Stony Brook University

Application of Risk Analysis based on Advanced Probabilistic Models

A Dissertation Presented

by

Naoshi Tsuchida

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Applied Mathematics and Statistics

(Quantitative Finance)

Stony Brook University

December 2013

Stony Brook University

The Graduate School

Naoshi Tsuchida

We, the dissertation committee for the above candidate for the

Doctor of Philosophy degree, hereby recommend

acceptance of this dissertation.

Svetlozar Rachev – Dissertation Advisor
Professor, Endowed Chair for Quantitative Finance,
Department of Applied Mathematics and Statistics

James Glimm – Chairperson of Defense
Distinguished Professor,
Department of Applied Mathematics and Statistics

Haipeng Xing
Associate Professor,
Department of Applied Mathematics and Statistics

Noah Smith – Outside Member
Assistant Professor,
College of Business

This dissertation is accepted by the Graduate School

Charles Taber
Dean of the Graduate School

Abstract of the Dissertation

Application of Risk Analysis based on Advanced Probabilistic Models

by

Naoshi Tsuchida

Doctor of Philosophy

in

Applied Mathematics and Statistics

(Quantitative Finance)

Stony Brook University

2013

Risk analysis is one of the central parts of modern finance theory. It covers various topics: modeling, measuring, managing, and forecasting risk and returns. The purpose of this dissertation is to describe the recent methodology of risk analysis and to show its practical applications.

Chapter 1 is dedicated to review the quantitative methodology of recent risk analysis, in which three components are presented in order to discuss its nature. The first component is the modeling of marginal distribution: The flaws of the Gaussian distribution and the alternative distributions based on the theory of the Levy process are discussed. The second component is the modeling of joint distribution: A copula model, a factor model, and independent component analysis are discussed. The third component is the definition of risk and its measurements: The idea of value at risk (VaR) is introduced.

In Chapter 2, we present a practical example of how to construct a portfolio based on the return model and risk measure. The result is tested by the method of backtesting. Consequently, it is found that (1) the ARMA-GARCH model with classical tempered stable (CTS) distribution provides better prediction than that with the normal and Student-t distribution, and (2) average VaR (AVaR) provides a better risk measure than variance. It is also suggested that the number of universe has effect on the portfolio return, and that it is effective to reduce stock universe to large capitalization stocks.

In Chapter 3, we analyze the distribution of returns on seven major Eurozone sovereign bonds (France, Germany, Greece, Ireland, Italy, Portugal, and Spain) and their co-movement. We investigate the ARMA-GARCH models based on different assumptions about the innovations: Gaussian, Student-t, CTS, normal tempered stable (NTS), and α-stable. For each of the five models, we apply four copula functions, and assess the forecasting performance of combinations

of these models. In addition, to find a forward-looking measure to detect the financial crisis of Greece, we analyze the evolution of the tail parameter over time.

In Chapter 4, we discusses the goodness of fitting of independent component analysis (ICA) using the sovereign CDS premiums of 11 Eurozone countries (Austria, Belgium, Finland, France, Germany, Greece, Ireland, Italy, Netherlands, Portugal, and Spain). Based on the log-likelihood, the Akaike information criterion and the Bayesian information criterion, we first show the fitness of ICA is as good as more complicated models such as the ARMA-GARCH and Student-*t* copula model. Also, using the structural default model based on CDS premium, we found that the characteristics of joint defaults based on ICA seem different from those based on copula.

