Title	統計的手法を用いた株価のクラスタリング―ファットテール性を有する株価収益率の相関構造の分析・知識形成
Author(s)	磯貝, 孝
Citation	
Issue Date	2015-06
Type	Thesis or Dissertation
Text version	ETD
URL	http://hdl.handle.net/10119/12872
Rights	
Description	Supervisor:Dam Hieu Chi, 知識科学研究科, 博士

Doctoral Dissertation

Clustering of Japanese stock returns:

Statistical analysis of the correlation structure of fat-tailed returns

Takashi Isogai

Supervisor: Dam Hieu Chi

School of Knowledge Science
Japan Advanced Institute of Science and Technology

June 2015

Abstract

In this research, we study the correlation structure of fat-tailed asset returns, specifically the Japanese stock returns. A structural analysis of the Japanese stock market with a new data-oriented classification of stocks is the key component, which contributes to the knowledge creation of portfolio investment and risk management. About 1,400 stocks listed on the First Section of the Tokyo Stock Exchange are selected for correlation clustering to find a more data-oriented classification than the standard sector classification. Advanced volatility models are applied to filter the fat-tailed stock returns to avoid the possible distortion of correlation due to volatility fluctuations when estimating the correlation matrix of returns. The correlation matrix is converted to a weighted undirected network. Then, the stock returns are clustered by recursive spectral clustering with a community detection technique developed in the complex networks theory. A new method for controlling the recursive clustering process is proposed there, which is generally applicable to other fat-tailed financial asset returns. The statistical comparison between the clustering result and the standard sector classification reveals some partial linkages between them. It has been proven that portfolio risks are more efficiently controlled with the new grouping by random portfolio simulations. Classification trees are built on the clustering result with various non-price data to explore the group properties; some informative variables are successfully identified. The clustering results are also applied to the analysis of dynamic changes in the correlation matrix. A multivariate volatility model with dynamic correlation is applied to a reduced size of sample portfolios to detect any differences in changing patterns across the market and between groups. It is confirmed that the correlation intensity changes over time; a higher level of correlation is observed during the crisis periods. This study contributes to the further development of knowledge science by proposing a wide range of combinations of analytical techniques in various fields of science to extract information from the complicated high-dimensional financial data.

Keywords: stock return, correlation, fat tail, volatility, portfolio, GARCH, copula, network clustering, complex networks, modularity, risk management

Preface

This dissertation is submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Knowledge Science in the School of Knowledge Science at Japan Advanced Institute of Science and Technology (JAIST). The research described in this dissertation was conducted under the supervision of Associate Professor Dam Hieu Chi of the School of Knowledge Science of JAIST.

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree. This dissertation is the result of my own work and includes nothing which is the research work achieved in collaboration with others, except where specifically indicated.

A version of Chapter 3 has been published as

Isogai, T. (2014). A comparative simulation analysis of unconditional VaR and ES calculation methods with truncated stable distribution. In the Japanese Association of Financial Econometrics and Engineering, editor, *Risk Management* (in Japanese), 121-171, ISBN978-4-254-29022-6 C3050, Asakura Shoten, Tokyo.

Chapter 3 of this dissertation comprises the first half of the paper.

The major parts of Chapter 4 and Chapter 5 have been published as the following papers:

Isogai, T. (2014). Clustering of Japanese stock returns by recursive modularity optimization for efficient portfolio diversification. *Journal of Complex Networks*, 2(4), 557-584, doi:10.1093/comnet/cnu023.

Isogai, T. (2013). Clustering of Japanese stock returns by recursive modularity maximization. In *Signal-Image Technology Internet-Based Systems (SITIS), 2013 International Conference on*, 569-576, doi:10.1109/SITIS.2013.94, IEEE.

A version of Chapter 6 has been published as

Isogai, T. and Hieu Chi D. (2014). Building classification tree on Japanese stock groups partitioned by network clustering. In *Proceedings of Asian Conference on Information Systems*, 56-63, ISBN:978-4-88686-089-7, IEEJ.

I am responsible for the whole process of research including the data collection, analysis, and manuscript edits.

Takashi Isogai
June 2015

Acknowledgements

I would like to thank sincerely my supervisor, Associate Professor Dam Hieu Chi, for his helpful advice and comments on my research work during the whole period of my study. I would also like to thank Professor Ho Bao Tu and Professor Takashi Hashimoto for their informative comments and guidance for further improvement of this dissertation. I am also deeply indebted to the Committee members, Professor Naoshi Uchihira (JAIST), Professor Kazunori Miyata (JAIST), and Associate Professor Tomoshiro Ochiai (Otsuma Women's University) for their valuable comments and advice. Many thanks also to Associate Professor Ayumu Sugiyama (Yamanashi Eiwa College) for his kind advice and help while he was at JAIST; Dr. Toshinao Yoshiba (Bank of Japan) for his informative suggestions; Mr. Takafumi Kawasaki (PhD student at JAIST) for joining discussions on various topics. Thanks to everyone of Dam laboratory. Finally but not least, I would like to express my gratitude to my wife Junko, whose support and encouragement helped me greatly.

Contents

Contents ix

List of Figures xiii

List of Tables xv

Nomenclature xviii

1 Introduction 1

1.1 Motivation and backgrounds . . . 1

1.2 Scope and objectives . . . 2

1.3 Overview of dissertation . . . 5

2 Literature review 9

2.1 Clustering financial asset returns . . . 9

2.1.1 Clustering time series data . . . 9

2.1.2 Correlation clustering and graph clustering . . . 10

2.1.3 Community detection in complex networks . . . 12

2.2 Volatility models for fat-tailed asset returns . . . 13

2.2.1 Volatility modeling by GARCH . . . 14

2.2.2 Multivariate GARCH model . . . 14

3 Fat-tailedness of asset returns and correlation 19

3.1 Modeling fat-tailed asset returns . . . 19

3.1.1 CLT and normal distribution . . . 19

3.1.2 GCLT and stable distribution . . . 21

3.1.3 Truncation of stable distribution . . . 23

3.1.4 Fitting truncated stable distribution to stock returns . . . 24

3.1.5 Conditional and unconditional modeling approaches . . . 27

3.2 Fat-tailedness and correlation of asset returns . . . 29

4 Clustering of Japanese stock returns **31**
4.1 Data-oriented classification of stock returns 31
4.2 Stock returns data and correlation 32
4.2.1 Stock returns data 32
4.2.2 Problem of standard sector classification 33
4.2.3 Fat-tailedness and correlation of stock returns 33
4.3 Modeling stock returns by multivariate GARCH model 36
4.3.1 Processing return series by GARCH filtering 36
4.3.2 CCC–GARCH for stock portfolio 38
4.4 Clustering by recursive network clustering 43
4.4.1 Clustering by modularity maximization 44
4.4.2 Resolution limit and recursive division of network 45
4.4.3 Post-pruning tree structure 47
4.5 Statistical significance and stability of clustering 49
4.5.1 Statistical significance of clustering 49
4.5.2 Stability of clustering 51
4.6 Comparison of identified groups and sector classification 53
4.7 Technical discussion about clustering 56
4.7.1 Static correlation matrix and volatility spillover 56
4.7.2 Contribution to modularity based clustering 58

5 Analysis of stock groups and application to risk control **61**
5.1 Properties of identified stock groups 61
5.1.1 Cyclical groups 61
5.1.2 Defensive groups 62
5.2 Application to efficient portfolio risk control 65
5.2.1 Comparative analysis by random portfolio simulation 65
5.2.2 Simulation result 70

6 Building classification trees of stock groups **73**
6.1 Classification tree of stock groups 73
6.1.1 Building classification trees 74
6.1.2 Non-price data for variables 76
6.1.3 Classification tree and hierarchical clustering 76
6.2 Findings from classification trees 78
6.2.1 Important variables 78
6.2.2 Comparison with standard stock price model 80

7 Dynamic correlation of Japanese stock returns **83**
7.1 Multivariate GARCH with DCC 83
7.1.1 Dynamic conditional correlation 84
7.1.2 Modeling correlation dynamics 85
7.2 Copula for modeling dependency of residuals 86
7.2.1 Copula 87
7.2.2 Gaussian copula and Student *t*-copula 89
7.3 Model fitting of Copula–DCC–GARCH 90
7.3.1 Group portfolios and market portfolio 91
7.3.2 Two-step MLE parameter estimation 92
7.3.3 DCC parameter estimation result 94
7.4 Dynamic changes in correlation intensity 99
7.4.1 A measure of correlation intensity 99
7.4.2 Dynamic changes in maximum eigenvalues 103
7.5 Impact of correlation changes on portfolio risk 109
7.5.1 Correlation intensity and volatility during turbulent periods 109
7.5.2 Impact study of correlation changes 109
7.5.3 Simulation results 115
7.6 Technical discussion about dynamic correlation 119

8 Conclusion **121**
8.1 Summary 121
8.2 Theoretical implications 123
8.3 Practical implications 126
8.4 Knowledge science 127
8.5 Future work 129

Publications **131**

References **133**

List of Figures

3.1 Performance and log returns of the Nikkei 20
3.2 Density and QQ plot of the Nikkei 21
3.3 Cumulative density of the Nikkei and estimated truncated stable distribution . 26
3.4 QQ plot of the Nikkei and truncated stable distribution 27
3.5 Random permutation of the Nikkei returns and random samples 28
3.6 A pair of sample fat-tailed returns 30

4.1 Heatmap of sample linear correlations of Japanese stock returns 35
4.2 Volatility synchronization index of Japanese stock returns 41
4.3 Hierarchical clustering by recursive modularity maximization 46
4.4 Mean standard deviation of correlations 48
4.5 Post-pruned groupings 49
4.6 Distribution of maximum modularities by random network simulation 51

5.1 Cyclical groups: G25 (Transportation equipment–1) 63
5.2 Cyclical groups: G26 (Transportation equipment–2) 64
5.3 Defensive groups: G19 (Electric power and gas, pharmaceutical, and foods) . 66
5.4 Defensive groups: G22 (Banks) 67
5.5 Framework of two-stage sampling for random portfolio formation 68
5.6 Distribution of NHHI 70
5.7 ES of sample portfolios 71

6.1 Classification tree and hierarchical clustering 77

7.1 Maximum eigenvalue of correlation matrix: Market 104
7.2 Maximum eigenvalue of correlation matrix: Cyclical groups–1 106
7.3 Maximum eigenvalue of correlation matrix: Cyclical groups–2 107
7.4 Maximum eigenvalue of correlation matrix: Cyclical groups–3 108
7.5 Maximum eigenvalue of correlation matrix: Defensive groups–1 110
7.6 Maximum eigenvalue of correlation matrix: Defensive groups–2 111
7.7 Maximum eigenvalue of correlation matrix: Defensive groups–3 112
7.8 Maximum eigenvalue of correlation matrix: Defensive groups–4 113

List of Tables

3.1 Normality tests 20
3.2 Parameter estimation of the truncated stable distribution 25
3.3 Goodness-of-fit tests for the truncated stable and normal distribution 25
3.4 Estimation results of stable index in previous studies 26

4.1 Mean return correlation by sector 34
4.2 Distribution of correlations before and after GARCH filtering 42
4.3 Mean return correlation by group 50
4.4 Optimized modularities by number of groups 52
4.5 Clustering stability test 53
4.6 Groups and standard sector classification 55
4.7 Over-expression by standard sector classification 57

6.1 List of variables of classification tree 76
6.2 Classification tree and over-expressing sectors 79
6.3 Variable importance of classification tree 82

7.1 DCC estimation results 96
7.2 Univariate ARMA–GARCH: Six cyclical groups (portfolio return indexes) . . 97
7.3 Univariate ARMA–GARCH: Eight defensive groups (portfolio return indexes) 98
7.4 Eigenvalues of the correlation matrix 102
7.5 Maximum eigenvalues and mean conditional volatilities 114
7.6 Impact of correlation changes: VaR 117
7.7 Impact of correlation changes: ES 118

Nomenclature

Greek Symbols

e	$\simeq 2.71$ (Napier's constant)
i	$\mathrm{i}^2 = -1$ (Imaginary unit)
π	$\simeq 3.14$

Subscripts

i, j	Subscript index

Other Symbols

β	The Beta coefficient in CAPM
CDF	Cumulative distribution function
$C_X(\cdot)$	Characteristic function of a random variable
$\stackrel{\mathrm{d}}{=}$	Equality in distribution
i.i.d.	Independent and identically distributed
PDF	Probability density function
$\Pr(\cdot)$	Probability function

Acronyms / Abbreviations

ARI	Adjusted rand index
ARMA	Autoregressive moving average
CAPM	Capital asset pricing model
CCC	Constant conditional correlation
CLT	Central limit theorem

DCC	Doynamic conditional correlation
ES	Expected Shortfall
FF	Fama and French
Fig.	Figure
GARCH	Generalized autoregressive conditional heteroskedasticity
GCLT	Generalized central limit theorem
HG	Hypergeometric
HHI	Herfindahl–Hirschman index
MLE	Maximum likelihood estimation
NHHI	Normalized HHI
PBR	Price book-value ratio
PER	Price earnings ratio
TOPIX	Tokyo stock price index
VaR	Value at Risk

Chapter 1

Introduction

1.1 Motivation and backgrounds

The correlation of financial returns is one of the key issues for portfolio risk management. When measuring and controlling the portfolio risk quantitatively, we need some numerical indicator to express the total amount of investment risks explicitly. The risk measures including the Value at Risk (VaR) and Expected Shortfall (ES) are widely used by many institutional investors including banks, insurance companies, and pension funds. The risk measure is regarded as one of risk communication tools, which is used for internal risk control as well as external regulatory control. The asset return correlation is also a key factor for investment decision making, considering balance between expected return and risk. It is, therefore, meaningful to develop a way to capture the correlation of asset returns appropriately.

The correlation can be expressed mathematically in many ways. The most frequently used of these measures is the Pearson's product-moment correlation coefficient (the Pearson's linear correlation). It is, however, well-known that this type of linear correlation may not be an appropriate measure when the sample data does not satisfy some conditions. Specifically, in the case of fat-tailed financial asset returns, the linear correlation may not be an appropriate measure. The point has been much debated, but the linear correlation measure is often used even when it may cause serious distortion to capture the actual correlation.

Another difficulty for the risk measurement and risk control, is the number of assets that are traded in financial markets. We often discuss the correlation between asset classes: e.g., stocks, bonds, and exchange rates. In such cases, the number of asset classes is limited in size, although the number depends on the definition of asset class. The correlation between individual asset returns within an asset class or between multiple asset classes generally involves a large number of pairs. The dimension of correlation matrix hence becomes very large. It requires advanced mathematical and statistical techniques to embed the correlation in a financial risk model. A combination of the fat-tailedness and large number makes the matter much more complicated.

In order to reduce the complexity, dimension reduction of the number of assets is one possible solution. Grouping similar assets together is an attractive method for that purpose, which is intuitively convincing. Suppose the case of stock investment. Stocks are categorized into several groups; the business sector classification is frequently employed as the standard classification in many stock markets. The dimension of correlation matrix can be reduced from the number of stocks to the number of groups. Such business classification also provides investors with screening conditions of individual stocks when examining their investment strategy.

The grouping approach is useful for categorization; however, it can be controversial whether the stocks in the same business sector have higher correlation. If that is not true, risk modeling and portfolio optimization based on the categorization may be distorted due to the unrealistic grouping of stocks. In that regard, the consistency between the sector classification and similarities of return fluctuation will be the key issue. It may be desirable to find a more data-oriented categorization of stocks that is consistent with the actual stock price movements. That is the motivation of our research.

We focus on the Japanese stocks as the typical example of financial assets with fat-tailed returns. This paper analyses the high-dimensional correlation structure of the Japanese stock returns to find a more data-oriented and flexible grouping than the Japan standard sector classification, which is based on the industry classification. Our research objective is to propose a new approach to detect the high-dimensional correlation structure of a large number of stock returns by applying complex networks theory.

We are also interested in the analysis of stock groups, since it is often mentioned that the correlation structure reflects investors' views on the way the market moves. The information about the group structure can give us some clues to understand more about the Japanese stock market. We aim to provide useful information on the Japanese stocks for efficient portfolio diversification that contributes to better portfolio risk control and investment decision making.

The knowledge created in this research will be useful for enhancing risk management of individual financial institutions as well as macro financial stability by reducing portfolio risk at micro level. Moreover, our research can contribute to the further development of knowledge science in that it combines analytical techniques in various fields of science to extract information from the complicated high-dimensional financial data.

More detailed explanation of the scope and objectives of this paper is described in the next section.

1.2 Scope and objectives

The main purpose of this research is to study the correlation structure of fat-tailed asset returns, specifically the Japanese stock returns, for better portfolio risk management. A structural analysis of the Japanese stock market with a new data-oriented classification of stocks in comparison with the existing business sectors is the key component of this research, which contributes to the knowledge creation of stock portfolio investment and risk management. It is expected that a deeper understanding of the correlation structure helps institutional investors including financial institutions enhance the efficiency of their risk management to maintain their asset quality; the macro level financial stability, which is fundamental requirement for solid economic growth, can also be enhanced accordingly.

A global trend of regulatory improvement has been intensified after the collapse of Lehman Brothers (the Lehman shock, hereafter), and further enhancement of portfolio risk management is gaining more importance.[1] The quantitative portfolio risk management comprises two parts: modeling individual financial asset returns and building the correlation structure of those financial returns for probabilistic risk evaluation. In this research, we focus on the Japanese stocks as a typical example of financial asset; however, the methods applied here can also be applied to other fat-tailed asset returns.

It has been well recognized that financial asset returns such as stock returns exhibit fat-tailed property in that large losses occur more frequently with much higher probabilities than those expected by the normal distribution. Intuitively, the risk of financial portfolio investment quantified by some risk measure, e.g., VaR, comprises two components: variance and covariance. Correlation plays a key role to calculate the risk of portfolio investment, since it directly affects the covariance part. It is, therefore, critical to deal with both the fat-tailedness and correlation structure of returns appropriately for precise risk evaluation. There are many things to be considered including the definition of correlation. For example, the most commonly used type of correlation: the Pearson's linear correlation can be significantly distorted for fat-tailed returns, showing a much higher degree of interdependence than actually exists, especially in crisis periods. In addition, correlation related issues become more complicated when the number of assets is large as in the case of stocks. The point is well understood by risk managers of financial institutions, although the simple risk modeling that cannot deal with such complicated issues have been implemented in many cases.

We focus on the Japanese stocks listed on the Tokyo Stock Exchange as the target of our analysis. More than 3,000 stocks are listed on the Tokyo Stock Exchange, with at least 1,700 of these listed on the First Section that includes larger stocks (blue chips). However, because handling a correlation matrix at such a large scale is challenging, some dimension reduction operation is required before carrying out the empirical analysis.

In order to reduce the dimension of the correlation matrix, factor models based on the Japanese standard sector classification, which is provided by the Tokyo Stock Exchange, are frequently used for quantitative and qualitative risk analyses as well as for investment decision making. It should be mentioned that the standard stock classification is based on industrial business sector classification.[2] The sector classification is, hence, not necessarily consistent with the comovement of the stock returns. Another problem is that the number of stocks in a sector is significantly unbalanced. Thus, finding a more data-oriented classification is meaningful and beneficial. It will contribute to measuring portfolio risks more precisely as

[1] "The Lehman shock" means the financial crisis after the collapse of Lehman Brothers as well as the event itself in this paper.

[2] The 33 industrial sectors are defined by the Securities Identification Code Committee.

well as controlling the risks better for an efficient diversification.

The stock price data are easily observable; therefore, the stock returns and their correlation structure can be analyzed by a probabilistic approach with mathematical models. Data-mining of the stock returns is of our interest, in which various kinds of analytical tools including complex networks models as well as standard econometric models are to be applied. More specifically, we first examine the fat-tailedness of return distributions by fitting the return data to a non-Gaussian distribution; then, an adequate approach to deal with the correlation structure of fat-tailed returns are examined.

Correlation clustering of stock returns is the main topic of this research. A new group classification of stocks offers a fundamental framework for the analysis of transactions and price developments in the stock market. We aim to develop a method to find homogeneous groups of stocks, the returns of which are highly correlated; hence providing a new data-oriented classification of stocks. The main technical issue is the method of clustering high-dimensional returns with fat-tailedness. The classification needs to be flexible in changing the size of groups to respond to various needs of modeling. The clustering result will be utilized for describing the stock market movement in a reduced dimension.

It is informative to compare the data-oriented classification, which is more directly linked to actual asset returns, with the existing sector classification to understand the group properties of the new classification. We are also interested if it is really beneficial to us in that portfolio risk would be reduced compared with the standard classification. It is possible to prove its effectiveness by some numerical simulations that are appropriately designed.

A major drawback of the clustering based classification is that it is difficult to know the properties of the identified groups. The groups have been recognized by correlation clustering as mentioned above; the data used is just stock price data. It may be possible to get additional information of the groups by applying some machine learning approach to the data that has not been used during the clustering. The group properties are identified by applying classification tree analysis. We build classification trees that provide sorting rules to reproduce the stock groups identified by correlation clustering of the Japanese stocks. When building the classification trees, various types of non-price data are tested for the identification of effective variables. It is also informative to examine if the selection of variables is consistent with the variables included in standard stock price models. Such information may be of some help to develop a stock price model for specific stock groups.

The clustering algorithm that employs the complex networks approach requires a large scale of correlation matrix that covers the whole stock market as an input. There is a concern that the correlation matrix changes dynamically as many previous research works reported.

Specifically, what happened to the correlation matrix during crisis periods is our interest. Significant changes during crisis periods can have a large impact on the risk of stock portfolios. What happened to the correlation structure during the crisis periods has not been clarified well because of the difficulty to handle such a high-dimensional fat-tailed returns, although such information is also informative from a systemic risk viewpoint. There is only a limited number of research works on this issue with little information about empirical analysis of the Japanese stock market.

The clustering result enables further study on the dynamic changes of the correlation structure of stock returns. The concept of dynamic conditional correlation (DCC) is introduced to model the changes in the correlation matrix of stock returns over time with parsimonious parameters. The correlation matrix is separated into sub-matrices based on the stock groups identified. The within-group correlation and between-group correlation of the stock market are analyzed to know if there is any difference between the groups. The two types of dynamic correlation models should reveal any differences of correlation changes during the observation period. The information gained from our study should be highly useful for risk management.

1.3 Overview of dissertation

The remainder of this paper is organized as follows. Chapter 2 is "Literature review." In Chapter 2, we review previous research works, focusing on the two areas: clustering financial asset returns and volatility modeling for fat-tailed asset returns, which are the main components of our study.

Chapter 3 is "Fat-tailedness of asset returns and correlation." In Chapter 3, we expailn our approach to compute the correlation of fat-tailed returns. At first stage, we confirm the fat-tailedness of return distributions, which is a prerequisite for further research. Specifically, the return data of the Nikkei 225 Stock Index (the Nikkei), which is a price-weighted equity index of 225 stocks in the First Section of the Tokyo Stock Exchange, are analyzed as an example of stock returns. The log-return of the Nikkei is fitted to a fat-tailed distribution, namely, the Levy-Stable distribution, to examine the fat-tailedness quantitatively. The estimation of the Levy-Stable distribution is technically difficult, but the estimation result has good information value supported by statistical theory. We also show that the linear correlation would be significantly distorted, if it were calculated for such heavily fat-tailed returns. This problem affects the result of correlation clustering based on a set of simple linear correlations.

In addition to the fat-tailedness of stock returns, the clustered volatility of stock returns is also serious problems, making it difficult to compute the correlation of returns. Dynamic volatility fluctuations of stock returns are believed to produce a fat-tailed distribution. We adopt the generalized autoregressive conditional heteroscedastic (GARCH) model that can handle the volatility effect explicitly for modeling stock returns. Applying the GARCH model, the stock returns are separated into volatilities (scale factors) and standardized residuals. The residuals are ensured to be independent and identically distributed (i.i.d.), which makes the modeling of a correlation structure much easier.

A high dimension of the correlation matrix is another issue to be addressed when modeling the correlation structure. The estimation of a high-dimensional correlation structure has been a challenging issue of financial engineering. We adopt the second best approach that does not require any dimension reduction process by common factor modeling. We introduce an assumption of the static correlation matrix in modeling the correlation structure. The way we incorporate the difficulty is described in more details.

Chapter 4 is "Clustering of Japanese stock returns by complex networks theory." In Chapter 4, we explain divisive hierarchical clustering of stock returns. Dividing the stock market into several homogeneous groups that share the same return trend seems to be a natural and convincing approach for dimension reduction; however, there are many technically difficult issues. It has been widely recognized that standard clustering methods such as k-means cannot be applied to correlation clustering of stock returns. We aim to find "homogeneous and balanced" groups of stocks in order to overcome the high-dimensionality problem, and contribute to more efficient portfolio risk management.

We employ an advanced technique for correlation clustering of stock returns, which was originally developed in the complex networks theory. Research works in many fields such as physics, biology, and social sciences have been continued to develop efficient algorithms and to apply the theory to real problems. The correlation matrix of stock returns is converted to a weighted network, which comprises nodes and edges that connect nodes. A node represents individual stock returns; an edge represents the degree (weight) of pairwise correlation of returns. The clustering method is based on a modularity maximization algorithm, which is frequently used for community detection in a network. The community detection in the context of complex networks is a group detection algorithm: a group of nodes in a network is identified as a set of nodes such that each node is densely connected internally. Unfortunately, community detection from a dense network has serious problem called as the resolution limit. When multiple smaller groups exist in a larger group, such smaller groups cannot be detected by standard community detection techniques. The problem is one of the hottest topics of complex networks theory, and many researchers propose solutions, although none of them has not solved the problem perfectly. It is not surprising if the stock market has such nested structures, since many stocks are correlated at multiple layers of a network. We adopt a recursive optimization approach to work around the problem. We also propose a method to control the recursive network division process to produce a group of stocks that are more balanced in size than the standard sector classification. It enables to generate homogeneous and balanced groups automatically.

Our study has originality in that the correlation structure of financial asset returns are modeled in the context of complex networks; there are not many similar research papers, and our research is the first one to deal with fat-tailedness directly in correlation clustering. As such, our correlation clustering method is a hybrid of complex networks theory and financial engineering, where analytical tools in both fields are efficiently combined.

Chapter 5 is "Analysis of stock groups and application to risk control." In Chapter 5, there are there components. We first compare the identified groups with the standard sector classification. Some statistical tests have been conducted to know how they are linked. The result reveals that many groups are over-expressed by multiple sectors in both the cyclical and defensive groups. Secondly, we examine the stability of clustering. The network adjacency matrix is compared with that of a randomized network to test the significance of the group identification statistically. The stability of grouping is a crucial issue for ensuring that the result is not obtained by chance. In addition, we delve into more detailed analyses of the individual groups to clarify the group properties from a micro viewpoint. A network visualization is also employed for an intuitive description of the links between individual stocks. Lastly, a set of simulation analyses for an application of the clustering to efficient portfolio risk control are conducted. In view of the application of the correlation clustering to portfolio risk management, random portfolio simulation is conducted to see how the homogeneous and balanced grouping contributes to risk reduction of the sample portfolio. What is important here is to form a homogeneous and balanced stock groups for minimizing concentration risk. Two cases are assumed in the numerical simulation: equally weighted portfolios diversified across 33 sectors and equally weighted portfolios diversified across the groups identified by the correlation clustering. We compare the size of risk expressed as ES between the two cases to confirm that the risk is better contained with the clustering result. The degree of concentration to specific sectors or groups is also alleviated in the latter case.

Chapter 6 is "Building classification trees of stock groups." In Chapter 6, the division process of recursive clustering at multiple layers is analyzed by building a set of classification trees with non-price data. The group clustering of stock returns is based only on the stock priced data. In that sense, we can say that the grouping is data-oriented one. On the other hand, we have more data with regard to the individual companies including the sector classification information, financial performance data, and price performance data of stocks. The grouping, therefore, can be used as the training data to build classification trees of stocks, in which other external data are selected and assigned to explain each split in the tree. The classification tree is built for every hierarchical division of the stock groups achieved in Chapter 4. We discuss some technical issues to build the trees including how to control the complexity of the trees. The variables that have higher importance to explain the splits in the trees are identified. The list of these variables is summarized to compare with the list of variables of the standard stock price model, namely, the Fama and French factor model. The result confirms the consistency between the classification tree analysis and the standard stock price model. The possible application of the classification tree is discussed there.

Chapter 7 is "An empirical study of dynamic correlation of Japanese stock returns." In Chapter 7, we extend the econometric model that are used in Chapter 3 to a more flexible and advanced one in order to incorporate the dynamic correlation issue. In Chapter 3, the correlation of stock returns is a priori assumed to be static due to the restriction of modeling to work around the high-dimensionality problem. It is, however, highly possible that the correlation changes over time. We are interested in the empirical issue, since it can affect the efficiency of portfolio risk control. Unfortunately, it is still difficult to estimate the parameters of the extended model, if the model simply includes every stock. The clustering result provides useful information to reduce the dimension of the correlation matrix. First, we estimate the dynamic correlation in each stock group with a reduced number of stocks; thus, changes in within-group correlation can be detected. Second, the groups are regarded as equally weighted portfolios to calculate individual factor returns; then, a sample portfolio that covers the entire market is formed to observe changes in cross-correlation between groups by fitting the DCC–GARCH model. Several statistical and mathematical techniques are employed for the extension of the multivariate GARCH model. DCC is built into the multivariate GARCH model, where the dynamics of correlation is parameterized parsimoniously. The joint distribution of residuals are modeled by the copula approach. The merits of using DCC and copula are described in details.

The time series of correlation matrices are generated from the estimation results of the extended model. The maximum eigenvalue of the correlation matrix is employed as the indicator of the intensity of correlation. The random matrix theory helps to identify which eigenvalue we should monitor. The time series of the maximum eigenvalue clearly shows how the correlation intensity changes during the crisis period including the Lehman shock and the Great East Japan Earthquake Disaster (the Great Earthquake, hereafter). The differences across the groups are also described there. The simulation analysis is conducted to evaluate quantitatively the impact of correlation changes on the risk of sample portfolios. The simulation results will be informative to provide solid quantitative information, although the answer is generally dependent on the specifics of the target portfolios.

Chapter 8 is "Conclusion." We summarize the research results in every chapter first. Findings and discussions regarding the research topics are described briefly there. The theoretical implication section includes evaluation of the statistical and mathematical models used in this research. We propose several new methods and approaches as well as empirical studies of the correlation related issues. From a viewpoint of "knowledge creation," the combination of the different fields of statistical and mathematical techniques of our interest. Our contribution to knowledge science is discussed. As for the practical implications, we discuss how the research results should be applied to the risk management of financial institutions. Quantitative risk measuring of portfolio risk, stress testing, and portfolio optimization are listed as the areas, in which our research results and proposals can be implemented. We also discuss the topics that are very important, but not dealt with this research. Finally, we offer possible directions of future work.

Chapter 2

Literature review

In this Chapter 2, we review previous research works. We focus on the two areas: clustering financial asset returns and volatility modeling for fat-tailed asset returns, which are the main components of our study.

2.1 Clustering financial asset returns

Mantegna and Stanley [65] is one of the earlier works to detect the amount of synchronization in the dynamics of stock prices in the US markets. In order to find driving forces under such synchronization, we need to summarize properties of multiple time series. In that sense, clustering of high-dimensional financial time series has been an important research topic as well as a practical issue for financial investment and risk management.

The clustering is generally defined as "those methods concerned in some way with the identification of homogeneous groups of objects"(Arabie et al. [5]). The goal of clustering depends on the need of the project, but understanding the data is definitely included as one of the primary purposes. It should be noted that "no correct" clustering exists in any application of the categorization problem. This is why the clustering is regarded as unsupervised learning. It is, therefore, required to understand the properties of data well before clustering. This contradictory situation complicates the problem in many cases.

2.1.1 Clustering time series data

There are many clustering methods proposed including traditional partitional and hierarchical algorithms. Warren Liao [107] surveyed and summarized major works regarding the clustering of time series data in various application domains. The survey covers general-purpose clustering algorithms commonly used in time series clustering studies. Rani and Sikka [84] also surveyed the clustering of time series in various areas including finance and economics. The traditional data clustering methods that have been frequently used include k-means, k-medoids (partitional), and hierarchical clustering (agglomerative and divisive). Kaufman and Rousseeuw

[55] and Everitt et al. [31] describe these methods as well as their application in details.

As mentioned in Warren Liao [107], time series clustering requires a clustering algorithm to form clusters given a set of unlabeled data; the choice of clustering algorithm depends both on the type of data and the purpose of clustering. Rani and Sikka [84] categorized time series clustering methods into three types depending on the data object: raw data, features extracted from the raw data, and models built from the raw data. Warren Liao [107] classified similarly, also mentioning the differences in terms of discrete-valued or real-valued, uniformly or non-uniformly sampled, univariate or multivariate, balance or unbalanced. For example, the data of financial asset are observable and real-valued; the dimension is univariate and multivariate (portfolio).

Both of the two survey papers covered a wide range of research works in each area including the clustering stocks and other financial assets. The distance (or dissimilarity) measures adopted are Euclidean distance and Kullback-Liebler divergence in many cases; the clustering algorithms are mainly k-means, hierarchical clustering, and Self Organizing Map (SOM). The log-likelihood is used for distance measure in the case of model based clustering, in which Autoregressive Moving Average (ARMA) and other time series models are used. For more details about individual research works, see Warren Liao [107] and Kaufman and Rousseeuw [55].

It is known, however, that traditional clustering methods do not work well, when they are applied to detect comovements of financial time series that have fat-tailed distributions. The clustering is comprised of the two parts: the initial data processing and application of clustering methods. Even if the clustering method works well for the purpose, the processed data or extracted features may not capture the fundamental factors of similarities. Then, the clustering result would be disappointing.

2.1.2 Correlation clustering and graph clustering

From a viewpoint of financial investment and risk management, the correlations between the asset returns are the crucial factor as mentioned in Chapter 1. In this regard, clustering of time series based on a correlation matrix has been of great interest. Tumminello et al. [102] is one of the key research works of clustering based on correlation matrices in a financial field. The correlation matrix is assumed to contain useful information on hierarchical aspects of the market. Tumminello et al. [102] discussed how to define and obtain hierarchical trees, correlation based trees, and networks from a correlation matrix. Tumminello et al. [102] assume that many complex systems observed in the physical, biological, and social sciences are organized in a nested hierarchical structure; the hierarchical structure of interactions among elements strongly affects the dynamics of complex systems. Some similar previous works on the stock market are also mentioned there including Mantegna [64], Bonanno et al. [13], and Miccichè et al. [69]. First, the linear correlations of returns are calculated; then, the hierarchical clustering are performed as filtering procedures of the correlation matrix to detect statistically reliable aspects of the correlation matrix. Correlation based networks are built based on the information available from the correlation matrix as well as the hierarchical clustering result. Three types of networks are generated: the minimum spanning tree, networks average linkage minimum spanning tree, and planar maximally filtered graph. A method to associate a hierarchically nested factor model to a hierarchical tree obtained from a correlation matrix is also discussed. The approach shown in Tumminello et al. [102] is convincing and efficient for the analysis of financial returns. The hierarchical division of the correlation matrix into nested groups plays a key role in knowledge discovery from the complicated stock return data in a large dimension.

The approach can be enhanced by incorporating other clustering techniques that are developed for the clustering of a correlation matrix. That kind of clustering is called as correlation clustering. The correlation clustering has been closely related with the network (graph) theory, since the network is usually represented by a weighted or unweighted adjacency matrix that can be converted from the correlation matrix. When considering the correlation clustering of financial time series data, preprocessing the data becomes crucially important in order to avoid possible distortion caused by the fat-tailedness as mentioned earlier.

Schaeffer [87] overviewed the definitions and methods for graph clustering—finding sets of related vertices in graphs. The survey is greatly informative in that many types of definitions for what is a cluster in a graph and measures of cluster quality are reviewed there. Graphs are the structures formed by a set of vertices (nodes) and a set of edges that are connections between pairs of vertices. Graph clustering is the task of grouping the vertices of the graph into clusters in such a way that there should be many edges within each cluster and relatively few between the clusters. The adjacency matrix will be diagonalized after the reordering operation of vertices and edges; therefore, matrix diagonalization is directly related with graph clustering. This means that some efficient algorithms can be applied to solve the problem that are expressed in the form of a set of linear equations. As for the definition of a cluster in graphs, there is no universally accepted one, and the variants used are numerous. Schaeffer [87] enumerated some measures for identifying clusters including distance and similarity measures, adjacency based measures, and connectivity measures.

The methods of graph clustering is categorized in a similar way as in the traditional clustering algorithms. They are hierarchical clustering that comprises divisive and agglomerative algorithms; divisive clustering that comprises methods based on graph cuts, maximum flow, spectral component, and betweenness of graphs. The choice of the methods again depends on the purpose of the clustering and available data. Vathy-Fogarassy and Abonyi [105] also provides informative survey about graph-based clustering algorithms. Two types of categories are presented: neigborhood-graph-based clustering and minimal spanning tree based clustering. The minimal spanning tree is a weighted connected graph in which the sum of the weights is minimal, while the nearest neighbor graph links each vertex to its nearest neighbor. They also introduced some combination of graph clustering algorithms and similarity measures.

Fortunato [35] surveyed various kinds of graph clustering methods that are useful to find a "community" in the network. A cluster is called as a community in the literature on complex networks. The methods covered there are not just the ones for simple clustering purposes but are also methods to detect some structures organized by a set of similar vertices. One of the methods of the divisive graph clustering is graph spectral clustering, which is directly linked to matrix algebra. The spectral clustering includes all methods that partition vertices in a graph into clusters by using the eigenvectors of the adjacency matrix. It consists of a transformation of the adjacency matrix into a set of points in an eigenvector space; the set of points is then clustered by applying some clustering methods. The Laplacian matrix has been used to obtain a bipartition of the graph with very low cut size. The eigenvector of the second smallest eigenvalue of the Laplacian matrix (the Fiedler vector), is the key element for the bipartition. The method works quite fast, since the eigenvector can be calculated efficiently even for a large scale of matrix. For more details about the graph spectral clustering and its application, see Von Luxburg [106] and Spielman [93].

2.1.3 Community detection in complex networks

Newman [73] states "Inspired by empirical studies of networked systems such as the Internet, social networks, and biological networks, researchers have in recent years developed a variety of techniques and models to help us understand or predict the behavior of these systems." The study of complex networks have been intensively continued and extended as mentioned by Newman [73]. The analysis of the complex networks is now more focused on the real networks than before.

Albert and Barabási [4] surveyed analysis of the real networks focusing on the statistical mechanics of network topology and dynamics of the complex networks as well as some empirical data. Methods and techniques that have been developed for the analysis of complex networks can be applied to the analysis of a correlation matrix in the context of finance. That is why we are interested in those methods. In the case of financial time series, nodes correspond to financial assets and edges correspond to pairwise correlations between assets.

In relation to the correlation clustering, the concept of community structure is the key issue. The community structure, i.e., the organization of the vertices in a network, is a group of vertices that have a high density of edges within them, with a lower density of edges between groups (Newman [73]). The concept of community is almost the same as a cluster in a network. In order to identify the community, it is helpful to define a quality function that assigns a goodness score to each partition of the network. The partitions hence can be ranked based on their score given by the quality function. The most widely used quality function is the "modularity" proposed by Newman and Girvan [77]. The modularity is defined as a sort of distance between the actual density of the edges and that of a randomized network. They also proposed a spectral bisection clustering approach with a modularity based quality function. The clustering algorithm based on the modularity is a hot topic still at the moment.

Fortunato [35] summarizes discussions regarding the comparison of clustering methods based on the modularity as well as the resolution limit problem that is related to the hierarchical structure of a network. The resolution limit problem is a major drawback of modularity based clustering methods. Clauset et al. [21] analyzed the hierarchical structure of networks to find that groups found in the network correspond to known functional units over multiple scales. They emphasize that issues related to the hierarchical structure are difficult tasks when using community detection-based approaches for clustering. We will mention this problem in more details in Chapter 4.

As for empirical studies based on correlation matrices, there are some recent studies that are associated with correlation-based networks in financial and economic contexts. Fenn et al. [33] studied correlations in foreign exchange markets by clustering exchange rate time series by modularity optimization; they apply community detection methods to gain useful insights regarding the dynamic community in the network. Piccardi et al. [83] also used a modularity-based network approach to study the correlation between stock prices in the US and Italy. They investigate the stability of communities detected, and test the significance quantitatively. In Japan, Iino and Iyetomi [49] investigated a large scale of the Japanese business transaction network to detect communities by maximizing modularity. They adopted a hierarchical recursive clustering approach to work around the resolution limit problem.

In our study, a recursive clustering that is similar to the one used by Iino and Iyetomi [49] is adopted with a different method to overcome the resolution limit problems; linkages between existing sectors and groups detected are examined statistically. We use the same type of network analysis approach as in the prior research, but we focus more on the fat-tailed properties of stock returns to build a correlation network. More detailed discussion regarding our method is described in Chapter 4.

2.2 Volatility models for fat-tailed asset returns

In the previous section, we surveyed the literatures on the clustering methods and applications. It turns out that the correlation clustering with modularity optimization is beneficial to our study. It has been also stressed that preprocessing the data in advance of the calculation of the correlation matrix of returns is crucial for successful clustering. It is, therefore, important to develop a model that can handle the fat-tailedness of returns with as less distortion effect on the correlation matrix as possible. We review the literature regarding the existing volatility models in this perspective as below.

It has been widely acknowledged that the distribution of many asset returns including stock prices and exchange rates are fat-tailed as mentioned before. Another well-known feature of those returns is volatility clustering, which means that “large changes tend to be followed by large changes, of either sign, and small changes tend to be followed by small changes” (Mandelbrot [63]). Many econometric models have been proposed to replicate such features, while there is no universally accepted explanation about the mechanism of these phenomena. The main purpose of these models is to establish a mathematical formula to describe the volatility fluctuation as an approximation to the unknown true data generating process. In that sense, they are often called collectively as the volatility models.

There are two major classes of volatility models, namely, the stochastic volatility (SV) models and the GARCH models. The SV models assume that the volatility is stochastic (randomly distributed), whereas the GARCH models assume that the volatility is deterministic.[1] The dynamics of volatility hence is modeled in totally different ways. The comparison of the two models is regarded as an empirical issue. There are many empirical researches that compare the performance of the two families of models including Kim et al. [56], Fleming [34], and Franses et al. [37]; however, many of them do not ensure any unconditional superiority of either model. The model selection generally depends on the situation; therefore, the two types of models have been widely used to describe the volatility dynamics still at present.

It is often mentioned that a family of GARCH models is generally easier to estimate than the SV models, since the GARCH models have only one stochastic variable, whereas both the volatility and standardized residual (error term) are modeled as stochastic in the SV models. This point will be attractive to multivariate modeling in a large size. We focus on the GARCH family, since we use the model in our study.

