Effective Portfolio Optimization Based on Random Matrix Theory*

by

Cheoljun Eom[1],Woo-Sung Jung[2,3], Taisei Kaizoji[4], and Yong H. Kim[5**], Jongwon Park[6]

* Preliminary draft. Prepared for presentation at the 16th joint conference of Korea-America Finance Association with four Korean finance associations in May 22-23, 2009. Please do not quote without the permission of authors.

1. Division of Business Administration, Pusan National University, Busan 609-735, Republic of Korea
2. Department of Physics and Basic Science Research Institute, Pohang University of Science and Technology, Pohang 790-784, Republic of Korea
3. Center for Polymer Studies and Department of Physics, Boston University, Boston, MA 02215, USA
4. Division of Social Sciences, International Christian University, Tokyo 181-8585, Japan
5. College of Business, University of Cincinnati, OH 45221, USA; **Corresponding author: yong.kim@uc.edu
6. Division of Business Administration, University of Seoul, Seoul 130-743, Republic of Korea

Effective Portfolio Optimization Based on Random Matrix Theory

Abstract

In this study, we investigate empirically whether the control of the correlation matrix via the random matrix theory (RMT) method can create a more efficient portfolio than the traditional Markowitz's model. The reasons for this improvement are also investigated. From the viewpoints of both the degree of efficiency and diversification, we find that the portfolio from the correlation matrix without the properties of the largest eigenvalue via the RMT method is more efficient than the one created from the conventional Markowitz’s model. Furthermore, we empirically confirm that the properties of the largest eigenvalue cause an increase in the value of the correlation matrix and a decrease in the degree of diversification, thus ultimately increasing the degree of portfolio risk. These results suggest that the properties of a market factor are negatively related to the degree of efficiency obtainable through the Markowitz's portfolio model. In addition, on the basis of the ex-ante test (using the expected stock returns and risk of the past period as well as actual data in the future period) we find that the performance of the observed RMT-based efficient portfolio is superior to that of the portfolio from Markowitz's model. These results demonstrate that the improvement of Markowitz's portfolio model via the control of the correlation matrix can be a source of significant practical utility.

1. Introduction

As Markowitz's (1952) seminal idea highlights, an increased emphasis on risk control in modern portfolio theory and practice characterizes the relevant finance and econometrics literature. Many studies have shown that portfolio optimization may generate substantial benefits in terms of risk reduction and performance evaluation in the investment industry (Evans and Archer, 1968; Elton and Gruber, 1977; Statman, 1987). However, the portfolio theory has been implemented with only a limited degree of success mainly because of the nature of the inputs required for portfolio theory. Specifically, the inputs of Markowitz's model applied to the portfolio theory are expected returns, expected risks, and expected correlation matrix; employing quadratic optimization function, this model should theoretically yield optimum portfolios. The problem lies in achieving accurate expectations of the three types of inputs required for this model (Elton and Gruber, 1974).

Although a great deal of attention has been paid in previous studies to the estimation of expected returns and risks, little attention has been paid thus far to the estimation of the expected correlations. That is because researchers must overcome a few problems before portfolio optimization can be practically applied. First, as the number of stocks involved in a portfolio typically counts in the hundreds or thousands, it is quite difficult to estimate every element of the correlation matrix. The correlation matrix determined by a portfolio containing N stocks has $N(N-1)/2$ elements. Therefore, in the extant literature, the simplest method of estimating expected correlation matrix using historical data has been to assume that the past correlation value is a useful estimate of its expected value in the future. Next, the characteristics of the financial time series are affected significantly by noises, outlier observations, missing data, and thin trading. If the length of the time series is not sufficiently long, measurement errors will exist in the calculated correlation matrix. This means that statistical uncertainties, such as measurement errors, will typically exist in the calculated correlation matrix. It is, therefore, necessary to control properly for these influential factors in estimating the correlation matrix for portfolio optimization.

Recently, several contributions have been made in the econophysics literature in terms of quantifying the degree of statistical uncertainty inherent to a correlation matrix. These findings have been obtained using the concepts and tools of random matrix theory (RMT) (Mehta, 1991). The RMT has been applied to financial time series as a means of removing noise in the correlation matrix and selecting statistically significant components (Laloux et al., 1999 and Plerou et al, 1999), and the effectiveness of the RMT has been verified by studies of the financial time series of a number of countries (Wilcox and Gebbie, 2004 and 2007; Garas and Argyrakis, 2007). Previous studies in which the RMT method has been applied to analyze the properties of a correlation matrix demonstrate that a considerable degree of randomness exists in the measured correlation matrix, and that the deviating eigenvalues from the random matrix remain stable over time. Additionally, several studies on the economic meaning of the properties of eigenvalues created by the RMT method indicate that the eigenvalues deviating from the random matrix have economic meanings, such as market and industrial factors (Plerou et al., 2002; Eom et al., 2008 and 2008). In particular, it has been relatively well established that the properties of the largest eigenvalue have the economic meaning of a market factor. In addition, the previous results regarding the RMT method are quite similar to the results drawn from the principal component analysis for finding the deterministic factors of stock prices (King, 1966; Meyers, 1973; Trzcinka, 1986; Brown, 1989). Based on eigenvalues, these studies also identified a number of common factors, which can commonly affect all stocks, and provided results revealing that the common factors have economic meanings---including those of market, industry, and company factors.

Several studies have explored the advantages of the RMT method in portfolio optimization

(E.g., Laloux et al., 2000; Rosenow et al., 2002; Sharifi et al., 2004; Tola et al., 2008; Eom et al., 2009). These studies have shown that the RNT-based correlation matrix is a better approximation of the portfolio elicited from the actual correlation matrix of the future period than the conventional approach using the correlation matrix of the past period as the estimated correlation matrix of the future investment period. In addition, these studies also demonstrate that the use of the RMT method improves the reliability of portfolio selection and enhances the accuracy of risk assessment, and the control of the correlation matrix provides many benefits in terms of stability.

Moreover, the interesting findings of the previous studies in which the RMT method was combined with Markowitz's model show that the portfolio constructed by the correlation matrix from the RMT method is located on the left side of those created by other correlation matrices. Recently, the study of Eom *et al*.(2009) generates empirical evidence that the risk of a portfolio elicited from the correlation matrix via the RMT method was smaller than that of a portfolio drawn from the conventional Markowitz model. Therefore, these findings have been recognized as a meaningful contribution, not only to the academics of portfolio theory, but also to the practices of the investment industry. However, sufficient efforts have yet to be exerted to explain the observed results systematically and to evaluate whether this approach is practically useful.

The principal objective of this study is first to empirically investigate whether controlling the correlation via the RMT method can generate a more efficient portfolio than one determined via the conventional Markowitz's portfolio theory at a given identical return. Furthermore, if a more efficient portfolio than one determined via the conventional Markowitz's model exists, we attempt to systematically provide possible explanations as to why a created portfolio is more efficient, on the basis of empirical evidence. Finally, in order to ascertain the practical utility of the observed more efficient portfolio, we employ an ex-ante test to confirm empirically the changes in the performance of a more efficient portfolio in terms of gradual improvements in the predictability of the expected return and expected risk (standard deviation).

The observed results are summarized as follows. We find robustly the existence of a more efficient portfolio with a lower risk than the portfolio determined by Markowitz's model from the viewpoint of efficiency. In particular, the portfolio created by the group correlation matrix, which has the properties of the eigenvalues except those of the largest eigenvalue among the eigenvalues deviating from the random matrix, is more efficient than those generated from other correlation matrices via the RMT method. Furthermore, using the group correlation matrix, the RMT method also proved to be an effective tool for improving the degree of diversification for investment weights among stocks in a portfolio---which has been a practical limitation to the application of conventional Markowitz portfolio theory. In addition, we show that the correlation matrix is affected directly by whether the property of the largest eigenvalue was reflected. The property of the largest eigenvalue causes an increase in the value of the correlation matrix, and thus a lower degree of diversification, which ultimately increases the degree of portfolio risk. Moreover, by considering the practical utility on the basis of an ex-ante test, we show that the performance index of the portfolio elicited from the group correlation matrix is higher than that of the portfolio created by the conventional Markowitz's model, as well as the equal-weighted method using the actual return and risk of a future period, by somewhat improving the predictability of return and risk of stocks in the past period. These results constitute empirical evidence that the group correlation matrix constructed by the RMT method has a high degree of utility for portfolio investments that are based on the Markowitz model.

The remainder of this paper is structured as follows. Section II presents a brief description of the analyzed data and describes the methods employed in this study. Following the stated research objectives, Section III provides the results of effective portfolio optimization from the conventional Markowitz portfolio theory and the RMT method. Section IV presents the results of an ex-ante test of a more efficient portfolio. Finally, we summarize the findings and conclusions of this study.

2. Data and Methods

2.1 Data

This study considers an efficient portfolio of individual stocks traded on the stock market, and thus we utilize the daily prices of stocks included in the market indices of Korea and the US. The following three steps are considered for the selection of stocks from each country. First, stocks with consecutive daily prices for the 18 years from January 1990 to December 2007 were selected. Second, stocks with outliers in the descriptive statistics of skewness (>|2|) and kurtosis (>30) were excluded. Third, stocks in sectors with four or less companies were excluded. The data selected according to these processes included 104 stocks from the KOSPI 200 of the Korean stock market, and 310 stocks from the S&P 500 of the US stock market. The stock returns, $R_{j,t}$, were calculated by the logarithmic changes of the prices, $R_{j,t} = lnP_{j,t} - lnP_{j,t-1}$, in which $P_{j,t}$ is the stock price on day t.

The period of this study is divided into the overall period and sub-periods. First, the final year (2007) of the entire period was established as the future period for the application of the estimated correlation matrix to the optimization function, and the past period (January 1990 ~ December 2006) was established as the estimation period for the correlation matrix. Second, the entire period was divided into three equal sub-periods to consider the effects of the changes in market status. The periods were defined with an identical method applied to the overall period of six years (January 1990 ~ December 1995, January 1996 ~ December 2001, January 2002 ~

December 2007). That is, the future period applied to the estimated correlation matrix is the final year (1995, 2001 and 2007), and the past period for the estimation of the correlation matrix is the previous five years (January 1990 ~ December 1994, January 1996 ~ December 2000, January 2002 ~ December 2006).

2.2 Random Matrix Theory Method

The RMT method is introduced as a means for controlling the correlation matrix with measurement errors in the financial time series. By the statistical properties of the correlation matrix created by the random interactions, if the length of the time series (L) and the number of data (N) are infinite, the eigenvalue(λ) probability density function of the correlation matrix, $P_{RM}(\lambda)$, is defined by Sengupta and Mitra (1999)

$$P_{RM}(\lambda) = \frac{Q}{2\pi} \frac{\sqrt{(\lambda_{+}^{RM} - \lambda)(\lambda - \lambda_{-}^{RM})}}{\lambda} \tag{1}$$

$$[\ \lambda_{\pm}^{RM} = 1 + \tfrac{1}{Q} \pm 2\sqrt{\tfrac{1}{Q}},\ Q \equiv \tfrac{L}{N} > 1\ ]$$

and the range of eigenvalues belonging to the random matrix are $\lambda_{-}^{RM} \le \lambda_i \le \lambda_{+}^{RM}$, in which λ_{+}^{RM} is the maximum eigenvalue, and λ_{-}^{RM} is the minimum eigenvalue.

