A network approach to optimal portfolio: Application to the Korean stock market

March 2017

Seok-won Ahn*, Yong-Choel Kim**, and Gabjin Oh*

Key words: correlation-based network, outlier and hub portfolio, diversification benefit, convexity of diversification.

JEL code: G11 - Portfolio Choice; Investment Decisions

*Division of Business Administration, Chosun University, Korea
**Lubar School of Business, University of Wisconsin-Milwaukee, Po Box 742, Milwaukee, Wisconsin 53201-0742, USA., email:ykim@uwm,edd

A network approach to optimal portfolio: Application to the Korean stock market

Abstract

Portfolios constructed from a sparse correlation-based network of stocks have rich and practical implications for stock index formulation and portfolio management. A hub portfolio consisting of approximately 20 stocks with a high degree of connections with remaining stocks shows a significantly positive correlation with the KOSPI index, and represents the total stock market with fewer stocks. An outlier portfolio containing stocks with small number of connections with other stocks shows minimum portfolio risk with about one-fifth of total number of stocks and higher risk-adjusted performance than Markowitz's random portfolio. A minimum variance portfolio (MVP) constructed with 50 outlier stocks in a sparse network of low interconnections suggests the existence of an optimal reference portfolio and the sensitivity of individual stocks to the MVP as an alternative risk measure similar to the market β risk.

Introduction

The benefits and gains from portfolio diversification based on mean variance theory (MVT) (Markowitz, (1952)[1], Sharpe (1964)[2]) and extended and adapted by numerous studies (e.g., Tobin (1958)[3], Evans and Archer (1968), Statman (1987) suggest that investors should fully diversify by including all stocks in the portfolio. However, many real-world portfolio selections and asset allocations diverge widely from the theoretical norms. Home bias (French and Porteba (1991), Huberman (1998), and others) and 1/n heuristics (e.g., Benartzi

---

[1] Markowitz (1952) proved the superiority of diversification mathematically for the first time, and proposed an optimal allocating weight concept, which is a trail blazer for the future portfolio theory.

[2] Sharpe (1964) proposed an optimal portfolio combination model and Capital Asset Pricing Model (CAPM). He classified the risk of a portfolio into idiosyncratic risk and systematic risk and considered the systematic risk to be a market risk that should be counted in the asset pricing model. He states that "…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."

[3] Tobin's separation theorem (1958) argues that there is an advantage of diversification of assets and investors are trying to do balanced investment in the portfolio based on MVT. He proposed the advantage of asset diversification 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."

and Thaler (2001) and Huberman and Jiang (2004)) are examples of practical asset allocation methods. The divergence between theoretical and practical real world portfolio diversification, optimal portfolio weighting and asset allocation are approached from many fronts.

On theoretical and statistical fronts, the Bayesian approach (Barry (1974), Bawa, Brown, and Klein (1979)) and Pastor and Stambaugh (2000))[4], the improvement of estimation and the measurement of risk and return based on the expanded asset pricing model (e.g.Fama and French (2014) suggested five-factor model)[5] are a few attempts to resolve the disparity between the MVT and empirical asset allocation and under-diversification. On another front, the insufficient diversification and sub-optimal asset allocation can be partly rationalized by real world and institutional restrictions. The transaction cost is one example that limits full diversification (Evans and Archer (1968) and Francis (1986), Meir Statman (1987))[6]. In addition, the empirical findings suggest some reconciliation of the gap between the MVT theory and practice[7] (DeMiguel, Garlappi, and Uppal (2009) and Kirby and Ostdiek (2012)). The behavioral approach is another rationale for under-diversification (e.g., Shiller (2000)). While the literature addresses the problem of mean-variance optimal weighting and under-diversification, it is an open question regarding how to reconcile the optimal MVT rule and practical investment decisions.

To address the issues, we start with the observation that movements of stock returns are inter-dependent and firms are connected with the other firms, sectors and industries. We propose a portfolio formation based on the correlation-based network and show that the "hub" portfolio (HP) exhibits economic characteristics similar to the KOSPI index portfolio and shows decreasing portfolio risk similar to random portfolios but at higher levels as the number of stocks increases. A surprising finding is the behavior of the "outlier" portfolio (OP). The OP shows minimum portfolio risk when the portfolio consists of approximately

---

[4] Barry (1974), Bawa, Brown, and Jobson Klein (1979) implemented the 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).

[5] Best and Grauer (1992) propose the method focused on reducing the estimation error of covariance matrix. In addition, moment restrictions given by factor structures of returns are exploited by MacKinlay and Pastor (2000) to reduce the estimation error.

[6] Evans and Archer (1968) and Francis (1986) observe that the benefit of the risk reduction effect is blurred when a portfolio contains approximately ten stocks. On the contrary, Meir Statman (1987) insisted that 30 or 40 stocks are needed to diversify enough by comparing the benefit of diversification with the transaction cost.

[7] DeMiguel, Garlappi, and Uppal (2009) show empirical results that the equal weight portfolio yields better performance than MVT supplementary techniques. Kirby and Ostdiek (2012) refute the results of DeMiguel, Garlappi, and Uppal (2009) using volatility timing and the reward-to-risk timing method; however, it is still an unstable estimation. These empirical works determine the weight allocation by calibrating the μ and σ. However, the results of papers are not consistent with each other as the parameter is vulnerable to the dataset and time-series and estimation error.

40-50 stocks, and the Sharpe ratio of the outlier portfolio exceeds the HP and the randomly selected portfolio with the same number of stocks.

Analyses based on correlation-based network models are investigated in the prior literature (Tse, Liu and Lau (2010), Bonanno et al. (2004), and Onnela et al. (2003), for example). Networks defined by links or connections are large and complex. For practical and meaningful analytical purposes, a filtering process is applied to reduce the complexity and size of the connection links to facilitate economic investigation.