More recently, the estimation of the realized volatility (RV) has garnered much interest with a growing volume of high-frequency trading data. The RV is fundamentally different from the two models in that the volatility is regarded as observable.[2]

2.2.1 Volatility modeling by GARCH

Bollerslev [9] is the seminal paper on the univariate GARCH model, which generalized the Autoregressive Conditional Heteroskedasticity (ARCH) process introduced by Engle [27] to allow for past conditional variances in the current conditional variance equation. In the GARCH model, the volatility of asset returns, which indicates the riskiness of the financial asset, is not constant (heteroskedasticity) as in the ARCH model but autoregressive, which enables to forecast its future value. The GARCH model can accurately replicate the time-varying volatility and volatility clustering to describe the dynamics of the dependency of conditional volatility.

There are many extensions of the GARCH model proposed including the integrated GARCH (iGARCH) by Engle and Bollerslev [28] and the exponential GARCH model (eGARCH) by Nelson [72]. As for the conditional distribution of the residual, extensions are also made from the normal distribution of the GARCH to other fat-tailed distributions such as the Student t distribution (Bollerslev [10]). Teräsvirta [96] well surveyed the family of univariate GARCH models including the above mentioned ones.

[1] For details of the SV models, see Taylor [95], Heston [46], and Shephard [88].

[2] McAleer and Medeiros [67] provides a good survey of major developments related to the RV.

2.2.2 Multivariate GARCH model

In addition to the volatility modeling by the univariate GARCH models, the comovements of financial returns is of great practical importance from a viewpoint of portfolio optimization and risk control. It is, therefore, important to extend the univariate GARCH models to the multivariate ones to incorporate the conditional covariances of asset returns. It should be noted that the conditional covariance matrix has to be positive definite at any time. That theoretical restriction complicates the estimation of the model; building a flexible but parsimonious model is therefore crucial.

The first generation of the multivariate GARCH is the VECH–GARCH model proposed by Bollerslev et al. [12]. The VECH–GARCH model is a natural multivariate extension of the univariate GARCH; the model is very flexible to cope with the interactions, but the covariance matrices are not ensured to be positive definite. The number of parameters increases rapidly as the number of assets increases.

The BEKK–GARCH model proposed by Engle and Kroner [29] is a restricted version of VECH–GARCH model. In the BEKK–GARCH model, the conditional covariance matrices are ensured to be positive definite by construction. The model, however, still suffers the high-dimensionality problem in terms of the number of parameters.

Alternatively, Engle et al. [30] proposed the FACTOR–ARCH model, in which some underlying factors are assumed to exist that generate the conditional covariances. The parsimonious structure with the common factors has a merit of dimension reduction. The factors are generally correlated in FACTOR-(G)ARCH type models; however, some extended versions with orthogonalized factors have been studied (Van der Weide [104] and Lanne and Saikkonen [61]).

CCC- and DCC-GARCH

The second type of multivariate GARCH model decomposes the conditional covariance matrix of returns into two parts: the conditional volatility and the conditional correlation of the residuals. Bollerslev [11] first introduced the constant conditional correlation (CCC) model, in which conditional correlation is assumed to be constant over time with only conditional volatility time-varying. CCC–GARCH is computationally attractive because the correlation matrix is constant and positive definiteness is ensured, although such assumption has often been regarded as too restrictive.

Engle [24] generalized the CCC model to make the conditional correlation matrix time-varying as in the DCC model. In DCC–GARCH, the positive definiteness of the covariance matrices depend on that of the correlation matrix of the residuals. Engle [24] introduced a dynamic proxy variable with a GARCH type structure to establish the positive definiteness of the correlation matrix. There are some other approaches to define the dynamic correlation matrix including the Varying Correlation (VC–GARCH) proposed by Tse and Tsui [101], which formulates the correlation matrix as a weighted sum of past correlations. The advantage of DCC–GARCH is that the dynamics of the correlation matrix are described by a pair of scalar parameters, assuming the same correlation dynamics for all assets. Hence, DCC-GARCH may be applied to large portfolios.

This benefit of DCC, however, becomes too restrictive when the assumption of the same correlation dynamics for all assets does not hold true, and thus many variants of DCC have been proposed. For instance, Billio et al. [8] proposed a block-diagonal structure to assume different correlation dynamics for every group of variables. The model is called as the Block DCC (BDCC) model. The number of parameters increases while more flexible settings are enabled. Cappiello et al. [20] proposed the asymmetric generalized DCC (AG–DCC) model, in which the dynamics of proxy variable depends on the level of the standardized residuals, to incorporates asymmetric effects. There are many other types of variants including STCC (Silvennoinen and Teräsvirta [89, 90]) in which the conditional correlation matrix varies smoothly between two extreme states; regime-switching DCC (Pelletier [81]) in which the conditional correlations follow a switching regime and are different across the regimes. Hafner and Franses [44] generalized DCC–GARCH as GDCC–GARCH so as to allow for asset-specific correlation sensitivities with different levels of the DCC parameters for individual assets. Aielli [1] proposed consistent DCC (cDCC) that includes a corrective step in the dynamics of the proxy variable to eliminate the estimation bias problem of the DCC parameters.

Moreover, there are also some research works on modifying the DCC type model to overcome the high-dimensionality problem. Engle and Kelly [25] introduce the Dynamic Equicorrelation model which assumes that all pairwise correlations are equal at every time period in order to estimate arbitrarily large covariance matrices with ease. Aielli and Caporin [2, 3] propose the GARCH clustering to reduce the complexity of a large scale DCC–GARCH, in which the model parameter matrices depend on the clustering of financial assets. They use a model-based clustering method, namly, the Gaussian Mixture algorithm for clustering with a vector of estimated univariate GARCH parameter as well as the correlation matrix of returns. This approach is quite similar to our clustering method in that data-driven group finding is implemented. The fundamental difference from our approach is that we do not use GARCH parameters but only use the correlation matrix.

A more complete review of multivariate volatility models is provided by Bauwens et al. [7] and Silvennoinen and Teräsvirta [91]. Caporin and McAleer [19] also provides an empirical comparative analysis with regard to the out-of-sample forecasting performance, which covers a wide range of multivariate GARCH models including BEKK, CCC, and DCC.

Copula–GARCH

Sklar [92] introduced the copula functions that enable to connect marginal distributions together to build a joint multivariate distribution. As mentioned earlier, the linear correlation coefficient is widely used as the measure of comovement. It is, however, well-known that the linear correlation is a reliable measure only for a linear dependence between the variables. The copula provides much wider options for measuring the dependency. Modeling the marginal distributions and the copula separately, rather than dealing with the multivariate joint distribution directly, provides greater flexibility to model randomness. The copula hence has been extensively used in finance. One of the earliest financial application of the copula functions was Frey and McNeil [38] that used the copula for modeling default dependencies.

There are many types of copula including the so-called implicit copula, which does not have a simple closed form, but are implied by multivariate distribution functions such as the normal and Student t distribution. For more details about mathematical definition and discussions, see Nelsen [71] and Demarta and McNeil [23].

In the context of multivariate GARCH, Jondeau and Rockinger [52, 53] developed a copula based method to incorporate time-varying skewness and kurtosis of marginal distributions with the multivariate GARCH model framework. Patton [80] applied the copula to the dynamic correlation analysis of exchange rates. Lee and Long [62] also studied a copula based multivariate GARCH, which permits modeling conditional correlation and dependence separately and simultaneously. As mentioned in Lee and Long [62], the copula based method can be widely applied to many multivariate GARCH models including DCC and CCC to link the marginals.

The separation of the marginals and joint distribution allows flexible modeling as well as more efficient estimation of parameters. In our study, Copula–DCC–GARCH is adopted to model the conditional correlation of the midle-sized of sample portfolios to observe the changes in correlation intensity over time. The flexible features of modeling of the Copula–DCC–GARCH, specifically the separation of the fat-tailedness of residuals from the tail dependence between them enables a more precise model building and parameter estimation. We try to incorporate the complexity of a large scale correlation structure not by building a more generalized complicated model that are difficult to estimate but by formulating the data structure to be organized in a reduced dimension by applying the clustering techniques, which enables to use a simple but more robust econometric model. Our proposal is a sort of hybrid approach that combines the different fields of tools; the point is that we put a priority on the data-oriented analysis. We will discuss this point further in Chapter 7.

Chapter 3

Fat-tailedness of asset returns and correlation

In this Chapter 3, first, we examine the fat-tailedness of asset returns statistically. Then, the difficulties of estimating the correlation of fat-tailed returns are discussed. Several methods of multivariate modeling of asset returns are also discussed from a viewpoint of risk measuring.

3.1 Modeling fat-tailed asset returns

3.1.1 CLT and normal distribution

In many fields of financial modeling, asset returns have been approximated by the normal distribution. In the normal distribution approximation, the central limit theorem (CLT) plays the key role. The CLT states that the distribution of the sum (or average) of i.i.d. probabilistic variables, each with a finite variance, converges to the normal distribution as the number of variables increases, no matter what the underlying distribution is.

In the context of financial asset returns, the CLT provides a theoretical foundation to approximate the distribution of returns by the normal distribution. When modeling financial asset returns, the normal distribution approximation enables easy handling of the variable in the model and computing probability under specific conditions. The risk is simply represented by the variance or standard deviation, since the normal distribution has only two parameters: the mean and variance. While financial theories and applications frequently assume that asset returns are normally distributed, substantial empirical data contradict this assumption. It has been widely recognized that the normal distribution is not necessarily appropriate; a more fat-tailed distribution than the normal should be used, considering repeatedly experienced a large scale of market disturbances in the past. The Lehman shock in 2008 was the recent historical event as a world-wide scale of financial market crash. The Great Earthquake in 2011 was another tail event, which disturbed the Japanese financial market significantly.[1]

[1] A tail event is an event that has a very low level of probability.

In this respect, we can test statistically if the normal approximation is appropriate assumption for financial modeling. Here, we select the Nikkei as a typical financial asset for modeling financial asset returns. Figure 3.1 shows the daily log returns of the Nikkei from January 2008 to August 2012.[2] Obviously, significant changes in volatility were observed during the Lehman shock in 2008 and the Great Earthquake in 2011.

The density of the returns of the Nkkei estimated by kernel smoothing as shown in Fig. 3.2 (the left pane) indicates that the return distribution is more leptokurtic and heavy-tailed than the normal distribution. The QQ plot (the right pane) also reveals that the distribution is significantly fat-tailed. The results of several normality tests (Table 3.1) also show that the distribution is far from the normal distribution.

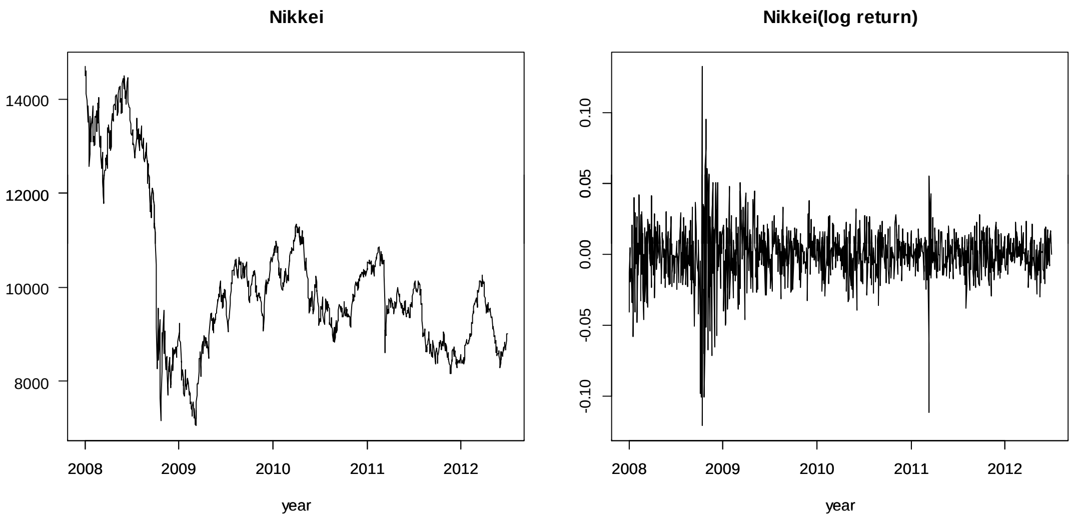


Figure 3.1 Performance and log returns of the Nikkei

Table 3.1 Normality tests

Kolmogorov-Smirnov	Anderson-Darling	Jarque Bera	Shapiro-Wilk
D = 0.074	AD = 14.07	X^2 = 2713, df = 2	W = 0.915
p-value= 9.12×10^{-6}	p-value = 5.45×10^{-7}	p-value < 2.2×10^{-16}	p-value < 2.2×10^{-16}

Note: The null hypothesis (H_0) is that the distribution is normal. When p-value is low enough, the null hypothesis can be rejected, suggesting that the distribution is not normal.

[2] A financial asset return is often measured in terms of logarithmic return.

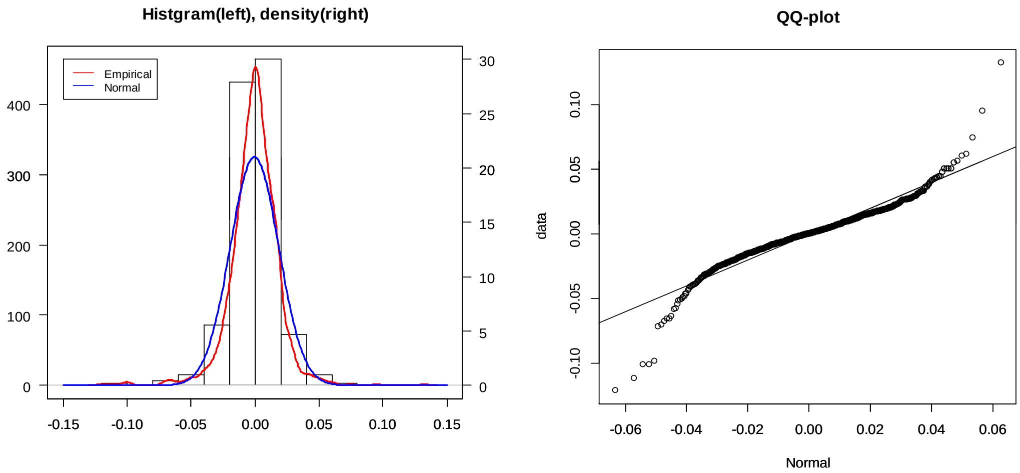


Figure 3.2 Density and QQ plot of the Nikkei

3.1.2 GCLT and stable distribution

We need some other methods for modeling financial returns, when it turns out that the normal distribution approximation is inappropriate. In order to overcome the fat-tailedness of the distribution, a more fat-tailed distribution than the normal distribution would be an alternative. There are, however, many fat-tailed distributions, including Student t distribution that belongs to the same family of elliptical distributions as the normal distribution. What is important here is the theoretical foundation for the approximation. In the case of the normal distribution approximation, the CLT provides a solid theoretical foundation. If we use the Student t distribution, we cannot rely on the CLT. We need to find an alternative distribution that has a solid theoretical foundation for the approximation.

It is known that the CLT is not valid without the finite variance assumption; the limit would then be a stable distribution (Generalized CLT, GCLT).[3] The normal distribution is a stable distribution with some specific values of distributional parameters. The stable distribution has long been studied in many fields since the study by Mandelbrot [63], and is often applied to the modeling of fat-tailed asset returns. The stable distribution has some convenient features in that it can well model asset returns with greater flexibility, replicating fatter tails and/or asymmetry observed. The stable distribution is also useful for time scaling of asset returns.

A random variable X is stable if and only if there are constants $C_n\,(C_n > 0)$ and D_n for all $n > 1$ such that:

$$S_n = X_1 + X_2 + \cdots + X_n, \quad S_n \stackrel{\mathrm{d}}{=} C_n X + D_n \tag{3.1}$$

where S_n is the sum of $X_1, \cdots, X_n$, which are independent copies of random variables X, and $\stackrel{\mathrm{d}}{=}$ denotes equality in distribution. When X is an asset return, the distribution remains the same stable distribution even after time scaling of the return from daily X to monthly S_n, although

[3] The stable distribution is also called the α-stable distribution, the Lévy alpha stable, and the Pareto Lévy stable distribution.

the location and scale of the scaled distribution are changed.[4]

The distribution of a stable random variable X is defined through its characteristic function, since a closed form of the density function of the stable distribution is not available except for some special cases mentioned later. The characteristic function can be equivalently converted to a probability density function (PDF) or a cumulative distribution function (CDF). The PDF of the stable distribution $f(x)$ is provided as the inverse Fourier transform of the characteristic function as below:

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp(-\mathrm{i}ux) C_X(u)\, du \tag{3.2}$$

where $C_X(u) = \mathrm{E}[\exp(\mathrm{i}uX)]$ is the characteristic function of the stable distribution. The characteristic function $C_X(u)$ is given by[5]

$$\begin{aligned} C_X(u) &= \mathrm{E}[\exp(\mathrm{i}uX)] \\ &= \begin{cases} \exp\Big(-\gamma^{\alpha}|u|^{\alpha}\Big[1 + i\beta\Big(\tan\frac{\pi\alpha}{2}\Big)(\mathrm{sign}(u))\Big(|\gamma u|^{1-\alpha} - 1\Big)\Big] + \mathrm{i}\delta u\Big), & \alpha \neq 1 \\ \exp\Big(-\gamma|u|\Big[1 + i\beta\frac{2}{\pi}(\mathrm{sign}(u))\log(\gamma|u|)\Big] + \mathrm{i}\delta u\Big), & \alpha = 1 \end{cases} \end{aligned} \tag{3.3}$$

where α is the index of stability, also called the tail index, β is the skewness parameter, γ is the scale parameter, δ is the location parameter, and sign(u) is defined as

$$\mathrm{sign}(u) = \begin{cases} -1, & u < 0 \\ 0, & u = 0 \\ 1, & u > 0. \end{cases}$$

A general stable distribution has four parameters $\theta(\alpha, \beta, \gamma, \delta)$ as shown in (3.3). The range of parameter α, β, and γ must satisfy the following conditions respectively:

$$0 < \alpha \leq 2, \quad -1 \leq \beta \leq 1, \quad \gamma \geq 0. \tag{3.4}$$

The two parameters α and β play a key role in determining the shape of the stable distribution. Specifically, the stability index α determines the rate at which the tails of the distribution taper off. The smaller value of α indicates that the distribution is more leptokurtic and heavy-tailed. When $\beta = 0$, the distribution is symmetric; when $\beta > 0$, the distribution is skewed to the right, which means the right tail is thicker than the left. The scale parameter γ can be any positive number. The location parameter δ shifts the distribution to the right if $\delta > 0$.

The other important properties with regard to stable distribution are as follows:

[4] The term stable distribution reflects the fact that the sum of i.i.d. random variables having a stable distribution with the same index α is again stable with the same value of α.

[5] There are multiple styles of parameterizations of a stable distribution. The definition of the parameter α is the same in any parameterization style; however, the other three parameters can be denoted differently depending on the style. For more details of the parameterization, see Nolan [78] and Misiorek and Weron [70].

Mean:	δ if $\alpha > 1$, otherwise it is undefined.
Variance:	$2\gamma^2$ when $\alpha = 2$ (the normal distribution); otherwise ($0 < \alpha < 2$) it is infinite.
Skewness, excess kurtosis:	0 if $\alpha = 2$; otherwise they are undefined.
Power law tail:	when $0 < \alpha < 2$, the tail density has asymptotic power law decay $x \to \infty \Rightarrow \Pr(X > x) \propto C \cdot x^{-\alpha},\ f(x) = \alpha C \cdot x^{-(\alpha+1)}$ where C is a constant.
Scaling:	when $0 < \alpha < 2$, the scaling constant in (3.1) is $C_N = N^{1/\alpha}$

There are three special cases where the closed forms of the density functions are available. When $\alpha = 2$ and $\beta = 0$, the closed form of the density function is the same as the normal distribution; it coincides with the fact that the stable distribution includes the normal distribution as a special case. The other two cases are the Cauchy distribution when $\alpha = 1$ and $\beta = 0$, and the Lévy distribution when $\alpha = 0.5$ and $\beta = 1$.

3.1.3 Truncation of stable distribution

The stable distribution is attractive for its flexibility in dealing with the fat-tailedness of asset returns. The feature of infinite variance when $\alpha < 2$, however, makes its practical application for modeling asset returns difficult; it may also be somewhat inconsistent with the perception of market participants. Further, the risk measure that is very sensitive to the tail of the return distribution becomes rather unstable, taking a very large value under such a circumstance.

Thus, when the stable distribution is used for modeling an asset return, some adjustments to the tail of the distribution have been frequently implemented. A simple but effective adjustment is "truncation" of the tail. The adjusted distribution is called a truncated stable distribution (Mantegna and Stanley [65]), the tail of which is bounded both below and above.[6]

Truncating the distribution at levels that correspond to the thresholds will result in zero probability of over-the-limit prices. The truncation points are a pair of constant values, which can be inferred from the properties of the financial asset or some market safety function like a circuit breaker, if any.[7] The truncated stable distribution always has a finite variance, which makes tail-sensitive modeling easier, while the distribution still has a feature of fat-tailedness.

[6] There are more flexible but complicated ways of tail adjustment of the stable distribution. The tempered stable distribution tails off smoothly with additional parameters to its characteristic function for controlling the decay of the density at the tail. The truncated stable distribution can be regarded as a special case of the tempered stable distribution. For more details about the tempered stable distribution, see Borak et al. [14], Koponen [57], and Svetlozar et al. [94].

[7] For example, a single-stock circuit breaker is triggered after a trade occurs at or outside of the applicable percentage thresholds; trading is halted thereafter. If not a circuit breaker or a similar function, the truncation points can be set at any points beyond the observed maximum loss (or gain) without upper limits.

The density function of the truncated stable distribution $f_{tr}(\cdot)$ is as follows:

$$f_{tr}(x) = \begin{cases} 0, & x > l_b \\ c_l \cdot f(x), & l_a \le x \le l_b \\ 0, & x < l_a \end{cases} \tag{3.5}$$

where $f(x)$ is the density function of the stable distribution as in (3.2), and c_l is a constant that satisfies $\int_{l_a}^{l_b} c_l \cdot f(x)dx = 1$. The truncation is symmetrical, if $l_a = -l_b$.

3.1.4 Fitting truncated stable distribution to stock returns

In order to model the fat-tailedness of asset returns by the stable distribution, the daily log returns of the Nikkei are fitted to the stable distribution. Here, the Nikkei has been chosen as a typical financial asset for modeling by the truncated stable distribution.

The parameters of the truncated stable distribution are estimated by fitting the distribution to the daily log returns of the Nikkei (at the close) during the period mentioned earlier including the two major events: the Lehman shock and Great Earthquake. The truncation points are set symmetrically at the levels that correspond to 20% change on a daily basis.[8] The truncation points are far beyond the maximum loss (gain) during the observation period.

The parameters of the truncated stable distribution can be estimated by maximum likelihood estimation (MLE) with numerical optimization. Specifically, the parameters $\theta = (\alpha, \beta, \gamma, \delta)$ are estimated by maximizing the likelihood function (3.6) with the restrictions of parameter values as shown in equation (3.4)

$$L(\theta | x_1, x_2, \dots, x_n) = \prod_{i=1}^{n} f_{tr}(\theta | x_i) \tag{3.6}$$

where $L(\cdot)$ is the likelihood function that is built using the density function (3.5) based on equations (3.2) and (3.3).

The maximum likelihood estimator (MLE estimator) $\hat{\theta}$ of θ is obtained as

$$\hat{\theta} = \arg\max \sum_{i=1}^{n} \log f_{tr}(\theta | x_i). \tag{3.7}$$

A numerical iterative method is employed for maximizing the log-likelihood, since an analytical method can hardly be applied.[9] The unique conversion of a set of parameter estimates during iterations with different initial values has been confirmed in order to avoid a local maxima problem.

[8] The maximum change of 20% is an ad-hoc setting for the truncation. The truncation is, however, conservative enough, since it seems unlikely to have such a stress level concurrently for every stock included in the Nikkei. The probability of the truncated region of the stable distribution is about 1%.

[9] We used a customized density function of the truncated stable distribution based on the “stabledist package” (https://www.rmetrics.org/) of R (http://cran.r-project.org/) in order to compute the log likelihood. As for the optimization of the log-likelihood, a new algorithm to control numerical optimization process is developed.

The results of parameter estimation are shown in Table 3.2. The stable index is smaller than 2; therefore, the tail of the distribution is more fat-tailed than in the normal distribution.[10] The skewness parameter is a small negative, which might be affected by the large negative shocks during the Lehman shock and Great Earthquake.

Table 3.2 Parameter estimation of the truncated stable distribution

	α Stability index	β Skewness	γ Scale	δ Location
Estimates	1.6555	−0.2005	0.0100	0.0006
Standard errors	(0.0460)	(0.1218)	(0.0302)	(0.0005)

Note: The negative truncation point l_a in (3.5) is −0.2, and the positive truncation point l_b is +0.2. The data periods are from January 2008 to August 2012.

Table 3.3 Goodness-of-fit tests for the truncated stable and normal distribution

(a) Truncated stable

	test stat	(p-value)
Kolmogorov-Smirnov test	D 0.016	(0.924)
Anderson-Darling test	AD 0.230	(0.979)

Note: The null hypothesis H_0 assumes the truncated stable distribution.

(b) Normal

	test stat	(p-value)
Kolmogorov-Smirnov test	D 0.075	(9.1×10^{-9})
Anderson-Darling test	AD 14.076	(5.5×10^{-7})

Note: The null hypothesis H_0 assumes the normal distribution.

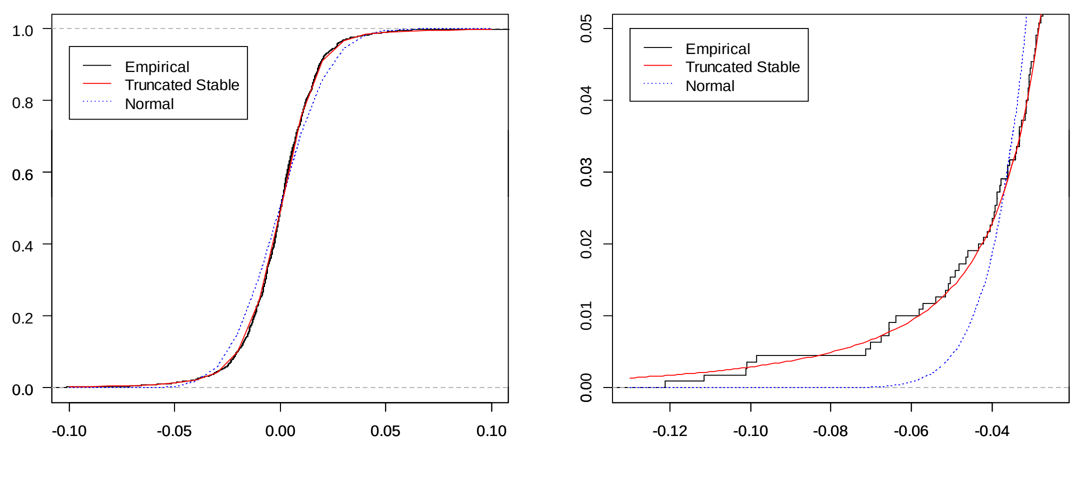
The results of the Anderson–Darling test in Table 3.3a show that the null hypothesis assuming the truncated stable distribution is not rejected, whereas one assuming the normal distribution is rejected in Table 3.3b. The CDF of the estimated truncated stable distribution and normal distribution as well as the empirical cumulative distribution function (ECDF) of the log returns of the Nikkei are plotted in Fig. 3.3 (left). The CDF of the truncated stable distribution fits fairly well to the ECDF even at the tails, while the CDF of the normal distribution does not fit the curve at all. More specifically, the enlarged figure of the negative tail in Fig. 3.3 (right) clearly shows a better fit of the truncated stable distribution than the normal distribution. The QQ plot in Fig. 3.4 also indicates a good fit of the data to the truncated stable distribution.

[10] The estimated values of the stable index in Table 3.2 are close to the findings of Kunitomo and Owada [59] in Table 3.4, although they fit the stable distribution to the daily returns of another Japanese stock index, the TOPIX (a capitalization-weighted index of all the companies listed on the First Section of the Tokyo Stock Exchange).

Table 3.4 Estimation results of stable index in previous studies

	α	Data	Data period
Kunitomo and Owada (2006)	1.6747	TOPIX, daily return	March 1990 – August 2005
Borak et al. (2005)	1.6411	DJIA, daily return	February 1987 – December 1994
Donalti (2010)	1.51	S&P500, daily return	March 1928 – October 2010
	1.56	Nasdaq100, daily return	October 1985 – August 2010

Note: The definition of the stable index is the same as in any notation formula; therefore, it can be compared between these studies.



(a) Cumulative density (b) Enlarged view of the left tail

Figure 3.3 Cumulative density of the Nikkei and estimated truncated stable distribution

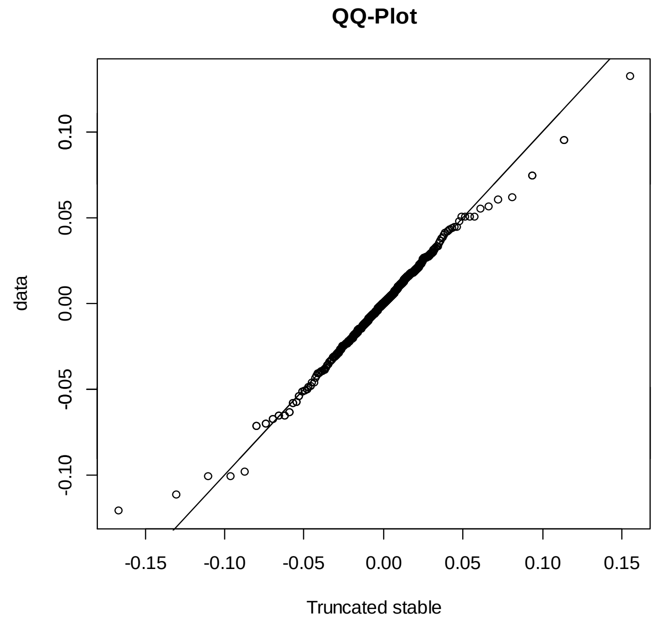


Figure 3.4 QQ plot of the Nikkei and truncated stable distribution

3.1.5 Conditional and unconditional modeling approaches

When estimating the parameters of the truncated stable distribution by using MLE as in equation (3.7), the i.i.d. condition of return distribution is implicitly assumed. It is, however, widely acknowledged that the time series of financial returns often exhibit the phenomenon of so-called volatility clustering: large fluctuations of returns tend to cluster together, resulting in persistence of high volatilities as observed in Fig. 3.1. The volatility clustering feature indicates that asset returns are not independent across time. The issue of possible long-range dependence in volatility has been frequently mentioned, although it is apparent that the returns themselves contains little serial correlation.[11]

It should be mentioned that a simple MLE estimator can be biased if the i.i.d. condition is not ensured, since the log-likelihood function cannot be expressed in the form of (3.6).[12] The changes of volatility can violate the i.i.d. assumption; therefore, both the CLT and GCLT do not hold true. This point is very important when modeling asset returns. The MLE is still valid for the parameter estimation, if the log-likelihood function is reconfigured to reflect the property of the inter-temporal dependence of asset returns.

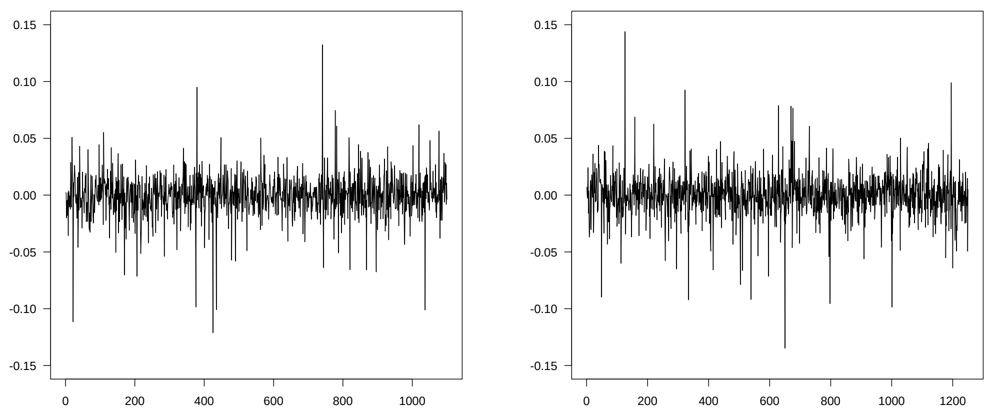
In this regard, there are two types of modeling approach: conditional and unconditional modeling. The two approaches are fundamentally different in that the conditional model captures the changes over time whereas the unconditional one does not. The conditional

[11] For more detailed information about dependence in asset returns, see Cont [22].

[12] We have conducted the Ljung-Box Q-test on a squared residual series of the daily returns. The null-hypothesis that autocorrelation with multiple lags exits cannot be rejected, which means that the independence condition may not be ensured.

model considers the dynamic features explicitly; those time-dependent changes are built-in as time-varying parameters of the distribution and/or some dynamic mechanics in the model. The advantage of the conditional model approach is its ability to capture nonlinear dynamics such as the time-dependent volatiles that are closely linked to the fat-tailedness and volatility clustering. On the contrary, the unconditional model does not consider the dynamic feature: the model assumes that the distribution of the asset return is stationary and not affected by time shift.

Figure 3.5 clarifies the meaning of the parameter estimation of an unconditional model by using MLE. The left pane of Fig. 3.5 shows the randomized return series of the Nikkei. The data is all the same as the one shown in Fig. 3.1(right), except that they don't include any time dependent information. The parameters shown in the Table 3.2 assumes an unconditional model that approximate the return distribution by the truncated stable distribution with density function shown as equation (3.5). The parameter estimation result never changes even if the order of the data is randomized, whereas the result would be different, if a conditional model were used. The right pane of Fig. 3.5 shows random samples that are drawn from the truncated stable distribution with the fitted parameters in equal size as the Nikkei data. No volatility clustering is observed there: the unconditional model cannot reconstruct any dynamic feature of stock returns.



Note: The left chart shows the randomized return series of the Nikkei. The right chart shows the sample return data that are randomly drawn from the estimated truncated stable distribution with the parameters listed in Table 3.2.

Figure 3.5 Random permutation of the Nikkei returns and random samples

Whether the dynamic features of asset returns should be considered or not depends on the purpose of modeling. If we focus on the time-dependent property, we should use the conditional model. The conditional model tends to be more complicated than the unconditional model; therefore, a choice of the unconditional model may be reasonable, unless the dynamic feature is not much concerned.

In fact, not just the conditional model but also the unconditional model has been widely used in finance. For example, the quantitative risk measurement of financial assets often assumes a model that approximate return distributions by some probabilistic distribution statically. If medium and longterm forecast of risks over some time horizon is the main target of the model, frequencies of a large loss can be estimated well by unconditional modeling, while the conditional model is useful for more responsive risk measurement. Such approaches are frequently observed in the regulatory implementation of financial risk modeling. Under the unconditional approach, it is important to find a distribution that best matches the fat-tailedness observed in actual returns, since an inadequate choice may cause serious problems, including the underestimation of risk.

3.2 Fat-tailedness and correlation of asset returns

The correlation of asset returns plays a key role for portfolio risk measurement and risk control as well as the volatilities of individual asset returns. The correlation of asset returns also has a critical role when we conduct correlation clustering of the Japanese stock returns in Chapter 4.

The most widely used statistical measure of correlation to indicate the degree of comovement of two variables is the Pearson's linear correlation. The Pearson's linear correlation $\rho_{X,Y}$ is defined as

$$\rho_{X,Y} = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\cdot\mathrm{Var}(Y)}} \tag{3.8}$$

where $\mathrm{Var}(\cdot)$ and $\mathrm{Cov}(\cdot)$ are variance and covariance operator, respectively. The Pearson's linear correlation, which is a scaled covariance, can take a range of values from $+1$ to -1.

The Pearson's linear correlation measures the strength of a linear relationship between paired data; therefore, it is also called as the linear correlation. In the context of financial asset returns, the use of the Pearson's linear correlation implicitly assumes that the two variables are normally distributed.

When the variables are not normally distributed or the relationship between the variables is not linear, the rank correlation measure such as the Spearman's rank-order correlation or the Kendall's τ correlation can be alternative measures of correlation. It is known that the rank correlation measures can be converted to the linear correlation measure with some distributional assumptions, which will be discussed later.

With regard to the normal distribution assumption of the linear correlation, the homoscedasticity that the variance of the distribution remains the same is a required property of the data. When the data show the heteroscedasticity with volatility fluctuations, the linear correlation may lead to a very different conclusion about the relationship. The existence of outliers can significantly distort the linear correlation, while the rank correlation measures mentioned above are far less sensitive to outliers.

As discussed earlier, the distribution of financial returns frequently shows fat-tailed features with many extreme values, and the volatilities change dynamically.

Figure 3.6 shows a pair of random noise series generated from the standard normal distribution. An extremely large value, which is drawn from the normal distribution with a variance of 10, is added to each series. The case simulates a financial disturbance. In this case, the Pearson's linear correlation that was almost 0 before the extreme values added increases up to 0.8 after they are added. This example reveals the difficulty of computing the Pearson's

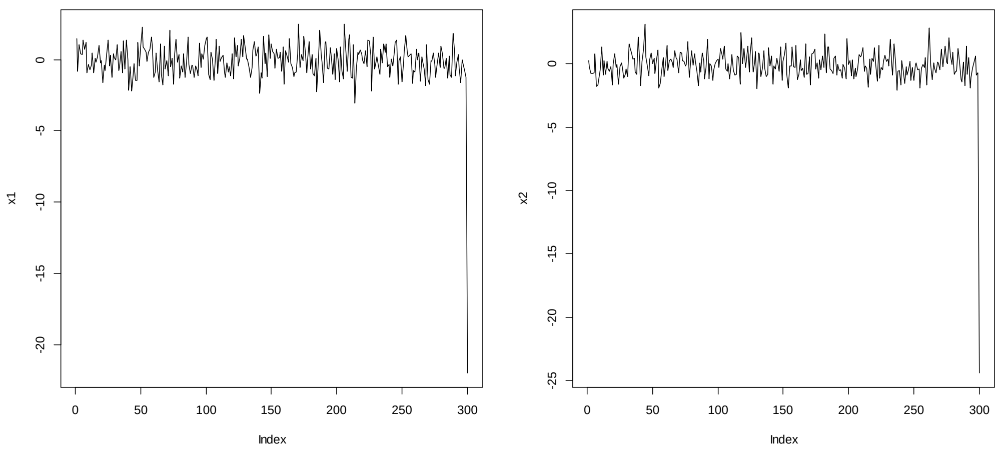


Figure 3.6 A pair of sample fat-tailed returns

linear correlation of fat-tailed asset returns that contain extreme values without filtering the data.

In relation to the model selection of conditional or the unconditional, the conditional model approach provides an useful solution in this regard. If volatility fluctuations can be properly controlled by the conditional model such as the GARCH model mentioned in Chapter 2, the standardized returns (residuals) can recover the homoscedasticity. In addition, if the volatility fluctuations are somehow related to the fat-tailedness of the distribution, the fat-tailedness can be removed or reduced by controlling volatilities. The i.i.d. assumption can also be recovered, which enables parameter fitting easier. It may be possible to get the same effect by the unconditional model, but the model would be more complicated.

As such, we adopt the conditional model, specifically the multivariate GARCH model, for preprocessing stock returns to deal with the fat-tailedness. More detailed specification of the multivariate model as well as the correlation structure will be discussed in Chapter 4 and 7.

Chapter 4

Clustering of Japanese stock returns by complex networks theory

In this Chapter 4, we analyze a high-dimensional correlation structure of the Japanese stock returns. A more data-oriented and flexible grouping than the Japan standard sector classification is identified by correlation clustering that employs a method based on the complex networks theory. The stock returns are filtered by an univariate GARCH model to separate volatilities from return series. A correlation matrix of the standardized returns is calculated, and an undirected network of the stock returns is built based on the correlation matrix. The stock returns are divided into several groups by a series of recursive spectral clustering with modularity optimization. We develop a new method to control the process of recursive clustering and determine the best group size. Our method based on community detection can be applicable for clustering other fat-tailed financial asset returns.

4.1 Data-oriented classification of stock returns

The Japanese standard stock classification, which is provided by the Tokyo Stock Exchange, is based on industrial business sector classification. The standard classification has been widely used for investment decision making and risk management. It should be, however, mentioned that the standard sector classification is not necessarily consistent with the comovement of the stock returns. Another point that should be mentioned is that the number of stocks differs significantly across the sectors.

It is very difficult to define the best grouping of the stock returns, since the preference of grouping depends on the situation. It is more productive to discuss what features of grouping are required for specific objectives. In the context of efficient portfolio diversification and risk control, a "homogeneous and balanced" stock grouping is a useful grouping. As for the homogeneity, it should be ensured that the grouping of stock returns is based on some similarity or distance measure. Specifically, the correlation of a pair of stock returns is frequently used as a distance measure, since the comovement of returns is of our interest. A homogeneous grouping, therefore, means that the stock returns in a group are highly correlated. A balanced grouping means that numbers of stocks are balanced across the groups.

If the grouping is homogeneous and balanced, it is easier to form a well-diversified portfolio with a reduced number of stocks compared with the total number of listed stocks. As such, the grouping will contribute to improving the framework of measuring and controlling portfolio risks. When modeling stock returns, a homogeneous group of stocks may provide hints to understand the common factor that drives the movement of returns. Even if the groups are all homogeneous, unbalanced groups are less useful due to possible concentration of stocks on specific groups. It may be possible to reconstruct a more balanced grouping by further divisions and mergers, keeping the homogeneity as much as possible. To reduce the concentration risk, diversification across different types of groups is essentially important; a reliable homogeneous and balanced grouping plays the key role in such operation. If the grouping is heterogeneous or unbalanced, it is more difficult to find a solution.

What is important here is a “data-oriented" classification of stock returns: partitioning a set of stock returns into homogeneous and balanced clusters in a way that comovement of returns are clearly observed. The classification needs to be flexible in changing the size of groups to respond to various needs of modeling. In the context of clustering problem, correlation clustering of stock returns can provide a reasonable solution. We can compare the clustering result with the standard classification to know if the standard classification reflects the correlation structure of stock returns, which reflects investors' views on the way the market moves. If the result of correlation clustering is preferable to the standard classification, we can use the grouping for portfolio diversification and risk control.

Technically, there are some issues to be discussed. First, how to calculate the correlation of stock returns, the distribution of which shows fat-tailedness as mentioned in Chapter 3. The second issue is the method of correlation clustering. The third issue is validation of clustering result. The last issue is a comparison of the standard classification and clustering result. These issues are discussed as below.

