just a hub company with a lot of interconnection of the equity network, but instead we analyze the broadest information of interconnection of stock market that represent the stock market characteristic in terms of diversification. We have constructed separate portfolio sets based on the number of connection (degree) and on the level of correlation among individual companies. We have observed that high correlation value is observed in hub portfolio with a few stocks. To overcome the limitation of previous stock index including KOSPI, network-based stock index was made. This can be useful stock index that reflect the interconnections among the individual stocks with the stocks that have high explanatory power to the others, not driven by the total market value.

Second, significantly lower risk than Markowitz's portfolio and convex pattern are shown on outlier portfolio. The stocks are selected from the outlier of the network which have low interconnections with others, which are less and even anti-correlated each other in portfolio. Moreover, the convex figure means that it has minimum point of risk of portfolio with fewer stocks. This result implies that when the network information is used, the risk of portfolio can be minimized with fewer stocks than Markowitz's portfolio.

Lastly, high performance by the Sharpe ratio can be obtained in outlier portfolio that shows low risk. When it comes to that low risk makes high Sharpe ratio, this result implies that the minimum point of risk can be an optimal point of portfolio, even when the transaction cost is considered. These outcomes have implication that this portfolio selection methodology can be used as a portfolio management tool in real investment.

Research in this thesis should also give an important implication to systemic risk. Systemic risk was larger and larger as the financial interconnections through the various financial commodities are formed intricately, however, it could be hedged or minimized reducing the interconnections among them. In this regard, it is controversial that it can be diversified or should be taken as a systematic risk in terms of portfolio risk diversification. In this state, the interconnection information is reflected to portfolio selection and can find the risk of portfolio is diversified than the risk considered as a systematic risk. Therefore, it can be suggested as follows: if the systemic risk is idiosyncratic risk, it can be diversified with network information among financial interconnections.

References

1. Anderson, Roy M., and Robert McCredie May. Infectious diseases of humans. Vol. 1. Oxford: Oxford university press, 1991.
2. Barry, Christopher B. "Portfolio analysis under uncertain means, variances, and covariances." The Journal of Finance 29.2 (1974): 515-522.
3. Bawa, Vijay S., Stephen J. Brown, and Roger W. Klein. "Estimation risk and optimal portfolio choice." NORTH-HOLLAND PUBL. CO., N. Y., 190 pp (1979).
4. Best, Michael J., and Robert R. Grauer. "The analytics of sensitivity analysis for mean-variance portfolio problems." International Review of Financial Analysis 1.1 (1992): 17-37.
5. Case, Karl E., and Robert J. Shiller. "Is there a bubble in the housing market?." Brookings Papers on Economic Activity 2003.2 (2003): 299-362.
6. Das, Sanjiv, et al. "Portfolio optimization with mental accounts." (2010): 311-334.
7. DeMiguel, Victor, Lorenzo Garlappi, and Raman Uppal. "Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy?." Review of Financial Studies 22.5 (2009): 1915-1953.
8. Evans, John L., and Stephen H. Archer. "DIVERSIFICATION AND THE REDUCTION OF DISPERSION: AN EMPIRICAL ANALYSIS*." The Journal of Finance 23.5 (1968): 761-767.
9. Fama, Eugene F. "Efficient capital markets: A review of theory and empirical work*." The journal of Finance 25.2 (1970): 383-417.
10. Francis, Ashish, et al. "Portfolio analysis." European Journal of

Operational Research. 1986.

11. Garas, Antonios, Panos Argyrakis, and Shlomo Havlin. "The structural role of weak and strong links in a financial market network." The European Physical Journal B 63.2 (2008): 265-271.

12. Jobson, J. D., Bob Korkie, and Vinod Ratti. "Improved estimation for Markowitz portfolios using James-Stein type estimators." Proceedings of the American Statistical Association, Business and Economics Statistics Section. Vol. 41. American Statistical Association Washington DC, 1979.

13. Jorion, Philippe. "Bayesian and CAPM estimators of the means: Implications for portfolio selection." Journal of Banking & Finance 15.3 (1991): 717-727.

14. Kirby, Chris, and Barbara Ostdiek. "It’s all in the timing: simple active portfolio strategies that outperform naive diversification." Journal of Financial and Quantitative Analysis 47.02 (2012): 437-467.

15. Lloyd-Smith, James O., et al. "Superspreading and the effect of individual variation on disease emergence." Nature 438.7066 (2005): 355-359.

16. MacKinlay, A. Craig, and Lubos Pastor. "Asset pricing models: implications for expected returns and portfolio selection." Review of Financial studies 13.4 (2000): 883-916.

17. Mantegna, Rosario N. "Hierarchical structure in financial markets." The European Physical Journal B-Condensed Matter and Complex Systems 11.1 (1999): 193-197.

18. Markowitz, Harry. "Portfolio selection*." The journal of finance 7.1 (1952): 77-91.

19. Paine, Robert T. The importance of species: perspectives on expendability and triage. Eds. Peter M. Kareiva, and Simon A.

Levin. Princeton University Press, 2003.