Basically, the correlation coefficient is calculated between two stocks and is transformed to an interconnection suitable for the network concept: a high correlation pair is positioned closely in a network, and vice versa. Mantegna and Rosario (1999) used the Minimal Spanning Tree approach (MST) to filter and investigate the financial market structure. The network approach, in general, has to filter the edges, which makes the network much simpler to analyze. However, the essential information of the network structure can also become lost in the filtering processes. Although the Planar Maximally Filtered Graph (PMFG) that is proposed by Tumminello, Michel et al. (2005) was applied to account for the shortcomings of the 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. Studies show that the stocks in the same industry are connected significantly in the network because of the strong correlation between them. However, these studies are limited to analyzing the structural properties of networks.

In portfolio choices, Garas et al. (2008) analyze the New York Stock Exchange (NYSE) from the network perspective. They found that weaker links (low correlation) contribute to the overall connectivity of the network, while strong links (high correlation) are clustered according to their industrial properties. In a similar vein, Chi K. Tse et al. (2010) proposed a method to create a stock index fund based on the interconnection of stocks in a correlation-based network.[8] Their approach is different from the previous network approach in that it makes a binary decision in connecting two stocks according to the correlation coefficient being larger than threshold value when the edges of the network are established. Moreover, it was the first time that the network approach was used in portfolio construction.

We follow the approach similar to Tse et al. (2010), but our network and portfolio selection criteria are extended. We form two portfolios, hub and outlier, based on the degree, i.e., the number of connections, and find the two portfolios show different profiles in terms of portfolio risk as the number of stocks included in the portfolio increases.

The risk of both portfolios degreases as the number of companies in a portfolio increases, showing the benefits in terms of diversification. However, the gains from diversification are significantly different for the two portfolios. The most noticeable result is that the portfolios selected from outliers in the network, outlier networks formulated with stocks low degree with the rest of stocks show substantially lower risk than either the HP or

---

[8] The index fund is composed of the stocks that have strong links with many other stocks, which show high correlation with the existing stock index.

random portfolio (RP). While the risk Markowitz RP shows a monotonic decreasing pattern (Meir Statman (1987)), the standard deviation of our network based portfolio shows a convex pattern with the number of assets, i.e., the portfolio shows minimum portfolio risk at approximately 50 stocks, and then it starts to increase. The risk of the minimum variance portfolio is lower than the systematic risk calculated from the traditional capital asset pricing model (CAPM). In other words, portfolios composed of stocks that have low interconnections with others show lower portfolio risk than the systematic risk of Markowitz's portfolio. It also shows higher performance in terms of the Sharpe ratio in the range of the number of stocks that show a convex pattern. The economic implications of our findings are useful and clear. The improvement of portfolio selections is substantial when both correlation and the network structure are factored in. Further investigation of the complex network structure is a promising research venue in optimal portfolio choices and possibly in other financial studies.

In this paper, we propose a novel method in portfolio selection built from interconnection among stock returns, and we argue that this method provides valuable information in the portfolio selection process compared to the Markowitz portfolio theory. The linear correlation between firms define weighted, undirected networks, so that the degree of interconnections provides a natural and insightful measure of the network characteristics. An HP return constructed from the approximately 50 firms with higher interconnection has a significantly positive correlation with the KOSPI index. In addition, the portfolio, built from firms with a small number of connections, has positive results in the diversification and sharp ratio.

Section 2 discusses the methodologies for portfolio formation. Section 3 describes the data and the empirical results and interpretation. The further economic analysis of portfolio is in section 4. Section 5 concludes.

2. Network and portfolio construction

The complex network in a financial market is constructed by correlation. Garas et al. (2008) observe that stock clusters are formed in accordance with their industry and found that the interaction between individual companies shows a different structure in a network according to their economic role. Tse et al. (2010) propose an approach called “winner-takes-all approach” in designing a network and argue that a network-based portfolio is representative of the market index. Our approach follows a similar network construction as in Tse et al. (2010). However, we use a sparse network based on the degree (number of links) and create two portfolios, hub and outlier, based on the degree. Furthermore, we experiment with the filter criteria for network construction and an optimal portfolio[9].
Since the fully connected network is complex and is impossible to analyze meaningfully, we use methods for removing links. One is to remove links with a weaker correlation than the pre-specified set value (parameter 1). To find a connected pair of stock returns, the

[9] There are two problems in Chi K. Tse, et al (2008)’s study 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 the portfolio is the second. By varying these two parameters, the performance of the portfolio is measured.

correlation coefficients are sorted in descending order and the pairs of stock returns with weaker correlations than a pre-set threshold value are filtered out. There will be isolated stock pairs or groups (=module) that are disconnected from the other groups because the weak correlation pairs are filtered out. We consider only the stocks that 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 described in Tse et al. (2010) and Garas et al. (2008)). The threshold value is the first parameter in the construction of the stock network. The pair of stocks removed in this fashion is considered un-connected and drops out from the analysis. The resulting network is sparse, can be analyzed and provides an economically meaningful network structure.

The construction of two portfolios, a hub and an outlier portfolio, is based on the number of degrees of a particular stock (a node in a network) in a network. The “degree” is the number of links with other stocks in a sparse network, i.e., the network after removing some links. Note that the links are not directional, i.e., it includes the links emanating from and converging to the node. The stocks with a high degree are the ones that a show higher correlation than the threshold value. As such, the degree is a measure of the intensity of connections in terms of similarity, influence or explanatory power of one stock to other stocks.

As a last step of selecting a portfolio from the correlation-based network, the degree - the number of links - of stocks is sorted in descending or ascending order. The 'hub portfolio’ consists of high degree stocks sorted by descending order and consists of stocks that are influential to the other stocks. Conversely, the 'outlier portfolio'[10] consists of stocks sorted by ascending order and consists of stocks with a low degree that are remotely connected with other stocks with weak correlations.[11]. These two stock portfolios show different characteristics as they are densely or loosely connected in a network structure (Garas et al. (2008)).

3. Empirical results

3-1. Data

We use daily returns of companies listed in KOSPI for the sample period from January 04, 2000 to December 31, 2012. There are 937 stocks listed on the Korean Stock Exchange for the total sample period.

The initial daily return of stock i on day t, $r_i(t)$, is defined as

---

[10] Mostly, the stocks that have many interconnections with others located in the hub of network. That's why we called the former as a 'hub portfolio'. Additionally, the stocks that have low degree position outside of the network, which is called the 'outlier portfolio'.