4.2 Stock returns data and correlation

4.2.1 Stock returns data

We focus on the Japanese stocks listed on the Tokyo Stock Exchange to find homogeneous groups of stocks, the returns of which are highly correlated. The number of stocks listed on the Tokyo Stock Exchange is more than 3,000; more than 1,700 stocks are listed on the First Section of the Tokyo Stock Exchange. The Fist Section covers most of the major companies in various industries. In order to exclude stocks that do not have a good enough level of liquidity, stocks that have complete daily price (at the close) data during the observation period from January 2008 to September 2012 are selected. The total number of stocks is 1,407 in 33 sectors. The period includes two major financial shocks: the Lehman shock in 2008 and the Great East

Japan Earthquake in 2011.[1] Daily closing prices are converted into natural logarithmic returns.

4.2.2 Problem of standard sector classification

In the context of homogeneous and balanced grouping, the number of stocks in each group is hopefully nearly equal as mentioned earlier. It is, however, apparent that the numbers of stocks are greatly different across the groups as shown in Table 4.1. The sectors including Electric Appliances and Retail Trade have more than 100 stocks, whereas the sectors including Air Transportation and Insurance have only less than five stocks. It is evident that there is a concentration of stocks on specific sectors. It is also apparent that the mean values of pairwise return correlations differ significantly across the groups as shown in Table 4.1. It may be possible to create subgroups with a higher level of mean correlations by dividing a large size group with a lower level of mean correlation.

Thus, there is a room for further improvement of the classification of stocks. That is why we are motivated to establish a more data-oriented grouping of stocks.

4.2.3 Fat-tailedness and correlation of stock returns

It is widely acknowledged that many financial asset returns including stock returns have fat-tailed return distributions as we discussed in Chapter 3. The fat-tailedness can possibly distort the correlation matrix of sample stock returns, since Pearson's linear correlation can be significantly distorted for fat-tailed returns, showing a much higher degree of interdependence than actually exists, especially in crisis periods when many assets tend to have larger volatilities.

Figure 4.1 shows the heatmap of a linear correlation matrix, in which stocks are grouped by sector as listed in Table 4.1 and are ordered by mean correlation: the group with the highest mean correlation is located on the top left. The heatmap shows that a high level of correlation is observed in many blocks. It seems that some sectors with higher mean correlations form homogeneous groups represented as dark diagonal blocks, although they are significantly unbalanced in size. It should be, however, noted that such observation may be unreliable, since the correlation matrix itself can be distorted by the fat-tailedness problem of the linear correlation.

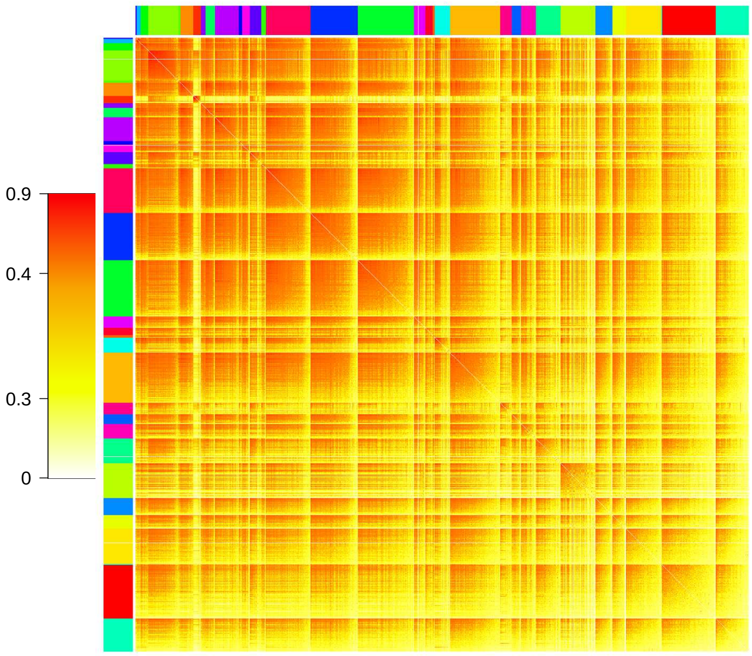
As mentioned in Section 3.2, the volatility fluctuations of stock returns should be properly controlled to deal with the fat-tailedness problem. The correlation matrix that is not affected by the fat-tailedness is essentially important to extract unbiased information on stock groups by applying correlation clustering.

[1] Data are downloaded from Yahoo Finance, Bloomberg and other sources.

Table 4.1 Mean return correlation by sector

Sector	Number of stocks	Mean correlation (standard deviation)	
Insurance	3	0.69	(0.04)
Marine Transportation	9	0.68	(0.11)
Securities	17	0.61	(0.08)
Banks	69	0.59	(0.10)
Mining	5	0.57	(0.07)
Iron and Steel	30	0.55	(0.10)
Electric Power and Gas	17	0.52	(0.17)
Rubber products	11	0.52	(0.08)
Nonferrous Metals	21	0.52	(0.12)
Transportation equipment	55	0.50	(0.10)
Oil and Coal products	8	0.48	(0.12)
Other Financing business	17	0.48	(0.13)
Land Transportation	27	0.45	(0.11)
Pulp and Paper	10	0.45	(0.09)
Machinery	103	0.45	(0.10)
Chemicals	108	0.43	(0.11)
Electric Appliances	129	0.43	(0.10)
Glass and Ceramics products	26	0.41	(0.10)
Warehousing and Harbour Transportation Services	17	0.40	(0.08)
Fishery, Agriculture, and Forestry	5	0.40	(0.09)
Real Estate	35	0.40	(0.13)
Wholesale Trade	115	0.40	(0.09)
Pharmaceutical	26	0.39	(0.10)
Precision Instruments	22	0.38	(0.10)
Textiles and Apparels	34	0.37	(0.10)
Foods	56	0.36	(0.09)
Construction	80	0.36	(0.10)
Other products	39	0.34	(0.10)
Metal products	31	0.34	(0.10)
Information and Communication	81	0.33	(0.07)
Air Transportation	2	0.32	(-)
Retail Trade	123	0.30	(0.09)
Services	76	0.27	(0.08)

Note: Sectors are sorted in a descending order of mean correlation, which is also used in sorting returns in Fig. 4.1.



Note: A colored point in the heatmap indicates a linear correlation of a pair of stock returns. Stocks are grouped by sector, and sorted in a descending order of mean correlation. Top and side color bars show sector partitions that also appear in Table 4.1. The labels of the left color key show the 0th, 25th, 75th, and 100th percentiles of the empirical distribution of correlations, respectively.

Figure 4.1 Heatmap of sample linear correlations of Japanese stock returns

4.3 Modeling stock returns by multivariate GARCH model

4.3.1 Processing return series by GARCH filtering

Volatilities of returns change over time, and shocks decay slowly (long memory effect), reflecting the autocorrelation of volatilities. This property of financial returns makes it difficult to calculate the degree of dependence of returns, since sample linear correlation can be distorted by a large scale of volatility changes.

In order to calculate a correlation matrix of fat-tailed stock returns that are less affected by volatility fluctuations, it is important to control volatilities of return series. If the returns are decomposed into deterministic volatilities and standardized random variables, it is possible to calculate a correlation matrix that is not affected by the volatility fluctuations. In that case, a correlation matrix of returns can be calculated as a correlation matrix of the standardized random variables, since the deterministic volatilities do not affect the correlation. That is clear for linear correlation defined by equation (3.8). It also makes sense for rank correlation, which depicts monotonic relationships between two variables, in that volatility fluctuations can change the rank orders of the variables.

Univariate GARCH

A GARCH model has been frequently used for such purposes (Bollerslev [9]). Volatilities are separated from return series by GARCH modeling, leaving filtered standardized residuals. Thus, pairwise correlations of the residuals are much less affected than the correlations of the sample returns. We adopt the GARCH filtering method for controlling volatilities of the stock returns to calculate a correlation matrix that is used for correlation clustering.

Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$ be a filtered probability space equipped with the filtration $\{\mathcal{F}_t\}$ of its σ-field $\mathcal{F}$ on a set Ω and probability measure $\mathbb{P}$ on $(\Omega, \mathcal{F})$. Consider an asset return as a stochastic process r_t, which is assumed to be described as

$$r_t = \mathrm{E}(r_t|\mathcal{F}_{t-1}) + \varepsilon_t \tag{4.1}$$

where $\mathrm{E}(\cdot|\cdot)$ denotes a conditional expectation operator with respect to the measure $\mathbb{P}$, $\mathcal{F}_{t-1}$ is the filtration (information set) at time $t-1$, which is generated by the observed series r_t up to and including $t-1$, and ε_t is the unpredictable residual (innovations).

In the context of conditional modeling, where the current term is assumed to be dependent on the previous period terms, the above mentioned process of the stock return can be decomposed into two parts: the conditional (time-varying) mean part and the conditional variance part. Assuming the conditional mean and volatility of r_t, equation (4.1) is written as

$$r_t = \mu_t + \sqrt{h_t} z_t \quad z_t \sim \mathcal{D}(0,1) \tag{4.2}$$

where μ_t is the conditional mean at time t, $\sqrt{h_t}$ is the volatility (as the square root of the variance), z_t is the i.i.d. standardized residual, and $\mathcal{D}$ is an arbitrary distribution with zero mean and unit variance.[2]

What is important here is that $\sqrt{h_t}$ is independent of z_t; therefore, shocks arising from changes in $\sqrt{h_t}$ can be identified and separated from residuals. The condition of i.i.d. of residual z_t is a fundamental requirement for statistical model building and parameter estimation. The conditional mean and volatility processes are modeled as below respectively.

Mean model

We adopt the ARMA model for describing the conditional mean process of stock return. The conditional mean μ_t in equation (4.2) is modeled by ARMA(P, Q) as[3]

$$\mathrm{E}(r_t|\mathcal{F}_{t-1}) = \mu_t = \mu + \sum_{i=1}^{P} a_i r_{t-i} + \sum_{j=1}^{Q} b_j \varepsilon_{tij} \tag{4.3}$$

equivalently,

$$r_t = \mu_t + \varepsilon_t = \mu + \sum_{i=1}^{P} a_i r_{t-i} + \sum_{j=1}^{Q} b_j \varepsilon_{t-j} + \varepsilon_t \tag{4.4}$$

where μ is a constant value; a_i and b_j are autoregressive and moving average coefficients, respectively.

Variance model

The variance h_t is modeled by the GARCH model. The GARCH(p, q) model is described as[4]

$$\begin{gathered} h_t = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j h_{t-j} \\ \omega > 0, \quad \alpha_i \geq 0, \quad \beta_j \geq 0 \end{gathered} \tag{4.5}$$

where ω is a constant value; α_i and β_j are moving average and autoregressive coefficients, respectively.[5] Equation (4.5) determines a time dependent structure of volatilities. The parameter α_i shows the sensitivity of h_t to previous shocks, while the parameter β_i represents the persistence of variance in the previous periods, which is related to the long memory effect of volatility.

A stationary condition with finite second moments is

$$\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j < 1. \tag{4.6}$$

[2] The conditional variance of r_t is h_t as $\mathrm{Var}\left(\mu_t + \sqrt{h_t} z_t | \mathcal{F}_{t-1}\right) = \mathrm{Var}\left(\sqrt{h_t} z_t\right) = h_t$.

[3] ARMA(AR order: P, MA order: Q)

[4] GARCH (GARCH order: p, ARCH order: q) as in Bollerslev [9]

[5] The description of the GARCH model here follows the notation of Bollerslev [9] with some modifications.

The combination of the mean and variance models are described as the ARMA(P, Q)–GARCH(p, q).

It should be mentioned that the volatility $\sqrt{h_t}$ is not probabilistic but deterministic, since it depends only on previously available values as shown in (4.5). This model setting is useful when calculating correlation of the stock returns in that any deterministic factor never affects the correlation of two probabilistic variables. Note that the correlation of returns is collapsed to the correlation of residuals as

$$\mathrm{corr}\left(r_{k\cdot t}, r_{l\cdot t}\right) = \mathrm{corr}\left(r_{k\cdot t} - \mu_{k\cdot t}, r_{l\cdot t} - \mu_{l\cdot t}\right) = \mathrm{corr}\left(\sqrt{h_{k\cdot t}}z_{k\cdot t},\ \sqrt{h_{l\cdot t}}z_{l\cdot t}\right) = \mathrm{corr}\left(z_{k\cdot t}, z_{l\cdot t}\right) = \rho_{kl\cdot t}$$

where $\mathrm{corr}(\cdot)$ is a linear correlation operator, k and l are stock IDs, and $\rho_{kl\cdot t}$ is a linear correlation coefficient.

4.3.2 CCC–GARCH for stock portfolio

Multivariate extension

In order to describe the joint process of stock returns in a portfolio, the univariate ARMA–GARCH model needs to be extended to a multivariate ARMA–GARCH model, in which r_t or z_t has a contemporaneous correlation between stocks.

Let (Ω, $\mathcal{F}$, $\{\mathcal{F}_t\}$, $\mathbb{P}$) be a filtered probability space equipped with the filtration $\{\mathcal{F}_t\}$ of its σ-field $\mathcal{F}$ on a set Ω and probability measure $\mathbb{P}$ on (Ω, $\mathcal{F}$). Now, consider multiple asset returns as a stochastic vector process $\boldsymbol{r}_t$. A vector of asset returns $\boldsymbol{r}_t$ is described as a multivariate extension of equation (4.2) in a vector form as

$$\boldsymbol{r}_t = \boldsymbol{\mu}_t + \boldsymbol{H}_t^{1/2}\boldsymbol{z}_t \tag{4.7}$$

$$\boldsymbol{\mu}_t = \mathrm{E}\left(\boldsymbol{r}_t | \mathcal{F}_{t-1}\right),\ \mathrm{E}\left(\boldsymbol{z}_t\right) = 0,\ \mathrm{Var}\left(\boldsymbol{z}_t\right) = \boldsymbol{I}_N$$

$$\mathrm{Var}\left(\boldsymbol{r}_t | \mathcal{F}_{t-1}\right) = \mathrm{Var}_{t-1}\left(\boldsymbol{r}_t\right) = \boldsymbol{H}_t^{1/2}\mathrm{Var}_{t-1}\left(\boldsymbol{z}_t\right)\left(\boldsymbol{H}_t^{1/2}\right)' = \boldsymbol{H}_t$$

where $\boldsymbol{z}_t$ is standardized residuals, $\mathrm{Var}(\cdot|\cdot)$ is a conditional variance operator, $\mathcal{F}_{t-1}$ is the filtration at time $t-1$, $\boldsymbol{H}_t$ is an $N \times N$ (N is the number of returns) symmetric positive definite matrix, which is a conditional variance–covariance matrix of $\boldsymbol{r}_t$, and $\boldsymbol{I}_N$ is an identity matrix of order N.[6]

Now, three factors must be considered to develop a multivariate model: the interactions of the individual mean processes and volatility processes as well as the correlation structure of the standardized residuals. For the interactions mentioned above, we follow the standard simplified settings that are frequently used in many previous research works in order to reduce the computational burden of the parameter estimation. The two sub-models are then implemented as the mean and volatility models.

[6] The description of multivariate GARCH models follows Bollerslev [11] and Ghalanos [40] with some modifications.

First, the conditional mean process is modeled separately for each stock return to allow us to estimate each ARMA(P, Q) model independently as

$$\boldsymbol{r}_t = \boldsymbol{\mu} + \sum_{i=1}^{P} \boldsymbol{A}_i \boldsymbol{r}_{t-i} + \sum_{j=1}^{Q} \boldsymbol{B}_j \boldsymbol{\varepsilon}_{t-j} + \boldsymbol{\varepsilon}_t \tag{4.8}$$

where $\boldsymbol{A}_i$ and $\boldsymbol{B}_j$ are $N \times N$ matrices.[7]

Secondly, the equation of the volatility dynamics is also extended to a multivariate one. As for the matrix process $\boldsymbol{H}_t$ in equation (4.7), there are generally two approaches, namely, modeling conditional covariance matrix $\boldsymbol{H}_t$ directly (e.g., VECH model or BEKK model) and modeling conditional correlation matrix indirectly with a correlation matrix (e.g., CCC and DCC model). We adopt the latter approach; therefore, equation (4.5) is transformed as a simple vector form of GARCH(p, q) model in which only the variance part of $\boldsymbol{H}_t$ is modeled explicitly as

$$\boldsymbol{h}_t = \boldsymbol{\omega} + \sum_{i=1}^{q} \boldsymbol{S}_i \boldsymbol{\varepsilon}_{t-i} \odot \boldsymbol{\varepsilon}_{t-i} + \sum_{j=1}^{p} \boldsymbol{T}_j \boldsymbol{h}_{t-j} \tag{4.9}$$

where $\boldsymbol{h}_t$ is the diagonalized matrix of $\boldsymbol{H}_t$ and $\odot$ denotes the Hadamard operator (the entry-wise product).

It is, however, difficult to estimate the parameters of a full-scale multivariate ARMA–GARCH model with cross interactions between the individual equations as described in (4.8) and (4.9), especially when the number of stocks is large. In order to work around the high-dimensionality problem, factor models are frequently used. The selection of factors itself, however, is a very difficult problem. More importantly, we are interested in the correlation structures of individual stocks; therefore, we employ a pairwise correlation approach rather than factor models.

Thus, we need to introduce some simplifications of the model for easier parameter estimation in a large dimension. In this regard, the cross effects in the mean process in equation (4.8) is omitted. The cross effects in the variance process in equation (4.9) are also omitted. Note that $\boldsymbol{A}_i$, $\boldsymbol{B}_j$ in equation (4.8) and $\boldsymbol{S}_i$, $\boldsymbol{T}_j$ in equation (4.9) are all diagonal $N \times N$ matrices, since no interaction between individual stocks is considered. These simplifications enable to estimate the univariate ARMA–GARCH model separately without considering any cross interaction between the individual mean and volatility equations.

While these model simplifications are frequently employed in the high-dimensional modeling of financial returns, such assumptions may be too restrictive in reality. As for the volatility, this model setting means that there is no volatility spillover effect between stock returns in that the past volatility changes of a stock have no effect on the current volatility of any other stocks. It should be reminded that this point is the major drawback of our multivariate ARMA–GARCH model, which may be a source of possible distortion of model estimation.

[7] The degree (P, Q) can take different values for every stock return, while the values and diagonal elements of $\boldsymbol{A}_i$ and $\boldsymbol{B}_j$ are determined empirically.

The third issue is the correlation of returns. We have two choices: unconditional and conditional correlation. The unconditional correlation approach is adopted for model simplification: the correlation matrix of z_t is fixed as non-varying $\boldsymbol{R}$ rather than $\boldsymbol{R}_t$. The choice of unconditional correlation means that we assume a static grouping of stock returns during the observation period.

The CCC by Bollerslev [11] is a typical unconditional correlation model, in which an $N \times N$ positive definite constant correlation matrix $\boldsymbol{R}$ is defined as

$$\boldsymbol{H}_t = \boldsymbol{D}_t \boldsymbol{R} \boldsymbol{D}_t = \left[\rho_{kl} \sqrt{h_{kk \cdot t} h_{ll \cdot t}}\right]_{k,\ l=1,\ \ldots,\ N} \tag{4.10}$$

where $\boldsymbol{D}_t$ is a diagonal matrix with the elements $\left(\sqrt{h_{11 \cdot t}},\ \ldots,\ \sqrt{h_{NN \cdot t}}\right)$, ρ_{kl} is the unconditional correlation of the returns between stock k and l, and $h_{kl \cdot t}$ is the conditional covariance of returns at time t. Note that the unconditional correlation of returns is the same as the unconditional correlation of standardized residuals as mentioned earlier.

$$\mathrm{corr}\,(r_k, r_l) = \mathrm{corr}\,(z_k, z_l) = \rho_{kl} \tag{4.11}$$

Thus, we focus on the correlation of standardized residuals, which are not affected by volatility fluctuation, rather than the sample correlation of returns.

Parameter estimation

The model simplification mentioned above enables two stage parameter estimation: the parameters of mean and volatility equations are estimated at the first stage; the unconditional correlation is calculated at the second stage.

The parameter sets to be estimated are the diagonal elements of $\boldsymbol{A_i}$, $\boldsymbol{B_j}$ in equation (4.8) and $\boldsymbol{S_i}$, $\boldsymbol{T_j}$ in equation (4.9). Each ARMA–GARCH model is estimated independently by using MLE at the first stage parameter estimation.

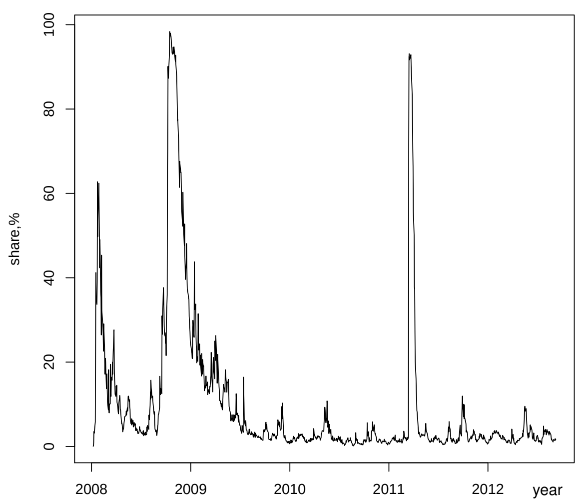
In order to build the log-likelihood function, we need to specify the distribution of z. We assume that it is one of the normal, the Student t, and the skewed t distribution. The lags in ARMA(P, Q)–GARCH(p, q) are also to be specified. The range of one to three lags are assumed.

The alternative models, therefore, exist with a combination of the distribution types, ARMA lags, and GARCH lags. An automated algorithm has been developed to select the best model in terms of the Akaike Information Criterion (AIC) with a choice of appropriate distribution and lag parameters. The Ljung-Box test is conducted for every model estimation to ensure that there is no autocorrelation in z for satisfying the i.i.d. condition. For more technical details of GARCH parameter estimation procedures, see Engle and Sheppard [26] and Silvennoinen and Teräsvirta [91].

Then, R is calculated as a set of pairwise correlations of z, which is calculated at the first stage of estimation.

As for the result of parameter estimation, the estimated GARCH lag parameters (p, q) are (1, 1), (1, 2), and (1, 3) in most cases. The ARMA lags P, Q are in the range of one to two in most cases, respectively. The result of Ljung-Box test shows that autocorrelation still exists for only less than 2% of the total stocks. The distribution of z is the Student t (68%) and the skewed t (32%), whereas only three stocks are the normal. The goodness-of-fit of the distribution has been checked by the Anderson–Darling test. We can say that the condition that z series are i.i.d. with zero mean and unit variance is almost satisfied.

As for volatility, Fig. 4.2 shows synchronization of the volatility $\sqrt{h_t}$. We define the Volatility Synchronization Index as the share of stocks, the volatility of which on a day is larger than the 95th percentile of the empirical distribution of individual volatilities during the data period. The index clearly shows market-wide synchronization of extreme volatilities in the two crisis periods, which can possibly strengthen the sample linear correlations of returns.



Note: The synchronization index takes 100% when all of the stocks take equal or greater than the 95th percentile of the individual empirical distribution of volatilities on a day. The higher the index, the more synchronized is the volatility fluctuation.

Figure 4.2 Volatility synchronization index of Japanese stock returns

Next, the unconditional correlation matrix R is calculated from z series. Here, we have a problem. Even after the significant part of volatility shocks has been removed from returns by GARCH filtering, z series still exhibit non-Gaussian features in many stocks: the fitted distribution of z is mostly non-normal. The degree of fat-tailedness of z differs significantly across stocks; therefore, sample pairwise linear correlation of z between stocks may induce distortion effects when calculating the unconditional correlation R. We, therefore, use rank correlation instead of linear correlation, specifically Kendall's τ coefficient, which is a measure of concordance between two variables. The Kendall's τ is defined as follows:

$$\tau(X, Y) = \Pr\{(X - X')(Y - Y') > 0\} - \Pr\{(X - X')(Y - Y') < 0\} \tag{4.12}$$

where (X', Y') is an independent copy of (X, Y) with continuous distributions. The Kendall's τ is less affected by outliers and is not changed by any monotonic transformation. Note that the Kendall's τ can be converted to the Pearson's linear correlation ρ as $\rho = \sin((\pi/2)\tau)$ for bivariate normal and other elliptical distributions (Greiner's relation). This transformation is sometimes used to get an initial value of correlation matrix estimation.

Table 4.2 shows the distribution of the pairwise correlations of the stock returns before and after GARCH filtering. Table 4.2a summarizes distributional information of the three types of correlation measures: the linear correlation of returns r before filtering as "Linear(a)," the linear correlation of residuals z after filtering as "Linear(b)," and the Kendall's τ of residuals z as "Rank." It clearly shows that the level of correlation has been contained by filtering at every quantile as theoretically expected. This means that the volatility effect actually increases the sample correlation of returns. The rank correlations of residuals are much lower than the linear correlation of residuals. The distribution of residuals is not normal as mentioned earlier; therefore, linear correlations of residuals are still at higher levels due to the non-normality, while the i.i.d. condition is ensured by GARCH filtering.

Table 4.2b shows the share of within-sector pairs to the total pairs at each quantile zone. There is no significant difference among the three correlation measures. The within-sector share tends to be higher at higher quantile zones. It suggests that the sector classification can be related to clusters of stock returns. We adopt the Kendall's τ to calculate the unconditional correlation matrix R mainly from a theoretical viewpoint as discussed above. The correlation matrix R is converted to an adjacency matrix for graph clustering at the next step.

Table 4.2 Distribution of correlations before and after GARCH filtering

(a) Distribution of pairwise return correlations

					Quantile					
Correlation	Mean	SD	Min	Max	25	50	75	90	95	99
Linear (a)	0.36	0.10	-0.11	0.91	0.29	0.35	0.43	0.49	0.53	0.60
Linear (b)	0.31	0.10	-0.27	0.89	0.25	0.31	0.37	0.44	0.47	0.55
Rank	0.23	0.07	-0.02	0.71	0.18	0.22	0.27	0.31	0.34	0.40

(b) Share of within-sector pairs (%)

	Quantile				
Correlation	50-75	75-90	90-95	95-100	99-100
Linear (a)	4.9	6.1	7.4	15.0	29.9
Linear (b)	4.8	6.2	7.8	16.0	32.5
Rank	4.5	6.2	7.9	16.8	33.3

Note: "Linear (a)" is linear correlation of returns r, "linear (b)" is linear correlation of residuals z after GARCH filtering, and "rank" is the Kendall's τ of the residuals z.

4.4 Clustering by recursive network clustering

Our goal is to achieve a more data-oriented grouping of stock returns than is possible with the sector classification. The fundamental requirements of such grouping are “homogeneous (in terms of return correlation) and balanced (in terms of size).”

The interaction between stock returns can be regarded as a correlation network whose adjacency matrix is constructed on the correlation matrix, which we calculated in Section 4.3. The vertices of the network are stocks that have undirected edges with weights equivalent to the return correlations between stocks. We apply a graph clustering method for an identification of groups of stock returns. Specifically, we focus on “modularity” based methods, which have been studied in the field of complex networks theory.

Modularity based methods are widely used for community detection and graph clustering, since they have a solid statistical foundation of forming homogeneous groups. It should be, however, noted that finding a clustering with the maximum modularity is NP-hard.[8] Thus, a large number of heuristic modularity maximization algorithms have been developed. A wide range of clustering methods are also implemented based on those heuristic modularity maximization algorithms. As a simple but effective method, we adopt a combination of modularity maximization and spectral clustering. We apply the method to divisive hierarchical clustering of the stock returns.

The procedure of clustering is as follows: first, the correlation matrix R is converted to a weighted adjacency matrix A. The diagonal part of R is converted from 1 to 0 for an application of modularity maximization with spectral clustering.[9] The non-diagonal part of R is preserved in A although less than the 5th percentile of the empirical distribution of correlations in every row (or column) of R is set to 0 for noise reduction. Negative values in R are also set to 0, since we focus only on positive correlations that correspond to the comovement of returns in the same direction.[10]

Correlation network or partial correlation network

Horvath [47] define an association network as a network (adjacency matrix) whose nodes correspond to vectors of features and whose adjacency matrix is based on an association measure between pairs of the vectors. We, therefore, need to specify the association measure when building the association network. The selection of the association measure depends on what we are interested in the association. In that context, a correlation network is a special case of an association network that uses a correlation coefficient as an association measure.

As for the linear correlation measure, there are two types of correlation coefficients: the standard correlation coefficient as defined equation (3.8) and the partial correlation coefficient.

[8] For the details of the mathematical proof and discussion, see Brandes et al. [15, 16].

[9] No self-linking of a node is assumed in the correlation network.

[10] The negative correlation between stocks is not considered as useful information for the correlation clustering to create homogeneous groups of stocks.

The standard correlation coefficient captures the overall relationship between the two variables, while the partial one captures the direct influence between the two eliminating the indirect influence via other variables.

The partial correlation matrix can be calculated from the inverse of the correlation matrix (precision matrix) on condition that the correlation matrix is invertible. The condition is usually satisfied in the correlation matrix of asset returns, since we assume positive definiteness of the correlation matrix. It should be mentioned that the partial correlation matrix depends on the linear correlation matrix. Horvath [47] also mentions that their construction is not robust with respect to noise; relatively small changes in the conditioning variables can lead to very different networks. It is also pointed out that multivariate regression models allow one to define a pairwise adjacency measure between variables; however, the method has similar problem as the partial correlation matrix.

The partial correlation networks are useful, if we are interested in such a direct bilateral relationship between the two variables. The point is whether we should exclude the indirect effect when calculating the correlation matrix or not. In the context of financial portfolio optimization and risk management, not the partial correlation matrix but the standard correlation matrix is used. This is clear if we think of the underlying common factors that drive asset price movements. Even if the two assets are not correlated directly but correlated by way of the common factors, such correlation is also meaningful to form a network structure. It is also difficult to compute the partial correlation for the rank correlation. Again, the selection depends on the research objective. We do not adopt that concept in our study, although the use of the partial correlation network is not excluded for other cases. For a more detailed discussion about the partial correlation network, see Horvath [47].

4.4.1 Clustering by modularity maximization

The modularity, which was introduced by Girvan and Newman [42], is one of the most frequently used quality functions for clustering and community detection in networks. The modularity is based on the idea that a random graph is expected to have no cluster structure. It is possible to detect clusters, if any, by comparing the actual density of edges in a subgraph and the density of the null model that preserves the same (weighted) degree distribution of the original graph.

Modularity Q of a weighted undirected network is defined as follows:

$$\begin{gathered} Q = \frac{1}{2W}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(A_{ij} - \frac{w_i w_j}{2W}\right)\delta\left(C_i, C_j\right) = \frac{1}{2W}\sum_{i=1}^{n}\sum_{j=1}^{n} B_{ij}\delta\left(C_i, C_j\right) \\ w_i = \sum_{j=1}^{n} w_{ij}, \quad 2W = \sum_{i=1}^{n} w_i, \quad -1 < Q < 1 \end{gathered} \tag{4.13}$$

where w_i,w_j is the sum of weights of stock i, j; A is the adjacency matrix; $2W$ is the sum of w_i over all n stocks; and $\delta(\cdot)$ is indicator function that takes 1 if both stocks are in the same class $\left(C_i = C_j\right)$, otherwise 0 (Newman [74, 75], Newman and Girvan [77]). The modularity takes the value between -1 and 1 with positive values indicating possible presence of some community structure.

The partition $C_{i,j}$ that corresponds to the maximum of Q on a given graph should be the best or at least very good partitioning; therefore, modularity maximization can be used as a quality function of clustering. There are some heuristic approaches that give fairly good approximations for the true maximization of Q, which has been proved to be an NP-hard problem. Specifically, we use the spectral bisection approach proposed by Newman [75], which is based on the properties of the spectrum of matrix B in equation (4.13).

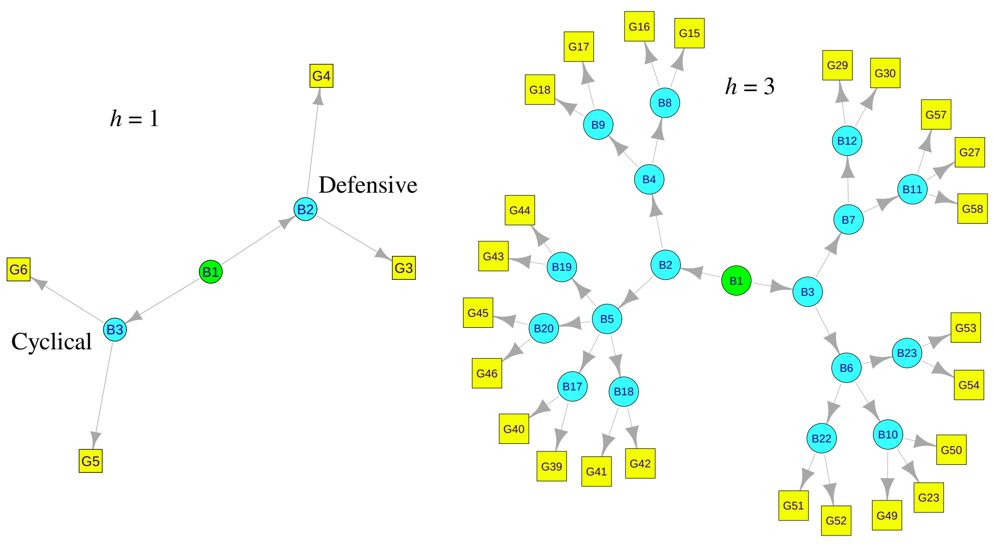
The clustering method works as follows: first, the eigenvector of B with the largest (positive) eigenvalue is computed. Secondly, sort the stock returns by the signs of the components of the eigenvector. Then, the vertices with positive components are all put in one group, the others in the other group. If B has eigenvalue 0, the maximum coincides with the partition consisting of the graph as a single cluster. We have to check if every bisection contributes to higher Q ($\Delta Q > 0$), before adopting it. Once the first bisection is adopted, the same procedure is applied to the two groups, while B is redefined in a smaller size. The process continues until any bisection gives no positive gain of ΔQ, when the best partitioning is achieved. For more details of the method, see Newman [75].

4.4.2 Resolution limit and recursive division of network

The correlation network of stock returns is almost a complete graph in which every stock is connected with edges weighted by pairwise correlations. In such a dense network, it is difficult to find the best grouping of stocks by a single operation of modularity maximization. We, therefore, use the recursive hierarchical clustering method with local modularity optimization. When partitioning restarts at a higher resolution, all inter-subnetwork edges that connect nodes over the border of subnetworks are removed. This operation transforms the adjacency matrix block diagonal. The modularity of the whole network is computed based on the transformed adjacency matrix that includes all subnetworks. The transformation of the adjacency matrix affects B matrix as well as W in equation (4.13), both of which result in changes in modularity Q of the network.

One of the difficult problems in application of modularity optimization to a real network clustering is a resolution limit problem (Fortunato and Barthélemy [36]). Many modularity based clustering methods do not work well when multilayers of subnetworks exist. Relatively small clusters cannot be detected due to the problem. It is difficult to know a priori whether a cluster is single or a combination of smaller weakly interconnected clusters. The resolution limit of modularity maximization results from the assumption of the model, the balance between the weighted sum of links in a group and the total weight of links of the whole graph. The null model of modularity assumes that every vertex can be linked without any horizontal access limit, which may be inconsistent with the real network. Small clusters tend to be merged during modularity optimization when large and small clusters coexist densely, since the plain null model does not adjust resolution for detecting both large and small clusters.

There are many methods proposed to work around the resolution limit problem including multiresolution models with tunable parameters (Arenas et al. [6], Granell et al. [43], Reichardt and Bornholdt [85]). The multiresolution approach assumes that communities can be detected at different scales of resolution. Many of these models, however, still have the resolution limit problem (Lancichinetti and Fortunato [60]). As mentioned in Lancichinetti and Fortunato [60], if the resolution limit problem is an inevitable feature of methods based on global optimization, it could be more easily circumvented by local optimization approaches. We extend the idea of the local optimization approach to work around the resolution limit problem without introducing any additional parameter for resolution adjustment. Our approach is a graph partitioning method based on recursive local modularity maximization. It works as follows: first, split the whole graph or network into small optimal subgraphs by the global modularity maximization. Then, the second round of partitioning starts from the beginning independently for each subnetwork created at the first stage partitioning. The procedure repeats recursively until all subnetworks are split into a minimal size of subnetworks. It should be noted that modularity is optimized locally in every subnetwork, and hence it cannot be maximized globally in the entire network with 1,407 stocks.



Note: The circles are intermediate subgraphs, the squares are terminal groups. h is the depth of graph layers, which works as a reset counter for recursive modularity maximization: h=1 means that B2 and B3 are identified from B1 by the first global modularity maximization, then four groups (G3, G4, G5 and G6) are identified by the second round modularity maximization after the first separation of subgraphs.

Figure 4.3 Hierarchical clustering by recursive modularity maximization

Figure 4.3 illustrates the actual divisive hierarchical clustering process, comparing the two stages at different recursive levels. The subdivision stops when the subgraph cannot be divided by modularity maximization even after the subgraph is reset as a whole graph. The reset is permitted only when the number of stocks of the subgraph is larger than 50 to avoid creating too many small groups. This can be regarded as pre-pruning of the tree structure of subnetworks. The process stops at the initial stage with just two groups. They are seemingly "cyclical" and "defensive" groups, as shown in Table 4.3, since there are significant differences in terms of TOPIX beta between the two groups. The numbers of groups identified (at layer depth h) are 2(0), 4(1), 11(2), 24(3), 40(4), and 42(5) when the process terminated. Figure 4.1 shows that stock returns are correlated marketwide beyond the border of sectors. It means that subgraphs can be interconnected at many layers, which causes the resolution limit problem: smaller groups are merged into larger groups. Our recursive method works fine to handle the situation for identifying these smaller groups at lower layers of subnetworks.

4.4.3 Post-pruning tree structure

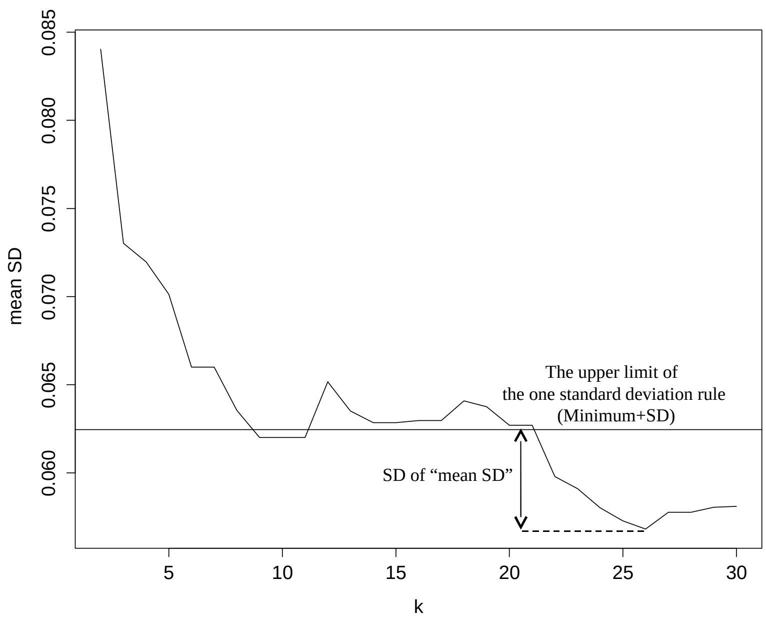
We implemented an evaluation procedure of the quality of grouping to control the processing order of the local modularity maximization. First, subgraphs are sorted in the order of division priority as mentioned below. Second, the subgraph with the highest priority is divided, and the total group counter is increased by the number of groups added by the division of the subgraph. Note that this is just a counting process to determine the order of subdivision, since divisive hierarchical clustering has already been completed. Continue these two steps until the group count matches or exceeds an upper limit k.

As for the setting of the division priority, we use a variance (or standard deviation) of numbers of stocks in individual groups as a control measure. Specifically, when determining the next subdivision, the variance of numbers of stocks after each possible subdivision is calculated for comparison; then, the division that minimizes the post division variance is selected as a prioritized one. The larger variance indicates an uneven group size distribution and possible higher concentration of stocks in specific groups. Our division sequencing rule is motivated to avoid such concentration. The relation between the variance and concentration can be clearly described by a measure of concentration such as the Herfindahl index (HHI), which is defined as the sum of squared shares of population:

$$HHI = \sum_{i=1}^{m} s_i^2 \tag{4.14}$$

where s_i is the share of group i and m is the total number of groups. If there is no concentration and every group has an equal share, HHI equals the minimum level of $\frac{1}{m}$. Our strategy of minimizing the variance of group sizes is, thus, linked to the diversification of stocks that comes with the lowest level of HHI.

The upper limit k of group size can be given externally as a setting of specific risk model, e.g., the number of risk factors. Otherwise, we need to determine k in advance of the counting process. Here, we focus on the homogeneity within a group. It is expected that the variance of correlation coefficients in a group can be decreased by subdivision of the group with local modularity maximization, reducing the heterogeneity of correlation coefficients in the group. Figure 4.4 shows how the mean of the standard deviations of correlations in individual groups changes as the upper limit of group size *k* changes. The range of *k* is set at less than 30, assuming a smaller number of groups compared with the standard sector classification, although *k* can be set at a much higher level if needed. The mean standard deviation drops sharply as *k* increases, and then remains at the same level (plateau) before it drops again.

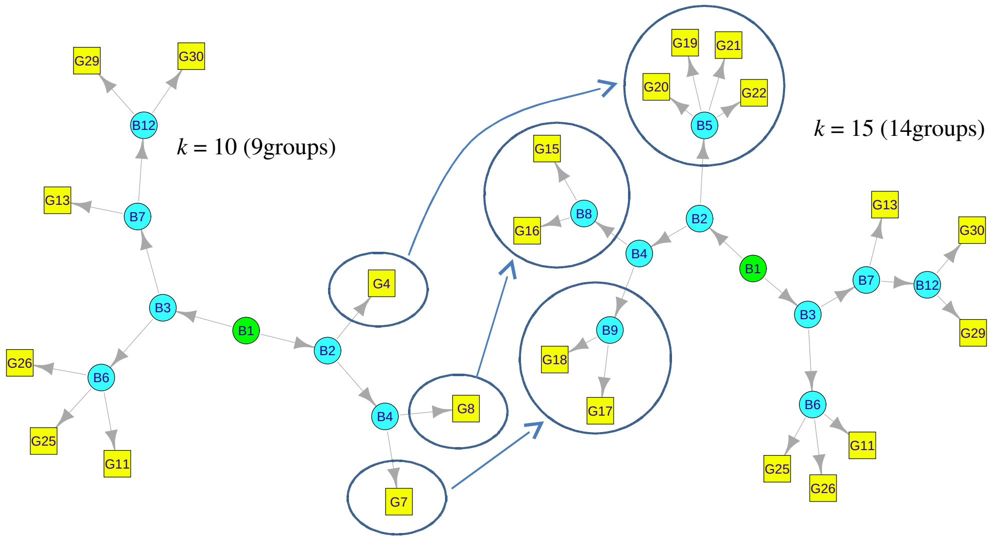


Note: For every group identified with an upper limit *k*, a standard deviation of the correlations in the group is computed; then the mean of the standard deviations (mean SD) are computed.