As in previous studies, we categorized the range into filtered, market, and group ranges on the basis of the eigenvalue range of the random matrix. First, the range at which the eigenvalue exceeds the maximum eigenvalue of the random matrix is defined as the filtered range, $\lambda_i > \lambda_{+}^{RM}$ (using script F, λ^F); second, the largest eigenvalue among those that exceed the maximum eigenvalue of the random matrix as the market range, $\lambda_1 \equiv \lambda^{Largest}$ (script M, λ^M); and third, the eigenvalue range except for the largest eigenvalue among those that exceed the maximum eigenvalue of the random matrix as the group range, $\lambda_i > \lambda_{+}^{RM}$, except for λ_1

(script G, λ^G). It has been relatively well established that the eigenvalues deviated from the random matrix have economic meanings such as market, industry, and company factors; and in particular, the properties of the largest eigenvalue give the economic meaning of a market factor. However, it was verified in previous studies have shown that economic meanings cannot be attached to the properties of the eigenvalue belonging to the range of the random matrix (See, e.g., Plerou et al., 2002; Eom et al., 2008 and 2008; King, 1966; Meyers, 1973; Trzcinka, 1986; Brown, 1989). Therefore, we exclude the random matrix from this study because it does not correspond with the primary objective of our study, which is to provide systematic explanations on the basis of empirical evidence.

2.3 Markowitz's Optimization Model

In the conventional portfolio theory, the optimization function for calculating the investment weight to minimize the risk at a given identical return can be expressed as is shown in Eq. 2. If an investment asset's expected return, $E(R_j)$, expected risk, σ_j, and correlation matrix, $\rho_{i,j}$, are provided as inputs to the optimization function, the investment weight, w_j, of each stock is yielded as output. The portfolio return, $E(R_p)$ and risk, σ_p are then numerically calculated using the investment weight. Accordingly, the Markowitz's model involves the determination of investment weights in order to create an efficient portfolio on the basis of the dominance principle as an investment rule---that we prefer less risk for a given return and prefer greater return for a given risk.

$$\sigma_p = \sqrt{\sum_i \sum_j w_i^p w_j^p \sigma_{i,j}} \quad (\sigma_{i,j} = \sigma_i \cdot \sigma_j \cdot \rho_{i,j}) \qquad (2)$$

Condition 1: $E(R_p) = \sum_j w_j^p E(R_j) = R_p^T \;\; [\, p = 1, 2, \ldots, 50 \,]$

Condition 2: $\sum_j w_j^p \equiv 1.0$

Condition 3: $w_j^p \geq 0.0$ $[\ j = 1, 2, \ldots, N\ ]$

As mentioned previously, our objective is to empirically assess and explain the existence of a more efficient portfolio using a variety of correlation matrices, which is an input to the optimization function shown in Eq. 2. Therefore, as in the empirical designs of previous studies, it can be assumed that the expected return, $E(R_j)$ and the expected risk, σ_j of individual stocks are known, and that only the correlation matrix from past periods is estimated (Laloux et al., 2000; Rosenow et al., 2002; Sharifi et al., 2004; Tola et al., 2008; Eom et al., 2009). The difference with the designs of previous studies is that we tested the portfolios under the assumption of the dominance principle. That is, every portfolio created by the correlation matrix compares the magnitudes of risk at an identical target return, R_p^T (condition 1 in Eq. 2). Then, a portfolio is generated with a minimum risk, σ_p, according to the minimization objective function for the identical target return, and the investment weight, w_j^p, for stocks in a portfolio are generated in the process. Finally, connecting the combination points $[\sigma_p,\ E(R_p)]$ of portfolio risks and returns created by varying the target return within a range, $p = 1, 2, ..., 50$, provides an efficient investment curve as a set of optimum portfolios. In turn, because every portfolio drawn has an identical return, a portfolio with lower risk becomes a more efficient portfolio. Conditions 2 and 3 show that short-selling is not permitted.

Five types of correlation matrices were applied to the optimization function of Eq. 2 in this study. The first of these were the correlation matrices using the original returns--the correlation matrix of the past period, $\rho^H \equiv C^H$ (using script H) used in the conventional portfolio theory, and the correlation matrix, $\rho^A \equiv C^A$ (script A) calculated using the actual data of the future period. The second type was the correlation matrix, C^H, measured during the past period and then decomposed into three types of correlation matrices via the RMT method. The third correlation matrix type applied to the optimization function of Eq. 2 was the type based on the eigenvalue range of the random matrix, which is a filter correlation matrix reflecting the properties of the eigenvalue that exceeds the eigenvalue range of the random matrix, $\rho(\lambda^F) \equiv C^F$. The fourth of these was the market correlation matrix that reflects only the properties of the largest eigenvalue, $\rho(\lambda^M) \equiv C^M$. Finally, the fifth type was the group correlation matrix reflecting the properties of the eigenvalues except for the properties of the largest eigenvalue, $\rho(\lambda^G) \equiv C^G$.

As mentioned above, we utilized three inputs of the optimization function: the first, the return, $E(R_j)$; the second, the risk, σ_j of individual stocks calculated using the actual data of the future period; and the third, the correlation matrix C^A in the future period, and the four types of correlation matrices, C^H, C^F, C^M, C^G, estimated in the past period. We then conducted a comparison of the five types of portfolio risks, σ_p^A, σ_p^H, σ_p^F, σ_p^M, σ_p^G, created under the given identical returns.

3. Results for Effective Portfolio Optimization

3.1 Existence of a More Efficient Portfolio via the RMT

This section presents empirical evidence as to whether a more efficient portfolio with a lower risk at a given identical return can be generated via a combination of the RMT method with

Markowitz's model. To ensure the robustness of the observed results, we utilized measurements from two perspectives---the degree of efficiency and the degree of diversification. The former involves determining whether the portfolio has a lower risk for a given identical return, and the latter is concerned with whether the investment weights among stocks in a portfolio are well distributed. Therefore, a more efficient portfolio should be satisfying from both the viewpoint of the degree of efficiency and that of the degree of diversification.

First, we present the observed results from the viewpoint of the degree of efficiency in Fig. 1 and Table 1. The actual stock returns and risks during the future period (2007), the four correlation matrices, C^H, C^F, C^M, C^G estimated in the past period (1990 ~ 2006) and the actual correlation matrix of the future period are provided as inputs for the optimization function of Eq. 2. In Fig. 1, the results are divided into those obtained from the Korean and US stock markets, as shown in Fig. 1(a) and Fig. 1(b), respectively. The X-axis denotes the portfolio risks, and the Y-axis indicates the portfolio returns as the identical target returns. In the figure, black circles indicate the portfolio created using C^A as inputs into Markowitz's model, C^H as cyan pentagrams, C^F as blue circles, C^M as green tangles, and C^G as red squares. According to our results, regardless of the country data, the portfolio created from C^G, which reflects the properties of the eigenvalues except for those of the largest eigenvalue among eigenvalues deviating from the random matrix, is in the left-most position. That is, the portfolio from C^G harbors a lower risk at a given identical return than those created from other correlation matrices. In order to obtain robust results in Fig. 1, we utilized an identical process for the stock of each country using three sub-periods (1990 ~ 1995, 1996 ~ 2001 and 2002 ~ 2007). These results are provided in Table 1. We utilized the average value of portfolio risks, $\overline{\sigma_p} = \frac{1}{50}\sum_{p=1}^{50}\sigma_p$. In Table 1, we also show that the risks, $\overline{\sigma_p^G}$, of the portfolio created from C^G are clearly smaller than those of the portfolios created from other correlation matrices in all sub-periods. Therefore, the findings can be said to reveal that the portfolio created using C^G is clearly more efficient than the ones created from other correlation matrices from the viewpoint of degree of efficiency as measured by the size of the portfolio risk.

Next, we provide the observed results from the viewpoint of degree of diversification in Fig. 2 and Table 2. In order to conduct a quantitative observation of how well the portfolio was diversified, we utilized two measurements of the intra-portfolio correlation, IPC_p, and the concentration coefficient, CC_p using the investment weights w_j^p calculated from the previous results of Fig. 1 and Table 1, as follows.

$$IPC_p = \frac{\sum_i^N \sum_j^N w_i^p w_j^p \rho_{i,j}}{\sum_i^N \sum_j^N w_i^p w_j^p} \quad [\ p = 1, 2, \ldots, 50\ ] \tag{3}$$

$$CC_p = (\sum_{j=1}^{N} (w_j^p)^2)^{-1} \tag{4}$$

The first measurement, IPC_p, quantifies the distribution level of the investment weights among stocks in a portfolio. The range of the IPC is $-1 \leq IPC \leq 1$. $IPC = -1$ denotes that the investment weights are well distributed among all of the stocks in the portfolio, and $IPC = +1$ shows that the investment weights are not distributed at all. Namely, the lower the IPC is, the higher the level of distribution for the investment weights among stocks in a portfolio will be. The second measurement, CC_p, as a complementary parameter to IPC_p, quantifies the degree of concentration of investment weights among stocks in a portfolio. The range of the CC is $1 \leq CC \leq N$. $CC = 1$ denotes that 100% of the investment is made on one stock in a portfolio, and $C = N$ means that an equal investment weight $w_j = \frac{1}{N} = w$ is applied to every stock in a portfolio. In other words, the higher the CC is, the lower the degree of concentration of investment weights among stocks in a portfolio---that is, the higher the level of distribution of investment weights---will be.

The correlation matrix is composed of the five types of correlation matrices, C^A, C^H, C^F,

C^M, C^G used in Fig. 1 and Table 1, and the investment weight, w_j^p is calculated from Eq. 2 using the five types of correlation matrices as an input of the optimization function, respectively. The observed results are provided in Fig. 2. Fig. 2(a) & (b) show the results using the Korean stocks, and Fig. 2(c) & (d) show the US stocks. The X-axis indicates the 50 portfolio cases, and the Y-axis presents IPC_p [Fig. 2(a) & (c)] and CC_p [Fig. 2(b) & (d)]. The figure is categorized as follows: black circles C^A, cyan pentagrams C^H, blue circles C^F, green tangles C^M and red squares C^G. The results illustrate that the IPC_p of the portfolio generated from C^G has the smallest value and its CC_p is the greatest, regardless of the country data. That is, the investment weights among stocks in a portfolio generated from C^G are not concentrated on a specific stock, and are well diversified among all stocks in a portfolio. In addition, the IPC_p (the CC_p) has a higher (lower) value as risk increases, $p:1 => p:50$. This means that the lower position among portfolios is better diversified than the portfolio in the higher position. These can be inferred in cases in which in order to achieve a higher return, we increase the investment weight of a specific stock with high return, and then these increase the risk of the portfolio, and such a distribution of investment weights results in a low degree of diversification. In order to achieve robust results in Fig. 2, we test identical processes for the stock of each country using three sub-periods. We present the results in Table 2. We utilized the average value, $\overline{IPC_p} = \frac{1}{50}\sum_{p=1}^{50} IPC_p$ and $\overline{CC_p} = \frac{1}{50}\sum_{p=1}^{50} CC_p$ for each sub-period. According to our results, the $\overline{IPC}$ (the $\overline{CC}$) of the portfolio created from C^G is smaller (larger) than the $\overline{IPC}s$ (the $\overline{CC}s$) from other correlation matrices. These results robustly confirmed that the investment weights among stocks in a portfolio created from C^G are well diversified.