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

$$r_i(t) = \frac{p_i(t) - p_i(t-1)}{p_i(t-1)}, \quad p_i(t) = stock \;\; price$$

where $p_i(t)$ is the stock price at day t.

Since the initial daily return is calculated from the price only, we adjusted the return data by the dividend. In our study, we use the daily return adjusted for the dividend payment. The statistical description of daily return is reported in appendix Table 1. The summary statistics of the return time series of firms listed in the KOSPI stock market shows time variations over the sample period of 2000-2012. While the means and the standard deviations vary over time, we observe a consistent pattern of positive skewness and kurtosis persists.

<insert Table 1. Summary description of correlation matrix>

The pairwise Pearson-correlation of daily returns of individual stocks traded in the Korean Stock Market are used as links of networks of stocks following Tse et al. (2010) who studied US data. The sample data covers 937 listed companies for 3217 business days, during the period from 2000 to 2012. We calculated the Pearson-correlation between individual companies with the historical time series of 1000 daily returns each day, and updated daily. In total, we have 2217 daily sets of correlation coefficients. To handle the delisted stocks, we exclude firms that have no return data for 10 consecutive days.

The cross correlation between stock returns, $x_i(t)$ and $x_j(t)$, is calculated as follows (Cohen et al. (2003).

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

where and $\overline{x_i}$ are means at time=t of times series over the period t=1 to t=T .

As we need to calculate correlations over rolling estimation period of 1,000 days, the number of stocks decrease to around 600 firms. Table 1 shows the pair-wise correlation over four year period from 2000 and rolling over one year. The mean of correlation coefficient ranges from 0.13 to 0.23 and has positive skewness, and high kurtosis. The maximum correlation coefficient is over 0.80 for all estimation periods with a negative minimum.

3-2. Network construction

Our empirical strategy is to construct correlation-based network first. Second, the various portfolio is formed based on the number of degrees of connections based on the close or remote connections between stock returns (Garas et al. (2008)). In this section we discuss network construction and show the characteristics of network. The empirical correlation matrix is a fully connected and weighted network. The range of average correlation coefficient for four-year estimation period is 0.13 to 0.23 from the table 1. To make the complex and fully connected network to a manageable network, we isolate links based on the correlation coefficient. Table 2 shows the number of firms and correlation threshold.

As a first step to construct network, we filter out a pair of firms with correlation coefficient is less than the threshold correlation coefficient. Second, since we are interested in the connections, not the strength of correlation, we take “winner-take-all” approach, i.e. any stocks remaining are considered to be connected. Now the relevant measure of network characteristics is not any more correlation coefficient but link or connections.

Network consists of firms with pair-wise correlation coefficient greater than the threshold value. For example, right most column in the panel A of table 2 include 358 out of 506 firms with correlation coefficient greater than 0.288. At the correlation threshold of 0.15, all 506 firms included in the network. By the “winner-take-all” approach, all remaining firms are considered to be linked. Further refinement to exclude firms with isolated islands are excluded to produce largest module, and the largest module is the basic network structure in our analysis.

<Insert Table 2>

Panel B of table 2 shows the characteristics of network biased on the connections. Average path length in network topology is the average number of steps along the shortest paths for all possible pairs of network nodes. The degree is defined as the number of links either originating from or conversing to the firm. There degree is second threshold in formulating portfolio. Some characteristics of network are typical measure of network structure: average shortest length, diameter, average degree and average clustering coefficient. Table reports the diameter and three most robust measures of network topology: average shortest length, average clustering coefficient and average degree. The diameter of the network is the longest geodesic, i.e., the longest shortest path between any two nodes in the network. A clustering coefficient is a measure of the degree to which nodes in a graph tend to cluster together.

3-3. Portfolio construction

This section discusses portfolio formation based on the network constructed with correlation threshold of 0.132, the second column form the left in table 2. We show later that portfolio based on other network structure show similar pattern. Two portfolios are constructed based on the degree, the number of connections in descend and ascending order. Hub portfolio (HP) starts with firms with the highest degree and augments with firms with the next degree. Outlier portfolio (OP) are formulated by sorting firms by ascending degree. We compare HP and OP with each other and to the third random portfolio proposed by Markowitz.

3-3-1. Comparison of Portfolio Risk

Table 3 and Figure 1show compares the standard deviations of the portfolio returns of three portfolios, the hub, outlier and random portfolios (Markowitz (1952). The random portfolio is Markowitz's randomly selected and naive portfolio, as described in DeMiguel, Garlappi, and Uppal (2009)[12]. In Figure 2, the risk (standard deviation) of the hub portfolio is considerably higher than that of the random portfolio, while the risk of the outlier portfolio (OP) is significantly lower than the random portfolio (RP), and even exhibits a convex pattern. Compared to the randomly selected portfolio, a higher risk is expected in the hub portfolio because stocks in the HP are highly correlated with other stocks, showing the least gains from diversification. Conversely, the stocks in the outlier portfolio would be superior to the other two portfolios from a diversification perspective, as it is composed of stocks that are not interconnected and are marginally or potentially negatively correlated with other stocks. Table 3 shows that the portfolio risk continuously declines as the number of firms increase in the portfolio. HP shows higher portfolio risk than RP regardless of the number stocks included in the portfolio. Surprising finding is that the portfolio risk of OP reaches the minimum at about 90 stocks, less than one-fifth of the total number of stocks include in the network, and show convex shape. Furthermore, OP exhibits lower portfolio risk than the RP irrespective of the number of stocks. The minimum portfolio risk and convex shape of OP remains intact as we use alternative network structure constructed with various correlation threshold.

<Insert table III and figure 2>

Figure 2 shows the risk of two portfolios, HP and OP, and the network-based portfolio and Markowitz's randomly selected portfolio. The graphical relation between the portfolio risk and the number of stocks of the portfolios constructed from the network filtering with the threshold correlation coefficient of 0.132. That is, pairs of stocks that have 0.132 or higher correlation are included in the construction network. The HP (OP) contains stocks starting from the highest (lowest) degree of connections. The Y-axis is the standard deviation and the X-axis is the number of stocks in respective portfolios. For easy and visual representation, we identify each portfolio by color: blue, red, and black represent hub, outlier, and randomly selected portfolios, respectively.