Figure 4.4 Mean standard deviation of correlations

We adopted "one standard deviation rule" to find the best *k*, which is frequently used for post-pruning a classification tree (Hastie et al. [45]). The horizontal line in Fig. 4.4 shows the level of the minimum plus one standard deviation. Possible choices are from *k* =10 to 20; we selected *k* =15 in-between of the range. The 14 groups are finally identified, as shown in the right pane of Fig. 4.5.

Table 4.3 shows a list of individual group IDs with numbers of stocks, mean correlations and TOPIX betas. The first 6 groups, which are all included in the subnetwork B3 as shown in Fig 4.5, are defined collectively as the cyclical group, since they have a higher level of TOPIX betas; the second group included in B2 is defined collectively as the defensive group with lower TOPIX betas. The levels of mean correlations are not so different between the two groups, whereas the levels of TOPIX betas are significantly different, as shown in Table 4.3. The individual group sizes in the cyclical group are larger than the defensive group,



Note: All subdivisions have been completed, as shown in Fig. 4.3. The subdivisions of the network represented by from B to G arrows are ordered by the priority of acceptance as explained in Section 4.4.3. k =15 (14 groups) means that the upper limit of group size is set at 15, and 14 groups exist when the last subdivision is accepted. G11, G25 and G26 under B6 correspond to B10, B22, and B23 in Fig. 4.3, respectively.

Figure 4.5 Post-pruned groupings

while the cyclical group has only 6 groups less than 8 groups in the defensive group. Such contrasts between the two major groups give us an impression that the cyclical group is more concentrated with some shared properties than the defensive group. Detailed analysis of the grouping is described in Section 4.6.

4.5 Statistical significance and stability of clustering

4.5.1 Statistical significance of clustering

Modularity is used as the quality measure to search for a better partitioning in recursive clustering; the value of modularity itself is not necessarily the final target. It is, however, still important to confirm statistically if the clustering result is truly significant or just a coincidence. For that purpose, z-score is frequently used to detect a significant deviation of modularity from the mean of random networks (Karrer et al. [54], Opsahl [79], and Piccardi et al. [83]).

We calculate the z-score of modularity as follows: first, the network that has been separated into subnetworks by recursive clustering is randomized by reshuffling edges within the individual subnetworks keeping the same edge weight distribution. The subnetworks structure is fixed so as not to be affected by the transformation of the adjacency matrix. Then the grouping with the highest modularity is identified by modularity maximization for every randomized subnetwork to calculate modularity of the network. This procedure is repeated 300 times to calculate the mean and standard deviation of the modularity distribution, from which the

Table 4.3 Mean return correlation by group

Group ID	Number	Correlation	(SD)	TOPIX beta	
G25	89	0.39	(0.06)	1.05	0.77 … Cyclical
G11	148	0.36	(0.07)	0.96	
G26	110	0.32	(0.04)	0.82	
G30	112	0.25	(0.05)	0.67	
G29	104	0.23	(0.05)	0.56	
G13	187	0.19	(0.08)	0.58	
G22	57	0.39	(0.12)	0.65	0.48 … Defensive
G15	69	0.26	(0.06)	0.63	
G21	57	0.26	(0.05)	0.58	
G16	124	0.25	(0.05)	0.47	
G17	95	0.24	(0.05)	0.40	
G19	72	0.22	(0.09)	0.46	
G18	99	0.21	(0.04)	0.31	
G20	84	0.18	(0.06)	0.37	
Total	1,407				

Note: The groups are sorted in a descending order of mean correlation. TOPIX beta is calculated by a robust MM-estimator for individual stocks with one factor (the TOPIX) linear model.

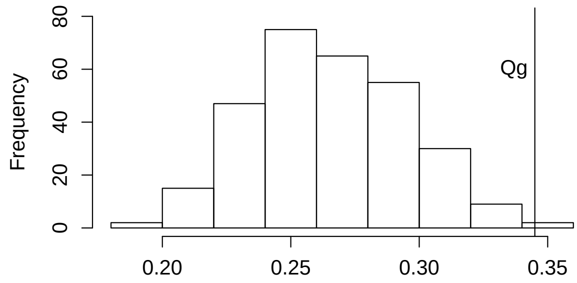
z-score of the clustering result is calculated.

Figure 4.6 shows the simulation result. The modularity of the clustering result (Qg for 14 groups) is 0.34 and its z-score is 2.55. In the case of real networks, it is often said that the values of modularity for networks with strong community structures typically fall in the range from about 0.3–0.7 (Newman and Girvan [77]). The modularity shows significance of the grouping compared with random networks as in Fig. 4.6b, although it is not at a high level. The clustering result depends on the groupings at lower resolutions with a smaller number of groups, which are created in the course of recursive clustering. Modularity of groupings at intermediate levels: five groups with three subnetworks and nine groups with five subnetworks show the significance of groupings as well. These results show that the grouping identified by recursive modularity optimization has some meaning, which is far from just a coincidence.

We also compare the clustering result with the one based on the linear correlation without GARCH filtering on the same condition. Figure 4.6b shows that the modularity level is not much different, but the z-score is lower than the one based on the rank correlation of filtered residuals. The result is supportive for our choice of the correlation matrix based on the rank correlation of filtered residuals; however, it should be noted that the theoretical consistency with the i.i.d. assumption of residuals is the main reason of the selection as mentioned in Section 3.2 and 4.3.

Table 4.4 shows optimized modularities by number of groups that are identified by recursive clustering. The z-score and p-value are calculated by simulation in the same way as in Fig.

4.6b. The modularity is higher than 0.3, except the case that the group number is 2; the null hypothesis is rejected at higher than the 90% confidence level in most cases. The low modularity for the first division of the cyclical and defensive groups means that the market-wide comovement of stock returns exist. A high level of covariance of stock returns between the two large groups contributes to the low modularity; however, z-score is high enough to reject the null hypothesis. The two large categories, therefore, are meaningful regardless of the low modularity.



(a) Optimized modularities ($Qsim$ for 14 groups)

(b) Z-score of modularity

		Qsim				Qg		
		Min	Max	Mean	SD	Qg	z-score	p-value
	5 groups	0.21	0.30	0.26	0.02	0.42	7.11	< 0.001
Rank cor	9 groups	0.21	0.37	0.29	0.04	0.43	3.75	< 0.001
	14 groups	0.19	0.34	0.27	0.03	0.34	2.55	0.005
Linear cor	14 groups	0.24	0.38	0.31	0.03	0.37	2.09	0.018

Note (a): The network is randomized by reshuffling links between vertices keeping the same edge weights distribution under the same subnetwork structure. It means that the network is randomized only at the current resolution or subnetwork level. The same procedure is repeated for 300 times.

Note (b): "Rank cor" indicates that the correlation matrix is based on Kendall's τ of residuals after filtering; "linear cor" indicates that correlation matrix is based on linear correlation of returns without any filtering. $Qsim$ is the modularity calculated for the simulation; Qg is the modularity of the recursive clustering result. z-score is calculated as $(Qg - \text{mean}(Qsim))/\text{sd}(Qsim)$, and p-value is calculated assuming normal distribution.

Figure 4.6 Distribution of maximum modularities by random network simulation

4.5.2 Stability of clustering

The result of divisive hierarchical clustering by recursive modularity maximization and post-pruning should remain stable even when the number of stocks changes. If the stocks

Table 4.4 Optimized modularities by number of groups

k	Number of groups	Q_g	z-score	p-value
2	2	0.039	210.783	0.000
3	3	0.368	8.646	0.000
4	4	0.265	1.751	0.040
5	5	0.418	7.110	0.000
6,7	6	0.539	7.647	0.000
8	8	0.398	3.582	0.000
9,10,11	9	0.425	3.750	0.000
12	12	0.337	1.481	0.069
13	13	0.383	3.729	0.000
14,15	14	0.342	2.550	0.006
16,17	16	0.302	1.193	0.116
18	18	0.372	1.775	0.038

Note: k is an upper limit of group numbers, which is set in advance of clustering as in Figure 4.4. Qg is the modularity of the recursive clustering result, which is calculated as described in Note (b) of Figure 4.6b.

that belong to a group do not belong to the same group when the same clustering method is applied to a reduced size of stock returns, the clustering method may have a stability problem. We have tested if such discrepancies occur by simulation. The grouping obtained by clustering the whole 1,407 stock returns is regarded as a reference set of partition, and groupings obtained by clustering random sampled stocks in various sizes are compared with the reference set. The upper limit of the number of groups is set as the same number: k =15. We simulated this process 100 times. The degree of agreement between the two sets of groupings results is measured by two methods: the Adjusted Rand Index (ARI) and Fisher's exact test. The ARI is a measure of agreement between two partitions, which is frequently used in clustering validation (Hubert and Arabie [48]). The ARI takes the value between -1 and 1; the ARI of randomly selected groups is expected to be 0. The ARI can be calculated, even if the sizes of two groups are different. The ARI is expected to increase as the number of sampled stocks increases. Fisher's exact test is a statistical test to determine if there are non-random associations between two categorical variables. The null hypothesis is that there is no correlation between the two results of clustering. The hypothesis test is based on a multivariate generalization of hypergeometric probability function. The labels of sample groups are changed so as to have as many overlaps as possible. The confidence level is set at 99%, and p-value is computed for every test. Table 4.5 summarizes the result of the clustering stability test. It shows that the ARI remains at higher levels; the mean value of ARI increases as the sample size increases. The null hypothesis is rejected for all Fisher's exact tests. The simulation result shows stability of our divisive hierarchical clustering method regardless of the size of stocks included in the sample. We also test if clustering results change for different sample periods. The data period is split into two subperiods: the first half (575 days from

January 2008) and the second half (570 days from May 2010) without overlap. Two adjacency matrices are built independently from 570 randomly sampled stocks; all other settings are the same as the simulation above. The ARI (=0.67) of the two clustering results is at high level. It seems that the group structure of stock returns has not been changed significantly between the first and the second half, although a more precise analysis with longer time periods is required for detecting any possible changes.

Table 4.5 Clustering stability test

Number of sample stocks	Mean ARI (SD)	Rejection of null hypothesis in Fisher's exact test
600	0.45 (0.05)	100/100
700	0.47 (0.06)	100/100
800	0.52 (0.06)	100/100
900	0.56 (0.05)	100/100
1000	0.59 (0.05)	100/100

Note: ARI is computed between a grouping of sampled stocks and the grouping of the whole stocks for 100 cases. k is set at 15 for all the cases. 100/100 of Fisher's exact test means rejection of the null hypothesis in all cases.

4.6 Comparison of identified groups and sector classification

We compare the 14 groups identified by the recursive clustering and the standard sector classification to examine if the standard sector classification is preserved in the grouping. If so, the grouping is at least partly a restructured form of some sector classification. We build a contingency table of group and sector. If the sector classification is well preserved in the 14 groups on the contingency table, these groups are well characterized by the sector information. We conduct hypergeometric test (HG test) for every pair of group and sector to determine which group is well characterized (over-expressed) by which sector.

The test assumes that the number of stocks that belong to a specific sector follows the hypergeometric distribution, when randomly sampled from a group without replacement. If the number of stocks that belong to a sector is larger than the expected value computed from hypergeometric distribution, the sector over-expresses the group. The p-value of the HG test is calculated as $p = 1 - cdf(q-1, m, n, k)$, where cdf is the CDF of hypergeometric distribution, q is the number of stocks of the sector in the group, m is the number of stocks of the sector, n is the number of stocks that belong to other sectors, and k is the number of stocks in the group. The confidence level is set at 99%. It should be noted that the threshold level should be adjusted conservatively, since multiple hypotheses HG tests are conducted for all sectors at the same time: the threshold is divided by the number of sectors (0.01/33) as the Bonferroni correction. For more details about the HG test and threshold level adjustment, see Tumminello et al. [103] and Horvath [47].

Table 4.6 summarizes the results of the HG tests. Sectors that over-express a group are listed for every group by row. The table shows that the 11 groups out of the 14 groups (except G16, G17, and G29) are over-expressed by one or more sectors.[11] Many groups are over-expressed by multiple sectors; they form communities of multiple sectors. Transportation Equipment, which includes automobile companies, over-expresses two groups in the cyclical group. Construction over-expresses two groups: one in the cyclical and the other in the defensive. On the contrary, in addition to the above-mentioned three groups that are not over-expressed, some groups are poorly over-expressed by sector, e.g., G30 that has 112 stocks is over-expressed only by Other Financing Services that has only 8 stocks. The sector classification seems to be informative to understand the group property; but only a limited number of sectors are meaningful.

[11] A stock group is identified and recognized by its unique ID (G$**$ as listed in Table 4.3 or B$**$ as shown in Fig. 4.5) hereafter.

Table 4.6 Groups and standard sector classification

Group ID		Over-expressing sectors							
Cyclical									
G25	(89)	Electric Appliances	(38)	Transportation Equipment	(16)	Precision Instruments	(7)		
G11	(148)	Iron and Steel	(16)	Nonferrous Metals	(11)	Marine Transportation	(6)	Securities	(9)
G26	(110)	Transportation Equipment	(28)	Machinery	(26)	Rubber Products	(6)		
G30	(112)	Other Financing business	(8)						
G29	(104)	-							
G13	(187)	Construction	(33)	Textiles and Apparels	(14)	Real Estate	(13)		
Defensive									
G15	(69)	Construction	(25)						
G16	(124)	-							
G17	(95)	-							
G18	(99)	Information and Communication	(15)						
G22	(57)	Banks	(44)						
G21	(57)	Information and Communication	(15)	Land Transportation	(9)				
G19	(72)	Electric Power and Gas	(16)	Pharmaceutical	(15)	Foods	(15)	Land Transportation	(7)
G20	(84)	Retail Trade	(40)	Foods	(14)				

Note: ( ) shows number of stocks in a group or a sector.

The same HG test with different grouping sizes has been conducted to examine if changes in the group size affect which sector over-expresses which group. The upper limit of group size (k) is set in a range from 5 to 30 by 5. Table 4.7 summarizes the results of the test: the individual results of three cases and the mean of all cases. The number of groups that are over-expressed by each sector is listed in the three columns. The number of groups over-expressed increases as the group size increases for some sectors; however, there is no significant difference between the different group sizes. The sectors such as Transportation Equipment, Construction, Electric Appliances, Pharmaceutical, Banks, Retail Trade, and Foods over-express one or more groups. On the contrary, the sectors in a small size (less than 10 stocks) such as Mining, Insurance and Fishery, Agriculture and Forestry over-express no group. It is not surprising that such small sectors have a limited power of over-expression. It should be mentioned, however, that other sectors such as Services, Wholesale Trade, Other Products, Glass and Ceramics Products, and Metal Products over-express very few groups, although these sectors are large enough in size. The definitions of those sectors are somewhat vague; hence, the stock returns may not show strong comovement due to their heterogeneity of companies included.

The result shown in Table 4.7 is almost consistent with that of Table 4.6. The fact that some sectors including Services and Wholesale Trade are not preserved in the grouping by modularity maximization is striking. Even if factors of those sectors are defined as the mean of stock returns, such factors may not contain any distinctive information.

4.7 Technical discussion about clustering

4.7.1 Static correlation matrix and volatility spillover in GARCH modeling

The goal of our research is to find a more data-oriented classification of Japanese stocks that contributes to building a more reliable risk control framework of stock portfolios. We focus on clustering of Japanese stocks, since the problem is a typical case of clustering fat-tailed asset returns with a high-dimensional correlation structure. If we can find a way to solve this problem, the same approach can be applicable to clustering other various types of financial assets.

The risk of portfolio returns comprises mainly two parts: variance and covariance of returns. The estimation of the correlation structure, which is a key component of precise risk calculation is widely recognized as a difficult problem due to the high-dimensionality and fat-tailedness of returns. A correlation matrix based on pairs of simple linear correlation can be significantly distorted. We mentioned this issue in Section 3.2, and propose to use GARCH filtering method to work around the problem of fat-tailedness. The stock returns are preprocessed by the ARMA–GARCH model to separate the volatility factor from returns. The estimation of as many as 1,407 ARMA–GARCH models is successfully performed, and the correlation matrix is built on the standardized residuals of those models.

There are many possible extensions of preprocessing return data, e.g., more flexible

Table 4.7 Over-expression by standard sector classification

Sector	Number of stocks	k=15 (14gruops)	k=20 (20groups)	k=30 (30groups)	Mean
Transportation Equipment	55	2	2	2	2.0
Land Transportation	27	2	2	2	1.8
Real Estate	35	1	2	2	1.6
Pharmaceutical	26	1	1	2	1.6
Construction	80	2	2	1	1.4
Information and Communication	81	2	2	0	1.4
Foods	56	2	2	2	1.2
Retail Trade	123	1	1	2	1.2
Machinery	103	1	1	2	1.2
Iron and Steel	30	1	1	2	1.2
Banks	69	1	1	1	1.2
Securities	17	1	2	1	1.2
Electric Appliances	129	1	2	1	1.2
Chemicals	108	0	2	1	1.2
Electric Power and Gas	17	1	1	2	1.0
Nonferrous Metals	21	1	1	1	1.0
Other Financing Business	17	1	2	1	1.0
Precision Instruments	22	1	1	0	1.0
Textiles and Apparels	34	1	1	2	0.8
Warehousing and Harbor Transportation services	17	0	0	2	0.8
Rubber Products	11	1	1	1	0.6
Marine Transportation	9	1	1	1	0.6
Oil and Coal products	8	0	1	2	0.6
Services	76	0	0	1	0.6
Air Transportation	2	0	0	1	0.6
Pulp and Paper	10	0	0	0	0.6
Fishery, Agriculture, and Forestry	5	0	0	0	0.4
Mining	5	0	0	0	0.4
Insurance	3	0	0	0	0.4
Other Products	39	0	0	1	0.2
Wholesale Trade	115	0	0	0	0.2
Glass and Ceramics products	26	0	0	0	0.0
Metal products	31	0	0	0	0.0

Note: The mean value is calculated for k=5,10,15,20,25, and 30 cases.

GARCH models that can make residual series normal. If we can convert residuals to the normally distributed series, then we can use linear correlations for risk measuring and aggregation. The computation burden will be greatly reduced.

An important caveat of the calculation of the correlation matrix is that the matrix is assumed to be static without any changes over time during the observation period. We used a CCC type multivariate GARCH model framework, which assumes the constant conditional correlation structure; therefore, the possibility of dynamic change of the correlation matrix is not considered. We think that our approach is acceptable as the first approximation of the correlation structure of stock returns with a view to applying it to the clustering. It is, however, required to examine how the correlation changes or does not change, especially during a crisis period, since it is crucial for portfolio risk control. In this regard, estimation of time-dependent correlation matrices is an important issue to be addressed. The DCC–GARCH model proposed by Engle [24] can be used to estimate such changing correlation structures, although it still has some constraints. We will discuss the dynamic correlation in more details in Chapter 7.

Another problem about GARCH modeling is the volatility spillover effect between different stocks. The problem may distort the estimation of correlation matrix; the degree of possible distortion needs to be considered for more precise estimation. We admit these technical limitations of our GARCH filtering approach; however, so far, our approach works fine to reduce the risk of misspecification of the correlation matrix due to the fat-tailedness of returns. Further refinement of the filtering method is considered to accommodate such time-varying correlation matrices and volatility spillovers in our future work.

4.7.2 Contribution to modularity based clustering

The clustering based on the correlation matrix is performed by recursive modularity maximization as described in Section 4.4.1. The most difficult problem in modularity optimization is the resolution limit problem. We have chosen the simple recursive optimization approach. In our implementation of the method, the resolution limit appears at almost every layer of the recursive optimization. This is an important finding, which suggests that a multilayered structure of communities actually exist regarding the stock return correlations. The existence of such structure is not surprising but convincing, since economic activities of the companies have complicated linkages at various channels. The investors are keen on such linkages to forecast future values of stocks, which are naturally reflected on the correlation structure of stock returns.

Our technical contribution here is a proposal of the controlling framework of recursive division of a network as described in Section 4.4.3. A series of modularity maximization is controlled as post-pruning of the tree structure of groups. The concept of creating homogeneous and balanced groups is implemented as ordering the subdivision of groups and thresholding the number of groups. This method is useful for clustering stock returns at various resolutions.

The whole set of stock returns is first clustered into two groups: the cyclical and defensive groups; then, clustered into much smaller 14 groups. They are compared with the standard sector classification by statistical tests to clarify the link between the groups and the sectors. It has been proved that the two categories are closely related, as shown in Table 4.6 and Table 4.7, although some of the groups identified are not necessarily linked to the existing sectors. Some sectors (e.g., Transportation Equipment and Electric Appliances) are proved to form communities, while other sectors (e.g., Wholesale Trade and Services) are not.

One important issue regarding the grouping is the stability or persistence of correlation structures. We have tested if the stability of clustering is affected by changes in sample size of stocks by simulation as described in Section 4.5.2. We do not have any significant problem regarding the persistence of identified groups. The test results ensure the stability of our clustering approach, at least in static applications. The other important issue is whether the grouping is dynamically stable or not. This is rather an empirical issue, which is related to a dynamically changing correlation structure. Our knowledge about the issue is very limited at the moment, and further development of analysis is required.

Chapter 5

Analysis of stock groups and application to risk control

In this Chapter 5, first, we delve into more details about the groups identified by recursive clustering in Chapter 4. The properties of individual groups are explored in the cyclical and defensive categories, respectively. Second, we also conduct random portfolio simulation based on the groups identified in order to examine if the grouping contributes to improving efficiency of portfolio risk control.

5.1 Properties of identified stock groups

The grouping of the stock returns is already linked to the standard sector classification in Section 4.6. Here, we investigate more details about the relationship between the groups and how individual stocks are grouped together. We confirm that recursive modularity optimization performs well in clarifying the multilayers of community structures of stock returns.

5.1.1 Cyclical groups

First, we focus on the cyclical groups, specifically the three groups (G11, G25, and G26) with higher mean correlations and TOPIX betas, as shown in Table 4.3. These three groups form a single block (B6) as a parent group, as shown in Fig. 4.5. A high level of mean correlations indicate that they form tightly connected communities. The block B6 includes stocks that belong to Transportation Equipment in G25 and G26, respectively (Table 4.6). We focus on these two groups to understand how Transportation Equipment stocks are divided into the two groups by clustering.

Figure 5.1 is a network representation of G25 that includes Electric Appliances, Transportation Equipment, Precision Instruments, and some other sectors; only the first three sectors over-express the group as listed in Table 4.6. The stocks of Transportation Equipment form a community with thicker edges that mean higher correlations. The stocks of this community are indicated as square vertices in Fig. 5.1. More specifically, the community includes global major car manufacturers such as Honda, Mazda, Nissan, Subaru, Suzuki, and Toyota. Some major automobile parts suppliers that have close business relations with specific automobile companies mentioned above are also included here. The second community in G25 comprises the stocks of Electric Appliances: Canon, Fanuc, Hitachi, Kyocera, Panasonic, and Sony; Precision Instruments: Nikon and Tokyo Seimitsu, and other sectors such as Machinery and Chemicals, which include companies that provide capital goods and materials with both Transportation Equipment and Electric Appliances companies. These stocks are indicated as circle vertices. Hence, G25 can be described as the group of leading exporting companies with some related companies.

Figure 5.2 represents G26 that includes another group of Transportation Equipment stocks. Compared with G25 with major automobile companies, this group comprises much smaller parts and components suppliers as well as some tire manufacturers (classified in Rubber Products). The stocks of this community are indicated as square vertices with relatively weaker linkages compared with the one in G25. Some companies that belong to Electric Appliances and Machinery sectors and have business relation with Transportation Equipment companies are also included as circle vertices. The other subgroup indicated as circle vertices comprises Machinery (Ma), Chemical (Ch), and Electric Appliances (EA) stocks; many of them are middle sized companies and less export-oriented compared with the companies in G25. G26 can be described as the group of parts and components suppliers for automobile companies and other middle sized companies that belong to Machinery and other sectors.

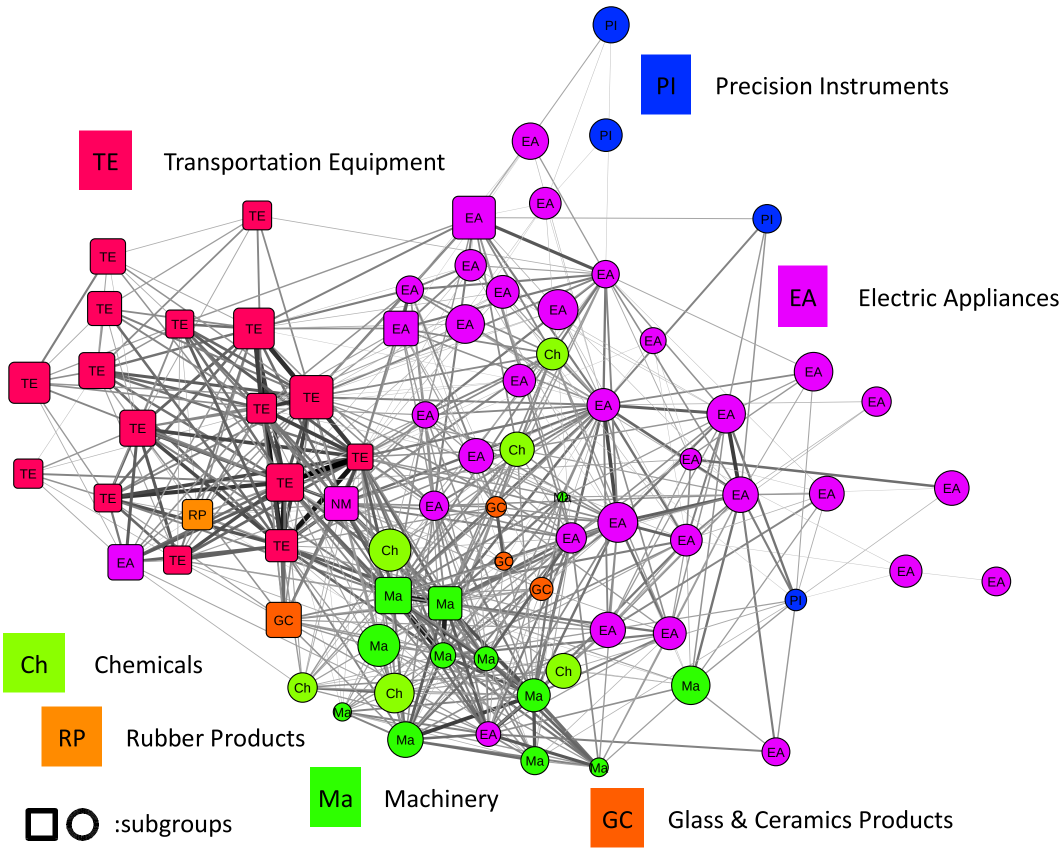
Thus, Transportation Equipment sector is separated into two homogeneous communities by modularity maximization: a community of large automobile companies and a community of smaller parts and components suppliers.

Similarly, G11 includes Iron and Steel, Nonferrous Metals, and Marine Transportation stocks, which are related to industrial materials and export- and import-related services. These sectors appear to be cyclical; therefore, it is not surprising to be in the same block with G25 and G26.

5.1.2 Defensive groups

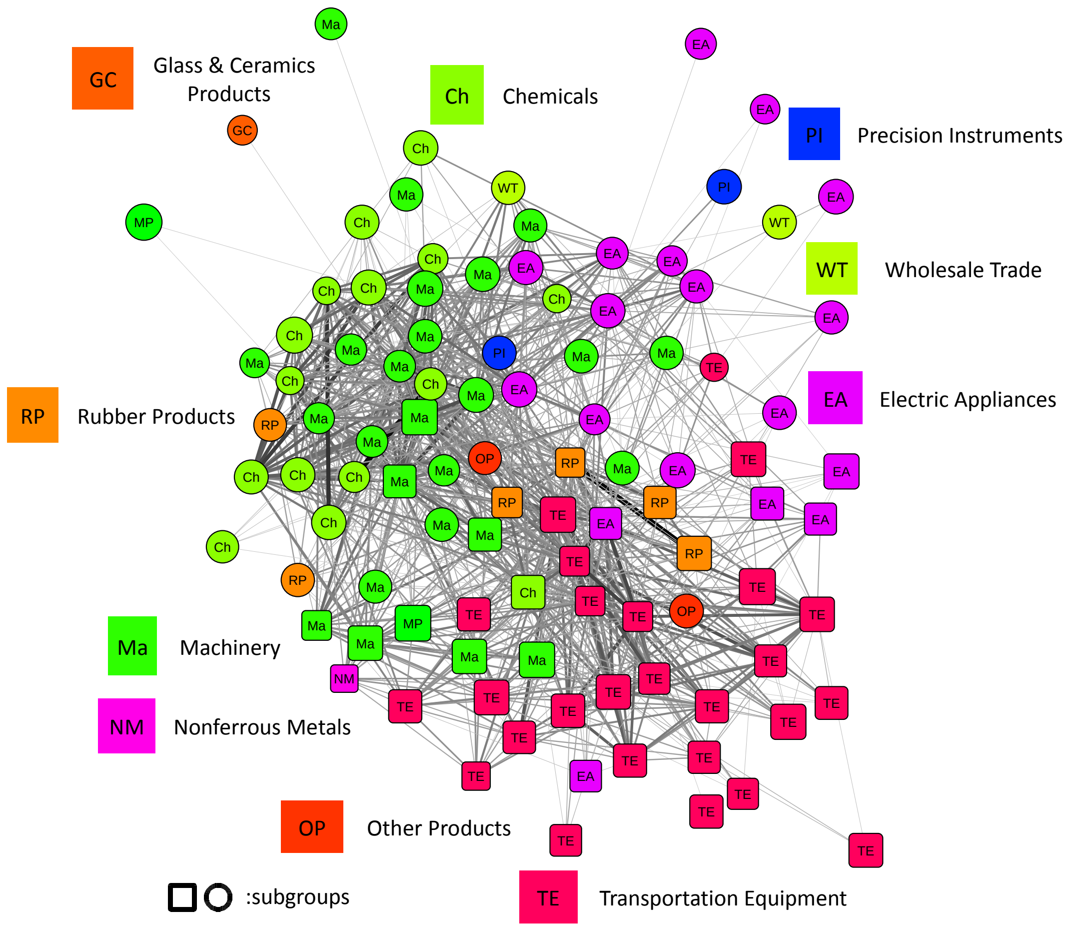
Secondly, we focus on the four defensive groups (G19, G20, G21, and G22) that belong to block B5 on Fig. 4.5. These groups are well over-expressed by several sectors, as shown in Table 4.6. Table 4.3 shows that G19 and G20 have a relatively lower level of correlations and TOPIX betas, while G21 and G22 have a relatively higher level of correlations and TOPIX betas. Here, we look into the details of two groups: G19 as a typical defensive group and G22 as a rather surprising case.

Table 4.6 shows that G19 is over-expressed by the four defensive sectors: Electric Power and Gas, Pharmaceutical, Foods, and Land Transportation, any of which is expected to have a lower level of TOPIX betas. As shown in Fig. 5.3, Electric Power and Gas (EP) stocks



Note: The size of a vertex shows its degree; the width of an edge shows the degree of correlation between vertices. The figure of a vertex: circle or square discriminates subgroups, as shown in Fig. 4.3 and Fig. 4.5. Some vertices with thin edges are intentionally omitted for clearer vision of communities. The same applies to the similar charts hereafter.

Figure 5.1 Cyclical groups: G25 (Transportation equipment–1)



Note: For details, see the note in Fig. 5.1.

Figure 5.2 Cyclical groups: G26 (Transportation equipment–2) form a tightly connected community, in which all major energy companies are included. Land Transportation (LT) stocks including major railway companies also form a tightly connected community. These two communities jointly form a subgroup indicated as square vertices. The second subgroup indicated as circle vertices includes stocks in Pharmaceutical (Ph) such as Astellas and Takeda; Foods (Fo) such as Ajinomoto and Nissin; Services (Sv), and other sectors. It seems that the second defensive subgroup is more consumer related, while the first defensive group is more industry related. Many of the companies in this group are leading companies in each sector with larger market capitalizations. The group, therefore, can be described as the major companies in the typical defensive sectors.

Again, Table 4.6 shows that G22 is over-expressed only by the Banks sector. Figure 5.4 shows that G22 is dominated by Banks (Ba); the share of Banks in G22 is nearly 80%. Note that stocks of other sectors with thinner edges are not displayed there. The group has a higher level of TOPIX betas; the group has a relatively higher level of correlations with the cyclical groups.

It seems to be strange to find financial institutions in the defensive groups. The important fact is that a small number of internationally operating large banks are not included in G22. They are included in the cyclical group (G11) together with other large securities companies. Most of the banks in G22 are smaller regional banks. Their business clients are mainly medium- and small-sized companies and households. They are less linked to large exporting companies than the large banks are; they are rather more linked to local economies. In addition, the banking sector is a strictly regulated industry; there is not much difference between regional banks in any district in Japan. It is, therefore, possible that the stock returns of those small banks are highly correlated, forming a single group in the defensive block.

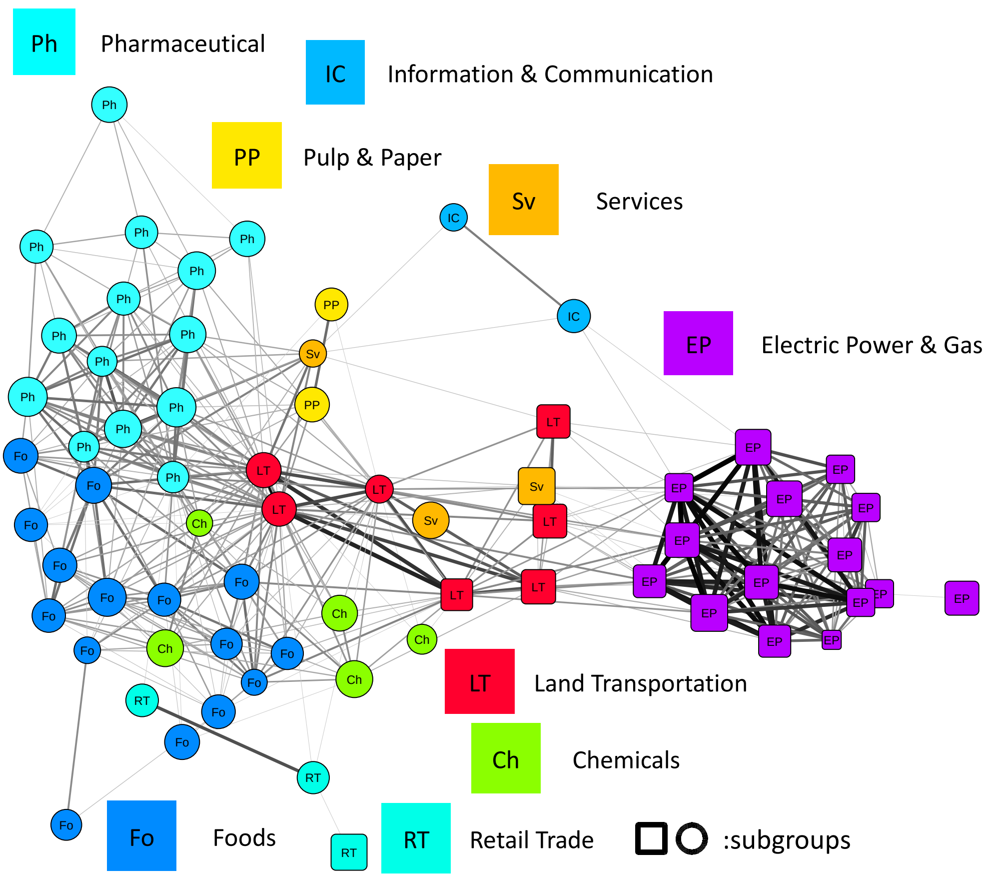
The two examples mentioned above reveal that our clustering method based on recursive modularity maximization works well to detect communities with more details, as shown in Fig. 5.1, Fig. 5.2, and Fig. 5.3. These clustering results are helpful for understanding the properties of the correlation structure within the stock groups as well as possible structural linkages between the groups.

5.2 Application to efficient portfolio risk control

5.2.1 A framework of comparative analysis by random portfolio simulation

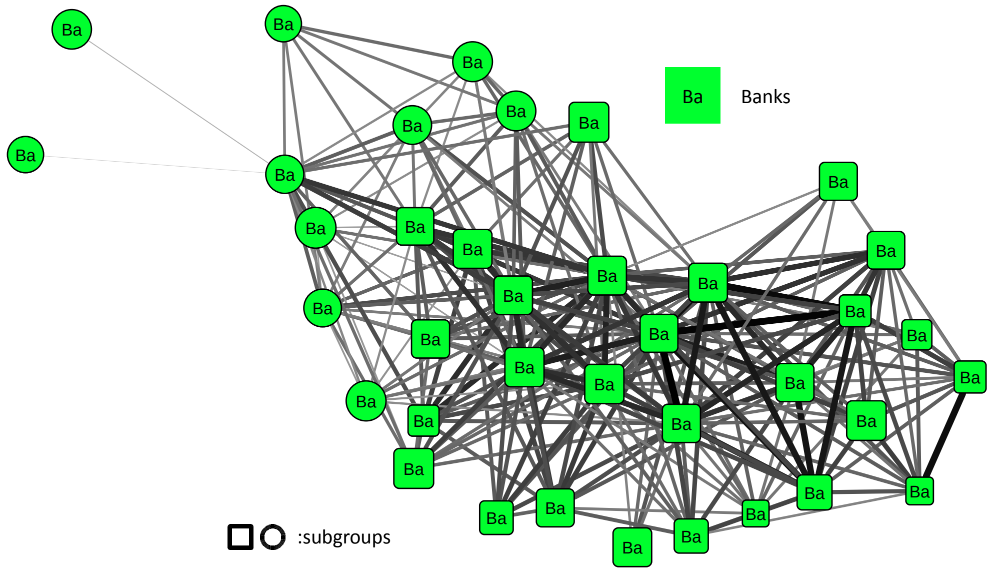
Our data-oriented clustering method can be applied to portfolio risk measurement and risk control of portfolio investment. The grouping of stocks identified by community detection is expected to contribute to improving the precision of risk measurement and increasing chances of risk reduction by diversification. As for risk control, investors can choose one or several companies from each group, better diversifying their portfolio and minimizing the loss margin.

It is, however, not easy to establish a method of evaluating the quality of grouping to determine how one is superior to the alternatives in a way that contributes to more efficient



Note: For details, see the note in Fig. 5.1.

Figure 5.3 Defensive groups: G19 (Electric power and gas, pharmaceutical, and foods)



Note: For details, see the note in Fig. 5.1.

Figure 5.4 Defensive groups: G22 (Banks)

risk control. What is important here are convincing objective criteria and a framework of such comparison. In order to prove if it is really useful, we have conducted a simulation analysis by sampling random portfolios as follows.

We propose a framework based on random portfolio simulation, in which multiple sets of stocks are sampled systematically to explorer the portfolio risk numerically through comparison between the two groupings, namely, the standard sector classification and the grouping achieved by recursive modularity maximization. In order to build sample portfolios, a risk-based portfolio construction approach that does not require any forecast of expected returns is adopted. It has been well recognized that estimating risks from the observed data of returns is more robust compared with forecasting average returns; therefore, minimum variance and equally weighted portfolios have prompted greater interest. In our simulation, the risk-based portfolio is also useful for reducing complicated components to link any findings with the quality of grouping that is of our interest.

The simulation is designed to generate two types of random stock portfolios using the standard sector classification and the clustering result. The simulation comprises two stages of sampling, as shown in Fig. 5.5: the first-stage creates a fixed size of subspace of stocks by sampling group IDs; the second stage creates sample portfolios by stratified sampling of individual stocks from the selected groups.

The first stage sampling means selection of risk factors (scenarios) from the whole set of risk factors. The sampling is designed to make a random selection of group IDs so as to make the sum of number of stocks as close to 20% of the total number as possible, but not greater than 20%. The level of 20% is arbitrarily chosen as a fixed share of small subspace that represents an investor's choice of risk scenarios. This sampling forms a biased subspace of stocks that include only a portion of risk factors, mimicking a process of determining a portfolio investment strategy. If the sectors or groups are not homogeneous and balanced, then concentration risk may increase at the first-stage sampling, since diversification may not work well under such situation.

Once the groups are identified, individual stocks are sampled proportionally from the selected groups so that the number of selected stocks per group reflects the actual number of stocks per group (stratified proportional sampling) and the total number of stocks in the portfolio is set at 30. The size of 30 is also arbitrarily chosen assuming a rather small size of portfolio.

The last thing to be considered is weighting of the portfolio. We use two types of settings: equal weighting and optimized weighting. The initial price is set at unit value for every stock, and the total portfolio value is also set at unit value with equal or optimized weights. The same weight, that is $\frac{1}{30}$, is attributed to all the stocks included in the equally weighted portfolio. The advantage of an equally weighted portfolio is that it has no concentration on any stocks in the portfolio. The risk of portfolio is, however, not necessarily minimized with the equal weighting. On the contrary, a portfolio with optimized weights is ensured to have the minimum risk, although such portfolio tends to be concentrated on stocks with less volatile returns. We focus on both of the two types of portfolios to observe how the two different groupings affect the risk amount of these portfolios.

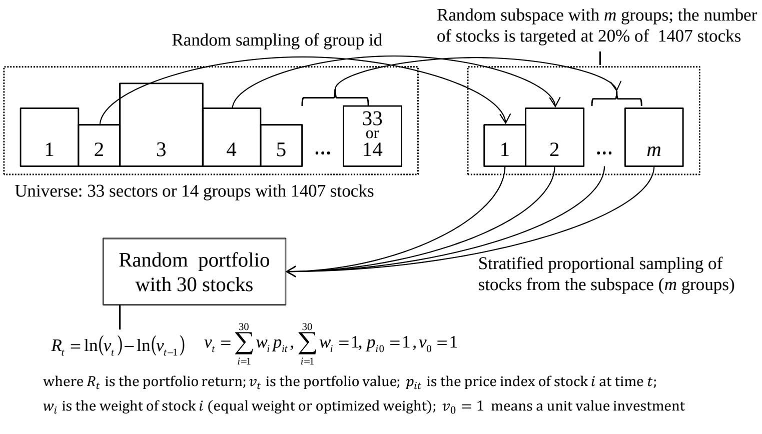


Figure 5.5 Framework of two-stage sampling for random portfolio formation

Now, we need to specify how to calculate and optimize portfolio risk. The risk of portfolio can be calculated from aggregated portfolio returns based on the same historical data mentioned in Section 4.2. As for risk measure, VaR and ES are the most frequently used, which are defined as below:

$$\begin{aligned} \mathrm{VaR}_p[X] &= -\inf\{x \mid \Pr[X \le x] > 1-p\}, \quad 0<p<1 \\ \mathrm{ES}_p[X] &= \mathrm{E}\left[-X \mid -X \ge \mathrm{VaR}_p[X]\right], \quad 0<p<1 \end{aligned} \tag{5.1}$$

where X is portfolio returns and p is the confidence level. We use ES as the portfolio risk measure, since ES is more reliable for detecting tail risks than VaR. The confidence level p is set at 99%, and ES is calculated by the Historical Simulation method in which the loss distribution is approximated by empirical cumulative distribution function (ECDF). For more details about risk measurement related issues, see McNeil et al. [68].