To Mariko Tsuchida

Contents

1 Introduction: Basic Concepts **1**

1.1 Heavy-tailed distribution 4
1.1.1 Gaussian distribution 4
1.1.2 Brownian motion 5
1.1.3 α-stable distribution 6
1.1.4 Lévy process 10
1.1.5 Classical tempered stable distribution 12
1.1.6 Normal tempered stable distribution 13
1.2 Risk measurement 15
1.2.1 Volatility 15
1.2.2 Value at risk 16
1.2.3 Practical computation of VaR 17
1.2.4 Time series model 21
1.2.5 Backtesting 23
1.2.6 Average value at risk 27
1.2.7 Backtesting revisit 29
1.3 Copula model 30
1.3.1 Gaussian copula 31
1.3.2 Tail dependency 32
1.3.3 Student-t copula 33
1.3.4 Estimation of copula parameters 34
1.3.5 Empirical copula 36
1.3.6 Archimedean copula 36
1.3.7 Skewed Student-t distribution 37
1.3.8 Multivariate normal tempered stable distribution . . . 39
1.4 Independent component analysis 39

1.4.1 Factor analysis in finance . . . 39
1.4.2 Principal component analysis . . . 41
1.4.3 Independent component analysis . . . 43
1.4.4 Theoretical background of ICA . . . 44

2 Mean-ETL Portfolio Selection under Maximum Weight and Turnover Constraints based on Fundamental Security Factors 47
2.1 Introduction . . . 48
2.2 Data . . . 49
2.3 Forecasting model . . . 50
2.3.1 Factor model . . . 50
2.3.2 Time series model . . . 53
2.3.3 Simulation . . . 54
2.4 Optimization approaches . . . 55
2.5 Comparison of forecasts and approaches . . . 57
2.6 Portfolio selection with turnover constraint . . . 62
2.7 Large market capitalization securities . . . 64
2.8 Result . . . 65
2.9 Conclusion . . . 65

3 Time Series and Copula Dependency Analysis for Eurozone Sovereign Bond Returns 69
3.1 Introduction . . . 70
3.2 Data . . . 72
3.3 Methodology . . . 73
3.3.1 Calculating the residuals from the ARMA-GARCH model 73
3.3.2 Fitting the marginal distribution to the residuals . . . 74
3.3.3 Applying the joint copula to the CDF of the residuals . 75
3.3.4 Backtesting . . . 76
3.4 Empirical results . . . 77
3.5 Tail heaviness and its application to risk prediction . . . 84
3.5.1 Tail parameters . . . 84
3.5.2 VaR spread . . . 87
3.6 Conclusion . . . 89

4 ICA-GARCH Analysis of Joint Default Probability of Eurozone Sovereigns 90
4.1 Introduction . . . 91
4.2 Data . . . 92
4.3 CDS return model based on ICA . . . 94
4.3.1 ICA model . . . 94
4.3.2 ICA-GARCH model . . . 95
4.3.3 AR-ICA-GARCH model . . . 96
4.3.4 ARMA-GARCH and copula model . . . 97
4.3.5 Empirical results of fitting . . . 97
4.4 Joint default model . . . 101
4.4.1 Default probability of single country . . . 101
4.4.2 Connecting default probability with CDS . . . 105
4.4.3 Joint default probability . . . 106
4.4.4 Joint default statistics . . . 107
4.4.5 Joint default results . . . 107
4.5 Conclusion . . . 108

5 Conclusion 112

List of Figures

1.1 α-stable distribution 9
1.2 CTS distribution 14
1.3 95% VaR 18
1.4 95% AVaR 28

2.1 Explanatory power of 21 fundamental factors 51
2.2 Sample correlation of factor and non-factor returns 52
2.3 Explanatory power with respect to the number of principal components 53
2.4 Simulated scenarios of the first hidden factors by normal Student-t and CTS distributions 54
2.5 Correlation of sample returns and correlation of simulated returns 56
2.6 Cumulative returns with weight constraint, without turnover constraint 63
2.7 Cumulative returns with weight and turnover constraint . . . 66

3.1 Time series of Cramér-von Mises statistics 81
3.2 Time series of tail-heaviness parameters 86
3.3 Daily VaR spread of Greece 88

4.1 Time series of log-likelihood, AIC and BIC during Period 2 . . 102
4.2 Time series of log-likelihood, AIC and BIC during Period 4 . . 103
4.3 Default probability during Period 2 104
4.4 Time series of $D_{\geq 1}$, D_1, and $D_{\geq 2}$ during Period 4 109
4.5 Time series of CPSD and SI during Period 4 110