20. Pástor, Ľuboš, and Robert F. Stambaugh. "Comparing asset pricing models: an investment perspective." Journal of Financial Economics 56.3 (2000): 335-381.

21. Sharpe, William F. "CAPITAL ASSET PRICES: A THEORY OF MARKET EQUILIBRIUM UNDER CONDITIONS OF RISK*." The journal of finance 19.3 (1964): 425-442.

22. Statman, Meir. "How many stocks make a diversified portfolio?." Journal of Financial and Quantitative Analysis (1987): 353-363.

23. Tobin, James. "Liquidity preference as behavior towards risk." The Review of Economic Studies 25.2 (1958): 65-86.

24. Tse, Chi K., Jing Liu, and Francis Lau. "A network perspective of the stock market." Journal of Empirical Finance 17.4 (2010): 659-667.

25. Tumminello, Michel, et al. "A tool for filtering information in complex systems." Proceedings of the National Academy of Sciences of the United States of America 102.30 (2005): 10421-10426.

26. Shefrin, Hersh, and Meir Statman. "Behavioral portfolio theory." Journal of financial and quantitative analysis 35.02 (2000): 127-151.

27. Shiller, Robert J. "Do stock prices move too much to be justified by subsequent changes in dividends?." (1980).

Appendix1

This table shows the weight of industry sector which shows what sector is dominant in portfolio. The row line means the code of industry and column one is first parameter. We can observe that the highest value is in finance and insurance at the table of hub portfolio. However, it is reserved completely in outlier portfolio by recording the last. Moreover, the ranking is comparably consistent over the parameter.

Table 2. Weight of industry sector in each portfolio[15)]

A. Hub portfolio composed of 100 stocks.

	C	F	G	H	J	K	M
10%	0.433	0.100	0.054	0.037	0.021	0.263	0.090
20%	0.453	0.094	0.056	0.037	0.020	0.248	0.090
30%	0.462	0.089	0.061	0.035	0.020	0.239	0.089
40%	0.468	0.089	0.063	0.034	0.020	0.233	0.088
50%	0.469	0.089	0.063	0.035	0.020	0.231	0.087
60%	0.473	0.087	0.063	0.035	0.018	0.227	0.090
70%	0.473	0.084	0.065	0.035	0.018	0.224	0.094
80%	0.476	0.084	0.065	0.034	0.019	0.224	0.090

B. Outlier portfolio composed of 100 stocks.

	C	F	G	H	J	K	M
10%	0.700	0.035	0.052	0.017	0.042	0.033	0.052
20%	0.676	0.037	0.064	0.024	0.035	0.029	0.057
30%	0.679	0.030	0.066	0.029	0.034	0.027	0.053
40%	0.679	0.026	0.073	0.032	0.035	0.023	0.055
50%	0.673	0.025	0.081	0.034	0.035	0.020	0.056
60%	0.671	0.024	0.086	0.035	0.037	0.018	0.055
70%	0.663	0.023	0.091	0.036	0.043	0.017	0.055
80%	0.661	0.023	0.093	0.036	0.044	0.016	0.053

15) C : Manufacturing / F : Construction / G : Whole, retail trade / H : Transfortation / J : Publishing, motion picture, broadcasting, telecommunications, and information service / K : Finance and insurance / M : Professional, scientific, and technical service

Appendix2

This appendix shows that how the risk of each portfolio variate before and after the financial crisis. The representative figure4 is illustrated in empirical result part, and the other figures are described as follows.