In our simulation, the optimized weighting is designed to minimize the ES of a sample portfolio. ES satisfies subadditivity and convexity; therefore, the weights can be optimized by linear programming to minimize ES. We adopt the transformed linear programming solution as introduced by Rockafellar and Uryasev [86] for minimization of ES with some additional constraints. Many of the minimum risk portfolios including the minimum ES portfolio and the traditional minimum variance portfolio tend to suffer from heavy concentration on specific stocks: the simply optimized weights form a less diversified portfolio. We set both the highest and the lowest limits of weights to avoid such problems: the lower bound is set at 20% of the equal weight and the higher bound is set at 500% of the equal weight. This constrained minimization works well to form a more diversified portfolio with significantly reduced ES compared with the equally weighted portfolio.

The two types of random portfolio formation are repeated 10,000 times, respectively, to calculate portfolio ES as well as a concentration measure to see how the portfolio is concentrated on specific groups chosen at the first stage of group sampling. In Section 4.4.3, we mentioned HHI in equation (4.14) as a useful measure of group concentration, which is implicitly involved in controlling the process of recursive clustering. It should be noted that the range of HHI depends on the number of groups m; it ranges from $\frac{1}{m}$ to 1. HHI, therefore, tends to be higher when the number of groups is larger. When we use HHI, in order to compare the degree of concentration between the random subspaces, HHI should be normalized as NHHI to remove the distortion effect that arises from significant differences in group size. NHHI is defined as follows:

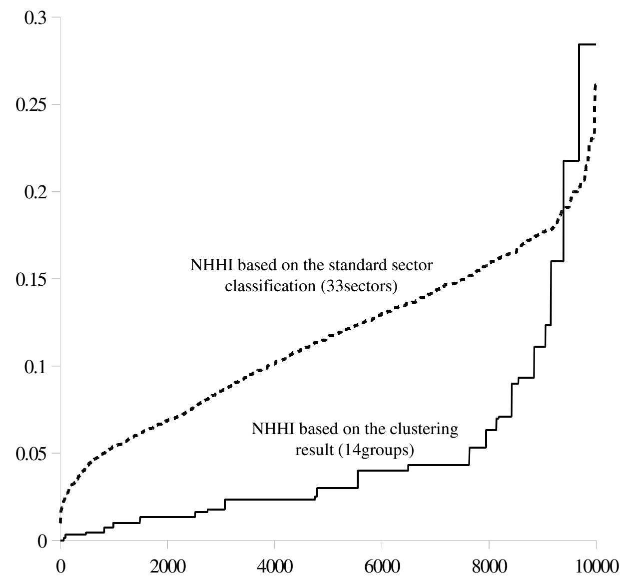
$$NHHI = \frac{\left(HHI - \frac{1}{m}\right)}{\left(1 - \frac{1}{m}\right)}. \tag{5.2}$$

NHHI ranges from 0 to 1 not being affected by the number of groups m.

The sampled group number m can be significantly different depending on the two sampling sources of 33 sectors and 14 groups, as shown in Fig. 5.5; therefore, such possible distortion effect should be properly controlled by using NHHI instead of HHI.

5.2.2 Simulation result

We begin by examining how the grouping affects the degree of concentration of a random portfolio. Figure 5.6 shows NHHI distributions; the two NHHIs are sorted in an ascending order for comparison between the clustering result and the standard sector classification. The curve of the NHHI based on the clustering is mostly located below the curve of the NHHI based on the standard sector classification, although the former NHHI increases rapidly at higher levels and surpasses the latter curve. The random portfolios are mostly less concentrated when using the clustering result for subspace sampling. Hence, we can say that the clustering result contributes more to reduce the degree of concentration compared with the standard sector classification. It is also true that more concentrated portfolios are formed based on the clustering result, but such cases are limited in number.



Note: NHHI are calculated in every run of the simulation, and sorted in an ascending order.

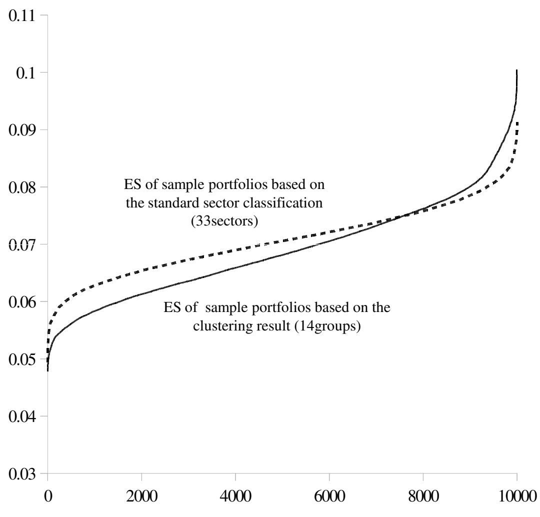
Figure 5.6 Distribution of NHHI

Secondly, we compare the ES distributions to see how the portfolio risks are different between the two cases. In equally weighted portfolios, as shown in Fig. 5.7a, the ES line of the clustering result is located below the ES line based on the sector classification at the lower ES values, whereas the curves cross at the higher ES value. It shows that the distribution of the ES of the sample portfolios based on the clustering result has wider dispersion than the sector classification.

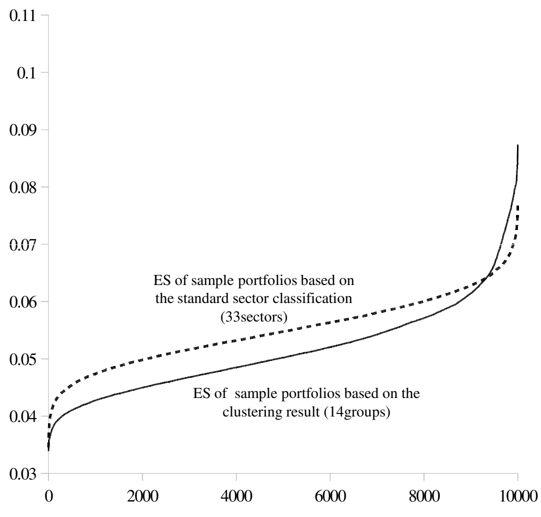
The random portfolios based on the clustering result can have lower risks than the sector classification, since more homogeneous and balanced groups are successfully formed by the two-stage sampling using the clustering result. Risks arising from higher concentration and correlation are better controlled when the clustering result is used during the two-stage sampling. It is also worth mentioning that the clustering result can also achieve higher ES values with higher concentration on riskier groups of stocks, although such cases are less frequent than the opposite cases. Thus, the clustering gives a wider range of opportunities for higher and lower portfolio risks.

In the minimum ES portfolios, as shown in Fig. 5.7, the ES curves are similar to those in the equally weighted portfolios as in Fig. 5.7a, although the risk levels are significantly reduced by the weight optimization. In addition to the risk reduction, the differences between the ES curves are wider in the optimized weight portfolios than in the equally weighted portfolios. It means that the weight optimization for minimizing portfolio ES works better with the grouping based on the clustering result compared with the standard sector classification. The subspace sampling based on the clustering result forms the more homogeneous and balanced groups than the standard sector classification as mentioned before, which can provide a wider range of diversification of stocks and flexibility of weight optimization.

In order to confirm the stability of the simulation result, the simulation is repeated with several different settings of subspace size and portfolio size. The robustness of the simulation analysis has been confirmed with similar results.



(a) ES of equally weighted portfolios



(b) ES of optimized weight portfolios

Note: ES is shown in natural logarithmic scale. ES is calculated for a random portfolio formed in every run of the simulation, and sorted in an ascending order. The share of individual stocks of the random portfolio is equally weighted in (a) or optimized for minimization of ES in (b).

Figure 5.7 ES of sample portfolios

The result of the simulation proves that the risk of random portfolio in terms of ES is better controlled in both equally weighted and optimized weight portfolios. This is just an example of application to portfolio formation; it is possible to apply the clustering result even to more sophisticated risk models for portfolio risk management. Our clustering method can provide reasonable approximation of fundamental factors of returns, which can contribute to building more precise risk modelling as well as a deeper understanding of the underlying processes of stock returns.

Chapter 6

Building classification trees of stock groups

In this Chapter 6, we try to link the clustering results that are based on the stock price data with non-price external data in order to explore how the hierarchical division process can be explained by other categorical and numerical variables. It will contribute to further understanding of the properties of stock groups partitioned by network clustering. In this regard, the classification tree works well; the hierarchical splits of groups are replicated by multiple sets of classification trees, specifying what type of non-price data are important to explain those splits. First, we build classification trees to identify the important variables. Second, we compare the identified variables with a standard stock price model to examine the consistency between the model and our findings from the classification tree analysis.

6.1 Classification tree of stock groups

We performed hierarchical divisive clustering of Japanese stocks listed on the Tokyo Stock Exchange in Chapter 4. The clustering method is based on a modularity maximization algorithm; a community detection method is applied to identify the homogeneous and balanced 14 groups, many of which are proved to be linked to the existing business sectors by the hypergeometric test.

Then, we have a new research question: "What is the meaning of the hierarchical divisions that are described in Fig. 4.5?" The recursive clustering of stock groups requires only stock price data; therefore, non-price data about individual companies may have some additional information to explain the group properties. In other words, it is possible to characterize those stock groups if the splitting process is well explained by some variables. It has been already confirmed that some sectors over-express some groups; the same type of analysis can be extended to a much wider range of quantitative and qualitative non-price variables.

In this regard, a classification tree analysis seems to work fine to infer how the divisions of the groups are explained by the external data including price performance indicators, financial information, and business information of listed companies. What is important here is the choice of non-price data that are expected to be informative to characterize the stock groups. If such variables are identified, we can compare the list of variables with the variables included in a standard stock price model. It may be possible that such comparison gives us some hint to build a localized stock price model with additional variables.

6.1.1 Building classification trees

We build classification trees that provide sorting rules to reproduce the stock groups identified by hierarchical network clustering of the Japanese stocks. The grouping becomes more convincing if such division rules are identified, since clustering itself gives few clues as to why and how such groups are formed.

Clustering is often regarded as unsupervised leaning: no class information on partitions of data is available. On the contrary, classification tree is regarded as supervised learning, in which class information are already available as a set of class labels that are expected to contain information on the true class structure.

As such, classification tree requires training data with known class labels for learning the splitting rules. We can use the network clustering results as the training data to build classification trees. The network clustering and the classification tree are combined for further understanding of the stock group structure. We aim to identify a list of informative variables to understand the group properties.

What is classification tree?

A classification tree, which is also known as decision tree, is built through binary recursive partitioning. First, the data is split into two groups; then, the same splitting process are iterated for every subgroup. The terminal branches of the tree are assigned to the discrete classes with known labels. Classification tree can provide a measure of confidence that the classification is correct.

In our context, building classification tree corresponds to splitting the covariate space, namely stock returns, into multiple partitions represented by group IDs that are categorical variables. The stock groups have already been identified by recursive network clustering; therefore, the class label is already available for every stock. The splitting process is to fit a constant model of the response variable in each partition.

We have external non-price data as response variables, which are listed later. Training data include group ID; and quantitative and qualitative (categorical) data including sector information for every stock. The best split is selected by some predetermined splitting criteria; the most influential variable for the split is assigned for the split. The ranking scores of influential variables in a tree are available as variable importance.

Splitting criteria and complexity control

There are several types of splitting rules. Here, we adopt the Gini impurity—diversity of a node—splitting criterion, which is widely used as the standard splitting criterion in various classification problems. The Gini impurity measure is defined as

$$\mathrm{I}(A) = \sum_{i=1}^{C} f(p_{iA}) = \sum_{i=1}^{C} p_{iA}(1 - p_{iA}) \tag{6.1}$$

where p_{iA} is the proportion of samples in node A that belongs to class i. The Gini impurity measure takes the minimum value of 0 when all samples in the node A fall into a single target group.

The best split of a group based on the Gini impurity criterion is the split that maximizes the change of the Gini impurity measure, $\Delta\mathrm{I}(A)$. It means that the split which most reduces the impurity or diversity in a group is selected. The algorithm searches the largest class and isolate it from the rest. It is known that the Ginni impurity criterion works well for noisy data. For more details about the splitting criteria, see Breiman et al. [17].

Variable importance

Once the split is determined, it is possible to assign the most influential variable from the data set to the split. It should be mentioned that not every variable is involved in a tree, although all the variables are examined to improve the goodness of fit of the tree. Then, the degree of importance is calculated for every variable included.

The variable importance is a predictor ranking as a primary or surrogate splitter based on the contribution to the construction of a tree. The rankings are relative to a given tree structure. It is calculated in terms of relative share based on the corresponding reduction of predictive accuracy when the predictor or variable of interest is removed. The variable importance can provide hints to understand the group properties; if there is a variable that appears frequently as highly important in many classification trees, such variable can possibly be a key factor that drives stock prices.

Controlling complexity of the tree

Controlling the complexity of a classification tree is critical, since too complex tree frequently causes an overfitting problem. It is important to keep a reasonable trade-off between the complexity and predictive accuracy of splits. We performed a cross validation of complexity parameters to find the best value with a enough level of predictive accuracy of those splits. Specifically, we set the complexity parameter value with the one standard deviation rule.[1] In

[1] The parameter is determined in a way that the estimate of predictor error as the minimum + one standard deviation.

other words, we choose the smallest tree in which cross-validation error is within one standard error of the minimum. This method is widely used to avoid the overfitting problem.

6.1.2 Non-price data for variables

When building the classification trees, various types of non-price data are examined. We consider the availability of data as well as theoretical rationale for the selection of variables. The non-price data of stocks listed in Table 6.1 are adopted after initial screening based on expert judgment. There are nine variables including price performance indicators, financial information, and business information about the listed companies.

It has been confirmed that the grouping is at least partly a restructured form of some sector classification; many of the groups are well over-expressed by one or more sectors. The 33 sector dummies, therefore, are also included to test if a split hinges on any sector information. The total number of variables is 42 (=9+33).

Table 6.1 List of variables of classification tree

Variable	Code	Type	Data
TOPIX beta	TPB	sensitivity	value
Yen/Dollar rate beta	EXB		
Market capitalization	MC	relative size	percentile
Price book-value ratio	PBR	price performance indicators	
Price earnings ratio	PER		
Return on assets	ROA	financial information	
Dividend rate	DR		
Overseas sales ratio	OSR	business information	
Foreign stockholding ratio	FSR		
33 sector dummies	S**		0 or 1
42 (9+33) variables			

Note: TOPIX beta and Yen/Dollar rate beta are estimated independently by single factor models with robust MM estimators. The variables from the price book-value ratio to foreign stockholding ratio are the mean values of quarterly data that are available during the data period. These values are converted to percentiles of an empirical cumulative distribution of each variable to avoid any disturbance due to scaling-related problems. The sector dummy (S**) is a dummy vector with value 1 for a specific sector, and 0 for other sectors like $(0,\cdots,0,1,0,\cdots,0)$. Data are downloaded from Yahoo finance, Bloomberg, and other sources.

6.1.3 Classification tree and hierarchical clustering

Multistage classification tree

The correspondence between the classification tree and hierarchical clustering is illustrated in Fig. 6.1. A classification tree is represented as the multistage binary trees as shown in Fig.

6.1a. The model comprises a series of divisions, which are dependent on quantitative and qualitative variables of individual stocks.

Figure 6.1b shows two classification trees, any of which corresponds to one part of the hierarchical groups as shown in Fig. 6.1a. Specifically, the cyclical group A is first divided into the group of B and C and the group of D; then, B and C are separated as individual groups. These divisions correspond to the left classification tree of Fig. 6.1b. The variable feature1 is assigned to the split from the group A to the group of B and C, and the group of D; then, another variable, feature2, is assigned to the split of B and C.

The classification tree is built as follows: the single variable is selected which best splits a group into two groups, and then the same process is applied separately to each subgroup recursively. The procedure is quite similar to that of the recursive clustering by modularity maximization, although the clustering includes more than two splits for a given group. We adopt the standard Gini impurity splitting criterion as mentioned in Section 6.1.1.

We build a classification tree for every split at each layer of the hierarchy. A vector of variable importance is calculated for every classification tree. The variables that appear frequently with high variable importance are of our interest.

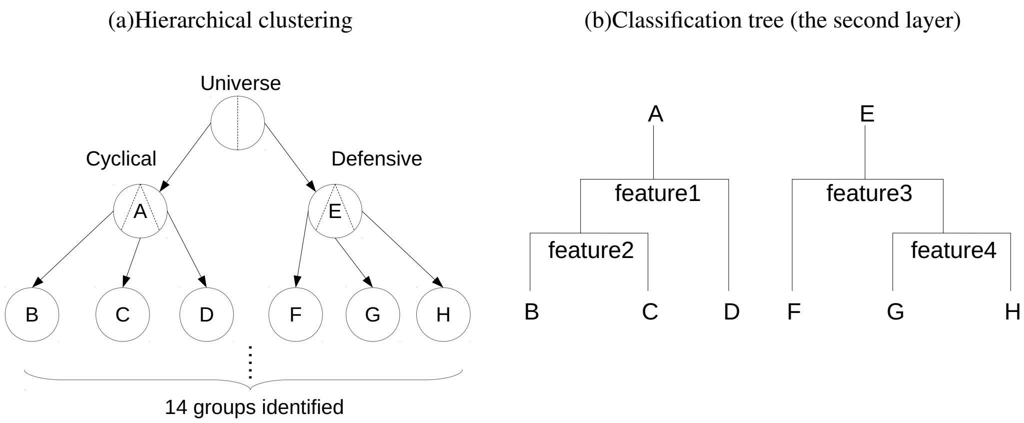


Figure 6.1 Classification tree and hierarchical clustering

Stock groups and over-expressing sectors

The classification trees with non-price variables, which helps clarify how the subnetworks and final groups are divided from the parent networks, are successfully built with a good level of accuracy as summarized in the first column of Table 6.2.[2] Table 6.2 shows a hierarchical

[2] We use R package "rpart"[97], which implements many methods of CART (Classification and Regression Trees [17]), to build and analyze classification trees.

structure of clustering: B1–B12 are intermediate subnetworks; G11–G30 are final groups of stocks. The labels are exactly the same as those in Fig. 4.5.

The top five selected variables for each division of subnetworks are shown in the first column of Table 6.2 together with the relative importance of variables in terms of percent share. The mean values of each variable in a subnetwork or group are also shown there. A vector of such values describes the group properties well, which helps understand the meaning of hierarchical splits. The second column of Table 6.2 shows sectors that over-express the final groups listed in the first column. As mentioned earlier, many of the final 14 groups are over-expressed by some sector categories, which are listed in the second column of Table 6.2.

The entire network (B1) is divided into two subnetworks: the cyclical (B3) and defensive (B2) groups. TPB (TOPIX beta) contributes most significantly to the first division, followed by OSR (overseas sales ratio) and MC (market capitalization). The mean values of TPB and OSR are considerably different between B3 and B2. The higher levels of TOPIX beta and overseas sales ratio can be well acknowledged as the common features of the cyclical groups.

Sector dummies are not included in the variables selected for the classification trees when a group is over-expressed by multiple sectors, whereas they are included when over-expressed by only a single sector (e.g., Banks, Construction, and Other Financing Services). It seems that the sector information is of some help in understanding the classification, but such cases are limited in number at the clustering resolution of 14 groups. It may be possible that the sector information is more often included as classification tree variables at a much higher resolution with a larger number of groups. Variables other than the sector dummies such as market capitalization and PBR (price book-value ratio) are included more frequently in the classification trees.

It seems that a list of variables with high importance scores depends on the group to be divided. The group properties can be described in more details by combining the list of important variables and over-expressing sectors. When we focus on a specific group, such information is helpful to understand the group structure. Furthermore, it may be also helpful for customizing the standard stock price model for the specific sector with improved model fitting.

6.2 Findings from classification trees

6.2.1 Important variables

In order to evaluate which of the variables have a higher level of importance, the aggregated variable importance scores as well as the aggregated frequency of appearance are calculated for the classification trees of B3–B12. Table 6.3 is a summary table that covers the divisions of both the cyclical and defensive groups. The mean importance score is calculated as an average of each variable importance in every classification tree. The variables that have the top three

Table 6.2 Classification tree and over-expressing sectors

B1 [1407]	AR: 0.80	TPB	OSR	MC	EXB	STR	
		(40)	(20)	(13)	(8)	(4)	
	B2 [657]	0.46	5.44	0.51	0.41	0.00	Defensive Groups
	B3 [750]	0.77	25.99	0.49	0.58	0.08	Cyclical Groups
B3 [750]	AR: 0.87	MC	FSR	TPB	PBR	EXB	
		(31)	(22)	(14)	(12)	(10)	
	B6 [347]	0.70	21.95	0.94	1.30	0.66	
	B7 [403]	0.29	8.02	0.60	1.02	0.50	
B6 [347]	AR: 0.74	MC	OSR	ROA	PBR	TPB	
		(20)	(17)	(12)	(12)	(9)	
	G11 [148]	0.54	26.12	1.38	1.18	0.96	Construction(33), Textiles & Apparels(14), Real Estate(13)
	G25 [89]	0.69	54.21	1.93	1.32	1.05	Electric Appliances(38), Transportation Equipment(16), Precision Instruments(7)
	G26 [110]	0.30	37.28	2.62	1.45	0.83	Transportation Equipment(28), Machinery(26), Rubber Products(6)
B7 [403]	AR: 0.73	DR	ROA	PER	EXB	PBR	
		(72)	(19)	(4)	(2)	(2)	
	G13 [187]	2.22	1.38	67.51	0.45	1.24	Iron & Steel(16), Nonferrous Metals(11), Securities(9), Marine Transportation(6)
	B12 [216]	2.41	1.78	26.32	0.54	0.88	
B12 [216]	AR: 0.72	MC	FSR	PBR	EXB	SOF	
		(63)	(9)	(9)	(8)	(6)	
	G30 [112]	0.64	10.94	1.00	0.60	0.07	Other Financing Business(8)
	G29 [104]	0.35	7.22	0.76	0.47	0.00	-
B2 [657]	AR: 0.89	MC	FSR	PBR	DR	EXB	
		(38)	(17)	(11)	(10)	(8)	
	B4 [387]	0.34	9.23	0.92	2.74	0.39	
	B5 [270]	0.73	16.91	1.30	2.03	0.43	
B5 [270]	AR: 0.69	SBA	ROA	MC	PBR	TPB	
		(32)	(17)	(16)	(13)	(7)	
	G20 [84]	0.00	4.69	0.38	1.50	0.37	Retail Trade(40), Foods(14)
	G22 [57]	0.77	1.04	0.38	0.89	0.65	Banks(44)
	G21 [57]	0.00	2.64	0.52	1.14	0.58	Information & Communication(15), Land Transportation(9)
	G19 [72]	0.00	3.77	0.71	1.50	0.46	Electric Power & Gas(16), Pharmaceutical(15), Foods(15), Land Transportation(7)
B4 [387]	AR: 0.79	MC	FSR	EXB	TPB	PBR	
		(49)	(23)	(11)	(9)	(4)	
	B8 [193]	0.69	12.95	0.46	0.52	0.96	
	B9 [194]	0.32	5.48	0.32	0.36	0.88	
B8 [193]	AR: 0.81	MC	SCO	FSR	TPB	EXB	
		(52)	(20)	(17)	(5)	(3)	
	G15 [69]	0.67	0.36	16.27	0.62	0.53	Construction(25)
	G16 [124]	0.41	0.03	11.16	0.47	0.42	-
B9 [194]	AR: 0.74	MC	DR	FSR	PBR	TPB	
		(57)	(12)	(11)	(8)	(7)	
	G18 [99]	0.367	3.445	4.473	0.861	0.316	Information & Communication(15)
	G17 [95]	0.651	2.515	6.572	0.898	0.401	-

Note: AR is an accuracy ratio defined as the ratio of the number of correctly grouped stocks to the number of total stocks. [ ] shows the number of stocks in a subnetwork or group. ( ) shows the relative importance score of each variable. As for the sector dummy variables (S**), STR is Transportation Equipment, SOF is Other Financing Business, SBA is Banks, and SCO is Construction. The sectors listed in the second column over-express a corresponding group in the first column. The sectors are identified by the hypergeometric test as in Table 4.6.

largest appearance counts are written in bold style.

The market capitalization, which implies the size of a company, plays a very important role with the largest count of appearance and the highest level of mean variable importance as shown in Table 6.3a. The market capitalization has the highest mean importance in both the cyclical and defensive groups as shown in Tables 6.3b and 6.3c.

The foreign stockholding ratio has the second highest mean importance. The mean importance of the foreign stockholding ratio is almost at the same level in both the cyclical and defensive groups, while the count is higher in the defensive groups. In Japan, a significant portion of stock trade is dominated by foreign investors; therefore, it is possible that stocks with a higher foreign stockholding ratio have different patterns of fluctuation compared to stocks with a lower ratio.

PBR also has larger counts, although the variable importance is relatively lower. PBR has the largest count in the cyclical groups.

The sensitivity variables, namely, TOPIX beta, Yen/Dollar rate beta—stock return sensitivity to a change in the exchange rate—have a middle range of counts and mean importance scores. Both TOPIX beta and Yen/Dollar rate beta have already been included in the first classification tree to separate the universe into the cyclical and defensive groups. The two variables are still effective with a lower level of importance at a later stage of subdivisions.

Other variables including the sector dummies appear at a lower frequency. One exception is that ROA has the same level of counts and mean importances as TOPIX beta and Yen/Dollar rate only in the cyclical groups.

The lists of variables are similar in the cyclical and defensive groups, although there are some slight differences in terms of the order of counts and mean importance scores.

6.2.2 Comparison with standard stock price model

The three variables: TOPIX beta, market capitalization, and PBR are confirmed to be effective. The selection of the three variables is consistent with the Fama and French (FF) three-factor model (Fama and French [32]), which is frequently used as the standard stock price model.

The FF three-factor model is built as an extension of the capital asset pricing model (CAPM), which has only a single market risk factor, with additional two factors: the size effect and value effect. The size and value effects mean that stocks with smaller capitalization tend to outperform stocks with larger capitalization; stocks with lower PBR tend to outperform stocks with higher PBR. In our classification tree context, the market risk factor corresponds to TOPIX beta. The size effect corresponds to the market capitalization, and the value effect corresponds to the PBR. In that sense, our splitting rules are connected with the FF three-factor model.

The FF three-factor model has been tested and proved to be effective for many stock markets including the Japanese stock market (Kubota and Takehara [58] and Pham and Long [82]). Our classification trees based on the clustering result also indicates the effectiveness of the FF three-factor model. Some other factors are also included in the classification trees. The foreign stockholding ratio is frequently adopted and has higher mean importance. As mentioned earlier, foreign investors have a larger share of trading in the Japanese stock market; therefore, it is not surprising that the foreign stockholding ratio is influential at many splits of the classification trees. The foreign stockholding ratio can be regarded as an additional regional factor of the FF three-factor model.

As for the sensitivity factor, the Yen/Dollar rate beta is also an important variable of the classification trees. It is included in the classification tree of the first split into the cyclical and defensive groups, and also in the classification trees of both the cyclical and defensive groups. It should be mentioned that TOPIX beta and Yen/Dollar rate beta are highly correlated; therefore, such correlation needs to be controlled when we consider including the two factors in a stock price model.

These findings suggest that it may be possible to build modified factor models of stock returns considering a flexible choice of the best fitting variables. There is another practical application of the classification tree. In our clustering, we excluded more than 300 stocks from the total stocks listed on the First Section of the Tokyo Stock Exchange. The number of stocks is reduced from over 1,700 to 1,407. Many of the stocks excluded do not have a full length of price data. Stocks that are newly listed or merged do not have enough price data, even if they have a sufficient level of liquidity. In such cases, it is difficult to calculate pairwise correlations from return series. If we apply the splitting rules built in the classification trees as well as the sector information, we can provide an alternative method to find a group even for stocks that have a limited length of return series data. This is very useful to complete clustering that covers a full size of stocks in the market.

Table 6.3 Variable importance of classification tree

(a) Cyclical+Defensive (14 groups)

Variable	Count	Mean importance
Market capitalization (MC)	**8**	**40.7**
Price book-value ratio (PBR)	**8**	**8.8**
Foreign stockholding ratio (FSR)	**6**	**16.3**
TOPIX beta (TPB)	**6**	**8.4**
Yen/Dollar rate beta (EXB)	**6**	**7.0**
Dividend rate (DR)	3	31.0
Return on assets (ROA)	3	16.0
Banks (SBA)	1	32.1
Construction (SCO)	1	20.2
Overseas sales ratio (OSR)	1	17.3
Other Financing Business (SOF)	1	6.2
Price earnings ratio (PER)	1	3.9

(b) Cyclical (6 groups)

Variable	Count	Mean importance
Price book-value ratio (PBR)	**4**	**8.5**
Market capitalization (MC)	**3**	**37.9**
Yen/Dollar rate beta (EXB)	**3**	**6.7**
Return on assets (ROA)	**2**	**15.4**
Foreign stockholding ratio (FSR)	**2**	**15.4**
TOPIX beta (TPB)	**2**	**11.5**
Dividend rate (DR)	1	71.7
Overseas sales ratio (OSR)	1	17.3
Other Financing Business (SOF)	1	6.2
Price earnings ratio (PER)	1	3.9

(c) Defensive (8 groups)

Variable	Count	Mean importance
Market capitalization (MC)	**5**	**42.4**
Foreign stockholding ratio (FSR)	**4**	**16.8**
Price book-value ratio (PBR)	**4**	**9.0**
TOPIX beta (TPB)	**4**	**6.8**
Yen/Dollar rate beta (EXB)	**3**	**7.2**
Dividend rate (DR)	2	10.7
Banks (SBA)	1	32.1
Construction (SCO)	1	20.2
Return on assets (ROA)	1	17.2

Note: Count is calculated as the sum of frequency of appearance. Mean importance is calculated as an average of each variable importance in every classification tree. The variables that have the top three largest counts are written in bold style.

Chapter 7

An empirical study of dynamic correlation of Japanese stock returns

In this Chapter 7, we extend the multivariate GARCH model that is used for volatility control of stock returns in Chapter 4. The static correlation of residuals is changed to the dynamic correlation. Two types of sample portfolios are formed: the market portfolio that covers the entire market and a set of group portfolios with a reduced number of stocks, which are selected from the stock groups identified by network clustering in Chapter 4. We explore the dynamics of correlation between every sample portfolio as well as the impact of changes in correlation on portfolio risk.

7.1 Multivariate GARCH with dynamic correlation

In Chapter 3, we have mentioned that the correlation of stock returns is a critical factor for portfolio risk measurement and risk control. The correlation matrix of the stock returns is converted to a network adjacency matrix for network clustering in Chapter 4. The method of clustering fat-tailed asset returns proposed in this study depends on the estimation of a large scale of correlation matrix.

It is, therefore, critically important to estimate the correlation matrix accurately from the stock price data. We used the multivariate GARCH model in Chapter 4. The GARCH model functions as a filter that separates volatility fluctuations from standardized residuals to eliminate the distortion effect of volatilities on the return correlation.

When building the multivariate GARCH model, we introduced two types of model simplification: the static correlation matrix and the assumption of no volatility spillovers between individual stocks to work around the high-dimensionality problem that may cause technical difficulties of model estimation. Here, we focus on the extension of the correlation matrix type from static to dynamic, since many previous research works regarding asset return correlation suggest that the dynamic correlation is more realistic as mentioned in Chapter 2. As for the volatility spillovers, we still keep the assumption mainly due to technical requirements for model settings. Incorporating the volatility spillover effects on the estimation of volatilities and correlation is a challenging topic that is beyond our scope.

We try to answer the following fundamental research questions by employing the segmented multivariate GARCH models with a dynamic correlation structure.

- Does the correlation matrix change over time?
- Are there any significant differences in the correlation dynamics between the groups?
- Does the correlation change significantly affect the portfolio risk?

We expect that the correlation matrix changes over time. The point would be the degree of changes, especially during turbulent periods including the Lehman shock and Great Earthquake. Even if the dynamic correlation is proved to be more realistic than the static one, we may still be able to rely on the risk measurement methods and network clustering results as reasonable approximations on condition that the degree is below some threshold.

7.1.1 Dynamic conditional correlation

The correlation of asset returns is one of the key issues for quantitative risk measurement and portfolio investment control. Among the many heavily debated issues related to this topic, dynamic changes in correlation remain particularly controversial. As discussed in Chapter 4, we use multivariate ARMA–GARCH model to describe the joint process of stock returns in a portfolio. Equation (4.8) as the mean equation and equation (4.9) as the volatility equation jointly form the multivariate ARMA–GARCH model.

The fundamental setting of the ARMA–GARCH model has not been changed: $\boldsymbol{A_i}$, $\boldsymbol{B_j}$ in equation (4.8) and $\boldsymbol{S_i}$, $\boldsymbol{T_j}$ in equation (4.9) are all diagonal matrices as assumed in Chapter 4, since no interaction of the mean or variance process between individual stocks is considered either here.

As for the correlation structure of the stock returns, we change the assumption from the unconditional correlation to the dynamic conditional correlation as mentioned above. We use "conditional" as the word that means "time varying" here. The most widely used implementation of dynamic correlation is the DCC by Engle and Sheppard [26].

The DCC is defined as

$$\boldsymbol{H}_t = \boldsymbol{D}_t\boldsymbol{R_t}\boldsymbol{D}_t = \left[\rho_{kl\cdot t}\sqrt{h_{kk\cdot t}h_{ll\cdot t}}\right]_{k,\ l=1,\ \dots,\ N} \tag{7.1}$$

where $\boldsymbol{D}_t$ is a diagonal matrix with the elements $\left(\sqrt{h_{11\cdot t}}, \cdots, \sqrt{h_{NN\cdot t}}\right)$, $\rho_{kl\cdot t}$ is the conditional correlation of the returns between stock k and l at time t, $\boldsymbol{R_t}$ is an $N\times N$ positive definite dynamic correlation matrix, and $h_{kl\cdot t}$ is the conditional covariance of returns at time t. Note that the conditional variances ($h_{kk,t}$) can be estimated separately as in equation (4.9). The conditional correlation matrix $\boldsymbol{R_t}$ must be inverted at every point in time. $\boldsymbol{R_t}$ should be constrained to be positive definite at every time t. As is the case of the static correlation of stock returns in equation (4.11), the dynamic correlation of stock returns is the same as that of i.i.d. standardized residuals as

$$\operatorname{corr}(r_k,\ r_l) = \operatorname{corr}(z_k,\ z_l) = \rho_{kl\cdot t}\,. \tag{7.2}$$

We, therefore, estimate the dynamic correlation matrices of standardized residuals by fitting the multivariate ARAM–GARCH model to historical data on stock returns.

7.1.2 Modeling correlation dynamics

DCC is more flexible than CCC; however, the number of parameters increases significantly when the number of stocks becomes large. Moreover, because the correlation matrix $\boldsymbol{R}$ can change depending on time t, ensuring that every correlation matrix satisfies the positive definite condition throughout the entire period is a challenge. Engle and Sheppard [26] satisfied this constraint by modeling a dynamic correlation process with the proxy variable $\boldsymbol{Q_t}$.[1]

The DCC model with time lags in conditional correlation is described as DCC(m, n). The proxy variable $\boldsymbol{Q}_t$ is modeled as

$$\begin{aligned}\boldsymbol{Q}_t &= \bar{\boldsymbol{Q}} + \sum_{i=1}^{m} a_i\left(z_{t-i}z_{t-i}^{'} - \bar{\boldsymbol{Q}}\right) + \sum_{j=1}^{n} b_j\left(\boldsymbol{Q}_{t-j} - \bar{\boldsymbol{Q}}\right) \\ &= \left(1 - \sum_{i=1}^{m} a_i - \sum_{j=1}^{n} b_j\right)\bar{\boldsymbol{Q}} + \sum_{i=1}^{m} a_i z_{t-i}z_{t-i}^{'} + \sum_{j=1}^{n} b_j \boldsymbol{Q}_{t-j}\end{aligned} \tag{7.3}$$

where a_i and b_j are non-negative scalars and $\bar{\boldsymbol{Q}}_t$ is the unconditional covariance of the standardized residuals (Engle and Sheppard [26]).

The parameter a_i shows the sensitivity of $\boldsymbol{Q}_t$ to previous shocks, while the parameter b_j represents the persistence of correlation in previous periods. The concept of the dynamic modeling is similar to the volatility process modeling in the GARCH model.

The correlation matrix $\boldsymbol{R}_t$ is then obtained by rescaling $\boldsymbol{Q}_t$ such that,

$$\boldsymbol{R}_t = \operatorname{diag}(\boldsymbol{Q}_t)^{-\frac{1}{2}}\,\boldsymbol{Q}_t\,(\boldsymbol{Q}_t)^{-\frac{1}{2}}\,. \tag{7.4}$$

A stationary solution and positive definiteness of $\boldsymbol{Q}_t$ as well as $\boldsymbol{R}_t$ is ensured by the following conditions:

$$a_i \geq 0, \quad b_j \geq 0, \quad \sum_{i=1}^{m} a_i + \sum_{j=1}^{n} b_j < 1. \tag{7.5}$$

For more details on the DCC–GARCH, see Engle and Sheppard [26] and Engle [24].

An inconsistency problem exists when estimating $\bar{\boldsymbol{Q}}$ in equation (7.3) with variance targeting. Aielli [1] pointed out that $\bar{\boldsymbol{Q}}$ is not the unconditional covariance matrix of z_t, as

[1] We follow the notation of the DCC–GARCH by Engle and Sheppard [26] with some modifications.

$\mathrm{E}\left[z_t z_t'\right] = \mathrm{E}\left[\boldsymbol{R}_t\right] \neq \mathrm{E}\left[\bar{\boldsymbol{Q}}\right]$. Aielli [1] proposed cDCC, which includes a corrective term for bias adjustment.[2]

The DCC parameters assumption that the dynamic processes of the conditional correlations are controlled only by a pair of scalar parameters: a_i and b_j is sometimes regarded as too restrictive, since such scalar values setting means that the same correlation dynamics is shared by all pairs of stock returns.[3] In this regard, more generalized DCC models have been studied as mentioned in Chapter 2. One of them is the Block DCC (BDCC)–GARCH model proposed by Billio et al. [8]. The BDCC–GARCH model has a block-diagonal structure to assume different correlation dynamics for every group of variables. The number of parameters increase while more flexible settings are enabled. Hafner and Franses [44] generalized DCC (GDCC–GARCH) so as to allow for asset-specific correlation sensitivities with different levels of DCC parameters for individual assets (vectorized DCC parameters).

The most important advantage of DCC is its compactness in terms of the number of parameters to be estimated. Such parsimonious parameterization of DCC is helpful for our empirical study. More complicated models including the ones above mentioned require much heavier computational burden for model fitting. We hence choose the most simple one, the unrestricted scalar DCC, to model the conditional correlation, even though improved estimation performance can be expected by applying more complicated models.

Thus, the stock returns are modeled by a combination of the ARMA(P, Q)–GARCH(p, q) and DCC(m, n), where the ARMA for the individual mean model, the GARCH for the individual variance model, and DCC for the correlation model. Now, all of the three models are constructed as conditional. The choice of CCC or DCC is an empirical issue, although DCC is often regarded as more realistic. We assume DCC here, since the dynamic changes in the correlation matrix are of our interest.

7.2 Copula for modeling dependency of residuals

The next point is the joint distribution of multivariate residual z, which is a key assumption for parameter estimation. In order to estimate the parameters of the ARMA–GARCH with DCC by using MLE, we need to build a likelihood function to be maximized. The probabilistic variable in equation (4.7) is z, which is a vector of the standardized residuals. We, therefore, need to specify the multivariate distribution of z as well as its dependency structure.

In Chapter 4, a series of univariate ARMA–GARCH models are fitted to individual stock returns when we estimate the parameters of the ARMA–GARCH with CCC by using MLE. The distribution of residual z is assumed to be one of the normal, the Student t, and the skewed t distribution in the log-likelihood function. In most cases, the Student t and the skewed t distribution are selected based on the AIC type criterion. It means that z still exhibits fat-tail

[2] Some studies have already addressed this issue, including Engle and Kelly [25].

[3] The limitations of the DCC–GARCH model (e.g., volatility spillover and bias problems) are discussed in more details in Caporin and McAleer [18].

features after the volatility fluctuations are removed by GARCH filtering. Then, we calculate pairwise correlations of z non-parametrically as an unconditional correlation matrix.

Modeling of the dynamics of the correlation matrix is required in DCC; the correlation or dependency between individual stock returns as well as the distribution of z should be modeled explicitly in a multivariate model setting. We have two ways to specify the distribution of z. The first option is a multivariate probabilistic distribution: e.g., the normal or Student t. The multivariate Student t distribution is a typical fat-tailed elliptical distribution. The other option is a copula approach. The copula is a function that joints univariate distributions to form a multivariate distribution proposed by Sklar [92].

The copula is a parametrically specified joint distribution generated from given marginals; therefore, the properties of the copula are similar to those of joint distributions. In other words, the copula expresses a multivariate distribution in terms of its marginal distributions. It enables to build a multivariate distribution by specifying univariate distributions as marginal distributions and a function that binds them together. It is also possible to estimate the parameters of marginal distributions and the dependence parameters of the copula separately.

The flexibility of copula has a significant advantage in that marginal distributions can have different parameters, respectively. The separation of the joint function and marginal distributions is apparently another advantage. If we use a multivariate distribution for z, namely multivariate Student t distribution, a correlation matrix and a shape parameter are required to be specified. The shape parameter represents the degree of tail dependence: comovements in the tails of the distributions. The problem is that the single shape parameter controls the fat-tailedness of every marginal distribution as well as the tail dependence between the marginals. The multivariate Student t distribution requires a single shape parameter shared by all stock returns, whereas the Student t-copula allows different values. When the degree of fat-tailedness significantly differs across the stocks in a portfolio, the shape parameter is set at some intermediate level, which is not necessarily consistent with the fat-tailedness of each marginal distribution. The tail dependence is also controlled by the same shape parameter. As such, a multivariate distribution approach is considerably restrictive to represent the fat-tailedness and tail dependence of stock returns. The problem becomes more serious when the number of stocks in a portfolio increases. The copula approach can solve the problem by decomposing the multivariate distribution into the joint function and marginal distributions. That is why we adopt the copula approach here.