List of Tables

1.1 Cumulative probability of VaR exceedances 25

2.1 Comparison of forecasts and approaches 59
2.2 Comparison of selection levels 67

3.1 The p-values for each marginal distribution and each country, using KS statistic 78
3.2 The p-values for each marginal distribution and each country, using AD statistic 78
3.3 Cramér-von Mises statistics for each joint copula 79
3.4 Number of rejected days in backtesting of each marginal model, using KS statistic 80
3.5 Number of rejected days in backtesting of each marginal model, using AD statistic 80
3.6 VaR violations, Christoffersen test, and average loss over AVaR: Daily result 82
3.7 VaR violations, Christoffersen test, and average loss over AVaR: Weekly result 83
3.8 Tail parameters for the entire period 85

4.1 Data periods and countries 93
4.2 List of models 98
4.3 List of log-likelihood 98
4.4 List of AIC 99
4.5 List of BIC 100

Acknowledgments

I would like to thank my advisor, Svetlozar T. Rachev, for his insightful guidance, invaluable advice, and warm encouragement. I also wish to thank my committee members, James Glimm, Haipeng Xing, and Noah Smith. I also thank Frank J. Fabozzi, Rosella Giacometti, John Guerard, Young Shin Kim, Andrew Mullhaupt, Ann Tucker and Xiaoping Zhou for research cooperation and helpful advice. In addition, I thank my friends, colleagues and staffs who have supported me during my time at the Department of Applied Mathematics and Statistics in Stony Brook University.

I acknowledge the tuition scholarship provided by Stony Brook University and the financial support provided by the Bank of Japan.

Finally, I would greatly appreciate the cooperation of my parents, Tamako Tsuchida and Kazunari Tsuchida, and my sister, Mariko Tsuchida.

Chapter 1

Introduction: Basic Concepts

Risk analysis is one of the central parts of modern finance theory. It covers various topics: modeling, measuring, managing, and forecasting risk and returns. Since the history of finance is the history of big losses, risk control has been one of the most important topics for both practitioners and researchers. The purpose of risk analysis is to understand the risk in a specific investment. However, the analysis is not limited to investment, but it is related to the other fields of finance such as asset and option pricing, economic analysis, regulations, and so on. This is because the effect of risk is not only limited to investors, but also involves all participants in the market and all participants of the global economy. Conversely, a specific investment cannot be separated from the effect of global trends. These interactions are easily seen in historical situations such as the Great Depression and the Lehman collapse. Ultimately, the purpose of risk analysis is to understand the entire of finance and economy, and might be to understand the world.

In finance, risk is usually classified into several categories. For example, the Basel II accord adopts three categories, namely, market risk (included to trading book issues), credit risk, and operational risk (Basel Committee of Banking Supervision, 2006). Among them, the market risk is related to the fluctuations of market prices, and its analysis is one of the major topics in quantitative finance, since huge amount of quantitative data is available. Meanwhile, credit risk and operational risk are defined as non-market risk, containing non-quantitative risk such as legal risk. However, a great part of these risks is now available to be analyzed by quantitative approaches. For

example, credit risk was not traded in the market before, but quantitative approaches enable market-traded credit products such as collateralized debt obligation (CDO) or credit default swap (CDS). Using these developments, it is now usual to estimate credit risk by quantitative approaches.

For quantitative understanding of risk in finance, we need to understand the distribution of returns at first. The Gaussian distribution is the primary candidate for this purpose, since it has many good properties such as computational simplicity and the central limit theory. However, many preceding studies reject this idea. The major reason is because the number of extreme losses is more frequent than that which Gaussian distribution predicts. This fact means that the empirical distribution has tail heaviness and skewness toward downside. The α-stable distribution is one of the distributions to describe the tail heaviness and skewness, and was the first non-Gaussian distribution which was adopted as the return distribution in early 1960s (Mandelbrot, 1962). Many distributions have been suggested since then. Among them, a classical tempered stable (CTS) distribution has the advantage that the option pricing based on the distribution is available (Boyarchenko and Levendorskiĭ, 2000).