The risk of a portfolio decreases as the number of companies in a portfolio increases, showing the benefits of diversification. However, the gains from diversification are significantly different for the three portfolios. The risk of the RP (black line) shows a monotonic decreasing pattern with regard to the number of stocks. The standard deviation reaches close to the minimum and the most diversification benefits are obtained at approximately 30~40 stocks, consistent with Statman, Meir (1987). Naturally, the maximum diversification gains or the minimum portfolio risk is observed with the largest number of

[12] Demiguel, Garlappi, and Uppal (2009) compared the naive portfolios to others with different methodologies for estimation of μ and σ for portfolio weighting. This study observed that the naive portfolio is superior to the others in terms of the 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.

stocks at the end of the x-axis. For the HP (blue line), the portfolio risk is the highest among the three portfolios, implying the least gains from diversification among the three. The monotonic pattern of decreasing portfolio risk as the number stocks in the portfolio increases is similar to the RP, albeit at a higher level of portfolio risk.

Most surprising, an intriguing result is the convex pattern of the portfolio risk of the OP (blue line). The portfolio risk of the OP reaches the minimum with a few stocks and is even lower than the risk of the fully diversified RP, i.e., the RP including all stocks. The outlier portfolio is composed of the stocks that are positioned in the periphery of the network. These stocks are positively correlated with fewer stocks and have lower degrees of connections than others. The results have both practical and theoretical implications. The reference portfolio in the calculation of a systemic or market risk of an individual stock is the market portfolio in the standard CAPM. The diversification pattern of the OP indicates that the reference portfolio, in determining the risk of individual stocks, should be the minimum variance portfolio.

For practical investors, the minimum variance portfolio (MVP) is more appealing in terms of transaction costs and behavioral aspects. Behavioral bias and asset allocation of investors such as the 1/n rule may be compatible with network based portfolio elections and need further investigation in the future.

The same pattern of increasing diversification gains as the number of stocks increase in the portfolio persists as the network filtering threshold lowers from 70% to 10% (Appendix Figure 2-B). However, it is less transparent in the range of a higher threshold of pairwise correlation coefficients. As the threshold correlation coefficient becomes larger, many stocks remain as islands in a network, i.e., completely disconnected from other stocks, and the pool of stocks to be selected in the HP or OP is limited.

Figure 2-A shows that the portfolio risk of the OP is much lower than that of the random portfolio (RP), which is opposite the existing portfolio theory insisting that the benefit of diversification is exhausted when a portfolio contains approximately 30 stocks, as reported in Statman, Meir (1987). In addition, Table 3 in the appendix indicates the number of stocks included in the portfolios that have the minimum standard deviation of the portfolio. From the table, we can identify the detailed number of stocks and the risk value. The noticeable point of the table is that the number of stocks that have the minimum variance is very stable under the first parameter.

3-3-2. Sharpe ratio of portfolio

The ranking of the portfolio risk of the three portfolios is the HP, RP and OP, from the highest to the lowest. Next, we measure the portfolio's performance using the Sharpe ratio[13].

---

[13] 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 it can be ignored. Furthermore, this study compares the performance of each portfolio, it does not affect the results.

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

Table 4 and the Figure 3 show that the risk-adjusted performance of the outlier portfolio (OP) is much higher than that of the hub portfolio (HP) and random portfolio (RP). The rank order of the Sharpe ratio is the exact reverse of the rank order of the portfolio risk - the ranking form the highest Sharpe ratio to the lowest is OP, RP and HP. The OP shows a moderate return, but the low return is compensated by a significantly lower risk than the other portfolios and shows a convex pattern with a minimum risk with approximately 50 stocks. On the other hand, the poor performance of the HP is because the return is not high enough compared to the risk, and its risk is disproportionately higher than that of the outlier.

We find that the overall Sharpe ratio of the outlier portfolio is superior to both the HP and RP. The next question is the existence of an optimal portfolio by investigating the number of stocks in the vicinity of the convexity of the OP. Convexity in the relationship between the OP risk and the number of companies in Figure 2A and 2B means that there is a minimum point of risk. The point of the convexity of the OP is not fixed as the cutoff threshold of the minimum correlation coefficient changes. However, the Sharpe ratio of the minimum point of risk improves as the correlation threshold value (parameter 1) decreases.

In summary, we find important and strong rationale for the optimal portfolio selection with a small set of stocks, contrary to the RP that includes all stocks: an OP with a minimum portfolio risk with approximately 40-50 stocks is the optimal portfolio with the maximum diversification benefits and the highest Sharpe ratio.

3-3-3. Correlation with stock index

The hub portfolio is composed of the stocks that have high degree, i.e., the largest number of connections with others and the closest correlation with the others. As such, the total stock market can be identified by a small number of modules or clusters. To measure how the network-based-portfolio is correlated with the stock index, we used the Pearson-correlation with the KOSPI return. As in Figure 1, the hub portfolio shows very high correlation with the index even with a few stocks. The result is consistent with the Tse et al. (2010) study, which shows that their portfolio with relatively few highly connected stocks represents similar characteristics of the S&P 500 index. The result of outlier portfolio is high enough, but it is not high enough with a few stocks.

By varying the two threshold values, correlation and the degree, we constructed diverse portfolio sets with complicated correlation structures are constructed. However, the sets are too complex to explain in tables, instead, summary results are shown in the tables in the appendix, and the general picture is described in the figures with color maps in the Figure 4. Figure 4 shows the correlation of the KOSPI and the network based portfolios (HP and OP) with the correlation value at the right hand side of each graph (Figure 4). The pairwise correlation between the KOSPI against the HP and OP changes with the cutoff threshold correlation coefficient in the X-axis and the number of companies in the portfolio in the right Y-axis.