7.2.1 Copula

The concept of the copula of an arbitrary distribution is a function to connect the marginal distributions to a joint distribution. An n-dimensional copula $C(u_1, ..., u_n)$ is an n-dimensional distribution in the unit hypercube $[0, 1]^n$ with uniform margins.

Sklar [92] showed that every joint distribution function $F(x_1, ..., x_n)$ of a vector of variables $\boldsymbol{X} = (X_1, \ldots, X_n)$ with marginal distribution functions $F_1(x_1), ..., F_n(x_n)$, can be represented as

$$F(x_1, ..., x_n) = C(F_1(x_1), ..., F_n(x_n)) \tag{7.6}$$

under absolutely continuous margins (Sklar's Theorem).

Considering that $x_1, ..., x_n = F_1^{-1}(u_1), ..., F_n^{-1}(u_n)$, the copula is obtained uniquely as

$$C(u_1, ..., u_n) = F\left(F_1^{-1}(u_1), ..., F_n^{-1}(u_n)\right) \tag{7.7}$$

where $F_i^{-1}(\cdot)$ is the quantile function of the i-th marginal distribution.

The joint distribution of $\boldsymbol{X}$ can be described by the marginal distributions $F_i(x)$ and the copula $C(\cdot)$. If $F(\cdot)$ is the multivariate normal (Student t) distribution, then $C(\cdot)$ is the Gaussian (Student t) copula.

The joint density function of $\boldsymbol{X}$ is obtained as

$$f(x_1, ..., x_n) = c(F_1(x_1), ..., F_n(x_n)) \prod_{i=1}^{n} f_i(x_i) \tag{7.8}$$

where $f_i(\cdot)$ are the marginal densities of x_i and $c(\cdot)$ is the density function of the copula given as follows:

$$\begin{aligned} c(u_1, ..., u_n) &= \frac{\partial^n}{\partial u_1 \dots \partial u_n} C(u_1, ..., u_n) \\ &= \frac{\partial^n F\left(F_1^{-1}(u_1), ..., F_1^{-1}(u_n)\right)}{\partial\left(F_1^{-1}(u_1)\right) \cdots \partial\left(F_n^{-1}(u_n)\right)} \times \frac{d\left(F_1^{-1}(u_1)\right)}{du_1} \times, \cdots, \times \frac{d\left(F_n^{-1}(u_n)\right)}{du_n} \\ &= \frac{\partial^n F\left(F_1^{-1}(u_1), ..., F_1^{-1}(u_n)\right)}{\partial\left(F_1^{-1}(u_1)\right) \cdots \partial\left(F_n^{-1}(u_n)\right)} \times \frac{1}{\frac{dF_1\left(F_1^{-1}(u_1)\right)}{d\left(F_1^{-1}(u_1)\right)} \times, \cdots, \times \frac{dF_n\left(F_n^{-1}(u_n)\right)}{d\left(F_n^{-1}(u_n)\right)}} \\ &= \frac{f\left(F_1^{-1}(u_1), ..., F_n^{-1}(u_n)\right)}{\prod_{i=1}^{n} f_i\left(F_i^{-1}(u_i)\right)}. \end{aligned} \tag{7.9}$$

The copula density function is the ratio of the joint density to what it would have been under the independence assumption as in (7.9). The parameters of the copula can be estimated by using MLE as mentioned below.

It is known that any comonotonic transformation of a vector of variable $\boldsymbol{X}$ does not affect the copula.[4] It means that the same copula can be applied before and after such transformation. For example, the copula for $\boldsymbol{X}$ that distributes as the multivariate normal $F_{G_a}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, is the same as that of $F_{G_a}(0, \boldsymbol{R})$, where $F_{G_a}(\cdot)$ is the distribution function of the normal distribution; $\boldsymbol{R}$ is the correlation matrix implied by the covariance matrix $\boldsymbol{\Sigma}$. The transformation from the standardized residuals $\boldsymbol{z}$ to stock returns $\boldsymbol{r}$ is positive linear transformation as in equation (4.7).

[4] $\boldsymbol{X} = (X_1, \dots, X_n)$ is comonotonic when the condition

$$\Pr(X_1 \le x_1, \dots, X_n \le x_n) = \min_{i \in \{1,\dots,n\}} \Pr(X_i \le x_i)$$

is satisfied. The comonotonic transformation preservers the comonotonicity of the variable $\boldsymbol{X}$. The comonotonicity can be regarded as an extension of the concept of perfect correlation or dependence.

The transformation is comonotonic; therefore, the same copula can be applied to both z and $\boldsymbol{r}$. Intuitively, this relationship is analogous to equation (7.2), which denotes the equivalence of correlation between z and $\boldsymbol{r}$. Once we have estimated the copula for z; then we can also apply the same copula to describe the dependency of $\boldsymbol{r}$.

7.2.2 Gaussian copula and Student t-copula

Gaussian copula

There are many types of implementation of copula, which impose different dependence structures on the data. We focus on the copula that are derived from multivariate elliptical distributions: the Gaussian copula and Student t-copula. Both of them are implemented as multivariate distribution functions; therefore, the parameters of the copula can be easily estimated by using MLE. Further, this approach can be easily extended to a higher dimensional framework.

The Gaussian copula is defined as

$$C^{G_a}(\boldsymbol{u}\,|\,\boldsymbol{R}) = \boldsymbol{\Phi}_{\boldsymbol{R}}\left(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_n)\right) \tag{7.10}$$

where $\boldsymbol{R}$ is a correlation matrix, $\Phi(\cdot)$ is the CDF of the univariate normal distribution, and $\boldsymbol{\Phi}_{\boldsymbol{R}}(\cdot)$ is the CDF of the multivariate normal distribution.[5]

The density function of the Gaussian copula is derived from equation (7.9) and equation (7.10) as

$$c^{G_a}(\boldsymbol{u}\,|\,\boldsymbol{R}) = \frac{1}{\sqrt{|\boldsymbol{R}|}} \exp\left\{-\frac{1}{2}\boldsymbol{q}'\left(\boldsymbol{I}-\boldsymbol{R}^{-1}\right)\boldsymbol{q}\right\} \tag{7.11}$$

where $|\boldsymbol{R}|$ is the matrix determinant of $\boldsymbol{R}$ and $q = (q_1, ..., q_n)$ is defined such that $q_i = \Phi^{-1}(u_i)$ for $i = 1, \ldots, n$.

Student t-copula

The Student t-copula is defined as

$$C^{S_t}(\boldsymbol{u}\,|\,\nu, \boldsymbol{R}) = \boldsymbol{t}_{\nu\cdot\boldsymbol{R}}\left(t_\nu^{-1}(u_1), \ldots, t_\nu^{-1}(u_n)\right) \tag{7.12}$$

where $\boldsymbol{R}$ is a correlation matrix, ν is a shape parameter, $t_\nu(\cdot)$ is the CDF of the univariate Student t distribution, and $\boldsymbol{t}_{\nu\cdot\boldsymbol{R}}$ is the CDF of the multivariate Student t distribution.

The density function of the Student t-copula is derived from equation (7.9) and (7.12) as

$$c^{S_t}(\boldsymbol{u}\,|\,\nu, \boldsymbol{R}) = \frac{\Gamma\left(\frac{\nu+n}{2}\right)\left(\Gamma\left(\frac{\nu}{2}\right)\right)^n \left(1+\nu^{-1}\boldsymbol{q}'\boldsymbol{R}^{-1}\boldsymbol{q}\right)^{-(\nu+n)/2}}{\sqrt{|\boldsymbol{R}|}\left(\Gamma\left(\frac{\nu+n}{2}\right)\right)^n \Gamma\left(\frac{\nu}{2}\right)\prod_{i=1}^n\left(1+\frac{q_i^2}{\nu}\right)^{-(\nu+1)/2}} \tag{7.13}$$

[5] We follow the notation of the copula and the Copula–DCC–GARCH by Demarta and McNeil [23] and Ghalanos [40] with some modification.

where $\boldsymbol{q} = (q_1, ..., q_n)$ is defined such that $q_i = t_\nu^{-1}(u_i)$ for $i = 1, \dots, n$.

It should be noted that the Student t-copula can handle the tail dependence between two variables while the Gaussian copula cannot. For continuously distributed random variables with the Student t-copula $C^{S_t}(\boldsymbol{u}\,|\,\nu, \boldsymbol{R})$, the coefficient of tail dependence λ is given as

$$\lambda = 2t_{\nu+1}\left(\frac{-\sqrt{\nu+1}\sqrt{1-\rho}}{\sqrt{1+\rho}}\right) \tag{7.14}$$

where ρ is the off-diagonal element of $\boldsymbol{R}$ (Demarta and McNeil [23]).

As mentioned in Section 7.2, a vector of standardized residuals z still exhibits fat-tail features after the volatility fluctuation is removed by GARCH filtering. Considering the possible tail dependence of z, the Student t-copula is more suitable than the Gaussian copula; therefore, we use the Student t-copula for modeling stock returns by the multivariate ARMA-GARCH model with DCC.

Thus, stock returns are modeled by the multivariate ARMA–GARCH; the dynamic correlation structure of returns are modeled by the Student t–copula. We call the model as the Copula–DCC–GARCH model, hereafter (Lee and Long [62], Patton [80], and Ghalanos [40]). The procedure of maximization of the log-likelihood of the Copula–DCC–GARCH is described later.

7.3 Model fitting of Copula–DCC–GARCH

Dimension reduction of correlation matrix

When fitting the Copula–DCC–GARCH model to the individual stock returns, there are some issues to be considered. The first and most important point is the coverage of stocks to be included in a sample portfolio to observe the possible changes in the correlation matrix. A market portfolio that includes every stock listed would be the best choice with the widest coverage without any bias or concentration. It is, however, difficult to cope with such a large portfolio for correlation analysis. A 1,400 × 1,400 correlation matrix is too large for the Copula–DCC–GARCH model estimation. Some stock index portfolio such as the Nikkei can be an alternative set, but more flexible choice with a wider coverage would be preferred.

We do not aim to analyze the return correlations of any specific category of stocks. A more generalized viewpoint is appropriate, covering a wider range of sectors or companies, since the knowledge is expected to be applied to further enhancement of portfolio risk management. One possible solution is to build a reduced size of portfolio with selected representatives of stocks. It may be possible to infer the underlying dynamics of the correlation structure of the market portfolio by observing the correlation matrix of an appropriately compressed portfolio, if such portfolio can well approximate the entire market. The point is how to select such stocks to form a reduced size of portfolio.

A factor model approach provides another option for dimension reduction of the correlation matrix. A set of explicitly or implicitly defined factors are expected to explain the driving forces of stock returns. The CAPM model is a typical example of this approach. The major drawback of the factor model approach is that finding factor(s) itself is a hard task, especially when building a multi-factor model. Some statistical technique like the principal component analysis can generate such factors; however, it is still difficult to understand the meaning of those factors. Hence, we do not adopt the factor model approach as mentioned in Chapter 3.

As for the method to form a reduced size of sample portfolios, the standard sector classification is frequently used for categorizing stocks. The sample portfolios can be formed by selecting representatives from sectors on certain criteria. Such a sector classification approach, however, has a fundamental problem in that the sector classification is not necessarily consistent with the comovement of stock returns; the distribution of group size is significantly unbalanced as we mentioned in Chapter 4.

Against this background, a more data-oriented grouping of stock returns was studied in Chapter 4. A homogeneous and balanced groups of stocks were identified by applying correlation clustering based on the complex networks theory. We adopt the grouping to create sample portfolios to observe the dynamic correlation of the Japanese stock returns.

7.3.1 Group portfolios and market portfolio

Group portfolios

The first type of sample portfolio is a set of partial portfolios, which covers only specific groups identified by the correlation clustering. The large-scale single correlation matrix is separated into 14 diagonal blocks. The 14 sample portfolios are then created based on this grouping to observe within-group dynamic correlation. Once the 14 group portfolios are formed, 14 Copula–DCC–GARCH models are fitted to the data of the individual groups.

The number of stocks in these 14 groups is around 100 on average, which is still large for estimating the dynamic correlation of returns when using the Copula–DCC–GARCH model. Thus, a second round of dimension reduction is required to specify the individual stocks to be included in each sample portfolio.

To identify and select stocks in each group, we adopt the eigenvector centrality measure, which is frequently used in network analyses. The network centrality is one of structural characteristics of a node; an individual with a higher centrality measure is often more likely to be a leading individual according to network theory. [6]

The eigenvector centrality of a node is defined as an element of the eigenvector of a network adjacency matrix with the maximum eigenvalue. Here, a node corresponds to a stock, while a network corresponds to a group to which the stock belongs. If we denote the eigenvector centrality of stocks in a group as $\boldsymbol{x} = (x_1, \dots, x_n)$, $\boldsymbol{x}$ is a non-zero vector that is defined as

$$\boldsymbol{A}\boldsymbol{x} = \lambda \boldsymbol{x} \tag{7.15}$$

[6] Typical centrality measures include degree centrality, closeness centrality, and betweenness centrality.

where $\boldsymbol{x}$ is an eigenvector of the adjacency matrix $\boldsymbol{A}$ with the largest eigenvalue of $\boldsymbol{\lambda}$. The eigenvector centrality measure is designed to provide a higher score to a node that has more links to a node with many links. In the context of stock returns, the eigenvector centrality of a stock is higher when it is correlated more with a stock that is highly correlated with other stocks.

The centrality measure is generally assumed to take a positive number. The Perron–Frobenius theorem ensures that the eigenvector centrality measure takes a positive number. This theorem ensures that there is a unique eigenvector with the largest positive eigenvalue; further, the eigenvector is positive and any non-negative eigenvector of $\boldsymbol{A}$ is a positive multiple of the vector on condition that $\boldsymbol{A}$ is a non-negative regular matrix. For more detailed information on the eigenvector centrality measure, see Newman [76]. Our network adjacency matrix is designed to be a non-negative regular matrix as mentioned in Chapter 4. We can safely define the eigenvector centrality measure as shown in equation (7.15).

Thus, the stocks that have the 20 largest eigenvector centrality values are selected to form a sample portfolio for each group. The coverage of the total number of stocks selected is about 20% of the total stocks. The 14 individual group models are built on these stocks to observe "within-group correlations" between stock returns. Note that the selection of the stocks depends on the centrality measure; therefore, other centrality measure may suggest a different set of stocks.

Market portfolio

The second type of sample portfolio covers the entire market. To reduce the dimension of the whole large correlation matrix, an equally weighted stock portfolio is first created for each group. Note that each group portfolio includes all of the stocks that belong to the group at this stage. Then, the return index of each portfolio is calculated as the mean of the individual stock returns in each group. The 14 return indexes can be regarded as the underlying factors of the development of the stock market, since any stock could belong to one of these 14 groups. Lastly, a single sample portfolio is created as an equally-weighted portfolio of the 14 return indexes to observe market-wide or between-group (between-factor) dynamic correlation. We define this sample portfolio as the market portfolio.

7.3.2 Two-step MLE parameter estimation

Likelihood function

To estimate the parameters of the Copula–DCC–GARCH model by using MLE, the log-likelihood function of the model needs to be specified. Two approaches can be used to build the conditional joint distribution of return $\boldsymbol{r}_t$ in equation (4.7). The first approach assumes a multivariate distribution, e.g., the multivariate normal, to specify the density function to maximize the log-likelihood with respect to the model parameters. In the case of the normal distribution, the maximization process can be simplified by separating the first-stage estimation of the individual GARCH models from the second-stage DCC parameter estimation. However, because the assumption of a normal distribution might not apply in every case, we select an alternative approach based on the copula function to model the dependency structure of the residuals.

The conditional joint distribution of returns $\boldsymbol{r_t}$ is defined as

$$F\left(\boldsymbol{r}_t|\boldsymbol{\mu}_t,\ \sqrt{\boldsymbol{h}_t}\right) = C\left(F_1\left(r_{1\cdot t}|\mu_{1\cdot t},\ \sqrt{h_{1\cdot t}}\right),\ldots,\ F_N\left(r_{N\cdot t}|\mu_{N\cdot t},\ \sqrt{h_{N\cdot t}}\right)\right) \tag{7.16}$$

where N is the number of stocks, $F_i(\cdot)$ $(i = 1,\ \ldots,\ N)$ is the conditional distribution of the i-th margin, and $C(\cdot)$ is the N-dimensional copula. The $\mu_{i\cdot t}$ and $\sqrt{h_{i\cdot t}}$ are modeled by equations (4.8) and (4.9), respectively.

The marginal distribution of $z_{i\cdot t}$, which is an i.i.d. random variable, is assumed to be one of the (standardized) normal, Student t, and skewed t distribution. The parameter set θ_i to be estimated for the i-th marginal distribution depends on its type: θ_i includes ξ_i and ν_i for the skewed t, ν_i for the Student t, and none for the normal, where ν_i and ξ_i are the shape and skew parameters, respectively.

The dependence structure of the marginals is then assumed to follow the Student t-copula with the conditional correlation $\boldsymbol{R_t}$ and constant shape parameter $\bar{\nu}$. The dynamics of $\boldsymbol{R}_t$ are assumed to follow the dynamics described by equation (7.3) and (7.4) with the DCC parameter vectors $\boldsymbol{a}$ and $\boldsymbol{b}$.

The joint density of returns $\boldsymbol{r}_t$ is defined as a combination of copula density and density of the i.i.d. residual $z_{i\cdot t}$ based on equation (7.8):

$$f\left(\boldsymbol{r}_t|\boldsymbol{\mu}_t,\ \sqrt{\boldsymbol{h}_t},\ \boldsymbol{R}_t,\ \bar{\nu}\right) = c^{S_t}\left(u_{1\cdot t},\ \ldots,\ u_{N\cdot t}|\boldsymbol{R}_t,\ \bar{\nu}\right)\prod_{i=1}^{N}\frac{1}{\sqrt{h_{i\cdot t}}}f_{i\cdot t}\left(z_{i\cdot t}|\theta_i\right) \tag{7.17}$$

where $u_{i\cdot t} = F_i\left(r_{i\cdot t}|\mu_{i\cdot t},\ \sqrt{h_{i\cdot t}},\ \theta_i\right)$, $c^{S_t}(\cdot)$ is the Student t-copula density, $\bar{\nu}$ is the shape parameter of the Student t-copula, $\frac{1}{\sqrt{h_{i\cdot t}}}$ is the Jacobian of the variable transformation between r_t and z_t, and $\frac{1}{\sqrt{h_{i\cdot t}}}f_{i\cdot t}\left(z_{i\cdot t}|\theta_i\right)$ is the density function of $\sqrt{h_{i\cdot t}}z_{i\cdot t}$.

The log-likelihood function $LL(\boldsymbol{\theta}|\boldsymbol{r}_t)$ is given by the density function (7.17) as

$$\begin{aligned} LL(\boldsymbol{\theta}|\boldsymbol{r}_t) &= LL_R\left(\boldsymbol{R}_t, \bar{\nu}\right) \\ &\quad + LL_V\left(\left(\theta_1,\ \mu_{1\cdot t},\ \sqrt{h_{1\cdot t}}\right),\ \ldots,\ \left(\theta_N,\ \mu_{N\cdot t},\ \sqrt{h_{N\cdot t}}\right)\right) \\ &= LL_R\left(\boldsymbol{a}_1,\ \ldots, \boldsymbol{a}_N,\ \boldsymbol{b}_1,\ \ldots, \boldsymbol{b}_N,\ \bar{\nu}\right) \\ &\quad + LL_{V_1}\left(\theta_1,\ \theta_1^{AG}\right) +, \ldots, + LL_{V_N}\left(\theta_N,\ \theta_N^{AG}\right) \end{aligned} \tag{7.18}$$

where $\boldsymbol{\theta}$ is the whole parameter set, $LL_R(\cdot)$ is the Copula–DCC part with the DCC parameters $(\boldsymbol{a}, \boldsymbol{b})$ as in equation (7.3), and $LL_{V_i}(\cdot)$ is the univariate GARCH part with a set of univariate ARMA–GARCH parameters $\left(\theta_i,\ \theta_i^{AG}\right)$ for stock i $(i = 1, \ldots, N)$.

As such, the log-likelihood can be separated easily into two parts when maximizing $\sum_{t=1}^{p} LL(\cdot)$, where p is the length of the time series data: the joint Copula–DCC part and the individual univariate GARCH part. The two parts of the log-likelihood function can be safely maximized independently without any shared parameters between them. Thus, the individual ARMA–GARCH parameters as well as the distributional parameters for the residual z are estimated first for the individual stocks by maximizing LL_{Vi}; then, the Copula–DCC parameters are estimated by maximizing LL_R.

This two-step approach reduces parameter estimation burden greatly compared with modeling by the multivariate distribution such as the Student t distribution, which requires a complicated convergence search process for maximizing the likelihood. Another merits of using copula is that independent shape parameters can be assigned to the Student t and skewed t distributions of univariate GARCH models as well as the Student t-copula for DCC.

The estimation of the ARMA–GARCH for stocks in the group portfolios has already been completed in Chapter 4. We only estimate parameters of the ARMA–GARCH for the market portfolio. We need to estimate parameters of the Copula–DCC for both the group portfolios and market portfolio.

Data

The data on stock returns are basically the same as the data used for the estimation of ARMA–GARCH models in Chapter 4. The data period is daily from January 2008 to December 2013. The end of the period has been extended from September 2012 to December 2013. The period includes the two major financial shocks examined herein: the Lehman shock in 2008 and the Great Earthquake in 2011.

The stocks that have complete daily price data (at the close) for the given period are selected from those listed on the First Section of the Tokyo Stock Exchange. These selection criteria have been introduced to avoid any inconsistency in the time series when calculating the correlations. The number of stocks is 1,384 in 33 sectors.[7]

Price data are converted into daily log-returns.

7.3.3 DCC parameter estimation result

The stock return data are separated into the 14 different groups. They are also categorized as the cyclical and defensive. The 14 group portfolios are created by selecting 20 stocks from corresponding groups based on the network centrality measure. The market portfolio is also created. The DCC–GARCH model is then simply fitted to the 14 group portfolios and the market portfolio, independently.

When estimating the DCC–GARCH model for the group portfolios, the univariate ARMA–GARCH model is first fitted to the individual stock returns based on the two-step estimation approach described in Section 7.3.2. The ARMA–GARCH lags and residual distribution of

[7] The number of stocks has been updated due to the change of the data period.

the model should be determined to identify the model. The multiple models with different lag patterns and choices of residual distribution are then estimated by using MLE, and the model with the highest AIC is selected for every stock return. In the second step, the DCC lags are determined similarly by selecting the model with the highest AIC from the alternatives. Specifically, the Copula–DCC model is fitted to the standardized residuals to estimate the DCC model parameters by using MLE. The whole likelihood maximization process shown in equation (7.18) is thus completed.

Similar to the group portfolios, the univariate ARMA–GARCH model is first fitted to the individual stock return indexes defined in 7.3.1 when estimating the DCC–GARCH model for the market portfolio. The remaining estimation process is the same as that for the group portfolios.

Table 7.1 shows the estimation results of the DCC parameters.[8] In this table, group portfolios are identified by the group ID (e.g., G25). The group IDs are the same as in Table 4.3 in Capter 4. The results of the univariate ARMA–GARCH model are summarized in Table 7.2 for the cyclical groups and Table 7.3 for the defensive groups. The estimation results of the univariate ARMA–GARCH model for the group portfolios are omitted because of space limitations.

Estimation result

First, the DCC order (m, n) in equation (7.5) is almost $(1, 1)$ or $(1, 2)$ as shown in Table 7.3. The order m, which is the lag order for the parameters a_i in equation (7.3), is 1 in all cases. The parameter a_i indicates the degree of responses of $\boldsymbol{Q_t}$ to the past covariances of shocks in equation (7.3). The result that the order $m = 1$ means that the effect of the past shocks on $\boldsymbol{Q_t}$ and hence the correlation $\boldsymbol{R_t}$ do not last longer. The order n, which is the lag order for the parameters b_j, is 1 for the market portfolio; 1 or 2 for both the cyclical and the defensive group portfolios. The parameter b_j indicates the degree of persistence of $\boldsymbol{Q_t}$ as well as $\boldsymbol{R_t}$. The order $n = 1$ (or 2) corresponds to the DCC parameter $b1$ (and $b2$) in Table 7.1.

The DCC parameters $a1$ are all non-zero positive numbers with enough significance, but are all very small numbers (< 0.02) compared with $b1$ and $b2$. We calculate $b1 + b2$ as the sum of $b1$ and $b2$ to compare the relative persistence of the sample portfolios. $b1$ and $b2$ take relatively large numbers (higher than 0.9 for some of them including the market portfolio). The values of $a1$+$b1$+$b2$ are all below 1; it is confirmed that the condition of equation (7.5) is satisfied. Hence, we can say that DCC is more realistic in our models than CCC is, since CCC assumes that $a_i = b_j = 0$ in equation (7.3). These findings are similar to those of previous studies that have estimated DCC models as mentioned in Chapter 2.

The DCC parameter estimates, especially $b1$ and $b2$, vary widely between the groups, implying that the correlation dynamics may differ across them. Indeed, the parameter estimates vary even within the cyclical and defensive groups. We explore the pattern of correlation

[8] We used the R(http://cran.r-project.org/) package "rmgarch" (Ghalanos [41]) for the parameter estimation.

Table 7.1 DCC estimation results

Group ID	m, n	$a1$	Std error	$b1$	Std error	$b2$	Std error	$b1+b2$	Shape	Std error
Market	1, 1	0.018	(0.004)	0.939	(0.021)		-	0.939	14.2	(0.959)
Cyclical										
G25	1, 2	0.009	(0.002)	0.293	(0.130)	0.638	(0.131)	0.931	18.8	(0.945)
G11	1, 1	0.010	(0.001)	0.930	(0.015)		-	0.930	20.9	(1.257)
G26	1, 1	0.008	(0.002)	0.881	(0.039)		-	0.881	22.0	(1.279)
G30	1, 2	0.012	(0.002)	0.301	(0.114)	0.610	(0.112)	0.911	28.9	(2.282)
G29	1, 2	0.013	(0.003)	0.424	(0.092)	0.368	(0.132)	0.792	27.1	(2.063)
G13	1, 1	0.015	(0.003)	0.806	(0.064)			0.806	22.3	(1.997)
Defensive										
G22	1, 2	0.015	(0.002)	0.331	(0.071)	0.561	(0.072)	0.892	13.9	(0.779)
G15	1, 1	0.009	(0.002)	0.852	(0.049)		-	0.852	22.5	(1.904)
G21	1, 1	0.006	(0.001)	0.943	(0.020)		-	0.943	22.9	(1.607)
G16	1, 1	0.009	(0.001)	0.889	(0.024)		-	0.889	26.2	(1.860)
G17	1, 2	0.011	(0.002)	0.483	(0.144)	0.331	(0.160)	0.814	23.5	(1.807)
G19	1, 2	0.012	(0.002)	0.363	(0.141)	0.563	(0.140)	0.926	16.0	(0.873)
G18	1, 2	0.017	(0.002)	0.352	(0.143)	0.389	(0.131)	0.740	28.5	(2.683)
G20	1, 1	0.006	(0.001)	0.908	(0.031)		-	0.908	29.2	(2.554)

Note: m and n are the DCC order as in equation (7.3). $a1$, $b1$, and $b2$ are the DCC parameters in equation (7.3). "Shape" is the shape parameter of the Student t-copula.

changes in every group more in Section 7.4.2. The shape parameters of the Student t-copula range between about 14 and 29. These relatively high values means that the tail dependence of the standardized residuals seems to be limited, if any.

Tables 7.2 and 7.3 summarize the estimation results of the univariate ARMA–GARCH model for the market portfolio. The parameter set depends on the individual ARMA–GARCH lag degrees and distribution types of standardized residuals. The distribution is selected to be the skewed t in most instances with the Student t in one group based on the AIC. The estimates of the shape parameters of the skewed t and Student t show values below 10 in many of the defensive groups, but higher values in may of the cyclical groups. A lower shape value means that the standardized residuals still exhibit fat-tailedness even after the fat-tailedness of returns is reduced by adjusting the volatility by using GARCH as mentioned in Chapter 4. The important advantage of the copula based method is that it can handle such heterogeneities in marginal distributions very well.

Table 7.2 Univariate ARMA–GARCH: Six cyclical groups (portfolio return indexes)

Group ID	(P, Q) (p, q)	Parameter	Estimate	Std error	Cdist
G25	(0,1) (1,2)	ma1	0.056	(0.026)	sstd
		garch1	0.830	(0.036)	
		arch1	0.079	(0.030)	
		arch2	0.060	(0.038)	
		skew	1.080	(0.043)	
		shape	26.46	(19.986)	
G11	(1,0) (1,2)	ar1	0.053	(0.026)	sstd
		garch1	0.773	(0.040)	
		arch1	0.095	(0.036)	
		arch2	0.081	(0.043)	
		skew	1.140	(0.051)	
		shape	31.789	(26.417)	
G26	(0,1) (1,1)	ma1	0.066	(0.028)	sstd
		garch1	0.828	(0.026)	
		arch1	0.134	(0.021)	
		skew	1.142	(0.058)	
		shape	17.587	(8.342)	

Group ID	$(P, Q)(p, q)$	Parameter	Estimate	Std error	Cdist
G30	(1,0) (1,1)	ar1	0.072	(0.027)	sstd
		garch1	0.767	(0.032)	
		arch1	0.168	(0.027)	
		skew	1.254	(0.063)	
		shape	14.849	(6.582)	
G29	(2,0) (1,1)	ar1	0.102	(0.027)	sstd
		ar2	0.048	(0.027)	
		garch1	0.763	(0.030)	
		arch1	0.161	(0.025)	
		skew	1.197	(0.051)	
		shape	10.844	(3.632)	
G13	(1,1) (1,1)	ar1	0.561	(0.240)	sstd
		ma1	-0.452	(0.256)	
		garch1	0.732	(0.033)	
		arch1	0.192	(0.028)	
		skew	1.207	(0.055)	
		shape	11.002	(3.473)	

Note: $(P, Q)(p, q)$ is (AR order, MA order) (GARCH order, ARCH order) as in equations (4.8) and (4.9). ar1, 2 are the parameter estimates for the AR part; ma1, 2 for the MA part in equation (4.8). garch1, 2 are the parameter estimates for the GARCH part; arch1,2 for the ARCH part in (4.9). cdist is the conditional distribution of standardized residuals: "std" represents the Student t, "sstd" skewed t, and "norm" normal. The best model is selected from multiple alternatives by using an AIC-type criterion.

Table 7.3 Univariate ARMA–GARCH: Eight defensive groups (portfolio return indexes)

Group ID	(P, Q) (p, q)	Parameter	Estimate	Std error	Cdist
G22	(1,1) (1,1)	ar1	-0.870	(0.069)	std
		ma1	0.833	(0.076)	
		garch1	0.838	(0.052)	
		arch1	0.117	(0.036)	
		shape	8.956	(2.344)	
G15	(2,2) (1,1)	ar1	-1.963	(0.001)	sstd
		ar2	-0.985	(0.001)	
		ma1	1.964	(0.000)	
		ma2	0.987	(0.000)	
		garch1	0.798	(0.141)	
		arch1	0.154	(0.094)	
		skew	1.129	(0.048)	
		shape	12.031	(3.985)	
G21	(0,0) (1,1)	mu	0.000	(0.000)	sstd
		garch1	0.810	(0.050)	
		arch1	0.140	(0.035)	
		skew	1.133	(0.048)	
		shape	12.310	(4.084)	
G16	(0,0) (1,1)	mu	-0.001	(0.000)	sstd
		garch1	0.788	(0.033)	
		arch1	0.159	(0.028)	
		skew	1.142	(0.042)	
		shape	11.516	(3.725)	

Group ID	(P, Q) (p, q)	Parameter	Estimate	Std error	Cdist
G17	(2,2) (2,1)	ar1	-0.553	(0.005)	sstd
		ar2	-0.989	(0.003)	
		ma1	0.543	(0.003)	
		ma2	0.990	(0.000)	
		garch1	0.473	(0.182)	
		garch2	0.257	(0.155)	
		arch1	0.188	(0.037)	
		skew	1.122	(0.040)	
		shape	9.095	(2.502)	
G19	(1,0) (2,1)	ar1	-0.065	(0.027)	sstd
		garch1	0.325	(0.173)	
		garch2	0.444	(0.172)	
		arch1	0.191	(0.050)	
		skew	1.073	(0.044)	
		shape	7.130	(1.211)	
G18	(2,0) (1,1)	ar1	0.054	(0.028)	sstd
		ar2	0.085	(0.027)	
		garch1	0.747	(0.035)	
		arch1	0.173	(0.029)	
		skew	1.141	(0.044)	
		shape	8.481	(2.516)	
G20	(0,0) (1,1)	mu	-0.001	(0.000)	sstd
		garch1	0.787	(0.041)	
		arch1	0.171	(0.035)	
		skew	1.086	(0.040)	
		shape	7.998	(1.813)	

Note: (P, Q) (p, q) is (AR order, MA order) (GARCH order, ARCH order) as in equations (4.8) and (4.9). ar1, 2 are the parameter estimates for the AR part; ma1, 2 for the MA part in equation (4.8). garch1, 2 are the parameter estimates for the GARCH part; arch1, 2 for the ARCH part in (4.9). cdist is the conditional distribution of the standardized residuals: "std" represents the Student *t*, "sstd" skewed *t*, and "norm" normal. The best model is selected from multiple alternatives by using an AIC-type criterion.

To confirm the stability of the estimation result of DCC–GARCH model, we fit the same model to two sub-period data sets of the market portfolio that have almost the same numbers of trading days. We find that the parameter estimates differ little between the whole period and sub-period cases. The same check is then performed for the group portfolios and the results are similar.

7.4 Dynamic changes in correlation intensity

7.4.1 A measure of correlation intensity

The parameters of the DCC–GARCH model were estimated for the market portfolio and group portfolios presented in Section 7.3.3. In this section, we calculate $\boldsymbol{R}_t$ in equation (7.4). Because one instance of $\boldsymbol{R}_t$ exists at a time, the total number of correlation matrices is the same as the length of the return series (i.e., larger than 1,300). The dimension of $\boldsymbol{R}_t$ is 20×20 for every group portfolio and 14×14 for the market portfolio. It is difficult to observe the time series development of $\boldsymbol{R}_t$ as it is in matrix form. We hence need a further dimension reduction of $\boldsymbol{R}_t$.

The eigenvalues of the correlation matrix can be used as a vector of the proxy variables that indicate the correlation intensity on the corresponding axes. A larger eigenvalue indicates a stronger correlation. The positive maximum eigenvalue of $\boldsymbol{R}_t$ is the proxy for the correlation intensity on the first axis with the largest variance.[9] If the maximum eigenvalue is large enough, other eigenvalues may have limited influence on the correlation intensity of $\boldsymbol{R}_t$. In that case, the time series of the maximum eigenvalues approximate well the development of the correlation intensity between stock returns.

To answer the first research question: "Does the correlation matrix change over time?" presented in this chapter, we focus on the time series of the maximum eigenvalues of $\boldsymbol{R}_t$.[10] We first calculate a series of $\boldsymbol{R}_t$ by the using estimated DCC-GARCH model for the market and group portfolios. Then, the time series of the maximum eigenvalue of $\boldsymbol{R}_t$ are calculated for every sample portfolio.

Eigenvalues and random matrix theory

Table 7.4 summarizes the eigenvalues of the conditional matrix $\boldsymbol{R}_t$ and unconditional matrix $\overline{\boldsymbol{R}}$. The three largest eigenvalues (EV1, EV2, and EV3) are listed from the whole set. Note that the length of the corresponding eigenvector is normalized to one for all eigenvalues.[11] "Min"

[9] This indicates "the maximum amount of the variance of the variables which can be accounted for with a linear model by a single underlying factor" (Friedman and Weisberg [39]).

[10] The changes in a correlation matrix have two components: correlation intensity (eigenvalues) and direction (eigenvectors). We focus on correlation intensity to observe any dynamic changes, assuming that intensity has a larger influence on portfolio risk. When simulating the quantitative impact of correlation changes in Section 7.5, changes in both intensity and direction are considered with different $\boldsymbol{R}_t$.

[11] A symmetrical and positive definite matrix $\boldsymbol{R}_t$ has orthonormal eigenvectors.

and “max” represent the minimum and maximum values of the time series of the eigenvalues of $\boldsymbol{R}_t$, respectively. “Uncon” represents the eigenvalue of $\overline{\boldsymbol{R}}$.

The maximum eigenvalue (EV1) is much larger than the second and third eigenvalues (EV2 and EV3) for the market portfolio and for all individual group portfolios, suggesting that EV1 mostly determines correlation intensity. If so, we can now focus on the time series development of the maximum eigenvalue as a proxy for correlation intensity.

Random matrix theory provides a reliable measure for distinguishing informative eigenvalues from uninformative ones. The largest and the smallest eigenvalues of a Wishart matrix converges almost surely to the respective boundaries of the support of the Marčenko–Pastur distribution when the true covariance matrix is an identity matrix (Marčenko and Pastur [66] and Johnstone [51]). The Marčenko–Pastur distribution is a good approximation to the density of the eigenvalues of the correlation matrix of randomized returns.

We are, however, interested in which of the eigenvalues are meaningful by examining the largest eigenvalue of the correlation matrix of randomized returns. Importantly, we must know the threshold value that the maximum eigenvalue of the correlation matrix of randomized returns can take. If an eigenvalue of a correlation matrix is larger than the threshold value, we can safely say that it is meaningful.

To determine the threshold, we need to know the limiting distribution of the maximum eigenvalue of the randomized return correlation matrix with the same size as the sample correlation matrix. Johnstone [51] showed that the asymptotic distribution of the properly rescaled largest eigenvalue of the white Wishart population covariance matrix is the Tracy–Widom distribution, which provides the limiting distribution of the maximum eigenvalue while the Marčenko–Pastur distribution suggests the boundary of the distribution of eigenvalues. For more mathematical details on eigenvalues and the Tracy–Widom distribution, see Johnstone [51] and Tracy and Widom [98, 99, 100].

The distribution function of the Tracy–Widom distribution $F_\beta(\cdot)$ has three types of definitions depending on the value of β (1, 2, and 4). The distribution function $F_\beta(\cdot)$ is defined as

$$F_1(x) = \exp\left(-\frac{1}{2}\int_x^\infty q(y)\,dy\right)(F_2(x))^{\frac{1}{2}} \tag{7.19}$$

$$F_2(x) = \exp\left(-\int_x^\infty (y-x)q^2(y)\,dy\right) \tag{7.20}$$

$$F_4(x) = \cosh\left(\frac{1}{2}\int_x^\infty q(y)\,dy\right)(F_2(x))^{\frac{1}{2}} \tag{7.21}$$

where q is the unique solution to the ordinary differential equation called the *Painlevé* (type II) equation.

The selection of β depends on the assumption of the correlation matrix structure: $\beta = 1$ for Gaussian orthogonal ensemble, $\beta = 2$ for the Gaussian unitary ensemble, and $\beta = 4$ for the Gaussian symplectic ensemble. We set β as 1, which provides the most conservative (largest) quantile value (to be used as a threshold) compared with the other settings. For more exact and complete definitions, see Tracy and Widom [99].

We calculate the 99th percentile of the Tracy–Widom distribution ($\beta = 1$) to identify the non-random eigenvalues that are beyond this value (Table 7.4). The 99th percentile value of the Marčenko–Pastur distribution is also calculated for additional information. [12]

Table 7.4 shows that the minimum value of the maximum eigenvalues (EV1) of $\boldsymbol{R}_t$ during the data period is larger than the 99th percentile of the Tracy–Widom distribution in all sample portfolios as indicated by "$*$" in the EV1 column. This finding means that these maximum eigenvalues are all meaningful enough. Next, we find that the minimum value of EV2 is larger than the 99th percentile of the Tracy–Widom distribution only in one group portfolio, while the maximum value of EV2 is larger than the threshold only in four sample portfolios (the market portfolio and three group portfolios). This finding means that EV2 is only meaningful at certain points of time during the period.[13] Finally, the maximum value of EV3 is larger than the threshold only in one group portfolio. Hence, EV3 does not convey meaningful information.

[12] We use R package "RMTstat" to calculate the density and quantiles of the Tracy–Widom and Marčenko–Pastur distributions.

[13] G22 (one of the defensive groups), including Banks, has the largest maximum of EV1, which shows seemingly very strong correlations in regional banks. EV2 is only meaningful in G19 (one of the defensive groups), including Electric Power and Gas.

Table 7.4 Eigenvalues of the correlation matrix

Group ID	EV1				EV2				EV3				TW(99)	MP(99)
	Min	Uncon	Max		Min	Uncon	Max		Min	Uncon	Max			
Market	10.890	11.619	12.520	*	0.487	0.849	1.645	†	0.221	0.389	0.685		1.239	1.190
Cyclical														
25	12.947	13.377	14.126	*	0.757	0.872	1.047		0.549	0.602	0.730			
11	12.197	13.018	14.099	*	0.727	0.842	1.166		0.582	0.675	0.832			
26	10.687	11.130	12.367	*	0.654	0.775	1.011		0.611	0.683	0.805			
30	9.954	10.521	12.286	*	0.783	0.960	1.298	†	0.552	0.681	0.877		1.280	1.229
29	9.365	9.728	11.625	*	0.665	0.795	1.104		0.548	0.657	0.872			
13	8.878	9.329	11.144	*	0.832	1.014	2.327	†	0.617	0.733	1.024			
Defensive														
22	14.287	14.832	16.425	*	0.440	0.638	0.798		0.320	0.400	0.578			
15	9.749	10.659	11.664	*	1.149	1.282	2.927	†	0.717	0.852	1.016			
21	9.999	10.360	11.527	*	0.808	0.900	1.050		0.730	0.850	0.954			
16	10.040	10.368	11.863	*	0.708	0.869	1.063		0.628	0.674	0.827			
17	9.980	10.263	11.730	*	0.740	0.885	1.128		0.568	0.645	0.825		1.280	1.229
19	9.748	10.484	12.433	*	2.018	2.709	3.573	*	0.757	0.910	1.418	†		
18	8.416	8.686	11.180	*	0.653	0.927	1.216		0.606	0.776	0.882			
20	8.113	8.543	9.871	*	0.905	0.964	1.109		0.809	0.924	1.002			

Note: EV1, EV2, and EV3 are the maximum eigenvalue, second largest eigenvalue, and third largest eigenvalue, respectively. "Min" of EV1, -2, and -3 stands for the minimum value of the time series of the corresponding eigenvalue, "uncon" represents the eigenvalue of the unconditional correlation, and "max" the maximum value of the time series of the corresponding eigenvalue. TW(99) and MP(99) are the 99th percentile of Tracy–Widom distribution and Marčenko–Pastur distributions, respectively. "*" shows the case that the minimum eigenvalue during the period is higher than TW(99). "†" shows the case that the maximum eigenvalue is larger than TW(99).