The second step of risk analysis is to incorporate the dependency between assets. A joint distribution can completely describe dependency. The simplest joint distribution is the multivariate Gaussian distribution, which is used widely by practitioners. However, it is reported that asset returns increase their dependency under a market stress, referred to as tail dependency, and it is pointed out that the Gaussian distribution cannot capture such increasing dependency (For example, see Demarta and McNeil, 2005). After the financial crisis at the end of 2000s, it is proposed to use other joint distributions in order to avoid the flaw of tail dependency. A copula function is a standard method to incorporate the dependency of variables.

The factor analysis is an alternative to direct modeling of dependency by joint distribution. It is the great advantage of the factor analysis that it can reduce the number of dimensions (factors) and that we can easily impose meaning on each factor. The capital asset pricing model (CAPM; Sharpe, 1964; Lintner, 1965; Mossin, 1966), the arbitrage pricing theory (APT; Ross, 1976) and Fama-French three factor model (Fama and French, 1992, 1993) are the important factor models in finance. The mathematical background of these factor models is very classical, but recently novel methodology for these models has been proposed by information science. Independent component analysis comes from such new methodology, and can be a powerful tool to

analyze financial markets.

The third topic of risk measure is to provide measures for risk and ways to manage it, which is the central part of risk analysis. For example, volatility of returns is usually used practically. Value at risk (VaR) is also a common measure widespread in financial industry. VaR was originally developed by RiskMetrics, a division of J.P. Morgan, in order to grasp the whole risk of the portfolio at a glance. For its simplicity in computation and easiness in interpretation, VaR has been adopted as a regulation criterion by the Basel accord (Basel Committee of Banking Supervision, 1996). However, VaR has a flaw that it does not take the extent of loss into account. In order to avoid this problem, an idea of coherent risk measure is proposed (Artzner et al., 1999), and average value at risk (AVaR) is proposed to incorporate the extent of loss.

These three topics are just the basics of risk analysis, and a further research is necessary. For investors, it is necessary to determine the best portfolio based on these analysis. For economists and analysts, it is important to draw the strategy from risk analysis. For researchers and all participants in finance, modeling the actual market based on these models is important. In this dissertation, I have tried to tackle with a part of these problems.

In the rest of this chapter, the basic concepts related to the three topics aforementioned are introduced. In Section 1.1, the marginal distribution and its tail heaviness are discussed, which is the first topic. We also introduce the concept of Lévy process, which is an important class of distribution appearing in financial modeling. In Section 1.2, the measurement of risk is discussed, which is the second topic. We introduce the concept of VaR there. In Section 1.3, we discuss modeling of joint distribution directly by introducing the idea of copula there, which corresponds to the second topic. In Section 1.4, we discuss the representation of random vector, which also corresponds to the second topic. We there introduce the recent idea referred to as independent component analysis.

The following chapters are consisted as follows: In Chapter 2, I introduce the example of construction of actual portfolio, showing the examples of modeling and risk measures. This chapter is the revised version of Tsuchida, Zhou and Rachev (2012), whose first author is the author of this dissertation. Chapter 3 introduces the application of time-series models and copula approaches to European sovereign bonds market. Chapter 4 shows the empirical result that the ICA-based dependency model is as good as the copula-based dependency model, and compute the joint default probability of Eurozone

sovereign default based on the CDS premium. Finally, Chapter 5 is devoted to conclusion.