For HP, almost all of the colors are yellow, i.e., the correlation coefficients are significantly high (0.8 ~ 0.9) regardless of the parameter values 1 (correlation coefficient for network construction). The portfolio selected from the network structure explains and represents the market portfolio, consistent with the US study of Tse et al. (2010). It is not surprising, as stocks contained in the hub of the network are highly correlated with other stocks in the network.

For the OP, in Figure 1-B, the values are comparably lower than the HP (Figure 1-A). This is because the outlier portfolios are constructed with the stocks that have low degree (a few number of connections with other stocks) and have less explanatory power to the general stock market. Even so, it is still correlated with the KOSPI in the range of correlation coefficients of 0.7 ~ 0.8.

The KOSPI and other indexes, KOSPI 200, and KOSPI 100, are critically driven by the total market value of stocks and are not representative of individual stocks. On the other hand, the correlation-based network-based indexes are composed of fewer, highly connected stocks. As such, it reflects market trend better and represents the majority of stocks. Interestingly, even the OP shows a moderate to high correlation with the market indices. The results signal an alternative methodology of index fund portfolio formation and management.

4. Analysis of portfolio

The gain economic meaning of the portfolios constructed from the correlation-based network, we check the industry component and the influence of finance crisis on the characteristics of two portfolios, HP and OP, and random portfolio.

4-1. Industry sector

To further investigate the economic meaning of the risk and performance of network based portfolios (HP and OP), we examine the industry sector of the stocks included in each portfolio. For simplification, we report only the results with the number of stocks of 50 and 100[14], at which point the OP reaches the minimum point of risk and approximately healthy performance.

Manufacturing firms make up almost half of the total stocks in 16 industry sectors of the Korean stock exchange. We examine the weights of each industry in two portfolios (HP and OP) and compare them to the total stock market with a skewed and biased distribution towards manufacturing (table 5). Weight is the number of stocks in each industry sector divided by 50 or 100. Table 5 shows that in the HP, the weight of the manufacturing sector is dominant, and the finance and insurance sector comes next.

Alternatively, in the OP, the weight of the finance and insurance sector is the lowest compared with that of the HP. The contrast of the industry composition of the HP and the OP shows that more financial sector stocks are located in the hub of the network. The inclusion

---

[14] The table of the portfolio composed of 100 stocks is in Appendix 1.

or exclusion of the financial sector stock seems to be the main factor that affects the portfolio risk and the performance of the two portfolios.

To investigate how the financial sector affects the portfolio, more detailed investigation is needed. However, when there is large difference in risk and performance between the hub and outlier portfolio, it is reasonable to conclude that a significant diversification effect is not expected in a portfolio that includes financial sector stocks.

<Insert Table 5>

4-2. Comparison before and after financial crisis

The inclusion of the financial sector is an important factor in determining portfolio risk and performance. In addition, network-based portfolios are prone to instability in terms of portfolio risk and performance over time (ref). Furthermore, the sample period covers a financial crisis that has a dramatic shift of sectoral impact on the portfolio as well as an increase in systemic risk and uncertainty. In consequence, there is likely to have been a major shift in the characteristics of the network and portfolio risk and performance of portfolios before and after the financial crisis.

Figure 4-A in the appendix shows that the overall risk of the pre-crisis period (2004 to 2007) is comparably lower than that of the post-crisis period (2008 and 2012.). However, the risk diversification effect of the outlier portfolio is more prominent in the post-crisis period. Because the economic state was very unstable and had a huge uncertainty, the overall risk of the pre-crisis period was higher than that of the post-crisis period. The same pattern holds in the risk of the randomly selected portfolio.

The comparison of the Sharpe ratio shows there are significantly large gaps between the pre-crisis and post-crisis periods. Although, the risk of each period is not significantly different, the Sharpe ratio in the pre-crisis period is approximately 5 times larger than that in the post-crisis period. The post crisis economic recession and the retreat of investments from the stock market seems to be the explanation. As a result, the post-crisis return was much lower than the pre-crisis return. Although, the risk diversification effect is more significant in the post-crisis period, the overall portfolio risk is higher than in the pre-crisis period, as the distance cutoff point (parameter 1) changes (see Figure 4-B in Appendix).

In summary, a higher(lower) Sharpe ratio during the pre-crisis (post-crisis) period is a result of low (high) return and high (low) risk during the period, consistent with the general stock market boom during the pre-crisis period and the recession during the post-crisis period.

5. Conclusions

Two portfolios parsimoniously selected from sparse correlation-based network show rich, economically meaningful and practical guidance in the diversification strategy and

portfolio selection for investors and fund managers.
We investigate and contrast the risk and performance of RP and two separate portfolios (HP and OP) constructed from the correlation-based network. First, the HP contains 20 stocks of high degree (large number of connections with the remaining stocks) that are highly correlated with the KOSPI. While the KOSPI are value-weighted and market value driven, the equal-weighted and interconnected network based index is a strong alternative to the existing indexes. The HP contains a smaller number of stocks, is stable over time and is less likely to be affected by transaction costs or other institutional limitations.

The OP exhibits significantly lower risk than Markowitz's random portfolio and diversification gains show a convex pattern. The OP, the portfolio of stocks selected from the periphery in the network with low interconnections with others, shows a convex pattern of diversification, indicating a minimum point with fewer stocks. The network information in addition to the pairwise correlation is an important piece of information that needs to be investigated further. One concern of the network analysis is the complex nature. However, we show an economically rich implication of the sparse network considering only the degree – number of links – in a network constructed after removing stocks at the threshold correlation coefficient.

In summary, we propose a novel approach in portfolio selection built from interconnections among stock returns, and we show that network-based portfolio selection provides valuable information compared to the Markowitz portfolio theory. The linear correlation between firms define weighted, undirected networks, and the degree of interconnections provide a natural and parsimonious measure of the network characteristics. Furthermore, HP return, constructed from the approximately 50 firms with higher interconnection, has a significantly positive correlation with the KOSPI index. In addition, the outlier portfolio (OP), built from the firms with a small number of connection, has positive results in the diversification and Sharpe ratio. This paper is the first to show the existence of the minimum variance portfolio with smaller number of stocks, approximately one-fifth of the total number of stocks in the sample. Further investigation of alternative network structure and portfolio formation is likely to give rich and interesting insights and potential syntheses explaining the under-diversification resolving the discrepancies between theoretical and empirical investment processes.