Besides, we have been additively considering the number of the stocks with non-zero investment weights, $w_j^p > 0$, in an effort to robustly examine the degree of diversification of the stocks in a portfolio. The results obtained with the Korean and the US stocks are provided in

Fig. 3(a) and Fig. 3(b), respectively. In the figure, the Y-axis represents the ratio, $FR_k = \frac{fq_k}{N}$, of the number of stocks having non-zero weights, fq_k, among the overall stocks, N in each 50 portfolios. For example, the '1 (=k)' on the left of the X-axis represents the position of the number of stocks having non-zero investment weights in one among the 50 portfolios. Otherwise, the '50 (=k)' on the right indicates the position of the number of stocks having non-zero investment weights among all 50 of the portfolios. Therefore, the position of the number of stocks having no weight among all 50 portfolios is the '0 (=k)' on the left-most of the X-axis. The X-axis indicates the number of cases of stocks with non-zero weights, $w_j^p > 0$, among the 50 portfolios in Fig. 1, and accordingly the range of the X-axis is $0 \leq X(=k) \leq 50$. In Fig. 3, the bar indicates the FR_k for the portfolios created from the actual correlation matrix C^A of the future period, and the figure is categorized as follows: black circles C^A, cyan pentagrams C^H, blue circles C^F, green tangles C^M and red squares C^G.

According to our results, the $FR_{k=0}$=69% (89%) of the stocks in a portfolio generated from C^A for the Korean (not the US) data does not have investment weights for all 50 of the portfolios. In other words, only 31% (11%) of the stocks have a non-zero investment weight. Additionally, the cases using the other correlation matrices, C^H, C^F, and C^M, were not appreciably different from the results of C^A, with $FR_{k=0}$=70% (80%) on '0' for the Korean (the US) data. However, interestingly, there is no stock with $w_j^p = 0$ in the portfolio generated from the C^G for the Korean data [$FR_{k=0}$=0%]. That means, the degree of diversification has taken place for every stock. For the US stock market, the ratio of the stocks with $w_j^p = 0$ was significantly lower (approximately $FR_{k=0}$=5%). These results indicated that the group correlation matrix generated via the RMT method, as a major input of Markowitz's model, is an effective means of improving the degree of diversification for investment weights among stocks in a portfolio, which has previously been a practical limitation in the application of conventional portfolio theory.

For the reasons mentioned above, we robustly detected the existence of a more efficient portfolio with a lower risk than the portfolio determined by Markowitz's portfolio theory. Furthermore, from the perspectives of the degrees of efficiency and diversification, the portfolio created from the group correlation matrix is more efficient. Additionally, we note that even the use of a correlation matrix that reflects only a few of the properties of the eigenvalues deviating from the random matrix generates a more efficient portfolio than the portfolios generated from the conventional Markowitz portfolio theory.

3.2 Reasons for Existence of a More Efficient Portfolio

This section provides empirical explanations as to why the portfolio generated from C^G is more efficient than those generated from other correlation matrices. From the results presented in the previous section, factors that may possibly affect the existence of a more efficient portfolio are the correlation matrix and the investment weight, because all of the other conditions were identically applied. That is, the correlation matrix is considered an input of Markowitz's model, and the investment weight was used as an output. Thus, we attempted to determine the reasons that those two influential factors helped to generate a more efficient portfolio.

First of all, we present the results obtained from the viewpoint of investment weight in Fig. 4. In order to assess the effects of the investment weight on a more efficient portfolio, we elected to assign a identical investment weight, $w_i = \frac{1}{N} = w_j$, to all of the stocks within a portfolio. This is because the different investment weights of stocks calculated from Eq. 2 were considered in the previous section. If a C^G assigned with an identical investment weight still has the lowest portfolio risk as compared to the other correlation matrices, the investment weight cannot be considered to be affecting the existence of a more efficient portfolio. Specifically, we apply the process in the previous studies (Evans and Archer, 1968; Elton and Gruber, 1977; Statman, 1987) to analyze the effects of portfolio diversification. That is, we examine the reduction of a portfolio's risk, $\sigma_p = \sqrt{\sum_i^M \sum_j^M w_i w_j \sigma_i \sigma_j \rho_{i,j}}$ $(w_i = \frac{1}{N} = w_j)$ is reduced as the number of stocks within a portfolio, M, increases. This is similar to the objective function of Eq. 2. The only difference is that the number of stocks in a portfolio, M, varies, whereas the number of stocks in a portfolio in Eq. 2 was fixed at N. Additionally, we utilized a varying number of M stocks in a portfolio, $25 \leq M \leq 70$ $[M < N]$. In addition, in order to minimize the effects of selection bias, we conducted a simulation of 100 repetitions for each 46(=70-25+1) case, changing the number of stocks within a portfolio, respectively. Importantly, the types of stocks in a portfolio were not identical for each of the 100 simulations. The minimum number of stocks was set at $M = 25$ due to the use of the correlation matrix generated via the RMT method.

The results are provided in Fig. 4. The diversification effects of the portfolio were assessed by applying the four types of correlation matrices, C^H, C^F, C^M, and C^G, estimated in the past period (1990~2006) to the stocks in the future period (2007). The reference of comparison was the result from the actual correlation matrix, C^A, of the future period. In addition, as the C^A of the future period did not adopt the RMT method, we applied a wider range, $1 \leq M \leq 70$, to also confirm the original diversification effects of a well-known portfolio in finance (Evans and Archer, 1968). The X-axis represents the number of stocks within a portfolio in the range of $1 \leq M \leq 70$, in which $M = 1$, the left-most position of the X-axis, corresponds to an individual stock. The Y-axis is the portfolio risks. In accordance with the empirical design, we conducted 100 simulations for each portfolio with a specific number of stocks, and thus, we present the results with an error-bar graph using the average and standard deviations of the 100 risks for each of the portfolios. The figure is divided into black circles of C^A, cyan pentagrams of C^H, blue circles of C^F, green tangles of C^M and red squares of C^G, according to the correlation matrices. Fig. 4(a) is the results attained with the Korean stocks, and Fig. 4(b) for the US stocks.

As expected, the results show that increasing the number of stocks within a portfolio exponentially reduces the risk of the portfolio. Moreover, as the number of M stocks in a portfolio increases, the portfolio created from C^G evidenced lower risk than those created from other correlation matrices, regardless of the country data used. That is, as the risk of the portfolio using C^G was smallest despite the assignation of identical investment weights to all stocks within a portfolio created from other correlation matrices, it is difficult to claim that the investment weight is a deterministic influential factor in whether or not a more efficient portfolio exists. It is believed that the investment weight performs a pivotal role in providing better diversification for a more efficient portfolio, rather than having effects of a deterministic factor on the existence of a more efficient portfolio.

Next, we present the results observed from the viewpoint of the correlation matrix on the existence of a more efficient portfolio in Fig. 5. We assessed the probability density function of frequency for the five correlation matrices used in the previous process, C^A, C^H, C^F, C^M, and C^G. The results are provided in Fig. 5. In the figure, the results indicate the probability distributions of five correlation matrices. In Figs. 5 (a) & (b), the yellow bar is the probability distribution of C^A in the future period, and the probability distributions of C^H in the past period are indicated with cyan pentagrams, C^F with blue circles, C^M with green tangles, and C^G with red squares. Fig. 5(a) shows the results of the Korean stocks, and Fig. 5(b) shows the results for the US stocks. According to the results, the center of the probability distribution of C^G is positioned to the left of all of the other correlation matrices. That is to say, C^G has the lowest average value. According to these results, the reason why the observed more efficient portfolio has a lower risk than the portfolios created from other correlation matrices is that C^G

has a smaller value than all of the other correlation matrices. Accordingly, these results show that the correlation matrix is a deterministic influential factor establishing the existence of a more efficient portfolio.

3.3 The Effects of Market Properties on a More Efficient Portfolio

It is interesting to observe from the results of Section 3.2 that among the correlation matrices generated via the RMT method, the only difference between the filtered correlation matrix, C^F and the group correlation matrix, C^G, is whether the properties of the largest eigenvalue are reflected. In contrast, the difference between C^F and C^G can be clearly seen in the previous results. Therefore, in this section, we discuss the additional analysis required to determine the properties of the largest eigenvalue, which is the difference between C^F and C^G, in order to identify the hidden deterministic factor behind the existence of a more efficient portfolio.

In the previous studies of the RMT method applied to finance, the correlation matrix, $C_i = \lambda_i \cdot V_i \cdot V_i^T$, can be divided into the components of the eigenvector, V_i, and the eigenvalue, λ_i. Therefore, components that can influence the difference between C^F and C^G are the eigenvector and the eigenvalue. We attempted to assess the possible explanations behind the difference of correlation matrices from two components.

First of all, we present the results observed from the properties of eigenvectors in the correlation matrix, as an influential factor. The results are provided in Fig. 6. In the figures, Fig. 6(a) & (c) display the eigenvector of the largest eigenvalue, λ_1 (red circles) and that of the second largest eigenvalue, λ_2 (black bar) for each stock using the Korean stock data, $N = 104$, and the US stock data, $N = 310$, respectively. The X-axis is the stocks within a portfolio. In addition, we observed each eigenvector for the eigenvalues deviating from the random matrix. 5 eigenvalues exceed the range of the random matrix for the Korean stock market, and 15 eigenvalues exceed the range of the random matrix for the US stock market. For the 5~15 types of eigenvalues, the observed results are shown as a box-plot in Fig. 6(b) for the Korean stock market, and in Fig. 6(d) for the US stock market.

The results show that the eigenvector with the properties of the largest eigenvalue (red circles) assigns a high value to every stock as a whole [Fig. 6(a) & (c)]. This means that high values are not assigned only to a few stocks. On the other hand, the eigenvector of the second largest eigenvalue (black bar) assigns a definitely high value to a few stocks, but a very small or no value to the majority of stocks. These results are robustly confirmed from the finding that the box-plot for the values of the first eigenvector are distributed in a narrow range, whereas the box-plot of other eigenvectors are scattered in a larger range [Fig. 6(b) & (d)]. Additionally, the center of the box-plot of the first eigenvector is positioned higher than those of other eigenvectors.

Our observed results are similar to the results of previous studies (e.g., Plerou et al., 2002; Eom et al., 2009) suggesting that the largest eigenvalue has market properties, and those others that exceed the range of the random matrix have industrial attributes [12-23]. A market factor represents a common factor in the stock market, which commonly affect every stock traded in the market. On the other hand, industrial factors affect only the stock group in a specific industry. Therefore, the value of correlation matrices among stocks are increased when the market factor properties are included in the properties of individual stocks. Conversely, when the market factor properties are removed, the value of the correlation matrix decreases because the common properties included in every stock are eliminated. According to these results, the C^G without the properties of the largest eigenvalue, as a market factor, has a smaller value than C^F, which evidences the properties of the largest eigenvalue.