1.1 Heavy-tailed distribution

1.1.1 Gaussian distribution

The Gaussian distribution is the most important among all distributions in finance, for both theoretical and practical reasons, since 1900 in which its first application in finance was proposed by Louis Bachelier (Bachelier, 1900). There are several advantage in the Gaussian distribution. For financial applications, especially, it is an important advantage that the average of many observations can be approximated by the Gaussian distribution whatever the original distribution is. This is referred to as the central limit theorem (CLT):[1]

Theorem 1 (Classical central limit theorem) *Suppose a scalar distribution with finite variance and independent random variables X_t, $t = 1, \ldots, N$ obeying the distribution. Let $\mu = \mathrm{E}\, X_t$ and $\sigma^2 = \mathrm{E}[(X-\mu)^2] < \infty$. Then the average of $X_1, \ldots, X_N$ converges to the Gaussian distribution[2] with mean μ and variance σ^2/N in distribution, that is,*

$$\sqrt{N}\left\{\frac{1}{N}\sum_{t=1}^{N} X_t - \mu\right\} \to \mathrm{N}(0, \sigma^2) \quad \textit{in dist.} \tag{1.1}$$

Conversely, the central limit theorem can be interpreted as a Gaussian random variable is the infinite sum of independent and identically distributed (i.i.d.) random variables. Of course, we can consider that these infinite variables are another Gaussian variables. Then, we can have an intuitive image that a Gaussian random variable is consisted of infinite numbers of infinitesimal Gaussian random variables. This image is also true in the case that the summation is finite: A Gaussian variable can be represented as the sum of any number of Gaussian variables. This property is summarized as *infinitely divisibility* and *stability*, which are later explained in Section 1.1.4.

[1] Also referred to as the classical CLT or the Lindeberg-Lévy CLT (Rachev et al.,2008)

[2] Hereafter we denote the Gaussian distribution of average μ and variance σ^2 by $\mathrm{N}(\mu, \sigma^2)$. This notation is sometimes extended to the multivariate Gaussian random variable.

1.1.2 Brownian motion

As we see, a Gaussian variable can be viewed as the finite or infinite sum of another Gaussian variables. Let us consider a single Gaussian random variable $r \sim \mathrm{N}(0,1)$. Suppose that we do not know its value at time $t = 0$, and we observe the value at time $t = 1$. Suppose we divide this into two identical Gaussian variables r_1^2 and r_2^2. Assume the value of r_1^2 can be observed at time $t = 1/2$, and the value of r_2^2 can be observed at time $r = 2/2 = 1$. For consistency, we impose the condition that r and $r_1 + r_2$ has the same distribution, that is, $r_1 + r_2 \sim \mathrm{N}(0,1)$. Then we obtain $r_1^2, r_2^2 \sim \mathrm{N}(0, 1/2)$

Next, let us divide r into four Gaussian variables r_1^4, r_2^4, r_3^4, and r_4^4, and they can be observed at times $t = 1/4, 2/4, 3/4, 4/4$, respectively. Then we obtain $r_1^4, r_2^4, r_3^4, r_4^4 \sim \mathrm{N}(0, 1/4)$. This result matches the twofold case by defining $r_1^2 = r_1^4 + r_2^4$ and $r_2^2 = r_3^4 + r_4^4$.

In this way, we can increase the number of divisions. These divisions are not limited to twofold but can be threefold or any other numbers, and they are consistent with each other. From these examples, it is intuitively straightforward that $r_t \sim \mathrm{N}(0, t)$ where r_t is the sum of these divisions up to time t. Also, it is intuitively clear that r_t and $r_s - r_t$ is independent where $s > t > 0$, since each division is independent.

This intuition can be formalized to the definition of Brownian motion.[3]

Definition 2 (Brownian motion) *Suppose a scalar continuous stochastic process B_t defined over the probability space $(\Omega, \mathcal{F}, P)$. If B_t satisfies the following conditions, then B_t is referred to as a Brownian motion:*

1. *$B_0 = 0$ almost surely (a.s.),*
2. *Independence of increments: For any $0 \leq u \leq t < s$, $B_s - B_t$ is independent of B_u,*
3. *Sample continuity: For any $\omega \in \Omega$, the sample path $B_t(\omega)$ is continuous, and*
4. *Normality of increments: For any $0 \leq t < s$, $B_s - B_t \sim \mathrm{N}(0, s - t)$.*

Along with the Gaussian distribution, the Brownian motion is an important building block of mathematical finance. For example, a cumulative log

[3]This definition is equivalent to that based on a Wiener measure, which is also common.