References

Anderson, Roy M., and Robert McCredie May. Infectious diseases of humans. Vol. 1. Oxford: Oxford university press, 1991.

Barry, Christopher B. "Portfolio analysis under uncertain means, variances, and covariance." The Journal of Finance 29.2 (1974): 515-522

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).

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.

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.

Das, Sanjiv, et al. "Portfolio optimization with mental accounts." (2010): 311-334.

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.

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.

Fama, Eugene F. "Efficient capital markets: A review of theory and empirical work*." The journal of Finance 25.2 (1970): 383-417.

Francis, Ashish, et al. "Portfolio analysis." European Journal of Operational Research. 1986.

French, Kenneth R., and James M. Poterba. 1991. Investor diversification and international equity markets. American Economic Review 81: 222–26.

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.

Huberman, Gur. 1998. Familiarity breeds investment. Working paper, Columbia University School of Business.

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.

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

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.

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

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.

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

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

Paine, Robert T. The importance of species: perspectives on expendability and triage. Eds. Peter M. Kareiva, and Simon A. Levin. Princeton University Press, 2003.

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

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

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

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

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

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

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

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.

Table I: Summary description of correlation matrix.

The sample of stocks for the study consists of all common domestic stocks that were traded on the KOSPI during between January 1, 2000, and December 31. 2012. We use the stock return to estimate the correlation coefficient between companies using four years by shifting one year. There are 937 stocks in our samples. We provide summary statistics of correlation matrix for each sub-period. The average value of correlation coefficients regardless of sub-periods shows the positive value in range of 0.13 to 0.23 and has a positive skewness.

Year	# of Companies	Mean	Median	Min	Max	Standard deviation	Skewness	Kurtosis
2000 ~ 2003	619.00	0.21	0.20	-0.08	0.87	0.10	0.66	4.22
2001 ~ 2004	611.00	0.18	0.17	-0.09	0.82	0.09	0.84	4.49
2002 ~ 2005	613.00	0.14	0.13	-0.09	0.84	0.08	0.94	4.83
2003 ~ 2006	629.00	0.13	0.12	-0.10	0.87	0.08	1.03	5.72
2004 ~ 2007	630.00	0.13	0.12	-0.12	0.88	0.07	0.89	5.45
2005 ~ 2008	641.00	0.21	0.20	-0.07	0.86	0.09	0.48	3.51
2006 ~ 2009	642.00	0.22	0.22	-0.06	0.86	0.10	0.46	3.41
2007 ~ 2010	644.00	0.22	0.21	-0.08	0.85	0.10	0.52	3.58
2008 ~ 2011	649.00	0.23	0.22	-0.10	0.86	0.10	0.53	3.53
2009 ~ 2012	654.00	0.15	0.14	-0.11	0.85	0.08	1.04	5.54

Table II: Summary description of stock network

Panel A:
The sample of stocks for the study consists of all common domestic stocks that were traded on the KOSPI during between January 1, 2000, and December 31. 2012. We use the correlation coefficient (parameter 1 is defined by the ratio of the remained links to whole links in the network) to construct the stock networks. There are 937 stocks in our samples. We provide summary statistics of stock networks created with whole periods 3217 days. For each network we compute the following network characteristics: # of Companies is the number of survival stocks for whole period (2000 ~ 2012) in the KOSPI market; Threshold is correlation coefficient to construct the network from all-to-all connected network; # of Links is the number of remained connections in network; # of Companies in module is the number of stock in modules; # of Companies in largest module is the number of stocks in the largest module; # of module is the number of modules in network. The constructed networks constructed by different threshold value (correlation coefficient) are more complicated because its characteristics depend on the number of node (company) and connection (link). The summary statistics of networks for whole period is reported in Table III. In Table III, the network characteristic can be attributed to the parameter 1 of filtering. The correlation value of parameter 1 (from 90% to 10%) increases monotonically, whereas network measure such as the number of links and number of node included in modules decreases. For overall parameters, the average value of correlation coefficients and the ratio of stock to involve in the winner-take-all approach (ref) are 0.189 and 92.89%. The average ratio of links between stocks is 50.2%.

# of Companies	506	506	506	506	506	506	506	506	506
Threshold(correlation)	0.109	0.132	0.150	0.167	0.184	0.202	0.222	0.248	0.288
# of Links	230486	204932	179380	153826	128274	102720	77168	51614	26062
# of Companies in Module	506	506	506	503	495	480	459	424	358
# of Companies in Largest Module	506	506	506	503	495	480	459	422	354
# of Module	1	1	1	1	1	1	1	2	3

Panel B: Summary description of network characteristics for largest module.

The sample of stocks for the study consists of all common domestic stocks that were traded on the KOSPI during between January 1, 2000, and December 31. 2012. We use the correlation coefficient (parameter 1 is defined by the ratio of the remained links to whole links in the network) to construct the stock networks. There are 506 stocks in our samples. In Panel, we provide summary statistics of stock networks for largest module. The module is defined by the group with more than two stocks. For each network we compute the following network characteristics: # of Companies is the number of stocks in largest module; # of Links is the number of remained connections in the largest module network; Average shortest length (ASL) is the average value of shortest path between stocks in largest module network; Diameter is defined by the maximum value among shortest path lengths between stocks in largest module; Average degree is the average value of the number of connections of individual company to other companies.; Average clustering coefficient (ACC) is defined by average value of individual companies, $ACC = \sum_{i=1}^{n} CC_i$ The summary statistics of the largest module network for whole period is reported in Table IV. Table IV presents summary statistics of stock network for whole period 2000~2012. In Table IV, the network characteristic can be attributed to the parameter 1 of filtering. The correlation value of parameter 1 (from 90% to 10%) increases monotonically, whereas network measure such as the number of links and number of node included in largest modules decreases. The average shortest path and diameter of network increases as threshold value increases, whereas the average degree and average clustering coefficient decreases.