Figure 4. Comparison the portfolio risk before and after financial crisis

B. 10% ~ 70% of correlation pairs are remained - Parameter 1

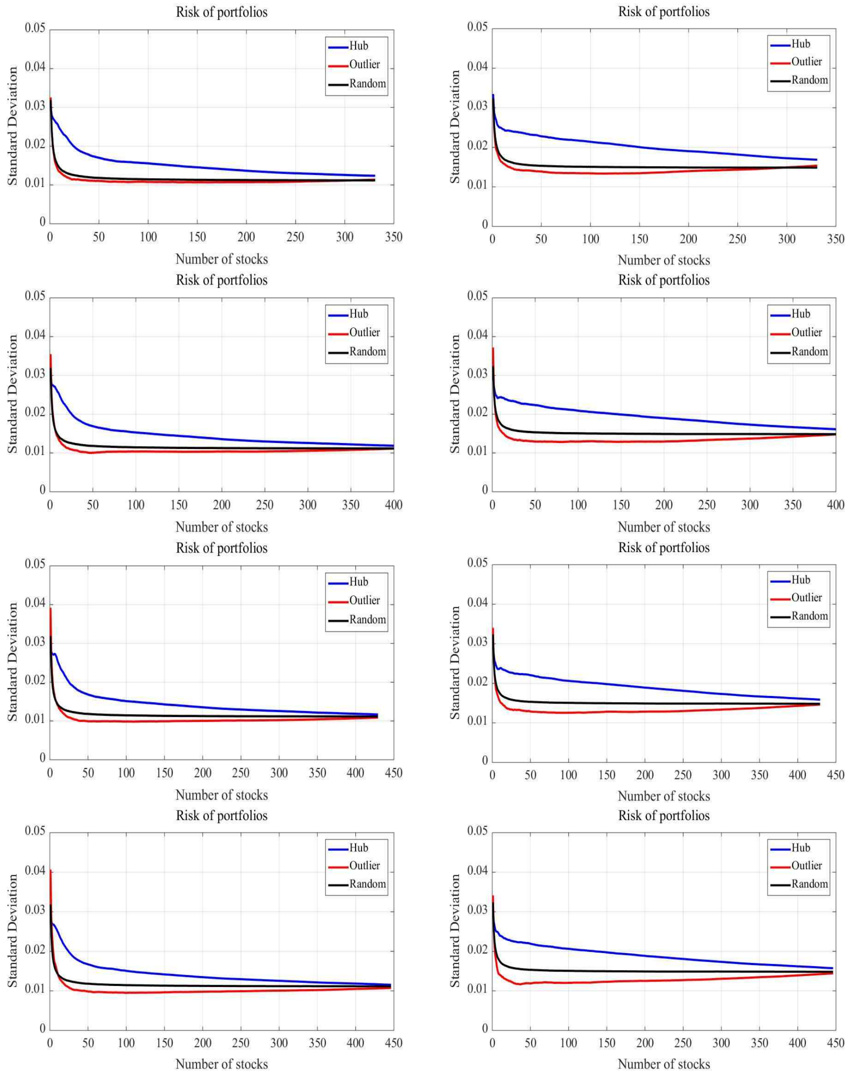
1. Before crisis (2000 ~ 2007) | 2. After crisis (2008 ~ 2012)

10%

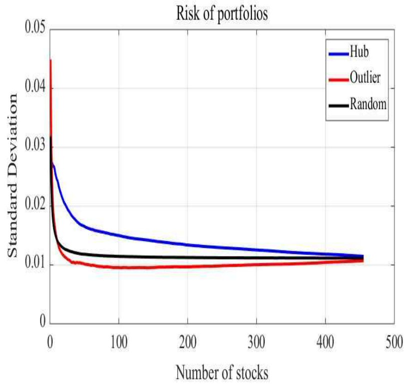
20%

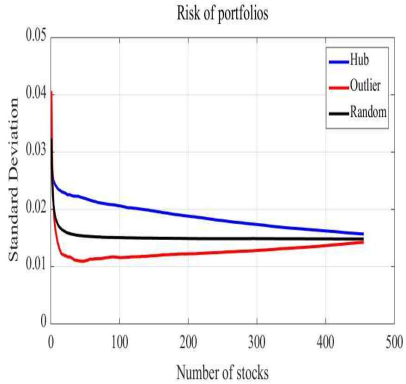
30%

40%

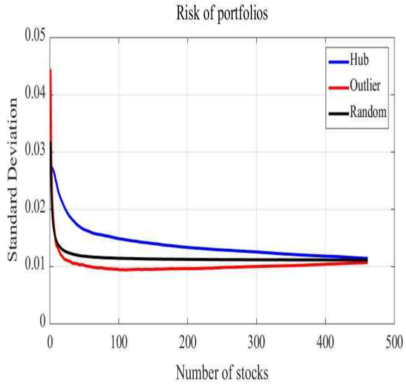


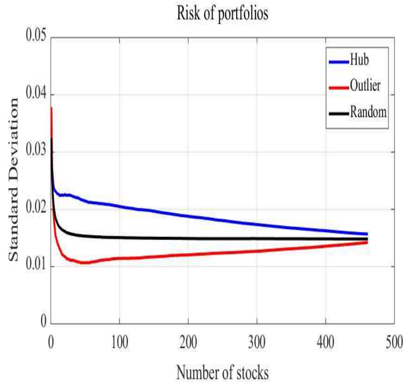
50%


Risk of portfolios
Standard Deviation
Number of stocks
Hub
Outlier
Random
0
0.01
0.02
0.03
0.04
0.05
100
200
300
400
500

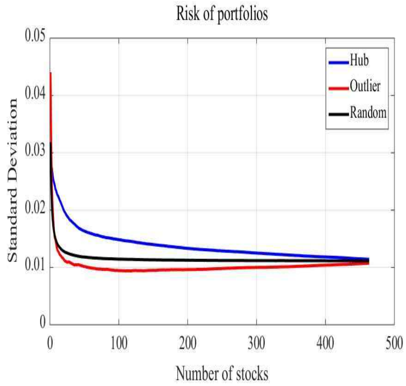

Risk of portfolios
Standard Deviation
Number of stocks
Hub
Outlier
Random
0
0.01
0.02
0.03
0.04
0.05
100
200
300
400
500

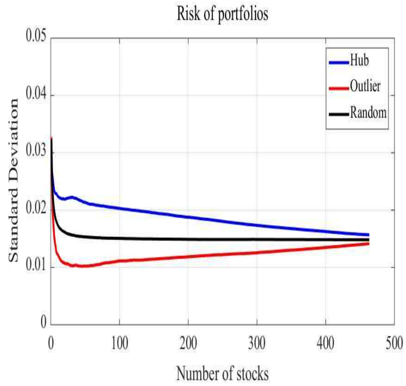
60%


Risk of portfolios
Standard Deviation
Number of stocks
Hub
Outlier
Random
0
0.01
0.02
0.03
0.04
0.05
100
200
300
400
500