Next, in order to assess the properties of the eigenvalue that can determine the properties of the correlation matrix, we present results observed from a time series reflecting the properties of each eigenvalue in Fig. 7. The time series, $R_i^{E(i)} = \sum_{j=1}^{N} V_{i,j} \cdot R_{j,t}$, is used to assess the properties of each eigenvalue in previous econophysics and finance studies, in which, $V_{i,j}$ indicates the eigenvector with the properties of the i eigenvalue, $i = 1, 2, ..., N$, and R_j^T, $t = 1, 2, ..., T$ denote the actual returns of stock j.

The results are shown in Fig. 7. As in the process of the previous section, the data utilized for analysis were divided into Korea [Fig. 7(a) & (b)] and the US [Fig. 7(c) & (d)]. In the figures, Fig. 7(a) & (c) provides the probability distribution of the time series created for the eigenvalues deviating from the random matrix. That is to say, 5 eigenvalues exceed the range of the random matrix from stocks of the Korean stock market, and 15 eigenvalues for the US stock market. In Fig. 7(a) & (c), the circles (black) are the time series that reflect the properties of λ_1, squares(colour) for λ_2~λ_5, triangles(colour) for λ_6~λ_9, diamonds(colour) for λ_{10}~λ_{13}, and pentagram(colour) for λ_{14}~λ_{15} [Korea: λ_1~λ_5 and the U.S.: λ_1~λ_{15}]. Fig. 7 (b) & (d) show the volatility, which was calculated from the standard deviation, $\sigma_{E(i)} = \sqrt{\frac{1}{T-1} \sum_t [R_t^{E(i)} - \overline{R_t^{E(i)}}]^2}$ for the time series, $R_t^{E(i)}$. The X-axis is the category of the time series in descending order of eigenvalue size, and the Y-axis is the volatility of the time series.

The results indicate that among the eigenvalues that exceed the range of the random matrix, the probability distribution of the time series, $R_t^{E(i=1)}$, calculated from the largest eigenvalue evidences the widest degree of dispersion [Fig. 7(a) & (c)] and the highest volatility [Fig. 7(b) & (d)]. On the other hand, as the magnitude of other eigenvalues become smaller, the degree of dispersion of the probability distribution of the other time series, $R_t^{E(i)}$, $i = 2, 3, ..., N$ becomes narrower, and the size of the volatility decreases exponentially. Accordingly, the properties of the largest eigenvalue cause an increase in the portfolio risks, as well as the value of the correlation matrix between stocks within a portfolio. These results show that the reason why the portfolio created from C^G has a lower risk than the portfolio created from C^F at a given return is that in C^F, the properties of the largest eigenvalue are included. Therefore, since it has been well established in previous studies that the properties of the largest eigenvalue have the economic meaning of a market factor, these results indicate that the market factor properties cause an increase in the value of the correlation matrix, thus lowering the degree of diversification, and increasing the degree of portfolio risk.

4. Results for an Ex-ante test of a More Efficient Portfolio

In this section, using the *ex ante* test, we attempted to determine empirically whether a more efficient portfolio elicited from the group correlation matrix has practical utility in the field of finance. The optimum portfolio elicited from Markowitz's model is affected not only by the expected correlation matrix, but also by the prediction accuracy of the expected return, $E(R_j)$ and the expected risk (standard deviations), σ_j, of stocks in a portfolio. In Section 3, however, we assumed that the expected return and risk of stocks in a portfolio were $E(R_j) = R_j^A$ and $\sigma_j = s_j^A$, in which R_j^A and s_j^A are the return and risk calculated from the actual data, respectively. In other words, we assumed that the expected return and risk of a portfolio's stocks were given, and focused only on comparing the performance of the portfolio created by the various correlation matrices. Accordingly, in order for a more efficient portfolio observed by the group correlation matrix to have practical utility, the changes in portfolio performance in the future period must be evaluated according to the gradual improvement of the predictability of return and risk of stocks from the past period.

4.1 An Empirical Design for an Ex-ante Test

In order to achieve the research objective mentioned, we should select a means for predicting the return and risk of stocks. A number of methods have been introduced in finance for the prediction of a stock's return and risk with various levels of predictability. However, if we select one of the prediction models, the observed results will be influenced not only by the prediction method we select, but it will also be difficult to assess the changes in portfolio performance in accordance with gradual improvements in predictability. Accordingly, in order to appropriately consider the stated objective, we assume that if we utilize a accurate prediction model, there will be no difference between the return and risk predicted from a past period and the actual return and risk of a future period. Therefore, with the initial prediction value as the starting point, when predictability is improved gradually at a specified interval until it approaches the actual value of the future period, we can assess the changes in the portfolio performance observed in a more efficient portfolio by C^G in the previous section.

We now present a design for the gradual improvement of predictability, as follows. The first step involves calculating the difference between the actual and predicted return and risk of stocks within a portfolio. We selected simple average return $\overline{R_j} = \frac{1}{T}\sum_{t=1}^{T} R_{j,t}$ and risk (standard deviation) $s_j = \sqrt{\frac{1}{T-1}\sum_{t=1}^{T}[R_{j,t} - \overline{R_j}]^2}$, as initial prediction values calculated from a past period, and the average return, R_j^A, calculated in the future period is utilized as the actual value. Then, the difference between the two values $D_j^R = R_j^A - \overline{R_j}$ is calculated. Similarly for the risk, the standard deviation of the past period, s_j , and that of the future period, s_j^A, are calculated, as well as their difference: $D_j^S = s_j^A - s_j$. In the second step, a constant value is calculated for the improved predictability required to cause the initial prediction to gradually converge to the actual value. We divided the process of improving predictability into 10 intervals. The difference, $D_j^R = |D_j^R| \times I_j$ consists of magnitude $|D_j^R|$ and sign I_j [$I_j = 1$ if $D_j^R > 0$, $I_j = -1$ if $D_j^R < 0$], and $|D_j^R|$ is divided into ten intervals. Similarly for D_j^S, $|D_j^S|$ is also divided into 10 intervals. Therefore, the constant values for improving predictability are $d_j^R = \frac{|D_j^R|}{10}$ and $d_j^S = \frac{|D_j^S|}{10}$ for return and risk, respectively. In the third step, using the initial prediction value as the starting point, the constant value for improving predictability is accumulated in order to calculate the predicted value until it approaches the actual value. Accordingly, the prediction values of the improved predictability are $R_j^{H(k)} = R_j^{H(k-1)} + d_j^R \cdot I_j$ ($k = 2, 3, ..., 10$), and we establish the simple average return as the initial prediction value, $R_j^{H(1)} = \overline{R_j}$ ($k = 1$). The prediction value of risk $s_j^{H(k)} = s_j^{H(k-1)} + d_j^S \cdot I_j$ ($k = 2, 3, ..., 10$) is also calculated in the same way, and the initial risk prediction value is $s_j^{H(1)} = \overline{s_j}$ ($k = 1$).

Additively, we utilized the mean squared error (MSE) to quantify the degree of gradual improvement of predictability from the initial prediction value to the actual value. As the simple average return and risk were established as the initial prediction values of stocks in a past period, the degree of prediction errors for the actual value in the future period are $MSE_R^{H(1)} = \sqrt{\frac{1}{N}\sum_{j=1}^{N}[R_j^A - R_j^{H(1)}]^2}$ and $MSE_S^{H(1)} = \sqrt{\frac{1}{N}\sum_{j=1}^{N}[s_j^A - s_j^{H(1)}]^2}$. The prediction error calculated from the initial prediction value becomes the maximum prediction error---that is to say, the maximum return prediction error, $MSE_R^{H(1)} = MSE_R^{MAX}$ and the maximum risk prediction error, $MSE_S^{H(1)} = MSE_S^{MAX}$. We then calculated the degree of improvement of predictability, MSE_k, as the average value of the degree of improvement of the return prediction error, MSE_k^R, and the risk prediction error, MSE_k^S. That is,

$$MSE_k = \frac{MSE_k^R + MSE_k^S}{2} \tag{5}$$

where, $MSE_k^R = \frac{MSE_R^{MAX} - MSE_R^{H(k)}}{MSE_R^{MAX}}$ $[MSE_R^{MAX} = MSE_R^{H(1)}]$

$MSE_k^S = \frac{MSE_S^{MAX} - MSE_S^{H(k)}}{MSE_S^{MAX}}$ $[MSE_S^{MAX} = MSE_S^{H(1)}]$

in which the ranges of improvement of return and risk prediction errors are $0\% \leq MSE_k^R \leq 100\%$ and $0\% \leq MSE_k^S \leq 100\%$, respectively. Accordingly, the degree of improvement of overall prediction error is also $0\% \leq MSE_k \leq 100\%$. For example, as $k = 1$ corresponds to the starting point, $MSE_{k=1} = 0\%$, which indicates no improvement in predictability, and then predictability improves continuously up to $k = 10$, at which the perfect prediction $MSE_{k=10} = 100\%$ is reached, becoming identical with the actual data. We divided the process with which predictability improves from the initial value to the actual data into 10 intervals, and MSE_k measured at each interval from Eq. 5 is shown in Fig. 8 with a bar graph, as well as in Table 3.

Next, we present the process for combining the correlation matrix and the predicted return and risk of the stocks in a portfolio from a past period in order to determine the portfolio return and risk of a future period. The investment weights of stocks in a portfolio, $F(R_j^{H(k)}, s_j^{H(k)}, C^G) = w_j^{p,G(k)}$, $p = 1, 2, ..., 50$, are calculated using $R_j^{H(k)}$, $s_j^{H(k)}$, and C^G predicted in the past period, as the input data of Markowitz's model. 50 investment weights are calculated in accordance with the given target return. In order to calculate the portfolio return, $R_p^{G(k)}$ and risk, $\sigma_p^{G(k)}$, the calculated investment weights are combined with the R_j^A and s_j^A values of stocks in a portfolio from a future period as follows.

$$R_p^{G(k)} = \sum_{j=1}^{N} w_j^{p,G(k)} \cdot R_j^A \qquad [k = 1, 2, ..., 10, \quad p = 1, 2, ..., 50] \tag{6}$$

$$\sigma_p^{G(k)} = \sqrt{\sum_{i=1}^{N} \sum_{j=1}^{N} w_i^{p,G(k)} \cdot w_j^{p,G(k)} \cdot s_i^A \cdot s_j^A \cdot C^G} \tag{7}$$

We selected three comparative portfolio performances in order to evaluate the usefulness of portfolio performance derived from the group correlation matrix. The first is the portfolio performance of the future period drawn from the historical correlation matrix, C^H, used in the conventional Markowitz's model. Using $R_j^{H(k)}$, $s_j^{H(k)}$ and C^H predicted in the past period, the investment weight, $F(R_j^{H(k)}, s_j^{H(k)}, C^H) = w_j^{p,H(k)}$ of the stocks in a portfolio is calculated, and then the calculated investment weight is combined with the actual return and risk of the future period to generate the portfolio return and risk by $R_p^{H(k)} = \sum_{j=1}^{N} w_j^{p,H(k)} \cdot R_j^A$ and $\sigma_p^{H(k)} = \sqrt{\sum_{i=1}^{N}\sum_{j=1}^{N} w_i^{p,H(k)} \cdot w_j^{p,H(k)} \cdot s_i^A \cdot s_j^A \cdot C^H}$, respectively. The second is the portfolio return, $R_p^E = \frac{1}{N}\sum_{j=1}^{N} R_j^A$ and risk, $\sigma_p^E = \frac{1}{N}\sum_{j=1}^{N} s_j^A$ calculated by the equal-weighted average method using the actual data of the future period. The third comparison is the portfolio performance elicited from Markowitz's model using the actual data of the future period. Using R_j^A, s_j^A and C^A in the future period, the $F(R_j^A, s_j^A, C^A) = w_j^{p,A)}$ of stocks in a portfolio is calculated, and then the calculated investment weights are combined with R_j^A and s_j^A, and the portfolio return and risk are $R_p^A = \sum_{j=1}^{N} w_j^{p,A} \cdot R_j^A$ and $\sigma_p^A = \sqrt{\sum_{i=1}^{N}\sum_{j=1}^{N} w_i^{p,A} \cdot w_j^{p,A} \cdot s_i^A \cdot s_j^A \cdot C^A}$, respectively.