Panel : Summary statistics of network characteristics of largest module for during 2000 ~ 2012									
# of Companies	506	506	506	503	495	480	459	422	354
# of Links	230486	204932	179380	153823	128263	102694	77121	51528	25906
Average shortest length	1.100	1.200	1.300	1.398	1.491	1.595	1.703	1.833	2.075
Diameter	2	2	3	3	4	4	5	6	6
Average degree	455.506	405.004	354.506	305.811	259.117	213.946	168.020	122.104	73.181
Average clustering coefficient	0.941	0.911	0.891	0.870	0.856	0.830	0.807	0.780	0.743

Table III: Portfolio risk of three portolfios

The sample of stocks for the study consists of all common domestic stocks that were traded on the KOSPI during between January 1, 2000, and December 31. 2012. We use the correlation coefficient (parameter 1 is defined by the ratio of the remained links to whole links in the network) to construct the stock networks. There are 506 stocks in our samples. The stock networks are constructed by the connection structure between stocks. Based on the network quantity, we establish the three portfolio types, including the random portfolio (RP), hub portfolio (HP), and outlier portfolio (OP). HP (OP) is defined by the sorted correlation coefficients in descending (ascending) order. The portfolio risk of HP and RP decreases as the number of stocks in portfolio set increases, whereas the OP shows a convex pattern of diversification, indicating a minimum point with fewer stocks. Especially, the OP exhibits significantly lower risk than Markowitz's random portfolio.

Panel : Portfolio risk (%) of three types, hub portfolio (HP), outlier portfolio (OP) and random portfolio (RP)																
	Number of stocks in portfolio constructed by network using parameter 1 (80%)															
	10	30	50	70	90	110	130	150	170	190	210	230	250	270	290	310
HP	2.57	2.36	2.21	2.10	2.04	1.98	1.93	1.89	1.87	1.84	1.80	1.78	1.77	1.75	1.73	1.70
OP	1.71	1.31	1.20	1.25	1.23	1.24	1.25	1.25	1.27	1.26	1.27	1.28	1.30	1.32	1.33	1.35
RP	1.84	1.66	1.61	1.60	1.59	1.58	1.58	1.57	1.57	1.57	1.57	1.57	1.57	1.57	1.57	1.57

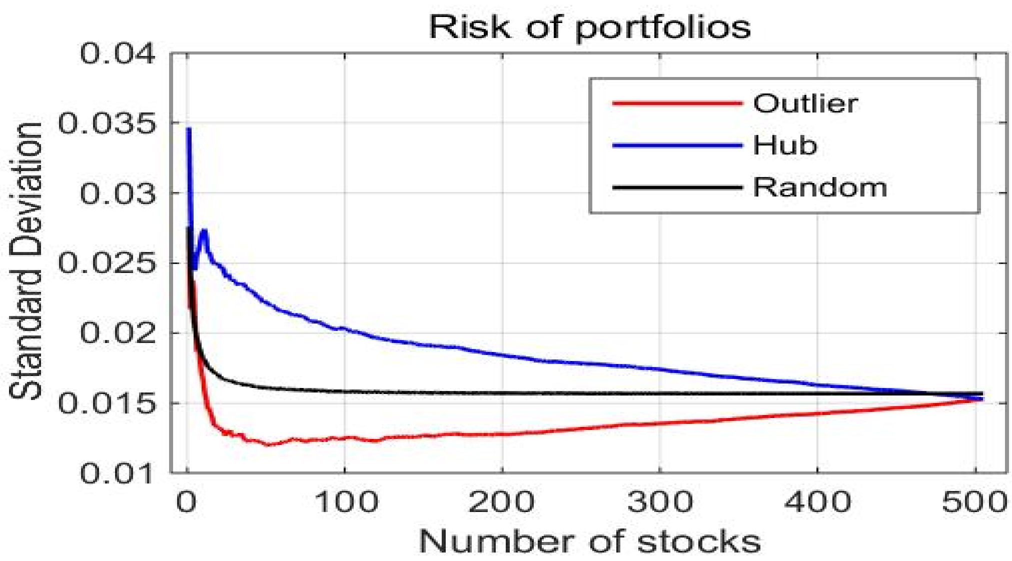


Figure II. Standard deviation of three portfolio vs. number of stocks for three portfolio types formed from price returns. The red, blue, and black solid lines correspond to the outlier, Hub, and random portfolio, respectively. This generally shows that the portfolio risk decreases as the number of stocks in portfolio increases. Especially, the outlier portfolio shows a convex pattern of diversification, indicating a minimum point with fewer stocks

Table IV : Sharp ratio of three types for sub-periods

The sample of stocks for the study consists of all common domestic stocks that were traded on the KOSPI during between January 1, 2000, and December 31. 2012. There are 937 stocks in our samples. The stock networks are constructed by the connection structure between stocks. Based on the degree value of individual companies, we establish the three portfolio types, including the random portfolio (RP), hub portfolio (HP), and outlier portfolio (OP). HP (OP) is defined by the sorted correlation coefficients in descending (ascending) order. The sharp ratio of OP around the minimum variance portfolio area shows a maximum value compared to the other portfolios (HP and RP).

Panel A : Sharp ratio (%) of three type Portfolio, hub portfolio (HP), outlier portfolio (OP) and random portfolio (RP)																
	Number of stocks in portfolio constructed by network using parameter 1 (80%)															
	10	30	50	70	90	110	130	150	170	190	210	230	250	270	290	310
HP	5.06	4.83	4.5	4.42	4.97	4.83	5.01	5.2	5.31	5.33	5.4	5.64	5.78	5.95	6.07	6.21
OP	5.17	7.54	8.02	7.85	7.96	8.28	8.58	8.58	8.55	8.54	8.46	8.46	8.3	8.19	8.04	8.02
RP	6.29	6.99	7.05	7.09	7.01	7.03	7.1	7.15	7.23	7.27	7.27	7.28	7.26	7.25	7.26	7.29