Risk of portfolios
Standard Deviation
Number of stocks
Hub
Outlier
Random
0
0.01
0.02
0.03
0.04
0.05
100
200
300
400
500

70%


Risk of portfolios
Standard Deviation
Number of stocks
Hub
Outlier
Random
0
0.01
0.02
0.03
0.04
0.05
100
200
300
400
500


Risk of portfolios
Standard Deviation
Number of stocks
Hub
Outlier
Random
0
0.01
0.02
0.03
0.04
0.05
100
200
300
400
500

Appendix3

This appendix shows that how the performance of each portfolio variate before and after the financial crisis. The representative figure4 is illustrated in empirical result part, and the other figures are described as follows.

Figure 5. Comparison the portfolio performance before and after financial crisis

B. 10% ~ 70% of correlation pairs are remained - Parameter 1

1. Before crisis (2000 ~ 2007)

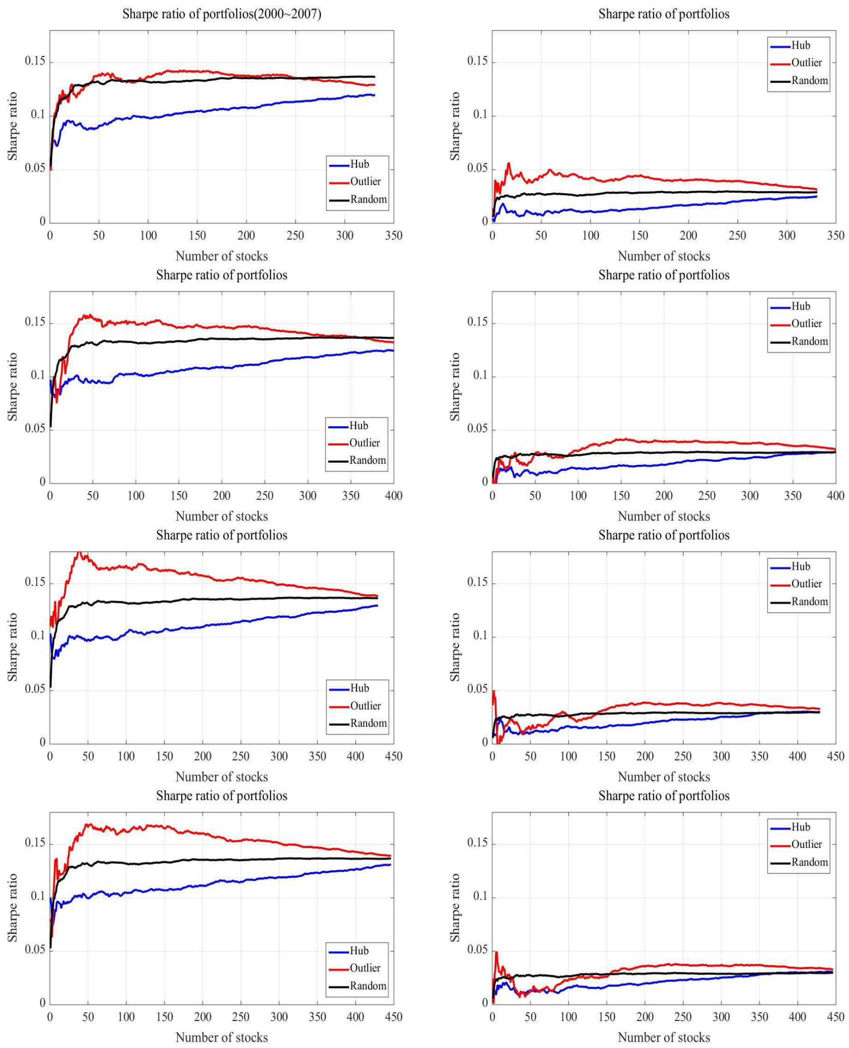
2. After crisis (2008 ~ 2012)