7.4.2 Dynamic changes in maximum eigenvalues

The time series of the maximum eigenvalue of a conditional correlation matrix $\boldsymbol{R}_t$ reveals that the correlation intensity changes dynamically in the market portfolio and group portfolios. This means that both the between-group correlation and the within-group correlations of stock returns change over time. We next describe the changes in the market portfolio as well as the cyclical and defensive group portfolios in more details to answer the second research question: "Are there any significant differences in the correlation dynamics between the groups?" of this chapter.

Market portfolio

Figure 7.1 depicts the time series development of the maximum eigenvalue of $\boldsymbol{R}_t$ of the market portfolio. Recall that the market portfolio comprises 14 equally weighted index returns calculated from the 14 individual group portfolios. The correlation matrix $\boldsymbol{R}_t$ means the between-group or market-wide factor correlation of the market portfolio. The top chart shows the maximum eigenvalue of $\boldsymbol{R}_t$ as a proxy variable for correlation intensity. The bottom chart shows the mean volatility, calculated as the mean of individual return index volatilities estimated by using the univariate GARCH model. The two dotted vertical lines indicate the trading date closest to the Lehman shock and Great Earthquake in that order.

Figure 7.1 clearly shows that the between-group correlation as well as the mean volatility changes over time. We see that the mean volatility increased significantly after both events, whereas the peak levels were similar. The market-wide factor correlation also intensified during crisis periods. Moreover, the maximum eigenvalue peaked after the Great Earthquake, while persistence of increased correlation intensity was also observed.

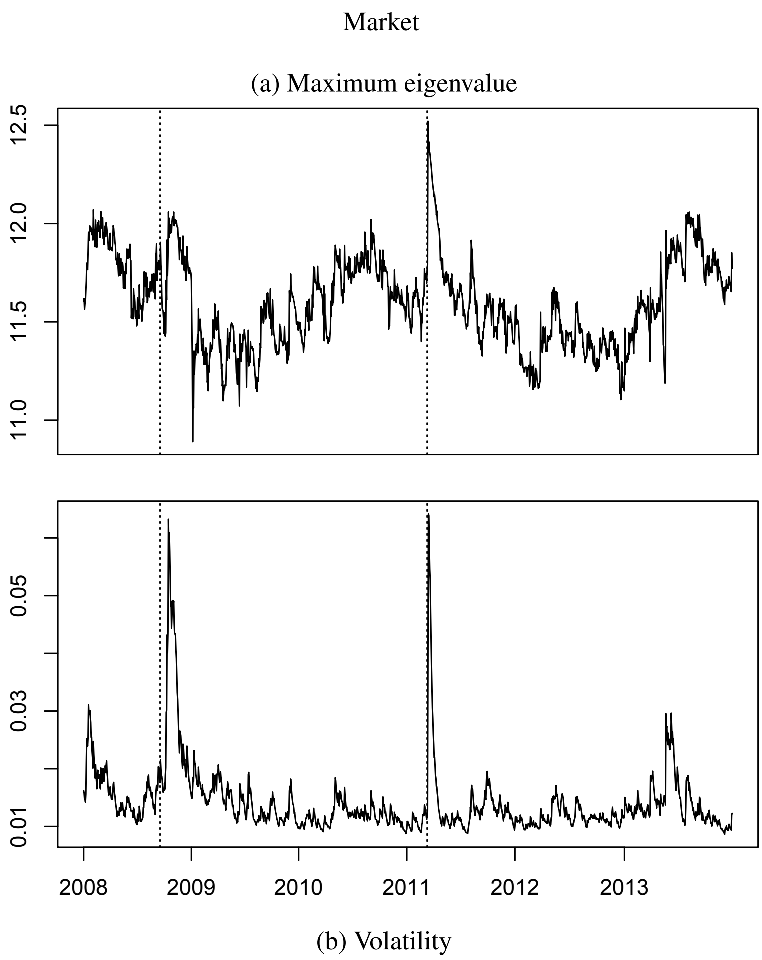
The trends of the maximum eigenvalue after the Lehman shock are complicated. This value increased considerably after the shock and remained at relatively high levels before dropping sharply. Nonetheless, the peak level after the Lehman shock is lower than that after the Great Earthquake. Further, an upward trend is also observed after the sharp decrease, which lasted for around two years, although a lack of information prevents us from clarifying the background of these movements.

Cyclical group portfolios

Figures 7.2 (G25 and G11), 7.3 (G26 and G30), and 7.4 (G29 and G13) depict how the correlation intensities of the cyclical groups change over time.

There are two charts for every group as in the market portfolio. The top chart shows the maximum eigenvalue of $\boldsymbol{R}_t$ as a proxy variable for correlation intensity. The bottom chart shows the mean volatility, calculated as the mean of the conditional volatilities of individual stock returns estimated by using the univariate GARCH model.

Overall, the correlation intensity changes dynamically in every group. Sharp increases in



Note: The maximum eigenvalue of the correlation matrix (top) and mean volatility (bottom).

Figure 7.1 Maximum eigenvalue of correlation matrix: Market within-group correlation intensity are observed after the Lehman shock and Great Earthquake, with sharp increases in volatility observed as well. The differences between the two events, however, vary by group. While persistence of increased correlation intensity as well as mean volatility is observed in many groups, the degrees of persistence differ.

In group G25, G11, and G26, for example, mean volatility is much higher after the Lehman shock than it is after the Great Earthquake. These groups include stocks in Electric Appliances and Transportation Equipment, both of which are more export-oriented sectors.[14] The larger increases in volatility suggest that the stock returns in these groups were affected more by the overseas shock. Further, the maximum eigenvalues increased significantly after both events; however, their peak levels are not necessarily higher after the Lehman shock compared with after the Great Earthquake. In G26, the maximum eigenvalue is the highest after the Great Earthquake, whereas the peak levels in G25 and G11 are similar for the two events.

By contrast, in G30, G29, and G13, which are relatively less export-oriented, mean volatility increased markedly after both events. The maximum eigenvalue also increased after both events in G30 and G29, while the degree of increase after the Great Earthquake was limited in G13.[15] In summary, the pattern of changes in correlation intensity seems to be significantly different by group.

Defensive group portfolios

Figures 7.5 (G22 and G15), 7.6 (G21 and G16), 7.7 (G17 and G19), and 7.8 (G18 and G20) depict how the correlation intensity of the defensive groups changes over time. The maximum eigenvalue of $\boldsymbol{R}_t$ changes dynamically in every group as observed in the cyclical groups. Sharp increases in within-group correlation intensity are observed after both events; the mean volatility also increased significantly after both. Comparing the changes after the two events, the peak levels of the maximum eigenvalues are higher after the Great Earthquake than they are after the Lehman shock in many groups. This trend seems to be more evident in the defensive groups, which are less export-oriented, than in the cyclical groups.

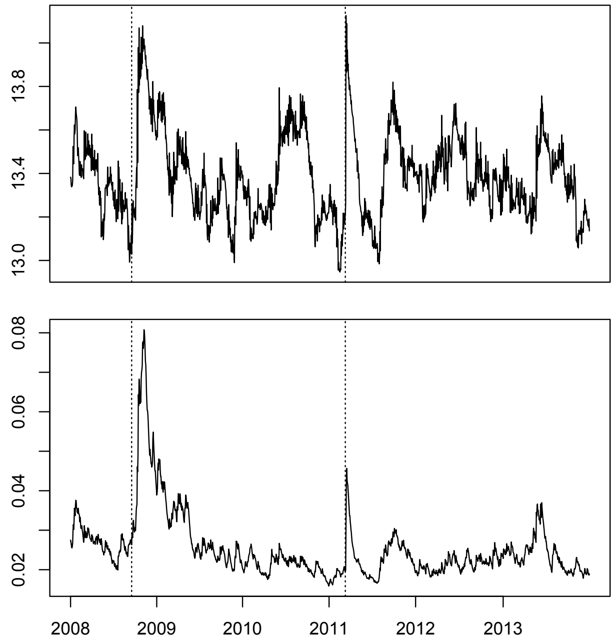
The persistence of increased correlation intensity as well as mean volatility is observed in many groups to different degrees. In G15, the correlation intensity increased after the Lehman shock, whereas the maximum eigenvalue decreased significantly after the Great Earthquake. G15 includes many construction companies like G13 in the cyclical category; hence, the same type of temporal correlation breakdown with a greater impact occurred at that time. In G19, typical defensive group, which includes Electric Power and Gas, Pharmaceutical, and Foods, a sharp increase in correlation intensity is observed after the Great Earthquake. In G21 (Information and Communication; Land Transportation) and G20 (Retail Trade; Foods), sharp increases are also observed after the Great Earthquake. Note that the mean volatilities of these

[14] For more details on the correspondence between the groups and business sectors, see Section 4.6.

[15] The stock prices of some construction companies in G13 showed an unusual pattern after the Great Earthquake; even sharp increases were partially observed, which seemingly contributed to the lower correlation intensity at that time.

G25

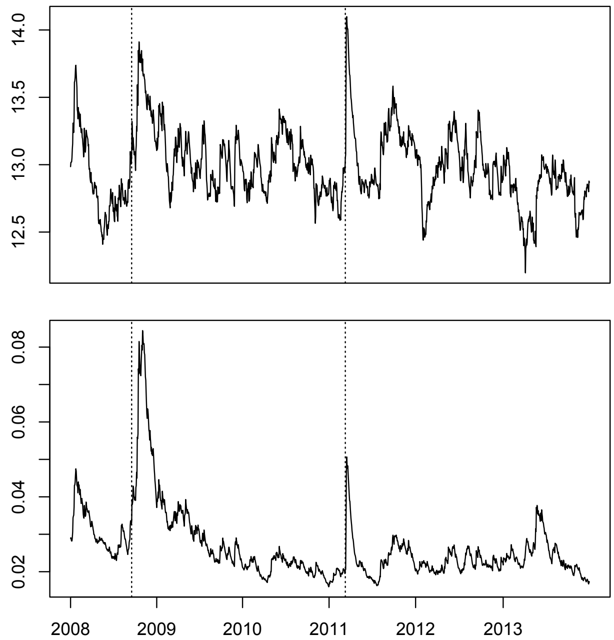
(a) Maximum eigenvalue



(b) Volatility

G11

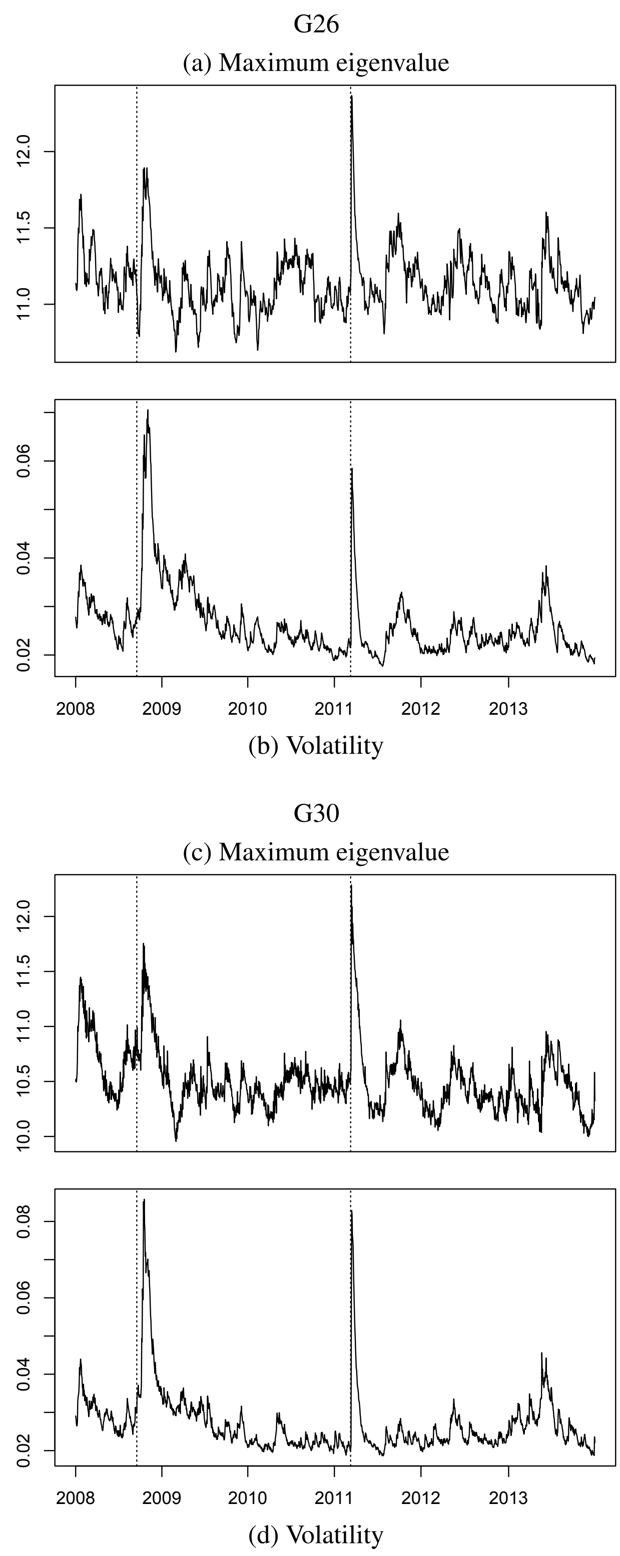
(c) Maximum eigenvalue



(d) Volatility

Note: The maximum eigenvalue of the correlation matrix (top) and mean volatility (bottom).

Figure 7.2 Maximum eigenvalue of correlation matrix: Cyclical groups–1

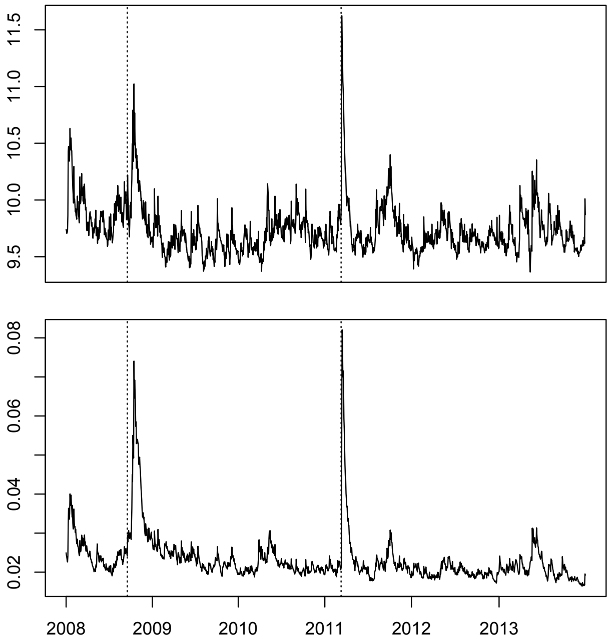


Note: The maximum eigenvalue of the correlation matrix (top) and mean volatility (bottom).

Figure 7.3 Maximum eigenvalue of correlation matrix: Cyclical groups–2

G29

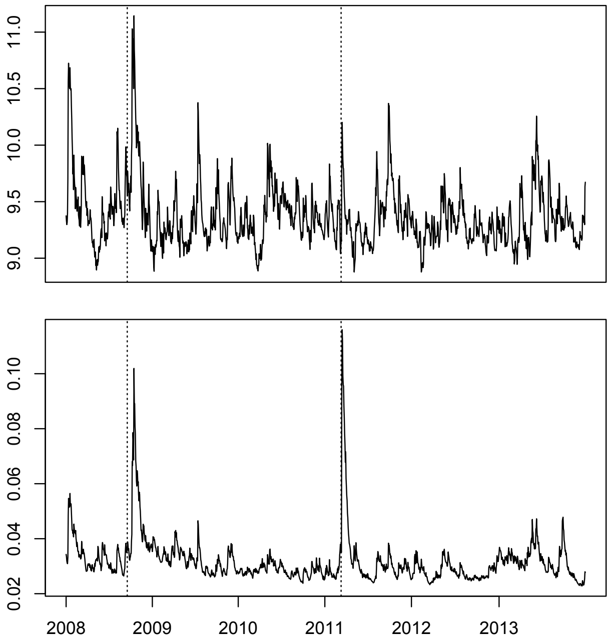
(a) Maximum eigenvalue



(b) Volatility

G13

(c) Maximum eigenvalue



(d) Volatility

Note: The maximum eigenvalue of the correlation matrix (top) and mean volatility (bottom).

Figure 7.4 Maximum eigenvalue of correlation matrix: Cyclical groups–3

groups have similar peak levels after both events, excluding G22 (Regional banks) and G15.

Hence, the combination of the observations from the cyclical and defensive groups confirms that within-group correlation intensity changes over time with a significant increase in crisis periods accompanied by a sharp rise in volatility. Further, we find significant differences in the changes in correlation intensity as well as volatility across the groups.

7.5 Impact of correlation changes on portfolio risk

Having confirmed that both within-group and between-group correlation intensities changes over time, then we are now interested in answering the third and the last research question: "How large is the impact of correlation changes on portfolio risk?" of this chapter.

To evaluate the influence on the risk of the sample portfolios, we conduct a numerical simulation analysis. The simulation focuses on the changes in correlation intensities and their influence on the portfolio risk measures: VaR and ES.

7.5.1 Correlation intensity and volatility during turbulent periods

Table 7.5 summarizes the maximum eigenvalues of the conditional correlation matrix $\boldsymbol{R}_t$ and mean volatilities of both the market and the group portfolios. This table also compares the relative changes in the maximum eigenvalue and mean volatility. The maximum eigenvalues are calculated as the mean values during the 20 trading days after the Lehman shock and Great Earthquake to smooth fluctuations. Moreover, the maximum eigenvalue of the unconditional correlation matrix $\overline{\boldsymbol{R}}$ is used as the benchmark for the relative comparison. The mean of the mean volatilities over the whole period is calculated as the unconditional mean volatility, which is also calculated as the benchmark.

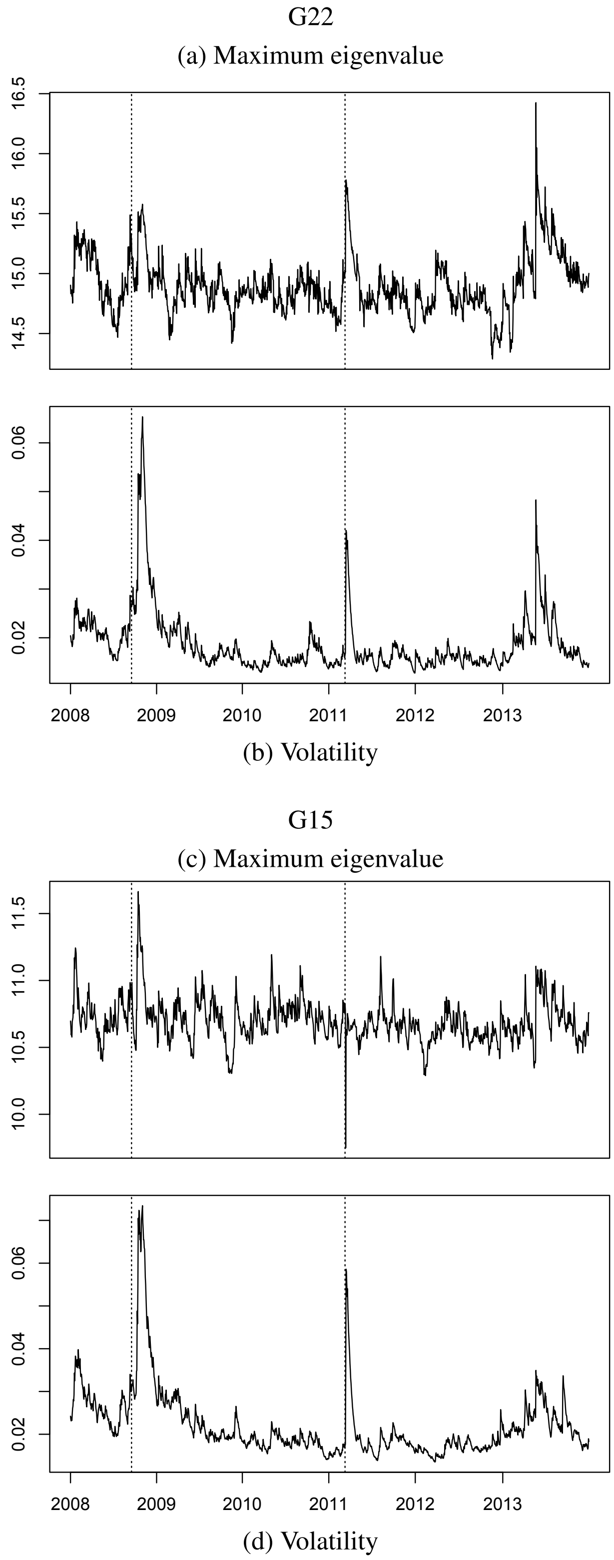
We can make two main observations here. First, the changes in the maximum eigenvalues from the unconditional one are relatively small compared with those of mean volatility, while the fluctuation in correlation intensity is also much smaller than that of mean volatility for the sample portfolios.[16] Second, the changes in the maximum eigenvalues are significantly different across the sample portfolios, while the changes in mean volatilities also differ across the sample portfolios, but not to a significant degree. At an event level, the changes are much larger after the Great Earthquake in terms of both the maximum eigenvalue and mean volatility.[17]

7.5.2 Impact study of correlation changes

We next present the results of a numerical simulation conducted to compare quantitatively the impact of correlation changes on the risk amount of the sample portfolios: the market

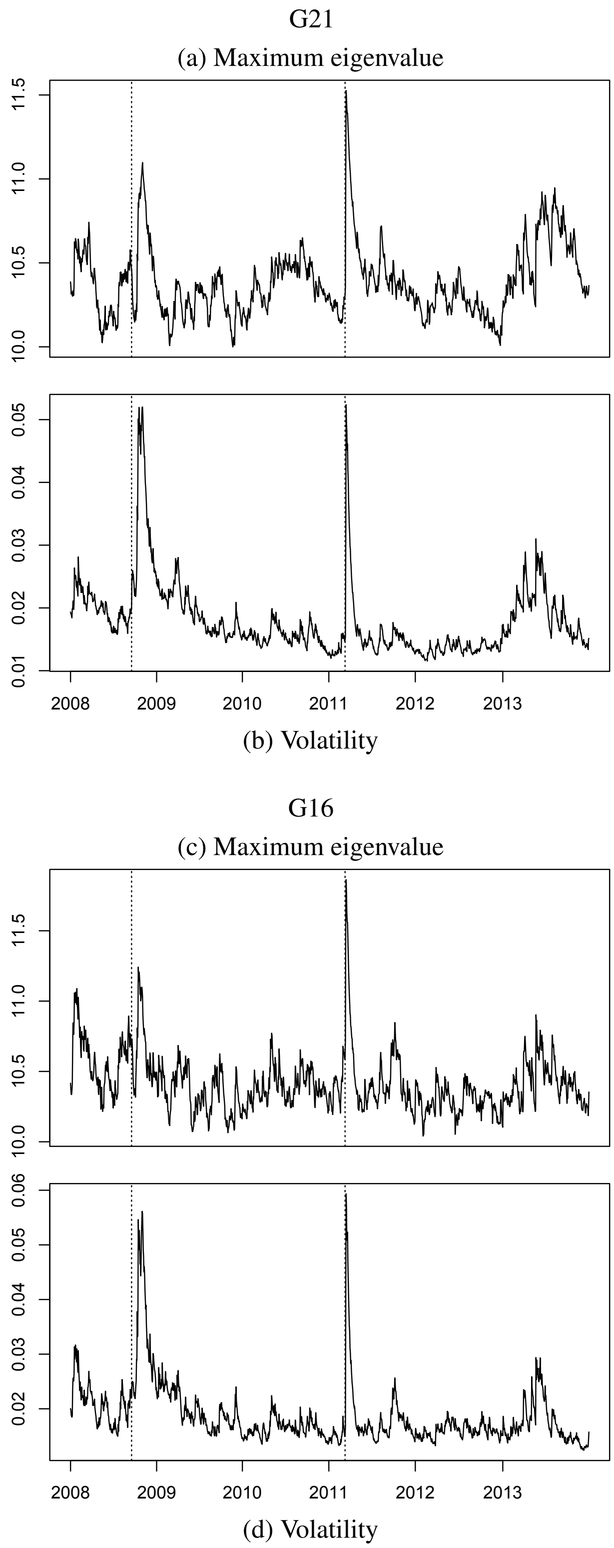
[16] Because the maximum eigenvalue and the mean volatility are measured by using different scales, the same changes in the two factors may influence the risk amount of the sample portfolio to a different degree.

[17] Note that because the sharp rise in volatility was slightly delayed after the Lehman shock, whereas it occurred immediately after the Great Earthquake, this lag might overemphasize the changes after the later event.



Note: The maximum eigenvalue of the correlation matrix (top) and mean volatility (bottom).

Figure 7.5 Maximum eigenvalue of correlation matrix: Defensive groups–1

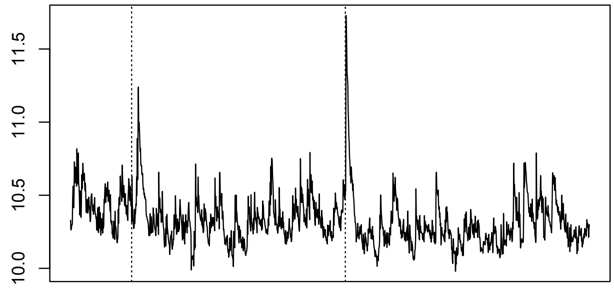


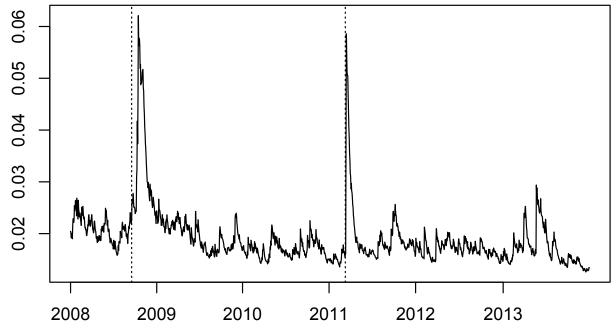
Note: The maximum eigenvalue of the correlation matrix (top) and mean volatility (bottom).

Figure 7.6 Maximum eigenvalue of correlation matrix: Defensive groups–2

G17

(a) Maximum eigenvalue

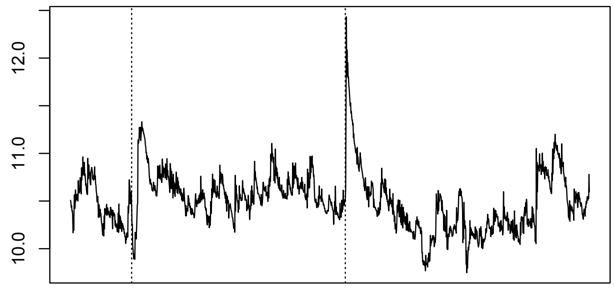


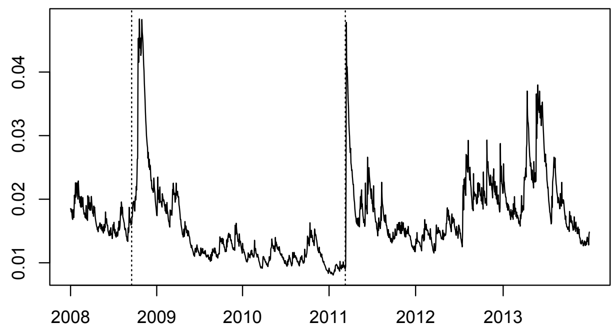


(b) Volatility

G19

(c) Maximum eigenvalue

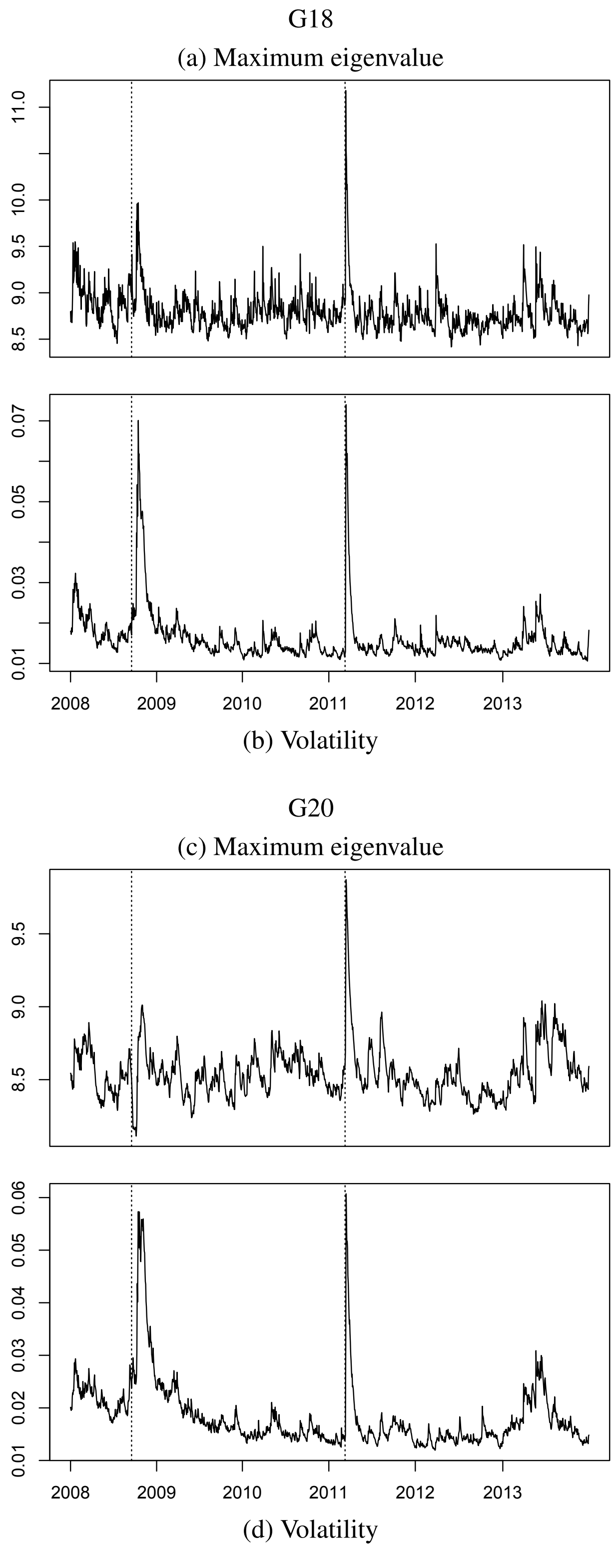




(d) Volatility

Note: The maximum eigenvalue of the correlation matrix (top) and mean volatility (bottom).

Figure 7.7 Maximum eigenvalue of correlation matrix: Defensive groups–3



Note: The maximum eigenvalue of the correlation matrix (top) and mean volatility (bottom).

Figure 7.8 Maximum eigenvalue of correlation matrix: Defensive groups–4

Table 7.5 Maximum eigenvalues and mean conditional volatilities

	Maximum eigenvalue					Mean volatility				
Group ID	Uncon (a)	L (b)	E (c)	$\frac{b-a}{a}$(%)	$\frac{c-a}{a}$(%)	Uncon (a)	L (b)	E (c)	$\frac{b-a}{a}$(%)	$\frac{c-a}{a}$(%)
Market	11.61	11.67	12.21	0.54	5.16	0.014	0.027	0.035	86.78	143.20
Cyclical										
G25	13.38	13.33	13.75	-0.37	2.77	0.025	0.034	0.033	36.00	32.00
G11	12.99	13.26	13.66	2.08	5.19	0.027	0.046	0.038	72.15	42.60
G26	11.14	11.17	11.71	0.26	5.15	0.026	0.036	0.041	34.72	55.44
G30	10.50	10.93	11.53	4.11	9.82	0.027	0.044	0.053	63.22	97.11
G29	9.74	10.13	10.46	4.01	7.37	0.023	0.038	0.049	63.24	111.05
G13	9.37	10.05	9.52	7.27	1.56	0.032	0.050	0.072	56.64	126.78
Defensive										
G22	14.90	15.05	15.45	0.98	3.69	0.019	0.030	0.030	61.51	60.34
G15	10.70	10.84	10.62	1.32	-0.73	0.022	0.037	0.040	67.22	78.88
G21	10.39	10.35	11.09	-0.40	6.74	0.018	0.028	0.034	56.35	87.27
G16	10.42	10.60	11.15	1.77	7.02	0.019	0.029	0.037	52.19	97.12
G17	10.33	10.52	10.93	1.87	5.82	0.019	0.031	0.037	64.40	94.09
G19	10.51	10.28	11.56	-2.14	10.02	0.017	0.023	0.029	39.05	74.50
G18	8.80	9.15	9.53	4.01	8.26	0.016	0.032	0.039	97.81	138.21
G20	8.54	8.32	9.23	-2.53	8.03	0.018	0.032	0.037	71.50	101.19

Note: "Uncon" of the maximum eigenvalue column denotes the maximum eigenvalue of the unconditional correlation matrix $\overline{\boldsymbol{R}}$. "Uncon" of the mean volatility column is the mean of the volatilities estimated by CCC-GARCH. L and E denote the Lehman shock and Great Earthquake, respectively. The maximum eigenvalues of L and E are calculated as the means of the maximum eigenvalues of the conditional correlation matrix $\boldsymbol{R}_t$ during the period of 20 trading days after the two events. The mean volatilities of L and E are calculated as the means of the conditional volatility during 20 trading days after the two events.

portfolio and the 14 group portfolios. Three factors are required when calculating portfolio risk: volatility, correlation, and the distribution of the probabilistic variable. The timing of the evaluation and confidence level should also be specified.

The timing here is set as the two trading days after the Lehman shock and Great Earthquake.[18] The estimated conditional volatility and correlation on these two trading days are specified in the simulation. To randomly sample the residuals, we use the Student *t*-copula, which has the same parameters, namely the conditional correlation matrix and degree of freedom, as those estimated by the DCC–GARCH model presented in Section 7.3.3. A sample data set is generated by randomly sampling 100,000 draws from the Student *t*-copula for every sample portfolio. These data are then converted into individual returns by the quantile functions of the marginal distributions; finally, the portfolio returns are calculated by applying the conditional volatilities. Note that the initial portfolio value is normalized to 1; therefore, the portfolio value after one day holding period is calculated as 1+ the sum of individual simulated returns. The portfolio risk measures (VaR and ES) defined as equation (5.1) are calculated by using the historical simulation method (for more details, see Section 5.2). The confidence level is set at 99%.

We calculate the risk measures for the sample portfolios with both unconditional and conditional correlations to observe the differences between these two cases. The case with the unconditional correlation is regarded as the benchmark. For comparison, VaR and ES are also calculated with the Gaussian copula and unconditional correlation, the combination of which is the most naive assumption when measuring risk.

7.5.3 Simulation results

Table 7.6 and 7.7 summarize the simulation results for VaR and ES, respectively. These results show that the VaR and ES of the sample portfolios increased in many cases when calculated using conditional correlation. This finding means that changes in correlation intensity can have non-negligible positive influences on portfolio risk.

In Table 7.6, the VaR values of the market portfolio with the conditional correlation is about 1.2% and 3.1% (b→a and e→d, respectively) larger than those with the unconditional correlation after the Lehman shock and Great Earthquake, respectively. The ES of the market portfolio is also about 0.3% and 3.7% larger as shown in Table 7.7. The impact is larger after the Great Earthquake than after the Lehman shock, as shown by the larger maximum eigenvalue, implying that the larger maximum eigenvalues after the Great Earthquake contribute to the larger impact.

In the case of the cyclical and defensive group portfolios, a similar tendency is observed as that in the market portfolio, although the impact differs considerably by group because of the different correlation intensities. The VaR and ES values with the conditional correlation

[18] The selected trading days are October 15, 2008 and March 16, 2011, when the maximum eigenvalue peaked for most of the sample portfolios.

are larger than those with the unconditional correlation in most groups. The VaR value with the conditional correlation is larger by about 11% (G18, after the Great Earthquake) at the maximum but is similar (G13 and G15, both after the Great Earthquake) at the minimum. The ES value is also larger by about 13% (G18, after the Great Earthquake) at the maximum and similar (G13 and G15, both after the Great Earthquake) at the minimum. These higher levels of the maximum eigenvalues seemingly contribute to the larger risk amount compared with the unconditional case, although the impacts differ significantly in both cyclical and defensive groups.[19]

When we use the Gaussian copula, the differences (between VaR and VaR* in Table 7.6; ES and ES* in Table 7.7) become larger, since this copula underestimates the risk without considering the tail dependence of the residuals unlike the Student *t*-copula. The tail dependency of returns is confirmed; however, the differences from the unconditional cases are still not large in many of the sample portfolios. The tail dependence is rather limited for the GARCH-filtered standardized residuals, as shown by the higher level of the shape parameters listed in Table 7.1.

[19] In some cyclical groups (G11, G30, and G29), the order of the maximum eigenvalues is not consistent with that of the risk amount, although the differences in VaR and ES changes are not large. This inconsistency is probably because of the differences in the random sample sets as well as the smaller eigenvalues and direction of the eigenvector of the conditional correlation contributing to the risk differently.

Table 7.6 Impact of correlation changes: VaR

	Unconditional	Lehman shock						Great Earthquake					
		Conditional		Unconditional		Changes		Conditional		Unconditional		Changes	
Group ID	Max eigv	Max eigv	VaR (a)	VaR (b)	VaR* (c)	(b→a) %	(c→a) %	Max eigv	VaR (d)	VaR (e)	VaR* (f)	(e→d) %	(f→d) %
Market	11.61	12.06	0.124	0.123	0.121	1.21	2.61	12.37	0.128	0.124	0.122	3.12	4.66
Cyclical													
G25	13.38	13.98	0.121	0.118	0.117	1.95	3.11	13.86	0.081	0.080	0.080	1.05	1.86
G11	12.99	13.98	0.140	0.134	0.133	4.97	5.14	14.12	0.089	0.086	0.085	3.57	3.86
G26	11.14	11.88	0.109	0.105	0.103	4.30	5.67	12.37	0.098	0.093	0.092	6.19	7.20
G30	10.50	11.76	0.146	0.138	0.137	5.27	6.51	11.77	0.140	0.134	0.132	4.40	6.11
G29	9.74	11.02	0.125	0.117	0.115	7.65	8.71	11.37	0.137	0.128	0.126	7.60	8.80
G13	9.37	11.24	0.164	0.150	0.146	9.50	12.40	10.13	0.167	0.167	0.163	-0.43	2.31
Defensive													
G22	14.90	15.51	0.110	0.107	0.107	2.81	2.67	15.78	0.088	0.085	0.085	3.37	3.28
G15	10.70	11.66	0.124	0.116	0.116	6.34	6.23	10.63	0.097	0.098	0.098	-0.29	-0.60
G21	10.39	10.86	0.087	0.084	0.084	3.49	3.38	11.53	0.093	0.087	0.087	6.68	7.38
G16	10.42	11.24	0.096	0.093	0.090	2.66	5.80	11.86	0.105	0.100	0.098	4.90	7.78
G17	10.33	11.24	0.110	0.105	0.103	5.21	7.10	11.65	0.104	0.099	0.098	5.57	6.70
G19	10.51	11.10	0.080	0.078	0.077	2.33	4.56	12.43	0.088	0.083	0.081	6.09	8.39
G18	8.80	9.97	0.115	0.107	0.107	7.52	7.12	10.54	0.124	0.112	0.112	11.00	10.95
G20	8.54	8.80	0.089	0.087	0.086	1.81	2.69	9.87	0.098	0.091	0.091	6.83	7.92

Note: “Conditional” and “Unconditional” represent the correlation matrix type. “Max eigv” stands for the maximum eigenvalue of the correlation matrix. The initial portfolio value is set to 1 for each portfolio. VaR is calculated at the 99% confidence level. VaR* is calculated from the estimation result of the DCC–GARCH model with the Gaussian copula.

Table 7.7 Impact of correlation changes: ES

	Unconditional	Lehman shock						Great Earthquake					
		Conditional		Unconditional		Changes		Conditional		Unconditional		Changes	
Group ID	Max eigv	Max eigv	ES (a)	ES (b)	ES* (c)	(b→a) %	(c→a) %	Max eigv	ES (d)	ES (e)	ES* (f)	(e→d) %	(f→d) %
Market	11.61	12.06	0.145	0.145	0.142	0.28	2.22	12.37	0.150	0.146	0.144	3.68	4.40
Cyclical													
G25	13.38	13.98	0.143	0.139	0.136	2.35	5.02	13.86	0.096	0.095	0.092	1.72	4.33
G11	12.99	13.98	0.165	0.158	0.156	4.58	5.99	14.12	0.106	0.102	0.100	3.69	5.24
G26	11.14	11.88	0.131	0.126	0.123	3.96	6.44	12.37	0.120	0.112	0.109	7.69	10.28
G30	10.50	11.76	0.175	0.166	0.162	5.60	8.04	11.77	0.169	0.161	0.157	4.82	7.29
G29	9.74	11.02	0.153	0.141	0.138	8.65	11.31	11.37	0.166	0.155	0.151	7.67	10.48
G13	9.37	11.24	0.201	0.182	0.175	9.99	14.76	10.13	0.202	0.202	0.194	-0.25	3.99
Defensive													
G22	14.90	15.51	0.135	0.130	0.129	3.46	4.75	15.78	0.108	0.104	0.103	3.82	5.13
G15	10.70	11.66	0.149	0.142	0.138	5.34	7.99	10.63	0.118	0.119	0.116	-0.40	1.85
G21	10.39	10.86	0.106	0.102	0.100	3.58	5.70	11.53	0.113	0.106	0.104	6.06	8.05
G16	10.42	11.24	0.116	0.112	0.108	3.60	7.71	11.86	0.129	0.121	0.117	6.38	10.40
G17	10.33	11.24	0.136	0.127	0.124	6.63	9.79	11.65	0.129	0.121	0.117	6.75	9.90
G19	10.51	11.10	0.100	0.097	0.093	3.08	7.50	12.43	0.110	0.103	0.099	6.86	10.66
G18	8.80	9.97	0.143	0.131	0.129	9.19	11.38	10.54	0.155	0.138	0.135	12.81	15.03
G20	8.54	8.80	0.109	0.106	0.103	2.79	5.64	9.87	0.120	0.111	0.108	8.14	11.26

Note: “Conditional” and “Unconditional” represent the correlation matrix type. “Max eigv” stands for the maximum eigenvalue of the correlation matrix. The initial portfolio value is set to 1 for each portfolio. ES is calculated at the 99% confidence level. ES* is calculated from the estimation result of the DCC–GARCH model with the Gaussian copula.