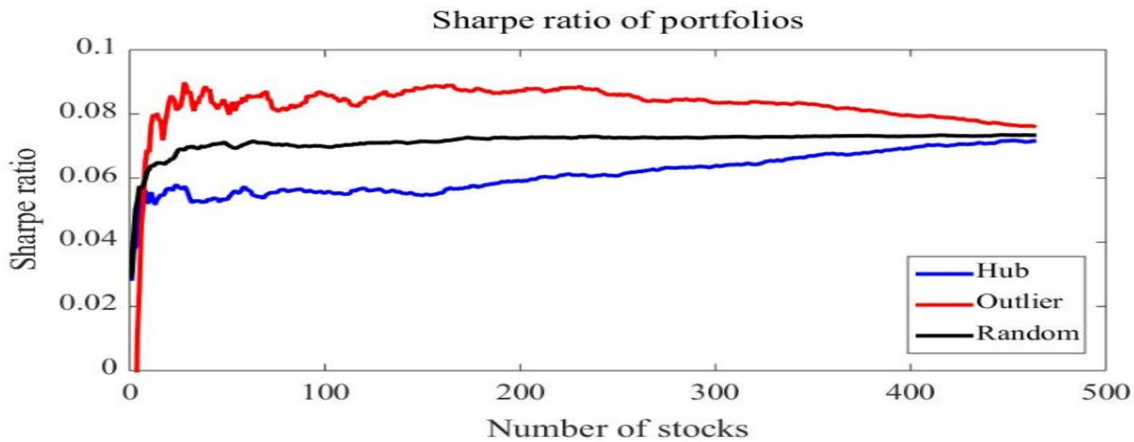
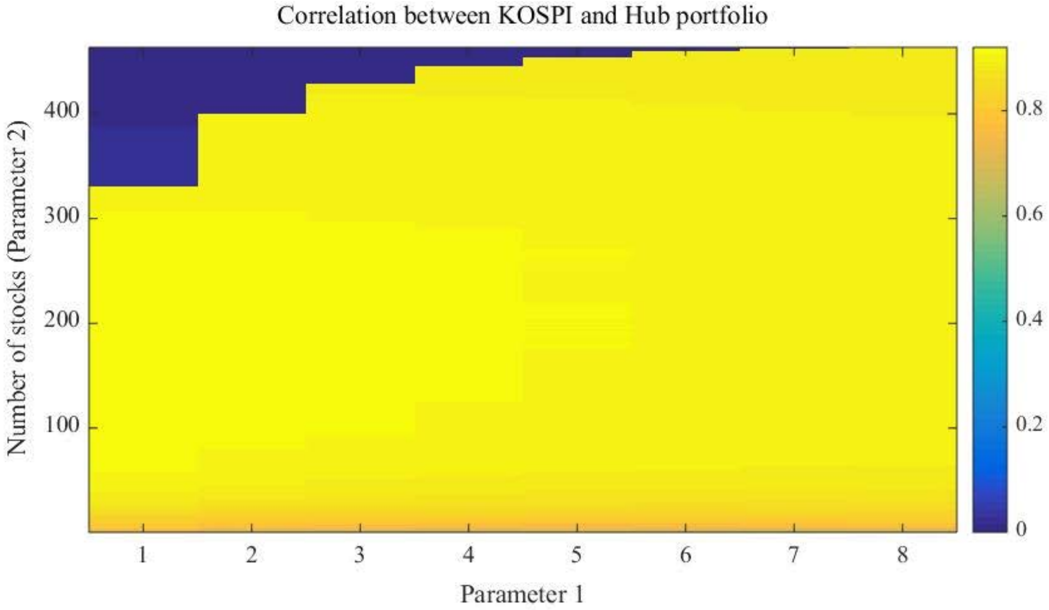


Figure 3. Sharp ratio vs. number of stocks for three portfolio types formed from stock returns. The red, blue, and black solid lines correspond to the outlier, Hub, and random portfolio, respectively. The portfolios are constructed by the connection information of network structure formed from the correlation matrix with 1000 days by shifting one day. The Sharpe ratio is measured by the excess return and standard deviation of portfolio return.

Figure 4. Correlation between KOSPI return and network-based portfolio return

A. Hub portfolio



B. Outlier portfolio

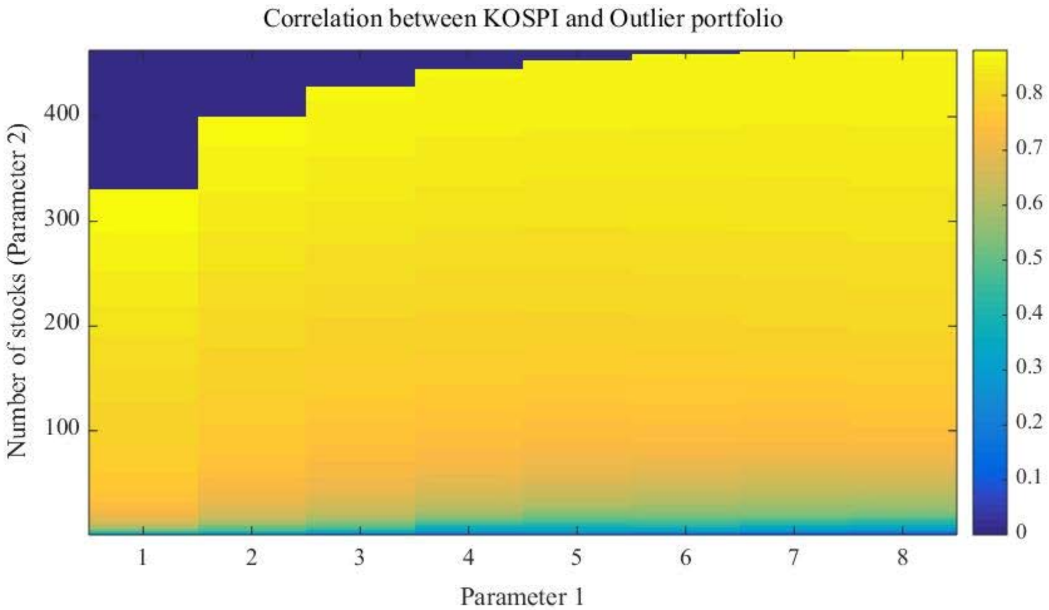


Table V. Weight of industry sector in each portfolio[11)]

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

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