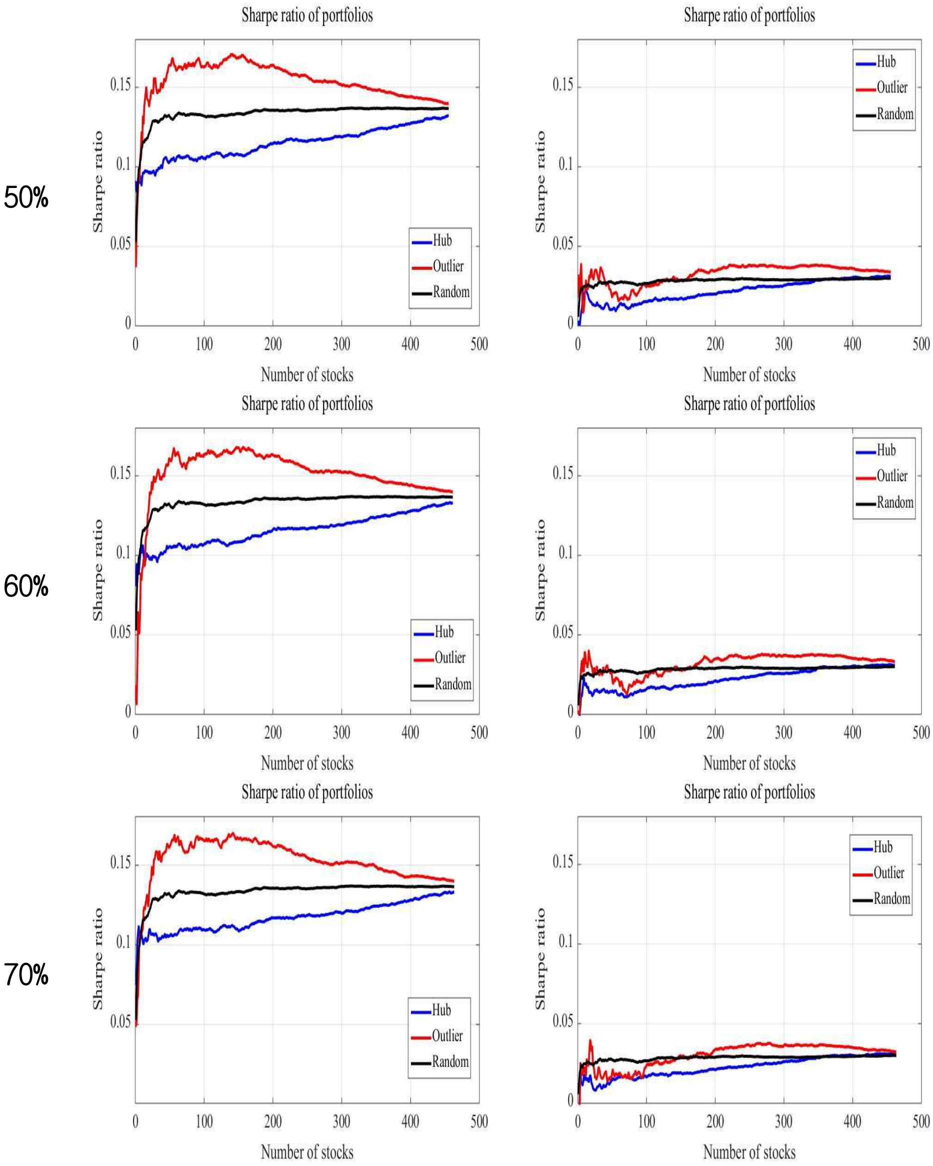
10%

20%

30%

40%




50%
60%
70%
Sharpe ratio of portfolios
Sharpe ratio
Number of stocks
Hub
Outlier
Random

Appendix4

This appendix shows the exact value of correlation between KOSPI and hub and outlier portfolio. Row-line indicates the first parameter and column line means the number of stocks (second parameter). As the number of stocks are increased, the correlation value is also larger, the the values are consistent over the first parameter.

Table 3. Correlation between KOSPI return and network-based portfolio return

A. Hub portfolio

	10%	20%	30%	40%	50%	60%	70%	80%
10	0.808	0.797	0.793	0.794	0.786	0.792	0.795	0.802
30	0.879	0.875	0.872	0.868	0.868	0.864	0.862	0.863
50	0.901	0.895	0.891	0.887	0.887	0.887	0.885	0.882
70	0.910	0.902	0.901	0.899	0.898	0.895	0.894	0.893
90	0.913	0.908	0.905	0.902	0.901	0.898	0.898	0.897
110	0.917	0.912	0.908	0.905	0.902	0.900	0.900	0.899
130	0.918	0.914	0.909	0.906	0.903	0.902	0.901	0.900
150	0.919	0.915	0.910	0.907	0.905	0.902	0.901	0.901
170	0.920	0.915	0.911	0.908	0.906	0.903	0.902	0.901
190	0.920	0.915	0.912	0.908	0.906	0.904	0.902	0.901
210	0.919	0.916	0.911	0.909	0.906	0.904	0.902	0.901
230	0.916	0.915	0.912	0.908	0.906	0.904	0.902	0.901
250	0.914	0.913	0.911	0.908	0.906	0.904	0.902	0.901
270	0.911	0.911	0.909	0.907	0.906	0.903	0.901	0.900
290	0.908	0.908	0.907	0.906	0.904	0.902	0.900	0.899
310	0.905	0.905	0.904	0.903	0.902	0.900	0.899	0.898
330	0.903	0.902	0.902	0.901	0.900	0.899	0.898	0.897
350		0.900	0.900	0.899	0.898	0.897	0.896	0.895
370		0.897	0.897	0.897	0.896	0.895	0.894	0.893
390		0.895	0.895	0.895	0.894	0.893	0.893	0.892
410			0.892	0.892	0.892	0.891	0.891	0.890
430				0.890	0.889	0.889	0.889	0.888