Finally, we create the portfolio performance index as a measure of portfolio returns per unit of portfolio risks in an investment. This is because the four types of portfolios mentioned above feature different risks and returns, and thus it is difficult to compare risks and returns based merely on their magnitude. Accordingly, we utilized the portfolio performance index, $PI_p = R_p/\sigma_p$. The index is quite similar to the Sharpe ratio (= $[E(R_p) - R_f]/\sigma_p$, in which R_f is the risk-free rate [25]). The difference is whether the portfolio return, R_p, or the portfolio excess return, $E(R_p) - R_f$, is utilized in the numerator. Thus, we conduct a comparison by calculating PI_p for the four portfolio types. First, $PI_p^{G(k)} = R_p^{G(k)}/\sigma_p^{G(k)}$ is calculated using the portfolio return and risk drawn from C^G, second, $PI_p^{H(k)} = R_p^{H(k)}/\sigma_p^{H(k)}$ is calculated using the portfolio return and risk from C^H, third, $PI_p^E = R_p^E/\sigma_p^E$ is calculated using the return and risk of the equal-weighted portfolio, and fourth, $PI_p^A = R_p^A/\sigma_p^A$ is calculated using the portfolio return and risk from C^A. The criteria of evaluation is that the higher the portfolio performance index, the higher the level of compensation of return for risk. That is, the higher the PI_p, the more efficient is the portfolio.

4.2. Results from an Ex-ante Test

According to the design mentioned above, the results conducted based on an *ex ante* test to determine whether a more efficient portfolio drawn from the group correlation matrix has practical utility are provided in Fig. 8 and Table 3. Fig. 8 provides the results for the entire period (estimating period 1990-2006/testing period 2007), and Table 3 indicates the results of the identical test process performed in three sub-periods (1990-1994/1995, 1996-2000/2001, 2002-2006/2007) to obtain the robustness of the results in Fig. 8.

In Fig. 8, the X-axis represents comparisons C1-C2, and the 10 intervals of gradual improvements of predictability, S1-S10. Stated differently, C1 and C2 are the PI_p^A from C^A and the PI_p^E from the equal-weighted average portfolio using the future period's actual data, respectively. Otherwise, S1-S10 indicate the 10 intervals with which predictability is improved gradually from the initial prediction value of the past period, S1, to the actual value, S10. In the figure, the yellow bar represents MSE_k, which indicates the level of improvement in predictability compared to the initial prediction value. In addition, the portfolio performance index calculated from C^G is marked with red circles, PI_{p50}^G, $1 \leq p \leq 50$ [and magenta hexagram, PI_{p25}^G, $1 \leq p \leq 25$] and the portfolio performance index calculated from C^H is denoted by blue squares, PI_{p50}^H [and cyan diamond, PI_{p25}^H]. Since we calculated the portfolio return and risk for 50 target returns via Markowitz's model, 50 portfolio performance indices are calculated. We also confirmed, from the previous results of Section 3, that the degree of diversification for the portfolio stocks was higher in the first half, $1 \leq p \leq 25$, than in the second half, $26 \leq p \leq 50$. Accordingly, we categorized the portfolio performance index into the results for the entire region, PI_{p50}, and the first half, PI_{p25}. In the figure, the average and standard deviation of each of the portfolio performance indices are represented by error-bar graphs. Fig. 8(a) and Fig. 8(b) show the results obtained using Korean and US stock market data for the entire period, respectively.

The results of the *ex ante* test empirically show that a more efficient portfolio derived from C^G has a higher portfolio performance index than the three comparisons. These results also show that a more efficient portfolio from C^G has a high degree of practical utility for portfolio investment on the basis of Markowitz's model. The results for each comparative portfolio performance can be summarized as follows.

The first comparison is the portfolio performance index, $PI^{H(k)}$, derived from the conventional Markowitz's model that utilizes C^H. We verified that the average value of the portfolio performance index drawn from C^G, $\overline{PI_{p50}^{G(k)}}$ (red circles) and $\overline{PI_{p25}^{G(k)}}$ (magenta hexagram), is higher on average than the portfolio performance index derived from C^H, $\overline{PI_{p50}^{H(k)}}$ (blue squares) and $\overline{PI_{p25}^{H(k)}}$ (cyan diamond), regardless of the degree of improvement in predictability $k = 1 \to 10$ of the return/risk of stocks in a portfolio. In other words, the portfolio performance index of a future period for a more efficient portfolio drawn from C^G is always higher than that of the conventional Markowitz's model, regardless of the level of predictability of the return and risk of stocks in a past period.

The second comparison is the portfolio performance index, PI^E, derived from the risk and return of the portfolio calculated via the equal-weighted averages method using the actual data of stocks in a future period. According to our results, the portfolio performance index calculated via the equal weighted average method in the future period for the Korean [Fig. 8(a)] and US [Fig. 8(b)] stock markets were PI^E=0.067 and PI^E=-0.019, respectively. Otherwise, the portfolio performance index derived from C^G in a past period is $\overline{PI_{p25}^{G(1)}}$=0.318 (t:21.44) and

$\overline{PI^{G(1)}_{p50}}$=0.187 (t:8.92) at $MSE_{k=1} = 0\%$ of the initial prediction value $k = 1$ for the Korean stock market, which is greater than PI^E, and the US stock market data yielded $\overline{PI^{G(2)}_{p25}}$=0.250 (t:6.73) and $\overline{PI^{G(2)}_{p50}}$=0.214 (t:10.10) at the second prediction value, $k = 2$, with a predictability improvement of $MSE_{k=2} = 21\%$ from the initial value---which was also greater than PI^E. These results revealed that the portfolio performance index derived from C^G is greater than that of the equal-weighted portfolio calculated using the actual stock returns and risks. That is to say, the reward-to-risk of the portfolio from C^G is higher than those of an equal-weighted portfolio.

The third comparison involved the portfolio performance index, PI^A, calculated from C^A in a future period. According to our results, the portfolio performance index derived from the actual data of a future period for the Korean stock market was $\overline{PI^A_{p25}}$=0.123 (t:9.31) and $\overline{PI^A_{p50}}$=0.150 (t:18.81). For the US stock market, it was $\overline{PI^A_{p25}}$=0.125 (t:13.04) and $\overline{PI^A_{p50}}$=0.148 (t:24.42). As confirmed previously, the portfolio performance indices, $\overline{PI^{G(1)}_{p\cdot}}$ for the Korean stock market and the $\overline{PI^{G(2)}_{p\cdot}}$ for the US stock market, derived from the past period's C^G are greater than $\overline{PI^A_p}$, even if based on the initial prediction value. However, as in the conventional Markowitz's model, the portfolio performance index elicited from the C^H of the past period does not evidence a significant difference from PI^A. In other words, despite the low predictability of the return and risk of the stocks in a past period, the portfolio performance index derived from C^G is greater than that of future period's portfolios in the *ex ante* test, and is clearly superior to the portfolio performance achieved via the conventional Markowitz's model.

Finally, we conducted a comparison between the first half, $1 \leq p \leq 25$, and the entire region, $1 \leq p \leq 50$, for the performance indices of the portfolio derived from C^G. With regard to the observation made from the Korean stock market data, the portfolio performance indices for the entire region were $0.187 \leq \overline{PI_{p50}^{G(k)}} \leq 0.381$, and those of the first half were $0.318 \leq \overline{PI_{p25}^{G(k)}} \leq 0.520$. The portfolio performance of the first half was superior to that of the entire region. Similarly, in the US stock market, the portfolio performance of the first half $-0.117 \leq \overline{PI_{p25}^{G(k)}} \leq 0.470$ was higher than that of the entire region $-0.077 \leq \overline{PI_{p50}^{G(k)}} \leq 0.396$. However, the portfolio performance indices calculated from the C^H of the conventional Markowitz's model evidence results that contrast with what was observed from C^G. In the case of the Korean stock market, the portfolio performance indices for the entire region were $0.026 \leq \overline{PI_{p50}^{H(k)}} \leq 0.153$, which were greater than those of the first half $0.046 \leq \overline{PI_{p25}^{H(k)}} \leq 0.124$. Additionally, for the US stock market, the portfolio performance indices for the entire region were $-0.026 \leq \overline{PI_{p50}^{H(k)}} \leq 0.195$, as compared with those of the first half $-0.042 \leq \overline{PI_{p25}^{H(k)}} \leq 0.174$. Accordingly, it can be noted from the observed results that the fact that the first half of the portfolio performance was higher than those of the entire region is a characteristic of a portfolio derived from C^G.

The above-mentioned findings are also confirmed in the results of an identical test conducted for the three sub-periods demonstrated in Table 3. Therefore, we robustly determined that the performance of a more efficient portfolio from the group correlation matrix is clearly superior to the portfolio performance of the conventional Marlowitz's model, even if it is combined with the stock returns and risks predicted *ex ante* in a past period. That is to say, the performance of a portfolio derived from the group correlation matrix has a higher level of compensation of return for risk than is observed from portfolios generated via other means. Based on these results, we surmise that our findings may contribute not only in terms of the academic aspects of the portfolio theory, but also in terms of the practical implications of efficient portfolio performance management.

5. Conclusions

The objective of the portfolio theory is to implement a methodology by which investors' portfolios can be optimized through diversification. Research efforts designed to generate more efficient portfolio selections have not only academic, but also practical, importance. The principal objective of this study was to provide empirical evidence as to whether a correlation matrix controlled via the RMT method might yield a more efficient portfolio, with a lower risk at a given return, as compared to the one created from the conventional Markowitz's model. Additionally, we intended to provide explanations regarding the existence of a more efficient portfolio on the basis of the empirical evidence. We also attempted to determine whether the observed more efficient portfolio has practical usefulness on the basis of the *ex ante* test of the return and risk predictability of stocks in a portfolio. We utilized stocks from Korea and the US stock market for various periods. The observed results can be summarized as follows.

First, we attempted to determine whether a portfolio from the correlation matrix generated via the RMT method is more efficient than the portfolios from Markowitz's portfolio theory in two perspectives—namely, that of the degree of efficiency and that of the degree of diversification. We robustly detected the existence of a more efficient portfolio with a lower risk than the portfolio generated by Markowitz's model from the viewpoint of efficiency. In particular, the portfolio created by the correlation matrix, which has the properties of the eigenvalues except for those of the largest eigenvalue among the eigenvalues deviating from the random matrix, proved more efficient than those generated from other correlation matrices via the RMT method. This correlation was designated as a group correlation matrix in this study. Furthermore, using the group correlation matrix, the RMT method is also an effective tool for improving the degree of diversification for investment weights among stocks in a portfolio, which has previously presented a practical limitation to the application of the conventional Markowitz's portfolio theory. Namely, the distribution level of investment weights among stocks in a portfolio is generally quite good.