7.6 Technical discussion about dynamic correlation

The dynamic correlation of the Japanese stock returns is modeled by using the DCC–GARCH model, which is fitted to two types of sample portfolios: the market portfolio and group portfolios. These sample portfolios are created based on the clustering results shown in Chapter 4. It is confirmed that the correlation matrix changes over time in both the market portfolio and the group portfolios. Significant differences in the patterns of changes between the sample portfolios are also identified with sharp increases in correlation intensity observed during crisis periods. These empirical findings suggest a number of discussion points.

Modeling issue of DCC–GARCH

The first point relates to technical limitations when modeling correlations. Although DCC–GARCH can model dynamic correlation and the copula-based two-step estimation procedure also works fine to incorporate the heterogeneity of individual return distributions efficiently, it remains difficult to evaluate the changes in the estimated correlation matrix. We adopted the maximum eigenvalue of a correlation matrix as the proxy measure for correlation intensity. The maximum eigenvalue reveals the dynamic changes in correlation intensity as a scalar indicator, which helps us follow the pattern of these changes. Nevertheless, however, the maximum eigenvalue is still not directly linked to the calculation of portfolio risk. More strictly, the changes in the eigenvector of a correlation matrix can also influence portfolio risk, demanding a simulation analysis of the quantitative impact of correlation changes on portfolio risk. As for the impact study, the simulation results depend on the modeling assumptions of DCC–GARCH: no volatility spillover is considered. Further, the correlation dynamics can be described differently by other more structural dynamic correlation models. To quantify the effect of volatility spillovers on dynamic correlation, we must estimate conditional correlation or covariance by using other types of multivariate models including the BEKK model. More generalized and complicated DCC models including the BDCC and AG-DCC would improve the estimation performance. It is also meaningful to test if the same result is obtained when other clustering methods are used to create sample portfolios. A higher level of coverage of stocks in a sample portfolio is another issue to be considered.

Practical application

The second discussion point is the practical aspects of DCC–GARCH. Since DCC–GARCH provides a consistent framework with which to combine conditional mean, volatility, and correlation to measure portfolio risk, it is thus flexible to capture any time-varying changes in those three factors. Despite the technical limitations of DCC–GARCH, its compact and flexible modeling framework with parsimonious parameters and easy parameter estimation procedure are valuable from a practical viewpoint of risk control. For example, the conditional approach is beneficial for stress testing portfolio risk by using possible combinations of extreme volatilities and correlation matrices. In brief, we need extreme but plausible scenarios for meaningful stress testing. The estimated historical time series of the conditional correlation matrix as well as conditional volatility can further provide a set of realistic combinations of volatility and correlation, which may not be available with a static correlation matrix in order to set the stress level. Moreover, the eigenvalues may be used to adjust correlation intensity when building the scenarios, although further study is required to clarify their quantitative impact.

Dimension reduction

Another discussion point is the high-dimensionality of the correlation matrix of stock returns. We adopt the reduced size of sample portfolios to monitor the whole stock market, since we are interested in a more general portfolio with wider coverage rather than a specifically targeted portfolio. The clustering algorithm is based on the correlation matrix of the whole universe of stock return data, which is calculated by assuming the use of CCC–GARCH. The volatility interaction between stocks is thus not considered. Using other methods to estimate the correlation matrix may lead to different group samples. This point is an important caveat to this study. Nevertheless, our dimension reduction method works well, even for a very large number of assets with fat-tailed returns, while the group size as well as the size of sample portfolios can be modified flexibly. The method of selecting representatives from a group based on a network centrality measure can also be easily applied to other groupings.

Chapter 8

Conclusion

8.1 Summary

In Chapter 3, the fat-tailedness of asset returns is modeled by the truncated stable distribution. It has been clarified that the return of the Nikkei has a fat-tailed probability distribution. It means that the correlation matrix comprised of the linear correlation of stock returns does not necessarily capture the true correlation structure, which significantly affects the risk of the stock portfolio. The volatility model like the GARCH model works well to separate volatility fluctuations from returns. The residuals are ensured to be i.i.d., the volatility of which is standardized.

In Chapter 4, a high-dimensional correlation structure of the Japanese stock returns is analyzed by the multivariate GARCH model with CCC. The assumption of static correlation is introduced a priori to work around the difficulty of high-dimensionality. We selected about 1,400 stocks with high liquidity that are listed on the First Section of Tokyo Stock Exchange. The pairwise correlations of daily log returns of those stocks are calculated using the multivariate GARCH model. The estimated correlation matrix is converted to a network of stocks with weighted edges that correspond to the pairwise correlation of returns. Then, correlation clustering by recursive modularity optimization is applied to detect the hierarchical group structure; the 14 groups of stocks are finally identified. The total groups are seemingly comprised of two large categories: the cyclical and defensive groups. The grouping is more homogeneous and balanced compared with the standard sector classification.

Our technical contribution here is the proposal of the controlling method of recursive clustering. The method has been proven to work well to post-prune the tree of hierarchical groups, enabling to balance the size of groups. The statistical significance of the grouping has been confirmed by the random network simulation. The stability of grouping is also confirmed by the sub-portfolio and sub-sample periods analyses.

The clustering result is compared with the standard sector classification to explore how these two groups are linked by statistical tests. The hypergeometric tests reveal that many groups are over-expressed by multiple business sectors. The current sector classification is partially effective to identify some of the groups; stocks in some sectors are grouped almost together, although multiple sectors included in a group even in such cases. The standard sector classification is separated into two types of sectors: over-expressing sectors and non-overexpressing sectors. The former includes Transportation Equipment; Construction; Electric Appliances; Pharmaceutical; Electric Power and Gas; Banks; Retail Trade; and Foods. The latter includes small-sized sectors with less than 10 stocks (Mining, Insurance; and Fishery, Agriculture, and Forestry) and large sized but vaguely defined sectors (Services; Wholesale Trade; Other Products; Glass and Ceramics Products; and Metal Products).

In Chapter 5, properties of 14 groups identified by recursive clustering in Chapter 4 are explored in more details. Some informative group structures are discovered. In the cyclical groups, a typical example is that the group of automobile manufacturers is separated from the group of parts and peripheral makers, while the two groups belong to the same transportation sector in the standard sector classification. In the defensive groups, it has been observed that stocks in Pharmaceutical; Electric Power and Gas; Foods; and Banks form a tightly linked sub-group structures. As for Banks, the regional banks collectively form a group, in which large internationally operating banks are not included. These findings mean that modularity optimization works well to detect communities as homogeneous stock groups.

We also perform random portfolio simulations to confirm if the identified grouping can contribute to improving the efficiency of portfolio risk control. We propose a simulation framework to evaluate the quality of grouping: the two stages sampling of stocks from the standard sector classification and the 14 groups is implemented. The sample portfolios are randomly formed by the sampling method with and without weight optimization. The simulation result shows that the risk of the sample portfolios based on the new grouping is better controlled in most cases.

In Chapter 6, we have built classification trees that provide splitting rules to reproduce the stock groups identified by hierarchical network clustering. The clustering result that is based on the stock price data are linked with the non-price external data in order to explore how the hierarchical division process can be explained by other categorical and numerical variables. The classification trees are proved to be valid to reproduce the clustering by modularity optimization. A classification tree is represented as multistage binary trees that split a group of stocks into two groups. The complexity of the tree is controlled by cross validation to avoid overfitting. The group properties are identified by the classification tree analysis. The variables that constitute a set of splitting rules are selected with the relative variable importance scores. It is expected that variable importance can provide some hints to understand the group properties.

The analysis of the aggregated variable importance reveals that TOPIX beta (the market factor), market capitalization, and price book-value ratio are included as significantly important variables. The selected variables are consistent with the frequently used factor model of stock returns. Some other factors are also included as variables that help clarify the properties of the Japanese stock market. Such information will be helpful to build localized models for stock returns. The classification tree can be applied to find a group even for the stocks that have limited price data due to low liquidity. In order to improve the goodness-of-fit of the classification trees, we need to search for new variables. Further research will be required to find such variables.

In Chapter 7, we extend the multivariate GARCH model that is used for the volatility control of stock returns in Chapter 4: the static correlation of residuals is changed to the dynamic correlation. A multivariate GARCH model with DCC is fitted to a reduced size of sample portfolios to avoid the high-dimensionality problem. Two types of sample portfolios are created: the market portfolio to observe between-group correlation and group portfolios to observe within-group correlation. We explore the dynamics of correlation of every group as well as the impact of correlation changes on portfolio risk.

The conditional correlation matrix is estimated by fitting the Copula–DCC–GARCH model to the return data of these sample portfolios. The intensity of correlation is approximated by the maximum eigenvalue of a correlation matrix. Most of the groups have the largest and predominant eigenvalue of the correlation matrix. It has been confirmed that the correlation matrix changes over time; the correlation intensity is apparently higher during the crisis periods. DCC, hence, is preferable to CCC for modeling the Japanese stock returns. The estimation result of the market portfolio shows that the factor covariances increased during the crisis periods. The estimation results of the group portfolios show that within-group correlations also increased; however, the responses significantly differ across the groups. We also conducted simulation analysis for comparative study of correlation changes on the risk of sample portfolio. The result shows that the dynamic changes in the correlation matrix can have non-negligible positive influences on the portfolio risk in terms of VaR and ES.

The Colula–DCC–GARCH provides a consistent framework of conditional volatilities and conditional correlations for measuring portfolio risk without any missing part. It is useful for stress-testing; the possible combinations of extreme volatility and correlation. The maximum eigenvalue of a correlation matrix is a key indicator of the intensity of correlation. It should be mentioned that our framework of DCC does not consider volatility spillovers between stocks, which can affect the correlation. The correlation dynamics can be described differently by other more structural dynamic correlation models.

Further topics for future research include a higher level of coverage of stocks in every group; extension of the model to incorporate the volatility spillovers; a more detailed analysis about the background of changes in correlation intensity.

8.2 Theoretical implications

In this research, we proposed a new approach to detect the high-dimensional correlation structure of the market-wide stock returns for reducing portfolio risk with effective diversification of investment. Our research work comprises different fields of science: statistics, financial econometrics, complex networks, machine learning, and other related mathematics. It should be emphasized that data-mining in financial topics is our fundamental research strategy. The process of our research is totally data-driven; conceptual discussions without explicit data are out of scope in any part of our research work. Technically, we do not necessarily stick to any field of science. It is desirable to combine effective tools from different fields of science, if they are helpful for the data analysis. What is important here is the way to apply analytical tools including mathematical models and algorithms. We always pay careful attention to theoretical assumptions and technical prerequisites for using such tools. Our research results include many findings that are meaningful and helpful for further research and real application. We also tried to extend the existing methods to cope with the difficulties encountered in this research. Our findings and technical contributions in each area are described as follows.

Fat-tailedness and financial model

The normal distribution approximation has been widely used in financial modeling and risk analysis. Statistically, there is a lot of evidence to reject such assumption as we show in Chapter 3. From a data-oriented viewpoint, it should be tested if the normal approximation can be accepted. The fat-tailedness of financial returns is observed in a wide range of financial assets including stocks as shown in Chapter 3. If a model that depends on the normal distribution approximation were used for a product with fat-tailed returns, the model estimation and observations would be distorted and misleading. We proposed the use of the stable distribution to model the fat-tailedness of returns, because it has a solid theoretical background. If we are interested in the data, we should be more careful to select an appropriate method to handle the data.

The correlation issue is closely related to the fat-tailedness of returns. The linear correlation is a useful tool to capture the commovement of two variables; therefore, it has been widely accepted as the standard correlation measure. We clarify the necessary conditions to use the linear correlation. It should not be used for any fat-tailed variables, since the linear correlation can be significantly distorted. The correlation plays a key role to calculate portfolio risks and optimize investment weights. This point should be better acknowledged by researchers and practitioners.

Multivariate GARCH model and correlation structure

We work around the fat-tailedness problem by applying ARMA–GARCH filtering to separate volatilities from returns, ensuring theoretically appropriate conditions for correlation calculation. A correlation matrix of returns is calculated in a theoretically consistent way, avoiding possible distortion effects due to the fat-tailedness. There are two discussion points regarding the design of multivariate volatility model: a high-dimensionality and correlation modeling issues. The high-dimensionality is frequently solved by introducing factor models, which can be applied to the multivariate volatility models. It is, however, sometimes difficult to define the common factors theoretically. The statistical method including the principal component analysis provides a method to extract common factors from the data set; however, the definition of the factors are not necessarily clear. We adopt the micro-based approach focusing more on the individual stock relations rather than linking them to the common factors.

The correlation modeling issue is related to the high-dimensionality. Estimation of a large sized correlation matrix itself is a technically hard task. The choice of static or dynamic correlation in modeling the correlation structure is conceptually and empirically complicated issue. It has been argued that the dynamic correlation (DCC) is more realistic than the static one (CCC). Theoretically, the static correlation is regarded as the special case of the dynamic correlation. In this regard, the extension of the correlation modeling from static to dynamic is implemented as Copula–DCC–GARCH. The dimension reduction is critically important, if the number of assets is large as in our case. We propose a solution in which the data size is reduced but the model is unchanged. The data-oriented grouping information of stocks by correlation clustering greatly helps reduce the data size by selecting representative stocks from every group. Such data operation has the similar effect as the more advanced models with additional parameters.

Lastly, the Copula–DCC–GARCH model has an advantage that the three factors–mean, volatility, and correlation–are all conditional. The model hence is flexible to capture any time varying changes of those three factors. The feature is theoretically preferable and more realistic.

Complex networks theory

A standard clustering method including *k*-means algorithm cannot be used for correlation clustering of time series data due to the difficulty to define an appropriate distance measure for stock returns. We, therefore, need to rely on correlation clustering that are based on a correlation matrix of stock returns. The complex networks theory is applied to the clustering of the correlation matrix of stock returns. The clustering is a sort of graph clustering, which divides a graph (network) into sub-graphs. The complex networks theory covers a very wide range of application. The community detection is an area that has growing interest among many researches. It focuses on the analysis of a real network to obtain some useful information on the relationship between the members of the network, whereas the classic network analysis focuses more on the theoretical properties of networks including network topologies.

The community detection technique has a relatively short history since early 2000's; therefore, its application is still limited. As for the financial application, our research is one of seminal studies. There are some previous studies in which the community detection is applied to correlation clustering of financial returns. Our research is distinctive in that the correlation matrix size is large, covering the major part of the Japanese stock market. The modularity optimization algorithm works fine, even for a much larger size of correlation matrix. The problem resides in the calculation of the correlation matrix. The fat-tailedness of returns should be appropriately handled as mentioned above. The second problem is the resolution limit problem. It has been much debated about how to solve the problem; however, the complete solution has not been found. We overcame this problem by recursive optimization that are systematically controlled. The proposed method can be applied to other similar problems.

As such, correlation clustering by modularity optimization is powerful tool, but we should be careful to apply the method to a real problem. First, we need to understand correctly how we build the correlation matrix: how to deal with the fat-tailedness of returns in our case. We also assume a static correlation matrix for simplicity. It should be examined carefully if such simplification is acceptable or not. Second, the technical limit of modularity optimization including the resolution limit problem should be well acknowledged to overcome the problem.

8.3 Practical implications

Data-oriented classification of stocks

The standard sector classification of the Japanese stocks is widely referred in business, although the sector classification is not necessarily consistent with the actual correlation of stock returns, since the standard classification is designed to reflect the type of business of the company. A stock, therefore, can be correlated higher with a stock that belongs to a different sector than the stock that belongs to the same sector.

We propose a data-oriented classification of stock groups, which is defined by correlation clustering of returns. The link between the two classifications is explored statistically to find that our grouping is at least partly a restructured form of some sector classification. We believe both the standard classification and our grouping are meaningful. The sector classification is so familiar to investors and researchers that we can understand the business property of a company with the sector information. The information is reliable and helpful as an initial clue to further research on the company. On the other hand, our grouping is designed to reflect actual return correlations between stocks; therefore, it is suitable for quantitative use, including risk measuring and portfolio optimization as we proved by the simulation analysis in Chapter 5.

It is, therefore, important to use the two classifications appropriately depending on the occasion. We do not insist that our grouping can replace the standard sector classification. It is rather desirable to use the grouping information as a supplementary tool, especially in quantitative calculation. The knowledge acquired through the correlation analysis would be greatly helpful for understanding the stock market structure and improving stock portfolio management.

Data-mining in finance

Our research is a typical example of data-mining in finance. We rely on various types of financial theories to handle the complicated structure of financial products. Many mathematical models have been developed to cope with the complexity. A high-dimensionality sometimes becomes an obstacle to apply such advanced financial theory. A large size of multivariate model can be conceptually designed, but the implementation of the model, including the parameter estimation, is often beyond our capability. The dimension reduction by clustering would be a practical solution in such cases. At least, it can provide some hints to work around the problem as we experienced in this research.

Efficient portfolio risk control

The risk of the stock portfolio depends on the size of volatility and distribution of residuals, the correlation of which can significantly affect the risk amount. Thus, the way to define and estimate the correlation is critically important for portfolio formation and its risk management. The Copula–DCC–GARCH model is a useful tool to estimate the dynamic correlation of returns. The copula approach is very flexible to include a larger size of assets in a portfolio. The DCC assumption has some drawbacks, but the merit of parsimonious parameterization would be attractive for practical use. The method can be applied not just to portfolio risk measuring but also to prompt portfolio optimization.

Stress testing

Another practical application of our research is stress testing. Financial institutions are required to conduct stress testing in order to maintain their asset quality. They assume various types of stress scenarios to evaluate quantitatively and qualitatively their asset quality. The stress scenarios should be "extreme but plausible" to make the stress testing meaningful. For example, if we have a stress scenario about the correlation of our portfolio, we need to create a stressed correlation matrix for stress testing. Normally, the correlation matrix is calculated during some observation period. It is, therefore, not easy to estimate a correlation matrix during the crisis period, even if we find some clues of the stressed correlation matrix in the history of time series data. The Copula–DCC–GARCH model enables to estimate a correlation matrix on a specific trading day as we did in Chapter 7. That will give us some hints to set the stress level of the correlation matrix. For a more detailed discussion of stress testing, see Isogai [50].

8.4 Knowledge science

This study has two contributions worthy of note to further development of knowledge science. The first point is to realize a commonly accepted or used practice from a different viewpoint or in a different context. There are many generally accepted practices in our society that are expected to be reliable and provide good conditions for us to achieve some specific purpose. Such practices have been formed based on various observations that reflect accumulated experience of many people. They have been widely used since we believe that it is efficient to employ such knowledge for the accomplishment of some specific task. People sometimes extend the use of such knowledge to other purposes that are beyond the border of possible application fields without deep consideration about the appropriateness of such an extended use. In that case, the knowledge may prevent us from achieving our goals, while we may be unaware of such adverse effects. As for the stock group classification, the standard sector classification has been widely accepted for both qualitative and quantitative analyses. It is implicitly assumed that stocks that belong to the same sector show strong comovements. Any analytical result, therefore, can be significantly affected if such comovements do not exist or unexpected comovements exist between stocks that belong to different groups.

As such, we should use an alternative classification if we are interested in the comovements of stocks. In this regard, we have successfully achieved a new data-oriented classification of stock groups that reflect the actual comovements between stocks. In other words, we need to develop a new practice (e.g., classification) when there is no reliable alternative available. The stock return data reveal a significantly different structure of stock groups that is only partially consistent with the standard sector classification. The meaning of those new findings about the commonly-accepted classification is that we discover the hidden real structure by precise data analysis. We believe that such discovery will extend the border of the application fields of accumulated knowledge. What is important here is just to confirm if any commonly-accepted practice or knowledge is really applicable to the target problem. We need to examine the consistency of some specific practice with the target data at the first stage of our analysis. Once any inconsistency is detected, we should carefully examine the problem before we continue our work further.

The second contribution is that our research strategy suggests an efficient approach of knowledge creation in the era of big data. This study employs a wide range of scientific approaches to process a large amount of financial data to extract useful information efficiently and automatically. The speed and flexibility of analysis is gaining more importance, considering the very fast speed of data accumulation. There are many difficult problems for which we cannot find solutions in a single area of science. The stock group classification is one of them. A high-dimensional correlation structure is beyond human understanding due to the complicated calculation process. It should be mentioned that an effective combination of technologies for a specific analytical issue is found by human inspiration in our study. More specifically, econometrics, complex networks, and machine learning are combined to deal with the classification problem of stock returns. The combination works well to identify inherent technical difficulties and also provide useful insights to solve the problem. It also gives us directions of further extension of our research.

The most important point is how to find the best mixture of technologies from many alternative sets. This is not an easy task. Mathematical discussion would not be helpful in this regard. It largely depends on researchers' views and experiences that cultivate their inspiration. In the case of the correlation structure of asset returns, for example, the knowledge is partly acquired from various experiences of investors and risk management experts, and often accumulated personally. A researcher or practitioner has his or her own knowledge background. It is natural to apply the methodology developed in the area which is familiar to them. It will be an efficient way to find a solution, while it can have a side effect of oversimplification. We occasionally simplify a complicated problem so as to solve the problem with the tools we already know. Such simplification may miss some important findings and lead to a misleading result. In that sense, we should be aware of the development of other fields of science as widely as possible. The linkage of theories and techniques across many sciences is essentially important. If we can enlarge our view, we will be able to have more chances to find such a good combination. We think that a new combination of existing techniques is also an innovation from a viewpoint of knowledge science. If we can solve the research problem better, there is no reason to stick to any specific field of science. Findings and improvements from research works with such new combinations can also have positive feedback to the original fields as we did in the community detection algorithm.

8.5 Future work

We summarize the direction of our future research as follows. The topic with the highest priority is a further extension of our clustering algorithm. The complex networks theory is rapidly expanding both in theoretical contexts and practical applications. We are keen on the recent development in this area. We also try to improve our recursive clustering methods for enhancing the modularity calculation part. Due to the complicated process of recursive division of an adjacency matrix, it will be possible to find a more efficient definition of modularity. Many researchers are proposing similar concepts of modularity, some of which may be applicable to that purpose.

Our correlation network is a weighted undirected network, which express the contemporaneous correlation between stock returns. From a viewpoint of short term forecasting of stock prices, a sort of causality analysis is also an interesting topic. The network should become directed one so as to handle the inter-temporal cross correlations. The volatility modeling should also be changed to a more complicated one with volatility spillover effects.

Another possible extension of the network is to adopt the dynamic correlation network, in which the adjacency matrix changes over time. The clustering based on the static network is not the best, but is acceptable as a practical alternative, since it can be updated periodically. The dynamic network, however, can depict how the network changes in response to the internal and external shocks. Such observation is greatly helpful for deeper understanding of the stock market and further development of risk control. Technically, the Copula–DCC–GARCH model can provide a way to calculate the dynamic correlation matrix.

As for the application of the Copula–DCC–GARCH model, parameter estimation of a larger size of portfolios is an important topic. The copula approach to model the multiple dependency is the key issue for the extension. The Student *t*-copula can handle a larger size of dependencies, although the selection of copula requires further discussions. The volatility spillover effects between stocks is a topic that we cannot handle in this research as mentioned above. It is a very important topic when we are interested in the contagion effects of the financial crisis. There is some advanced model in which the volatility spillover effect is implemented. The problem again is the high-dimensionality issue.

In relation to the classification tree analysis, we need to find new data for further improvement of the goodness of fit of the classification trees. The tree can provide helpful information when we need to find a group for a newly listed stock, if the properties of the stock are expressed by non-price data. We are also interested in building a localized stock price model, adding locally effective variables to the standard model.

For the practical implementation of our research results, the clustering method should be modified so as to match any request from the business side. The algorithm should become much more flexible for any type of clustering needs. It is also important to apply our methods to other financial products to test if the method works fine.

Publications

Journal articles (peer reviewed)

[1] Isogai, T. (2014). Clustering of Japanese stock returns by recursive modularity optimization for efficient portfolio diversification. *Journal of Complex Networks*, 2(4):557–584, doi: 10.1093/comnet/cnu023.

[2] Isogai, T. (2014). A comparative simulation analysis of unconditional VaR and ES calculation methods with truncated stable distribution. In The Japanese Association of Financial Econometrics and Engineering, editor, *Risk Management* (in Japanese), 121–171, ISBN978-4-254-29022-6 C3050, Asakura Shoten, Tokyo, Japan.

Conference proceedings (peer reviewed)

[1] Isogai, T. and Hieu Chi, D. (2014). Building classification tree on Japanese stock groups partitioned by network clustering. In *Proceedings of Asian Conference on Information Systems*, December 1-3, Nha Trang, Viet Nam, 56–63, ISBN:978-4-88686-089-7, IEEJ.

[2] Isogai, T. (2013). Clustering of Japanese stock returns by recursive modularity maximization. In *Signal-Image Technology & Internet-Based Systems (SITIS), International Conference on*, December 2-5, Kyoto, Japan, 569–576, doi:10.1109/SITIS.2013.94, IEEE.

Book part

[1] Isogai, T. (2009). Scenario design and calibration. In Quagliariello, M., editor, *Stress-testing the Banking System: Methodologies and Applications*, 18–36, ISBN:978-0521767309, Cambridge University Press, Cambridge, UK.

Presentations

[1] Isogai, T. (2014). Network clustering and dynamic correlation of Japanese stock. *The 8th International Conference on Computational and Financial Econometrics*, December 6-8, Pisa, Italy, 37 pages, http://www.cfenetwork.org/CFE2014/docs/BoA%20CFE-ERCIM%202014.pdf

[2] Isogai, T. (2013). Benchmarking of unconditional VaR and ES calculation method: A comparative simulation analysis with truncated stable distribution. *The 7th International Conference on Computational and Financial Econometrics*, December 14-16, London, UK, 23 pages, http://cfenetwork.org/CFE2013/docs/BoA.pdf

[3] Isogai, T. (2013). Detecting correlation structure of stock returns by network clustering. *Financial Networks and Systemic Risk*, July 17-19, Kyoto, Japan, poster, http://www.econophysics.jp/download/fnet13_kyoto/poster/Isogai_Takashi.pdf

[4] Isogai, T. (2013). Comparative performance analysis of various VaR and ES calculation methods by numerical simulation with truncated stable distribution (in Japanese). *The 21st annual conference of Japan Finance Association*, June 1-2, Tokyo, Japan, 29 pages

References

[1] Aielli, G. P. (2011). Dynamic conditional correlation: On properties and estimation. *SSRN Electronic Journal, http://www.ssrn.com/abstract=1507743*.

[2] Aielli, G. P. and Caporin, M. (2013). Fast clustering of GARCH processes via Gaussian mixture models. *Mathematics and Computers in Simulation*, 94:205–222.

[3] Aielli, G. P. and Caporin, M. (2014). Variance clustering improved dynamic conditional correlation MGARCH estimators. *Computational Statistics & Data Analysis*, 76:556–576.

[4] Albert, R. and Barabási, A.-L. (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74(1):47.

[5] Arabie, P., Hubert, L. J., De Soete, G., et al. (1996). *Clustering and Classification*. World Scientific, Singapore.

[6] Arenas, A., Fernández, A., and Gómez, S. (2008). Analysis of the structure of complex networks at different resolution levels. *New Journal of Physics*, 10(5):053039.

[7] Bauwens, L., Laurent, S., and Rombouts, J. V. (2006). Multivariate GARCH models: A survey. *Journal of Applied Econometrics*, 21(1):79–109.

[8] Billio, M., Caporin, M., and Gobbo, M. (2006). Flexible dynamic conditional correlation multivariate GARCH models for asset allocation. *Applied Financial Economics Letters*, 2(2):123–130.

[9] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327.

[10] Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. *The Review of Economics and Statistics*, 69(3):542–547.

[11] Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH model. *The Review of Economics and Statistics*, 72(3):498–505.

[12] Bollerslev, T., Engle, R. F., and Wooldridge, J. M. (1988). A capital asset pricing model with time-varying covariances. *Journal of Political Economy*, 96(1):116–131.

[13] Bonanno, G., Caldarelli, G., Lillo, F., and Mantegna, R. N. (2003). Topology of correlation-based minimal spanning trees in real and model markets. *Physical Review E*, 68(4):046130.

[14] Borak, S., Misiorek, A., and Weron, R. (2011). Models for heavy-tailed asset returns. In *Statistical Tools for Finance and Insurance*, pages 21–55. Springer, Berlin, Heidelberg.

[15] Brandes, U., Delling, D., and Gaertler, M. (2007). On finding graph clusterings with maximum modularity. In Gamkerlidze, R. V., editor, *Graph-Theoretic Concepts in Computer Science*, pages 121–132. Springer, Berlin, Heidelberg.

[16] Brandes, U., Delling, D., Gaertler, M., Görke, R., Hoefer, M., Nikoloski, Z., and Wagner, D. (2008). On modularity clustering. *IEEE Transactions on Knowledge and Data Engineering*, 20(2):172–188.

[17] Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J. (1984). *Classification and Regression Trees*. Wadsworth International Group, Belmont, CA.

[18] Caporin, M. and McAleer, M. (2013). Ten things you should know about the dynamic conditional correlation representation. *Econometrics*, 1(1):115–126.

[19] Caporin, M. and McAleer, M. (2014). Robust ranking of multivariate GARCH models by problem dimension. *Computational Statistics & Data Analysis*, 76:172–185.

[20] Cappiello, L., Engle, R. F., and Sheppard, K. (2006). Asymmetric dynamics in the correlations of global equity and bond returns. *Journal of Financial econometrics*, 4(4):537–572.

[21] Clauset, A., Moore, C., and Newman, M. E. J. (2008). Hierarchical structure and the prediction of missing links in networks. *Nature*, 453(7191):98–101.

[22] Cont, R. (2007). Volatility clustering in financial markets: Empirical facts and agent-based models. In *Long Memory in Economics*, pages 289–309. Springer, Berlin, Heidelberg.

[23] Demarta, S. and McNeil, A. J. (2005). The t copula and related copulas. *International Statistical Review*, 73(1):111–129.

[24] Engle, R. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*, 20(3):339–350.

[25] Engle, R. and Kelly, B. (2012). Dynamic equicorrelation. *Journal of Business & Economic Statistics*, 30(2):212–228.

[26] Engle, R. and Sheppard, K. (2001). Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH. *National Bureau of Economic Research*, w8554.

[27] Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4):987–1007.

[28] Engle, R. F. and Bollerslev, T. (1986). Modelling the persistence of conditional variances. *Econometric Reviews*, 5(1):1–50.

[29] Engle, R. F. and Kroner, K. F. (1995). Multivariate simultaneous generalized ARCH. *Econometric Theory*, 11(01):122–150.

[30] Engle, R. F., Ng, V. K., and Rothschild, M. (1990). Asset pricing with a factor-ARCH covariance structure. *Journal of Econometrics*, 45(1-2):213–237.

[31] Everitt, B. S., Landau, S., Leese, M., and Stahl, D. (2011). *Cluster Analysis*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Chichester, UK.

[32] Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56.

[33] Fenn, D. J., Porter, M. A., Mucha, P. J., McDonald, M., Williams, S., Johnson, N. F., and Jones, N. S. (2012). Dynamical clustering of exchange rates. *Quantitative Finance*, 12(10):1493–1520.

[34] Fleming, J. (2003). A closer look at the relation between GARCH and stochastic autoregressive volatility. *Journal of Financial Econometrics*, 1(3):365–419.

[35] Fortunato, S. (2010). Community detection in graphs. *Physics Reports*, 486(3):75–174.

[36] Fortunato, S. and Barthélemy, M. (2007). Resolution limit in community detection. *Proceedings of the National Academy of Sciences of the United States of America*, 104(1):36–41.

[37] Franses, P. H., van der Leij, M., and Paap, R. (2008). A simple test for GARCH against a stochastic volatility model. *Journal of Financial Econometrics*, 6(3):291–306.

[38] Frey, R. and McNeil, A. J. (2001). Modelling dependent defaults. Technical report, ETH, Eidgenössische Technische Hochschule Zürich, Department of Mathematics.

[39] Friedman, S. and Weisberg, H. F. (1981). Interpreting the first eigenvalue of a correlation matrix. *Educational and Psychological Measurement*, 41(1):11–21.

[40] Ghalanos, A. (2014a). *The rmgarch Models Background and Properties (Version 1.2-8)*. ftp://cran.at.r-project.org/pub/R/web/packages/rmgarch/vignettes/The_rmgarch_models.pdf.

[41] Ghalanos, A. (2014b). *rmgarch: Multivariate GARCH Models*. R package version 1.2-8 (http://cran.r-project.org/web/packages/rmgarch/index.html).

[42] Girvan, M. and Newman, M. E. J. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America*, 99(12):7821–7826.

[43] Granell, C., Gomez, S., and Arenas, A. (2012). Hierarchical multiresolution method to overcome the resolution limit in complex networks. *International Journal of Bifurcation and Chaos*, 22(07):1250171.

[44] Hafner, C. M. and Franses, P. H. (2009). A generalized dynamic conditional correlation model: Simulation and application to many assets. *Econometric Reviews*, 28(6):612–631.

[45] Hastie, T., Tibshirani, R., and Friedman, J. (2005). *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. Spinger, New York.

[46] Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies*, 6(2):327–343.

[47] Horvath, S. (2011). *Weighted Network Analysis: Applications in Genomics and Systems Biology*. Springer, New York.

[48] Hubert, L. and Arabie, P. (1985). Comparing partitions. *Journal of Classification*, 2(1):193–218.

[49] Iino, T. and Iyetomi, H. (2012). Subcommunities and their mutual relationships in a transaction network. *Progress of Theoretical Physics Supplement*, 194:144–157.

[50] Isogai, T. (2009). Scenario design and calibration. In Quagliariello, M., editor, *Stress-testing the Banking System: Methodologies and Applications*, pages 18–36. Cambridge University Press, Cambridge.

[51] Johnstone, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Annals of Statistics*, 29(2):295–327.

[52] Jondeau, E. and Rockinger, M. (2002). Conditional dependency of financial series: the copula-GARCH model. *FAME research paper*, (69).

[53] Jondeau, E. and Rockinger, M. (2006). The copula-GARCH model of conditional dependencies: An international stock market application. *Journal of International Money and Finance*, 25(5):827–853.

[54] Karrer, B., Levina, E., and Newman, M. E. J. (2008). Robustness of community structure in networks. *Physical Review E*, 77(4):046119.

[55] Kaufman, L. and Rousseeuw, P. J. (2009). *Finding Groups in Data: An Introduction to Cluster Analysis*. John Wiley & Sons, Inc., New York.

[56] Kim, S., Shephard, N., and Chib, S. (1998). Stochastic volatility: Likelihood inference and comparison with ARCH models. *The Review of Economic Studies*, 65(3):361–393.

[57] Koponen, I. (1995). Analytic approach to the problem of convergence of truncated Lévy flights towards the Gaussian stochastic process. *Phys. Rev. E*, 52(1):1197–1199.

[58] Kubota, K. and Takehara, H. (1997). Common risk factors of Tokyo stock exchange firms: In finding mimicking portfolios. *Advances in the Pacific Basin Capital Market Research*, 3:273–302.

[59] Kunitomo, N. and Owada, T. (2006). Empirical likelihood estimation of Levy processes. *Graduate School of Economics, University of Tokyo Discussion Paper*, (CIRJE-F-272).

[60] Lancichinetti, A. and Fortunato, S. (2011). Limits of modularity maximization in community detection. *Physical Review E*, 84(6):066122.

[61] Lanne, M. and Saikkonen, P. (2007). A multivariate generalized orthogonal factor GARCH model. *Journal of Business & Economic Statistics*, 25(1):61–75.

[62] Lee, T. H. and Long, X. (2009). Copula-based multivariate GARCH model with uncorrelated dependent errors. *Journal of Econometrics*, 150(2):207–218.

[63] Mandelbrot, B. B. (1963). The variation of certain speculative prices. *The Journal of Business*, 36(4):394–419.

[64] Mantegna, R. N. (1999). Hierarchical structure in financial markets. *The European Physical Journal B-Condensed Matter and Complex Systems*, 11(1):193–197.

[65] Mantegna, R. N. and Stanley, H. E. (2000). *An Introduction to Econophysics: Correlations and Complexity in Finance*. Cambridge University Press, Cambridge.

[66] Marčenko, V. A. and Pastur, L. A. (1967). Distribution of eigenvalues for some sets of random matrices. *Sbornik: Mathematics*, 1(4):457–483.

[67] McAleer, M. and Medeiros, M. C. (2008). Realized volatility: A review. *Econometric Reviews*, 27(1-3):10–45.

[68] McNeil, A. J., Frey, R., and Embrechts, P. (2010). *Quantitative Risk Management: Concepts, Techniques, and Tools*. Princeton Series in Finance. Princeton University Press, Princeton.

[69] Miccichè, S., Bonanno, G., Lillo, F., and N Mantegna, R. (2003). Degree stability of a minimum spanning tree of price return and volatility. *Physica A: Statistical Mechanics and its Applications*, 324(1):66–73.

[70] Misiorek, A. and Weron, R. (2012). Heavy-tailed distributions in var calculations. In Gentle, J. E., Härdle, W. K., and Mori, Y., editors, *Handbook of Computational Statistics*, volume 05, pages 1025–1059. Springer, Berlin, Heidelberg.

[71] Nelsen, R. B. (2006). *An Introduction to Copulas (Springer Series in Statistics)*. Springer-Verlag, New York.

[72] Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, 59(2):347–370.

[73] Newman, M. E. J. (2003). The structure and function of complex networks. *SIAM Review*, 45(2):167–256.

[74] Newman, M. E. J. (2004). Analysis of weighted networks. *Physical Review E*, 70(5):056131.

[75] Newman, M. E. J. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences of the United States of America*, 103(23):8577–8582.

[76] Newman, M. E. J. (2008). Mathematics of networks. In Durlauf, S. N. and Blume, L. E., editors, *The New Palgrave Encyclopedia of Economics*. Palgrave Macmillan, Basingstoke, UK, 2nd edition.

[77] Newman, M. E. J. and Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical Review E*, 69(2):026113.

[78] Nolan, J. P. (2015). *Stable Distributions - Models for Heavy Tailed Data*. Birkhauser, Boston. In progress, Chapter 1 online at academic2.american.edu/~jpnolan.

[79] Opsahl, T. (2009). *Structure and Evolution of Weighted Networks*. University of London (Queen Mary College), London.

[80] Patton, A. J. (2006). Modelling asymmetric exchange rate dependence. *International Economic Review*, 47(2):527–556.

[81] Pelletier, D. (2006). Regime switching for dynamic correlations. *Journal of Econometrics*, 131(1):445–473.

[82] Pham, V. T. L. and Long, T. (2007). Constructing Fama-French factors from style indexes: Japanese evidence. *Economics Bulletin*, 7(7):1–10.

[83] Piccardi, C., Calatroni, L., and Bertoni, F. (2011). Clustering financial time series by network community analysis. *International Journal of Modern Physics C*, 22(1):35–50.

[84] Rani, S. and Sikka, G. (2012). Recent techniques of clustering of time series data: A survey. *International Journal of Computer Applications*, 52(15):1–9.

[85] Reichardt, J. and Bornholdt, S. (2006). Statistical mechanics of community detection. *Physical Review E*, 74(1):016110.

[86] Rockafellar, R. T. and Uryasev, S. (2000). Optimization of conditional value-at-risk. *Journal of Risk*, 2:21–41.

[87] Schaeffer, S. E. (2007). Graph clustering. *Computer Science Review*, 1(1):27–64.

[88] Shephard, N. (1996). *Statistical Aspects of ARCH and Stochastic Volatility*, pages 1–67. Chapman & Hall, London.

[89] Silvennoinen, A. and Teräsvirta, T. (2005). Multivariate autoregressive conditional heteroskedasticity with smooth transitions in conditional correlations. Technical report, SSE/EFI Working Paper Series in Economics and Finance.

[90] Silvennoinen, A. and Teräsvirta, T. (2009). Modeling multivariate autoregressive conditional heteroskedasticity with the double smooth transition conditional correlation GARCH model. *Journal of Financial Econometrics*, 7(4):373–411.

[91] Silvennoinen, A. and Teräsvirta, T. (2009). Multivariate GARCH models. In Mikosch, T., Kreiß, J. P., Davis, R. A., and Andersen, T. G., editors, *Handbook of Financial Time Series*, pages 201–229. Springer, Berlin, Heidelberg.

[92] Sklar, M. (1959). Fonctions de répartition à n dimensions et leurs marges. In *Publ. Inst. Stat. 8*, pages 229–231. Université Paris, Paris.

[93] Spielman, D. A. (2007). Spectral graph theory and its applications. In *Proceedings of the 48th Annual IEEE Symposium on Foundations of Computer Science*, FOCS '07, pages 29–38, Washington, DC. IEEE Computer Society.

[94] Svetlozar, T. R., Young, S. K., Michele, L. B., and Frank, J. F. (2011). *Financial Models with Lévy Processes and Volatility Clustering*. John Wiley & Sons, Inc., Hoboken, New Jersey.

[95] Taylor, S. J. (1994). Modeling stochastic volatility: A review and comparative study. *Mathematical Finance*, 4(2):183–204.

[96] Teräsvirta, T. (2009). An introduction to univariate GARCH models. In Mikosch, T., Kreiß, J. P., Davis, R. A., and Andersen, T. G., editors, *Handbook of Financial Time Series*, pages 17–42. Springer, Berlin, Heidelberg.

[97] Therneau, T., Atkinson, B., and Ripley, B. (2013). *rpart: Recursive Partitioning*. R package ersion 4.1-1.

[98] Tracy, C. A. and Widom, H. (1994). Level-spacing distributions and the airy kernel. *Communications in Mathematical Physics*, 159(1):151–174.

[99] Tracy, C. A. and Widom, H. (1996). On orthogonal and symplectic matrix ensembles. *Communications in Mathematical Physics*, 177(3):727–754.

[100] Tracy, C. A. and Widom, H. (2009). The distributions of random matrix theory and their applications. In *New Trends in Mathematical Physics*, pages 753–765. Springer, Netherlands.

[101] Tse, Y. K. and Tsui, A. K. C. (2002). A multivariate generalized autoregressive conditional heteroscedasticity model with time-varying correlations. *Journal of Business & Economic Statistics*, 20(3):351–362.

[102] Tumminello, M., Lillo, F., and Mantegna, R. N. (2010). Correlation, hierarchies, and networks in financial markets. *Journal of Economic Behavior & Organization*, 75(1):40–58.

[103] Tumminello, M., Miccichè, S., Lillo, F., Varho, J., Piilo, J., and Mantegna, R. N. (2011). Community characterization of heterogeneous complex systems. *Journal of Statistical Mechanics: Theory and Experiment*, 2011(01):P01019.

[104] Van der Weide, R. (2002). GO-GARCH: A multivariate generalized orthogonal GARCH model. *Journal of Applied Econometrics*, 17(5):549–564.

[105] Vathy-Fogarassy, Á. and Abonyi, J. (2013). *Graph-Based Clustering and Data Visualization Algorithms*. Springer-Verlag, London.

[106] Von Luxburg, U. (2007). A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416.

[107] Warren Liao, T. (2005). Clustering of time series data–a survey. *Pattern Recognition*, 38(11):1857–1874.