B. Outlier portfolio

	10%	20%	30%	40%	50%	60%	70%	80%
10	0.633	0.556	0.508	0.428	0.415	0.440	0.397	0.388
30	0.738	0.693	0.660	0.629	0.607	0.593	0.581	0.577
50	0.764	0.725	0.710	0.687	0.668	0.661	0.653	0.648
70	0.777	0.748	0.730	0.722	0.709	0.701	0.698	0.700
90	0.788	0.764	0.747	0.735	0.729	0.733	0.730	0.733
110	0.793	0.773	0.759	0.750	0.744	0.745	0.744	0.748
130	0.801	0.782	0.772	0.763	0.760	0.760	0.759	0.762
150	0.809	0.790	0.780	0.775	0.771	0.772	0.774	0.775
170	0.818	0.796	0.787	0.783	0.781	0.781	0.784	0.786
190	0.827	0.803	0.795	0.790	0.790	0.791	0.793	0.797
210	0.836	0.810	0.800	0.797	0.798	0.800	0.803	0.804
230	0.845	0.818	0.809	0.806	0.805	0.807	0.809	0.811
250	0.853	0.826	0.816	0.812	0.812	0.813	0.815	0.818
270	0.861	0.834	0.823	0.819	0.818	0.820	0.821	0.824
290	0.869	0.841	0.831	0.826	0.826	0.827	0.827	0.829
310	0.875	0.849	0.838	0.833	0.832	0.832	0.833	0.834
330	0.882	0.856	0.845	0.839	0.838	0.838	0.838	0.839
350		0.862	0.851	0.845	0.844	0.843	0.843	0.844
370		0.868	0.857	0.851	0.849	0.849	0.848	0.849
390		0.874	0.862	0.857	0.855	0.854	0.854	0.854
410			0.867	0.862	0.860	0.859	0.858	0.859
430				0.866	0.863	0.863	0.862	0.863

Appendix5

This appendix shows the exact value of standard deviation of randomly selected portfolio, hub and outlier portfolio. Row-line indicates the number of stocks (second parameter) and 8 tables are described over the first parameter. For easy distintion, the values are multiplied by one hundred. the shaded room of tables shows the value of minimum point of standard deviation. Comparably lower risk than the others can be examined from this table.

Table 4. Comparison of risk of portfolios

A. 10% of first parameter

	10	30	50	70	90	110	130	150	170	190	210	230	250	270	290	310
Hub	2.46	2.15	2.02	1.94	1.9	1.86	1.81	1.76	1.72	1.68	1.65	1.62	1.59	1.56	1.53	1.51
Outlier	1.49	1.29	1.26	1.23	1.22	1.22	1.22	1.22	1.23	1.24	1.26	1.27	1.28	1.29	1.31	1.33
Random	1.6	1.41	1.37	1.36	1.35	1.34	1.34	1.33	1.33	1.33	1.33	1.32	1.32	1.32	1.32	1.32

B. 20% of first parameter

	10	30	50	70	90	110	130	150	170	190	210	230	250	270	290	310
Hub	2.5	2.12	1.99	1.91	1.86	1.82	1.78	1.75	1.71	1.68	1.64	1.61	1.59	1.56	1.53	1.51
Outlier	1.48	1.21	1.17	1.17	1.18	1.18	1.18	1.17	1.18	1.18	1.18	1.19	1.2	1.21	1.22	1.23
Random	1.6	1.41	1.37	1.36	1.35	1.34	1.34	1.33	1.33	1.33	1.33	1.32	1.32	1.32	1.32	1.32

C. 30% of first parameter

	10	30	50	70	90	110	130	150	170	190	210	230	250	270	290	310
Hub	2.47	2.09	1.97	1.89	1.84	1.8	1.77	1.73	1.7	1.67	1.64	1.61	1.58	1.56	1.53	1.51
Outlier	1.48	1.21	1.16	1.14	1.14	1.14	1.14	1.15	1.15	1.15	1.16	1.16	1.17	1.18	1.19	1.2
Random	1.6	1.41	1.37	1.36	1.35	1.34	1.34	1.33	1.33	1.33	1.33	1.32	1.32	1.32	1.32	1.32

D. 40% of first parameter

	10	30	50	70	90	110	130	150	170	190	210	230	250	270	290	310
Hub	2.45	2.08	1.96	1.88	1.84	1.79	1.76	1.73	1.69	1.66	1.63	1.61	1.58	1.55	1.53	1.51
Outlier	1.43	1.12	1.1	1.11	1.09	1.09	1.1	1.11	1.12	1.13	1.13	1.14	1.15	1.15	1.17	1.18
Random	1.6	1.41	1.37	1.36	1.35	1.34	1.34	1.33	1.33	1.33	1.33	1.32	1.32	1.32	1.32	1.32