Next, we attempted to determine empirically why the portfolio generated from the group correlation matrix is more efficient than those from other correlation matrices. We discovered that the correlation matrix was directly affected by whether the property of the largest eigenvalue was reflected. That is, the other correlation matrices have the properties of the largest eigenvalue, but the group correlation matrix excluded the properties of the largest eigenvalue. Therefore, we noted that the properties of the largest eigenvalue cause an increase in the value of the correlation matrix, lowering the degree of diversification, and ultimately increasing the degree of portfolio risk. In previous studies, it has been relatively well established that the properties of the largest eigenvalue have the economic meaning of a market factor. Therefore, these results indicate that the correlation matrix without the property of the market factor has a lower correlation matrix value and a lower degree of risk, and thus yields a more efficient portfolio.

Finally, we attempted to verify whether a more efficient portfolio derived from the group correlation matrix has practical utility, based on the results of an *ex ante* test. In other words, we conducted an empirical observation of the variation in the performance of the portfolio derived from the group correlation matrix as the level of predictability of the expected return and the risk of stocks in a portfolio changed. We found that the performance index of the portfolio derived from the group correlation matrix was higher than that of the portfolio generated via the equal-weighted method using the actual return and risk of a future period. Moreover, by somewhat improving the predictability of the expected return and the expected risk of the past period, the performance index of the portfolio derived from the group correlation matrix was higher than that of the portfolio generated by Markowitz's model using actual data. Therefore, we confirmed empirically that the group correlation matrix constructed by the RMT method has a high degree of utility for portfolio investment on the basis of Markowitz's model.

With regard to the research topic of portfolio selection, the ability to select a more efficient portfolio is not only an academic improvement of the portfolio theory, but also constitutes a practical improvement for effective asset allocation. These results imply that our research efforts have uncovered empirical evidence supporting the notion that the correlation matrix controlled via the RMT method has a high degree of utility in the development of investment strategies that apply the portfolio theory in practice, from both academic and practical viewpoints. In addition, our study results may be directly applicable to the real world, and the observations made in this study are expected to bring about applications and expansions of future studies of various portfolio types.

Acknowledgements

This work was supported by the Korea Science and Engineering Foundation (KOSEF) grant funded by the Korean government (MEST) (No. R01-2008-000-21065-0)

REFERENCES

Brown, S. T. (1989), "The Number of Factors in Security Returns," the Journal of Finance 44(5), 1247-1262.

Elton, E. J. and M. J. Gruber (1973), "Estimating the Dependence Structure of Share Prices: Implications for Portfolio Selection," the Journal of Finance 28(5), 1203-1232.

Elton, E. J. and M. J. Gruber (1977), "Risk Reduction and Portfolio Size: An Analytical Solution," the Journal of Business 50(4), 415-437.

Eom, C., G. Oh and S. Kim (2008), "Statistical Investigation of Connected Structures of Stock Networks in Financial Time Series," the Journal of Korean Physical Society, 53(6), 3837-3841.

Eom, C., W-S. Jung, T. Kaizoji, and S. Kim (2008), "Effect of Changing Data Size on Eigenvalues in the Korean and Japanes Stock Market," preprint available on web: arXiv.org:0811.4021.

Eom, C., W-S. Jung, T. Kaizoji and Yong H. Kim, J. Park (2009), "The Effect of Market Properties on Portfolio Diversification in Korean and Japanese Stock Market," preprint available on web: rXiv.org:xxxxxxx.

Evans, J. L. and S. H. Archer (1968), "Diversification and the Reduction of Dispersion: An Empirical Analysis," the Journal of Finance 23(5), 761-767.

Garas, A., and P. Argyrakis (2007), "Correlation Study of the Athens Stocks Exchange," Physica A 380, 399-410.

King, B. F. (1966) "Market and Industry Factors in Stock Price Behavior," the Journal of Business 39(1), 139-190.

Laloux, L., P. Cizeau, J. Bouchaud, and M. Potters (1999), "Noise Dressing of Financial Correlation Matrices," Physical Review Letters 83(7), 1467-1470.

Laloux, L. P. Cizeau, M. Potters, and J.-P. Bouchaud (2000), "Random Matrix Theory and Financial Correlations," International Journal of Theoretical and Applied Finance 3(3), 391-397.

Markowitz, H. (1952), "Portfolio Selection," the Journal of Finance 7(1), 77-91.

Metha, M. L. (1991), Random Matrices, second ed., Academic Press, New York.

Meyers, S. L.(1973), "A Re-Examination of Market and Industry Factors in Stock Price Behavior," the Journal of Finance 28(3), 695-705.

Plerou, V., P. Gopikrishnan, B. Rosenow, L. A. N. Amaral, and H. E. Stanley (1999), "Universal and Nonuniversal Properties of Cross Correlations in Financial Time Series," Physical Review Letters 83(7), 1471-1474.

Plerou, V., P. Gopikrishnan, B. Rosenow, L. A. N. Amaral, T. Guhr, and H. E. Stanley (2002), "Random Matrix Approach to Cross Correlations in Financial Data," Physical Review E 65, 066126.

Rosenow, B., V. Plerou, P. Gopikrishnan, and H. E. Stanley (2002), "Portfolio optimization and the random magnet problem," Europhysics Letters 59(4), 500-506.

Sengupta, A. M., and P. P. Mitra (1999), "Distributions of Singular Values for some Random Matrices," Physical Review E 60, 389-392.

Sharifi, S., M. Crane, A. Shamaie, and H. Rukin (2004), "Random Matrix Theory for Portfolio Optimization: a stability approach," Physica A 335, 629-643.

Sharpe, W. F. (1975), "Adjusting for Risk in Portfolio Performance Measurement," the Journal of Portfolio Management, 29-34.

Statman, M. (1987), "How Many Stocks Make a Diversified Portfolio," the Journal of Financial and Quantitative Analysis 22(3), 353-363.

Tola, V., F. Lillo, M. Gallegati, and R. N. Mantegna (2008), "Cluster Analysis for Portfolio Optimization," Journal of Economic Dynamics & Control 32, 235-258.

Trzcinka, C. (1986), "On the Number of Factors in the Arbitrage Pricing Model," the Journal of Finance 41(2), 347-368.

Wilcox, D. and T. Gebbie (2004), "On the Analysis of Cross-correlation in South African market Data," Physica A 344, 294-298.

Wilcox, D. and T. Gebbie (2007), "An Analysis of Cross-correlations in an Emerging Market," Physica A 375, 584-598.

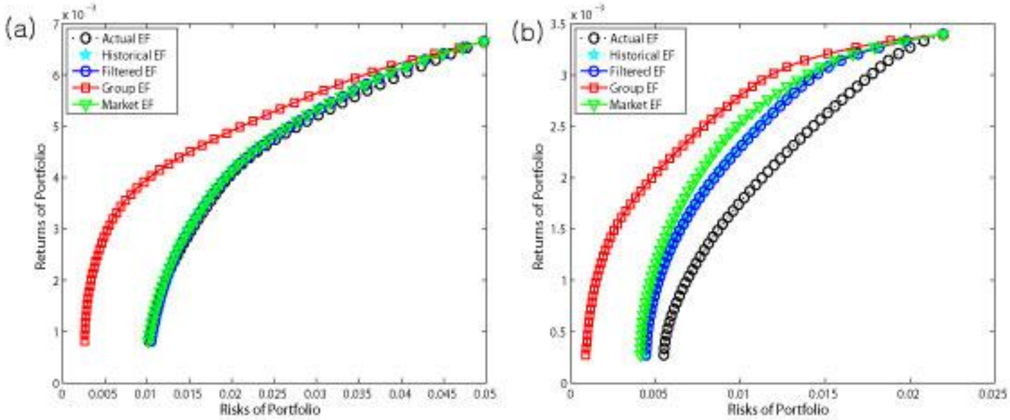

Fig. 1 (colour online): This figure indicates the results from the viewpoint of the degree of efficiency. Fig. 1(a) shows the results for the Korean stock market, and Fig. 1(b) for the US stock market. The X-axis denotes the portfolio risk, and the Y-axis indicates the portfolio returns. In the figure, black circles indicate the portfolio created from the actual correlation matrix, C^A, cyan pentagrams as the historical correlation matrix, C^H, blue circles as the filtered correlation matrix, C^F, green triangles as the market correlation matrix, C^M, and red squares as the group correlation matrix, C^G.

Portfolio Risks \ Sub-period		Sub-period 1 (99.1~95.12)	Sub-period 2 (96.1~01.12)	Sub-period 3 (02.1~07.12)
Korean Stock Market	Actual, $\overline{\sigma_p^A}$	0.0097*	0.0162*	0.0225*
	Historical, $\overline{\sigma_p^H}$	0.0110*	0.0188*	0.0215*
	Filtered, $\overline{\sigma_p^F}$	0.0110*	0.0191*	0.0216*
	Market, $\overline{\sigma_p^M}$	0.0107*	0.0187*	0.0215*
	Group, $\overline{\sigma_p^G}$	0.0055*	0.0095*	0.0149*
US Stock Market	Actual, $\overline{\sigma_p^A}$	0.0070*	0.0109*	0.0114*
	Historical, $\overline{\sigma_p^H}$	0.0077*	0.0106*	0.0101*
	Filtered, $\overline{\sigma_p^F}$	0.0078*	0.0107*	0.0102*
	Market, $\overline{\sigma_p^M}$	0.0076*	0.0104*	0.0095*
	Group, $\overline{\sigma_p^G}$	0.0059*	0.0071*	0.0060*

Notes:* significant at the 1% level.

Table 1: This table represents results of average values of portfolio risk, $\overline{\sigma_p} = \frac{1}{50}\sum_{p=1}^{5} 0\sigma_p$, to quantify the degree of efficiency for each sub-periods.