E. 50% of first parameter

	10	30	50	70	90	110	130	150	170	190	210	230	250	270	290	310
Hub	2.41	2.07	1.95	1.87	1.83	1.79	1.75	1.72	1.69	1.65	1.63	1.6	1.58	1.55	1.53	1.51
Outlier	1.43	1.11	1.06	1.07	1.07	1.07	1.07	1.08	1.1	1.11	1.11	1.12	1.13	1.14	1.15	1.16
Random	1.6	1.41	1.37	1.36	1.35	1.34	1.34	1.33	1.33	1.33	1.33	1.32	1.32	1.32	1.32	1.32

F. 60% of first parameter

	10	30	50	70	90	110	130	150	170	190	210	230	250	270	290	310
Hub	2.36	2.06	1.92	1.86	1.82	1.78	1.75	1.72	1.68	1.65	1.62	1.6	1.58	1.55	1.53	1.51
Outlier	1.38	1.1	1.05	1.05	1.05	1.05	1.06	1.07	1.08	1.09	1.1	1.11	1.12	1.13	1.14	1.15
Random	1.6	1.41	1.37	1.36	1.35	1.34	1.34	1.33	1.33	1.33	1.33	1.32	1.32	1.32	1.32	1.32

G. 70% of first parameter

	10	30	50	70	90	110	130	150	170	190	210	230	250	270	290	310
Hub	2.27	2.04	1.91	1.85	1.81	1.77	1.74	1.71	1.68	1.65	1.62	1.6	1.58	1.55	1.53	1.51
Outlier	1.29	1.06	1.02	1.02	1.03	1.04	1.04	1.06	1.07	1.08	1.09	1.1	1.11	1.13	1.14	1.15
Random	1.6	1.41	1.37	1.36	1.35	1.34	1.34	1.33	1.33	1.33	1.33	1.32	1.32	1.32	1.32	1.32

H. 80% of first parameter

	10	30	50	70	90	110	130	150	170	190	210	230	250	270	290	310
Hub	2.26	1.99	1.89	1.81	1.78	1.75	1.72	1.69	1.66	1.64	1.61	1.6	1.57	1.55	1.53	1.51

Appendix6(Table4 ~ Table5)

This appendix shows the exact value of Sharpe ratio of randomly selected portfolio, hub and outlier portfolio. Row-line indicates the number of stocks (second parameter) and 8 tables are described over the first parameter. For easy distintion, the values are multiplied by one hundred. Overall performance of outlier portfolio is superior to the other portfolios because of its convex pattern. Slightly lower performance is shown in minimum variance point, however, when it comes to transaction cost, it can be a optimal point of portfolio.

Table 4. Comparison of Sharpe ratio of portfolios

A. 10% of first parameter

	10	30	50	70	90	110	130	150	170	190	210	230	250	270	290	310
Hub	4.75	4.36	4.20	4.60	4.63	4.59	4.76	4.98	5.10	5.27	5.28	5.53	5.71	5.88	5.98	6.16
Outlier	7.44	7.71	8.38	8.35	8.17	8.24	8.36	8.55	8.30	8.09	8.02	7.98	7.88	7.69	7.50	7.30
Random	6.29	6.99	7.05	7.09	7.01	7.03	7.10	7.15	7.23	7.27	7.27	7.28	7.26	7.25	7.26	7.29

B. 20% of first parameter

	10	30	50	70	90	110	130	150	170	190	210	230	250	270	290	310
Hub	5.06	4.83	4.50	4.42	4.97	4.83	5.01	5.20	5.31	5.33	5.40	5.64	5.78	5.95	6.07	6.21
Outlier	5.17	7.54	8.02	7.85	7.96	8.28	8.58	8.58	8.55	8.54	8.46	8.46	8.30	8.19	8.04	8.02
Random	6.29	6.99	7.05	7.09	7.01	7.03	7.10	7.15	7.23	7.27	7.27	7.28	7.26	7.25	7.26	7.29

C. 30% of first parameter

	10	30	50	70	90	110	130	150	170	190	210	230	250	270	290	310
Hub	5.37	5.03	4.67	4.78	4.90	5.17	5.08	5.24	5.34	5.42	5.63	5.73	5.90	6.00	6.13	6.28
Outlier	5.55	8.12	8.01	7.91	8.67	8.15	8.26	8.68	8.74	8.85	8.70	8.57	8.69	8.46	8.47	8.34

D. 40% of first parameter

	10	30	50	70	90	110	130	150	170	190	210	230	250	270	290	310
Hub	5.70	5.06	4.93	4.97	4.95	5.32	5.21	5.26	5.41	5.49	5.69	5.90	5.93	6.07	6.20	6.29
Outlier	7.36	7.42	7.97	7.74	7.95	8.45	8.58	8.44	8.71	8.70	8.74	8.66	8.55	8.56	8.47	8.29
Random	6.29	6.99	7.05	7.09	7.01	7.03	7.10	7.15	7.23	7.27	7.27	7.28	7.26	7.25	7.26	7.29

E. 50% of first parameter

	10	30	50	70	90	110	130	150	170	190	210	230	250	270	290	310
Hub	5.63	4.94	4.87	5.00	5.02	5.26	5.28	5.30	5.42	5.62	5.78	5.92	6.01	6.16	6.22	6.33
Outlier	6.33	8.56	8.56	8.21	8.48	8.48	8.85	8.83	8.93	8.79	8.92	8.81	8.78	8.66	8.59	8.41
Random	6.29	6.99	7.05	7.09	7.01	7.03	7.10	7.15	7.23	7.27	7.27	7.28	7.26	7.25	7.26	7.29