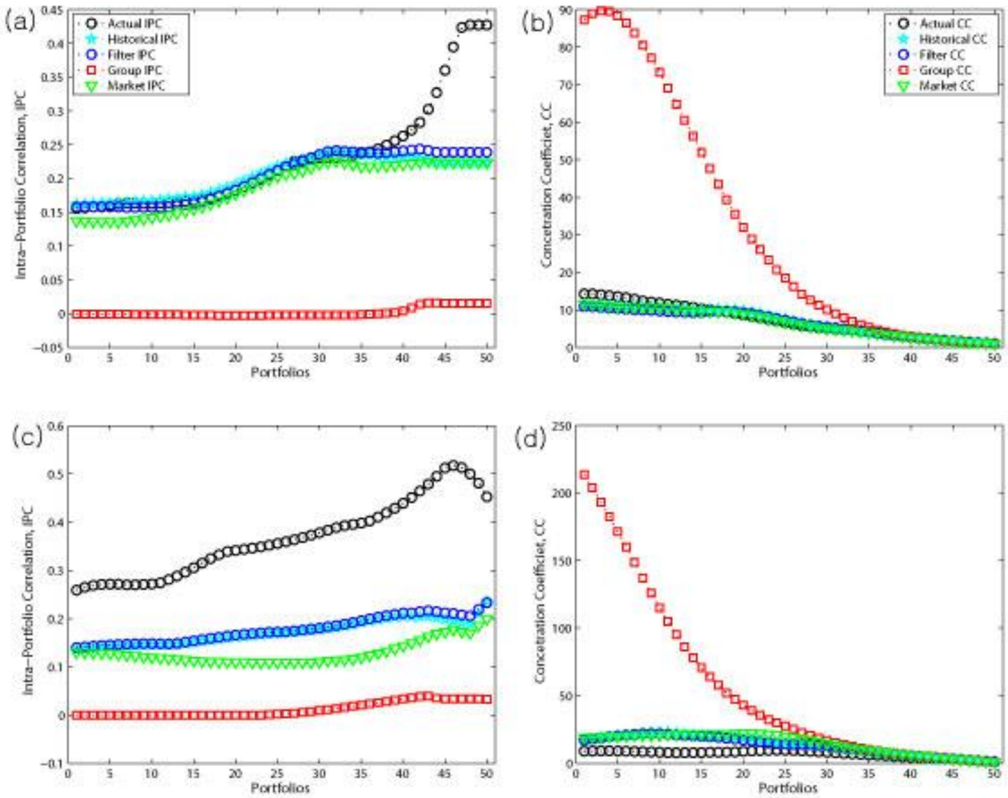

Fig. 2 (colour online): This figure denotes the results from the viewpoint of degree of diversification. In the figures, Fig. 2(a) & (b) are the results generated using the stocks of the Korean stock market, and Fig. 2(c) & (d) for the stocks in the US. The X-axis indicates the 50 cases of portfolio, and the Y-axis presents IPC_p [Fig. 2(a) & (c)] and CC_p [Fig. 2(b) & (d)]. The figure is categorized as black circles C^A, cyan pentagrams C^H, blue circles C^F, green triangles C^M and red squares C^G.

Correlation Matrix \ Sub-period		Sub-period 1 (99.1~95.12)		Sub-period 2 (96.1~01.12)		Sub-period 3 (02.1~07.12)	
		$\overline{IPC_p}$	$\overline{CC_p}$	$\overline{IPC_p}$	$\overline{CC_p}$	$\overline{IPC_p}$	$\overline{CC_p}$
Korean Stock Market	Actual, C^A	0.1173*	11.63*	0.1187*	8.69*	0.2304*	7.07*
	Historical, C^H	0.1778*	11.05*	0.2248*	7.81*	0.1926*	8.68*
	Filtered, C^F	0.1803*	10.25*	0.2375*	7.92*	0.1972*	8.28*
	Market, C^M	0.1572*	11.25*	0.2265*	7.91*	0.1875*	8.33*
	Group, C^G	-0.0060*	31.79*	-0.0037*	36.04*	-0.0017*	31.59*
The US Stock Market	Actual, C^A	0.015*	16.76*	0.095*	5.19*	0.368*	6.89*
	Historical, C^H	0.065*	13.35*	0.122*	5.90*	0.241*	9.23*
	Filtered, C^F	0.069*	13.64*	0.126*	5.61*	0.250*	9.09*
	Market, C^M	0.056*	13.88*	0.111*	6.06*	0.198*	10.09*
	Group, C^G	0.005*	71.43*	0.012*	41.84*	0.016*	52.01*

Notes:* significant at the 1% level.

Table 2: This table represents results of the average values of intra-portfolio correlation, $\overline{IPC} = \frac{1}{50}\sum_{p=1}^{50} IPC_p$, $p = 1, 2, ..., 50$ and those of concentration coefficient, $\overline{CC} = \frac{1}{50}\sum_{p=1}^{50} CC_p$, for stock of each country using three sub-periods.

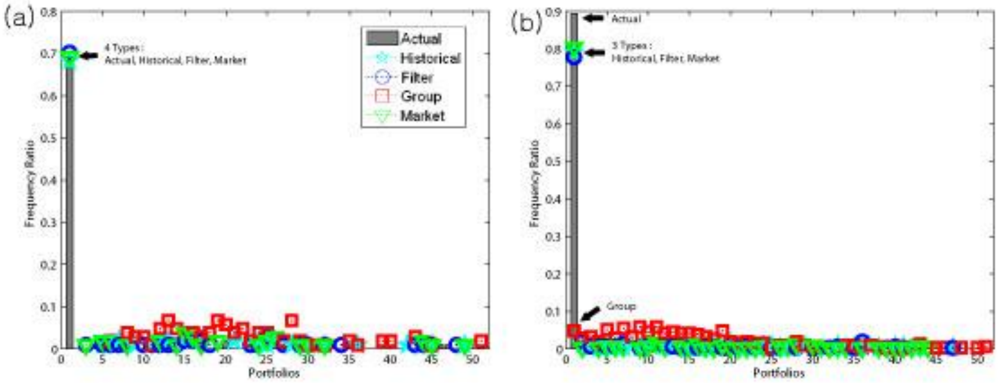

Fig. 3 (colour online): This figure shows the frequency of the stocks with non-zero investment weights, $w_j^p > 0$, in order to examine the degree of diversification of the stocks in a portfolio. Fig. 3(a) is results using the stocks of the Korean stock market, and Fig. 3(b) for the stocks in the US. The Y-axis represents the frequency ratio of the number of stocks having non-zero weight among the overall number of stocks in each 50 portfolios. The X-axis indicates the number of cases of stocks with non-zero weights, $w_j^p > 0$ among the 50 portfolios in Fig. 1, and accordingly, the range of the X-axis is $0 \le X(=k) \le 50$. In the figures, the bar indicates the frequency ratio for the portfolios generated from the actual correlation matrix C^A of the future period, and the figure is categorized as black circles C^A, cyan pentagrams C^H, blue circles C^F, green triangles C^M, and red squares C^G.

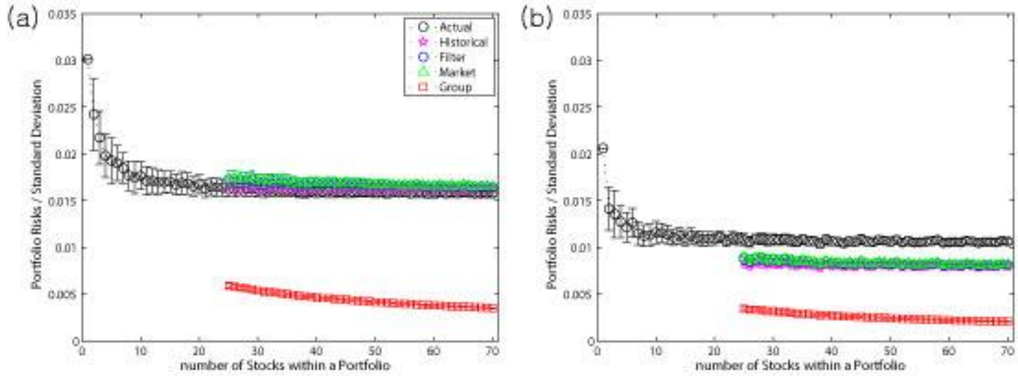

Fig. 4 (colour online): This figure represents the results of the diversification effects of the portfolio with an error bar graph using the average and standard deviations of 100 simulations. Fig. 4(a) shows the results of stocks of the Korean stock market, and Fig. 4(b) for the US stock market. The X-axis represents the M number of stocks within portfolio, $1 \leq M \leq 70$ range, in which $M = 1$ corresponds to an individual stock. Y-axis is the portfolio risks. The figure is divided into black circles of C^A, cyan pentagrams of C^H, blue circles of C^F, green triangles of C^M, and red squares of C^G.

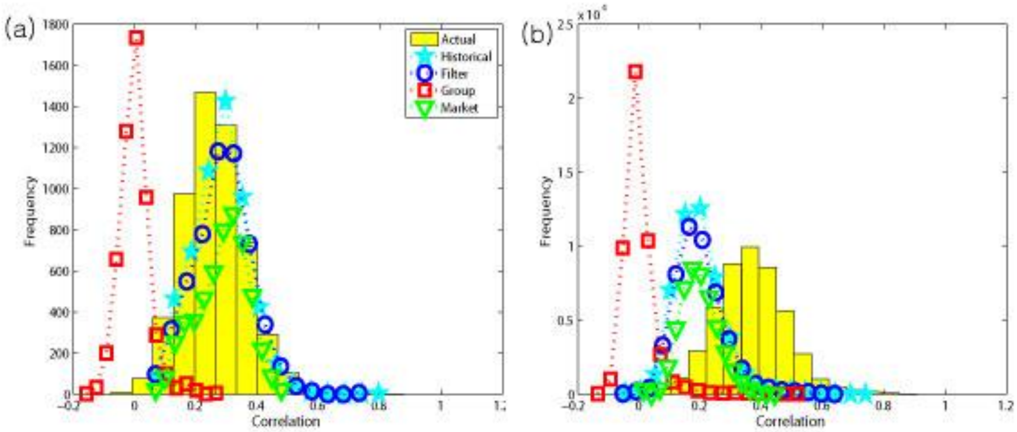

Fig. 5 (colour online): This figure shows the results from the probability distributions of five correlation matrices. Fig. 5(a) is the results of stocks of the Korean stock market and Fig. 5(b) for the US stock market. In Figs. 5(a) & (b), the yellow bar is the probability distribution of C^A in the future period, and the probability distributions of C^A in the past period are indicated with cyan pentagrams, C^F with blue circles, C^M with green triangles, and C^G with red squares.

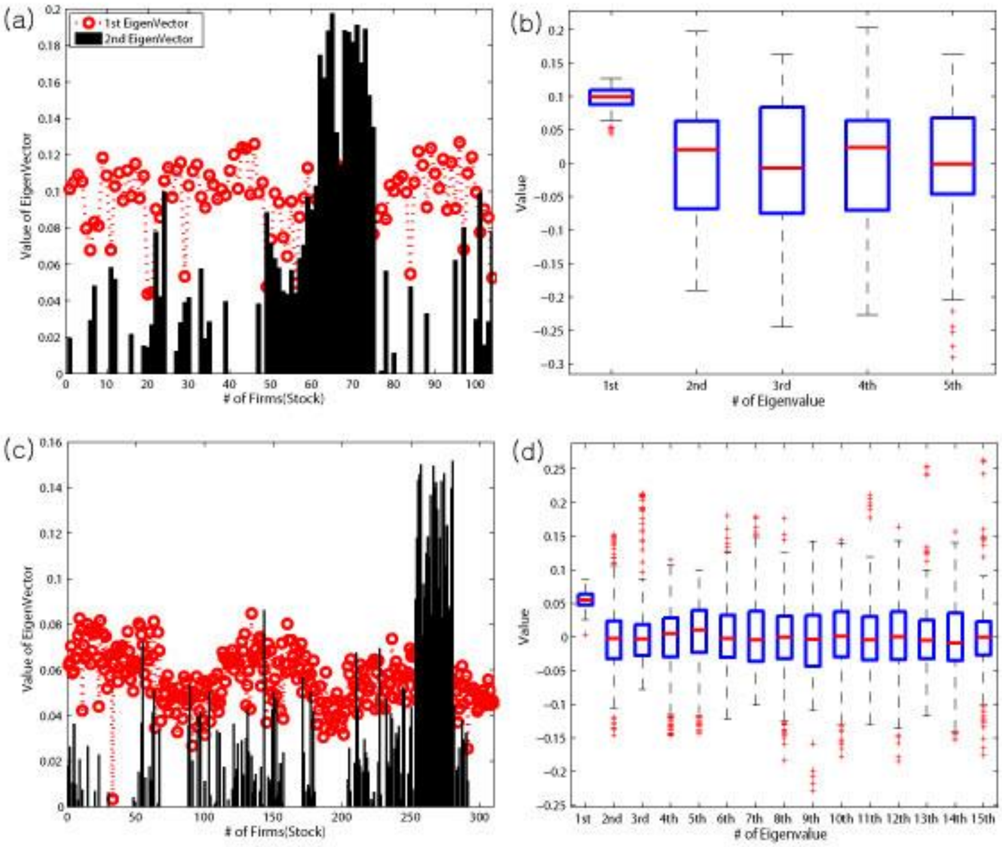

Fig. 6 (colour online): This figure represents the results from the properties of eigenvectors in the correlation matrix. In the figure, Fig. 6(a) & (c) show the eigenvector of the largest eigenvalue (red circles) and that of the second largest eigenvalue (black bar) for each stock using the Korean and US stock data, respectively. The X-axis is the stocks within a portfolio. In the box-plot of Figs. 6(b) & (d), there are eigenvalues that exceed the range of the random correlation matrices from stocks of the Korean stock market and the US, respectively. The X-axis differentiates the five types of eigenvalues.