F. 60% of first parameter

	10	30	50	70	90	110	130	150	170	190	210	230	250	270	290	310
Hub	6.14	5.12	5.23	4.94	5.12	5.36	5.23	5.36	5.51	5.69	5.88	5.94	6.00	6.18	6.25	6.33
Outlier	6.40	8.50	8.81	7.90	8.36	8.66	8.72	8.79	8.84	8.92	8.80	8.86	8.61	8.62	8.55	8.52
Random	6.29	6.99	7.05	7.09	7.01	7.03	7.10	7.15	7.23	7.27	7.27	7.28	7.26	7.25	7.26	7.29

G. 70% of first parameter

	10	30	50	70	90	110	130	150	170	190	210	230	250	270	290	310
Hub	6.01	5.23	5.21	5.46	5.39	5.47	5.52	5.45	5.61	5.79	5.93	5.93	6.09	6.18	6.29	6.42
Outlier	6.60	8.99	8.78	8.28	8.59	8.64	8.68	8.91	8.80	8.78	8.88	8.72	8.67	8.58	8.50	8.49
Random	6.29	6.99	7.05	7.09	7.01	7.03	7.10	7.15	7.23	7.27	7.27	7.28	7.26	7.25	7.26	7.29

H. 80% of first parameter

	10	30	50	70	90	110	130	150	170	190	210	230	250	270	290	310
Hub	5.31	5.51	5.34	5.49	5.58	5.50	5.60	5.46	5.68	5.87	6.03	6.09	6.08	6.25	6.35	6.42

Table 5. Robustness of Sharpe ratio over first parameter

		10%	20%	30%	40%	50%	60%	70%	80%
10	Hub	4.75	5.06	5.37	5.7	5.63	6.14	6.01	5.31
	Outlier	7.44	5.17	5.55	7.36	6.33	6.4	6.6	6.8
	Random	6.29	6.29	6.29	6.29	6.29	6.29	6.29	6.29
20	Hub	4.95	5.1	5.1	5.37	5.22	5.27	5.7	5.67
	Outlier	7.6	5.78	7.36	7.59	8.46	7.89	8.23	8.34
	Random	6.53	6.53	6.53	6.53	6.53	6.53	6.53	6.53
30	Hub	4.36	4.83	5.03	5.06	4.94	5.12	5.23	5.51
	Outlier	7.71	7.54	8.12	7.42	8.56	8.5	8.99	8.63
	Random	6.99	6.99	6.99	6.99	6.99	6.99	6.99	6.99
40	Hub	4.2	4.72	4.88	4.99	5.04	5.07	5.04	5.34
	Outlier	7.7	7.57	8.02	7.61	8.48	8.58	8.64	8.7
	Random	7.01	7.01	7.01	7.01	7.01	7.01	7.01	7.01
50	Hub	4.2	4.5	4.67	4.93	4.87	5.23	5.21	5.34
	Outlier	8.38	8.02	8.01	7.97	8.56	8.81	8.78	8.17
	Random	7.05	7.05	7.05	7.05	7.05	7.05	7.05	7.05
70	Hub	4.6	4.42	4.78	4.97	5	4.94	5.46	5.49
	Outlier	8.35	7.85	7.91	7.74	8.21	7.9	8.28	8.71
	Random	7.09	7.09	7.09	7.09	7.09	7.09	7.09	7.09
90	Hub	4.63	4.97	4.9	4.95	5.02	5.12	5.39	5.58
	Outlier	8.17	7.96	8.67	7.95	8.48	8.36	8.59	8.34
	Random	7.01	7.01	7.01	7.01	7.01	7.01	7.01	7.01
120	Hub	4.77	4.93	5.13	5.22	5.27	5.36	5.46	5.69
	Outlier	8.36	8.55	8.46	8.42	8.53	8.46	8.65	8.38
	Random	7.07	7.07	7.07	7.07	7.07	7.07	7.07	7.07
150	Hub	4.98	5.2	5.24	5.26	5.3	5.36	5.45	5.46
	Outlier	8.55	8.58	8.68	8.44	8.83	8.79	8.97	8.79
	Random	7.15	7.15	7.15	7.15	7.15	7.15	7.15	7.15
200	Hub	5.29	5.37	5.52	5.58	5.72	5.84	5.86	5.9
	Outlier	8.05	8.49	8.82	8.78	8.91	8.94	8.86	8.72
	Random	7.25	7.25	7.25	7.25	7.25	7.25	7.25	7.25
250	Hub	5.71	5.78	5.9	5.93	6.01	6	6.09	6.08
	Outlier	7.88	8.3	8.69	8.55	8.78	8.61	8.67	8.57
	Random	7.26	7.26	7.26	7.26	7.26	7.26	7.26	7.26
300	Hub	6.16	6.16	6.29	6.27	6.24	6.26	6.32	6.39