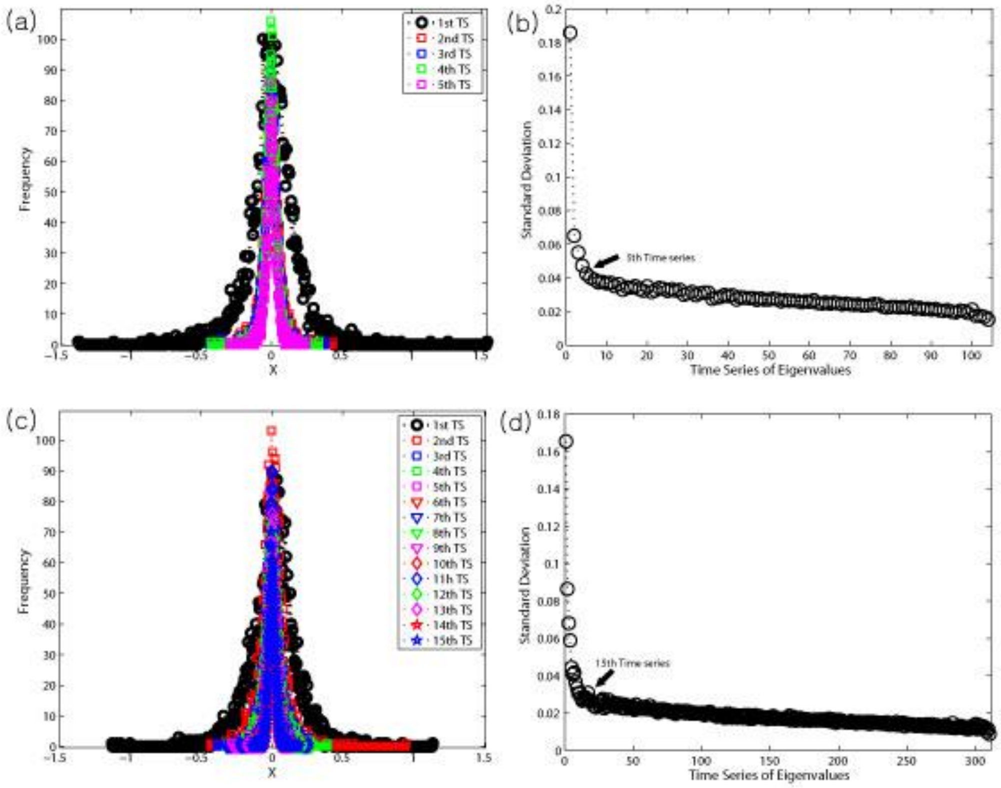

Fig. 7 (colour online): This figure shows the results from time series reflecting the properties of each eigenvalue. The data used for Korea [Fig. 7(a) & (b)], and the US [Fig. 7(c) & (d)]. Fig. 7(a) & (b) display the probability distribution of the time series created for the eigenvalues that exceed the range of random correlation matrices. In the figure, the circles (black) are the time series that reflect the properties of λ_1, squares(colour) for λ_2~λ_5, triangles (colour) for λ_6~λ_9, diamonds (colour) for λ_{10}~λ_{13}, and pentagram (colour) for λ_{14}~λ_{15} (Korea: λ_1~$\lambda_5$and the U.S.: λ_1~λ_{15}). Fig. 7(b) & (d) display the volatility (standard deviation) calculated for the time series. The X-axis is the category of the time series in descending order of eigenvalue, and the Y-axis is the volatility of the time series.

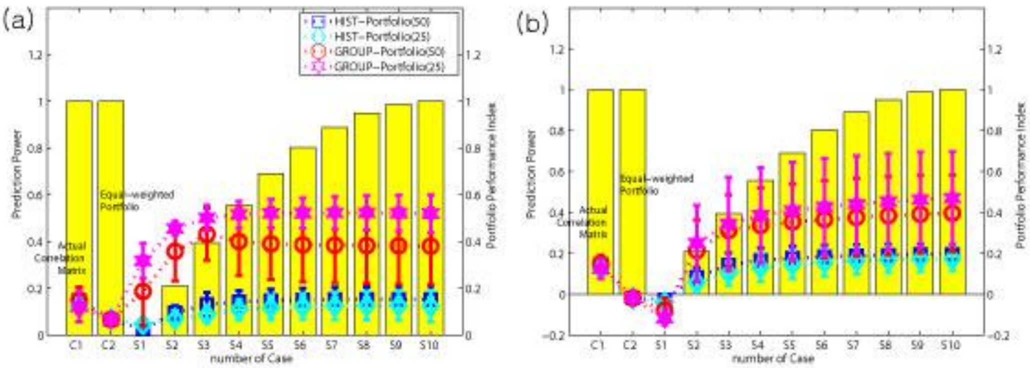

Fig. 8 (colour online): This figure shows the results based on an *ex ante* test to determine whether a more efficient portfolio from the group correlation matrix has practical utility. In the X-axis, C1 and C2 on the X-axis are PI_p^A from the actual correlation matrix and PI_p^E from the equal-weighted average portfolio using the future period's actual data, respectively, and S1-S10 are the 10 intervals of gradual improvements of predictability. In the figure, the yellow bar represents MSE_k. In addition, the portfolio performance index calculated from the group correlation matrix is indicated with red circles, PI_{p50}^G [and magenta hexagram, PI_{p25}^G] and those from the historical correlation matrix with blue squares, PI_{p50}^H [and cyan diamond, PI_{p25}^H]. In the figure, the average and standard deviation of each portfolio performance index are represented with error-bar graphs. Fig. 8(a) and Fig. 8(b) are the results obtained using the Korean and US stock market data for the entire period, respectively.

		Sub-Period 1				Sub-Period 2				Sub-Period 3			
		GROUP :$\overline{PI^G}$		HIST :$\overline{PI^H}$		GROUP :$\overline{PI^H}$		HIST :$\overline{PI^H}$		GROUP :$\overline{PI^G}$		HIST :$\overline{PI^H}$	
MSE_k		25P	50P	25P	50P	25P	50P	25P	50P	25P	50P	25P	50P
# using Data of Korean Stock Market													
1	0%	0.32	0.18	0.07	0.04	0.31	0.19	0.06	0.04	0.29	0.17	0.02	0.01
2	21%	0.46	0.37	0.11	0.12	0.45	0.35	0.09	0.11	0.42	0.33	0.06	0.09
3	40%	0.51	0.44	0.11	0.15	0.50	0.43	0.11	0.14	0.48	0.42	0.08	0.13
4	56%	0.53	0.41	0.13	0.16	0.51	0.40	0.12	0.15	0.50	0.40	0.10	0.14
5	69%	0.54	0.40	0.13	0.16	0.52	0.40	0.13	0.15	0.51	0.39	0.12	0.15
6	80%	0.54	0.40	0.13	0.16	0.52	0.39	0.13	0.16	0.51	0.39	0.12	0.15
7	89%	0.54	0.40	0.13	0.16	0.52	0.39	0.13	0.16	0.51	0.38	0.13	0.16
8	95%	0.54	0.40	0.13	0.16	0.52	0.39	0.13	0.16	0.51	0.38	0.13	0.16
9	99%	0.54	0.40	0.13	0.16	0.52	0.38	0.13	0.16	0.51	0.38	0.13	0.16
10	100%	0.54	0.39	0.13	0.15	0.52	0.38	0.13	0.16	0.51	0.38	0.13	0.16
ACTUAL :$\overline{PI^A}$		25P	0.16			25P	0.17			25P	0.12		
		50P	0.17			50P	0.18			50P	0.15		
EQUAL :$\overline{PI^E}$		0.07				0.07				0.07			
# using Data of the US Stock Market													
1	0%	-.15	-.10	-.07	-.04	-.15	-.10	-.04	-.03	-.13	-.09	-.03	-.02
2	21%	0.29	0.23	0.07	0.09	0.24	0.21	0.06	0.09	0.22	0.20	0.07	0.08
3	40%	0.40	0.34	0.11	0.16	0.34	0.31	0.11	0.15	0.34	0.30	0.12	0.15
4	56%	0.44	0.37	0.14	0.18	0.38	0.34	0.14	0.17	0.39	0.33	0.14	0.16
5	69%	0.46	0.39	0.15	0.19	0.40	0.35	0.15	0.18	0.41	0.35	0.14	0.17
6	80%	0.48	0.40	0.16	0.20	0.52	0.36	0.17	0.19	0.43	0.36	0.15	0.17
7	89%	0.49	0.41	0.16	0.20	0.43	0.37	0.17	0.20	0.44	0.37	0.16	0.17
8	95%	0.50	0.42	0.17	0.21	0.45	0.38	0.18	0.20	0.45	0.38	0.16	0.18
9	99%	0.51	0.43	0.18	0.21	0.46	0.39	0.18	0.20	0.46	0.39	0.16	0.18
10	100%	0.52	0.43	0.18	0.22	0.46	0.40	0.18	0.20	0.47	0.39	0.16	0.18
ACTUAL : $\overline{PI^A}$		25P	0.23			25P	0.19			25P	0.12		
		50P	0.25			50P	0.21			50P	0.15		
EQUAL : $\overline{PI^E}$		-0.04				-0.02				-0.02			

Notes:* significant at the 1% level.

Table 3 : This table represents the results from four types–the $\overline{PI^G}$, from the group correlation, $\overline{PI^H}$, from the historical correlation, $\overline{PI^A}$ from the actual correlation, and $\overline{PI^G}$ from the equal weighted portfolio for the stock of each country, using three sub-periods. MSE_k, $k = 1 \rightarrow 10$ indicates the level of improvement in predictability as compared to the initial prediction value.