return up to time t is frequently modeled as

$$X_t = \mu t + \sigma B_t, \tag{1.2}$$

whose distribution (with the information up to time $t = 0$) is a Gaussian distribution $\mathrm{N}(\mu t, \sigma^2 t)$. In this equation, the first and the second terms of the right hand side are referred to as the *drift* and the *diffusion* terms, respectively. Suppose a infinitesimal time period Δt starting at time t, then the instantaneous return during the period is $\Delta X_t = \mu \Delta t + \sigma \Delta B_t$, and its distribution is $\mathrm{N}(\mu \Delta t, \sigma^2 \Delta t)$. This is formally written as

$$dX_t = \mu dt + \sigma dB_t. \tag{1.3}$$

This is what is referred to as a stochastic differential equation, and its solution is given by eq.(1.2) almost surely. The theory of stochastic equations is a great field of mathematics, and its application is not only in finance but in a variety of sciences (Itô et al., 2012).

1.1.3 α-stable distribution

Regardless of its popularity, the Gaussian distribution is not a good approximation of the actual market. This was pointed out as early as in 1960s (Mandelbrot, 1962). There are several reasons for that. This is primarily because extreme large losses or profits occur more frequently in reality than in the Gaussian approximation. Secondarily, in addition, losses occur more frequently than profits. In mathematical terms, these two reasons are summarized as: The actual return distribution has *heavy tails* and negative *skewness*.

A heavy tail means that the tail values of the probability distribution function (PDF) is larger than those of the Gaussian distribution. The heaviness of both tails is measured by the value referred to as a *kurtosis* k, which is the standardized fourth moment minus three[4] as for the average, that is,

$$k = \mathrm{E}\left(\frac{X - \mu}{\sigma}\right)^4 - 3, \tag{1.4}$$

where $\mu = \mathrm{E}\, X$ and $\sigma = \sqrt{\mathrm{var} X}$. Since the Gaussian distribution has the value of zero as its kurtosis, a heavy-tailed distribution is usually defined by

[4]It is also common to define the kurtosis as the standardized fourth moment itself.

positive kurtosis. Such a distribution is also said to be *fat-tailed*, *leptokurtosis* or *leptokurtic*.[5]

Meanwhile, the skewness means that the PDF is asymmetric as for the average. Mathematically , its value s is defined as the third moment as for its average:

$$s = \mathrm{E}\left(\frac{X-\mu}{\sigma}\right)^3. \tag{1.5}$$

If the skewness is positive or $s > 0$, then the tail in the positive direction is heavier than that in the negative direction, which is said to be right-skewed. If the skewness is negative or $s < 0$, then the tail in the negative direction is heavier than that in the positive direction, which is said to be left-skewed. The latter is the case in asset returns.

The *α-stable distribution*, also referred to as the *stable Paretian distribution* (named after Vilfredo Pareto) or the *Lévy distribution* (named after Paul Lévy), was used as the model of asset returns by Bunoît Mandelbrot in 1962 (Mandelbrot, 1962, 1963; Fama, 1963) at first. Its definition is as follows:

Definition 3 (α-stable distribution) *Suppose a scalar random variable X and constants $\alpha \in (0, 2]$, $-1 \leq \beta \leq 1$, $\sigma \in \mathcal{R}^+$, and $\mu \in \mathcal{R}$. If the characteristic function $\phi_X(u)$ has the following form,*

$$\phi_X(u) = \mathrm{E}\exp(iuX) \tag{1.6}$$

$$= \begin{cases} \exp\left\{-\sigma^\alpha |u|^\alpha [1 - i\beta sign(u)\tan(\pi\alpha/2)] + i\mu u\right\} & \text{if } \alpha \neq 1, \\ \exp\left\{-\sigma|u|[1 + i\beta(2/\pi)sign(u)\log|u|] + i\mu u\right\} & \text{if } \alpha = 1, \end{cases} \tag{1.7}$$

then X is said to have an α-stable distribution.

It has four parameters, and is commonly denoted as $S_\alpha(\sigma, \beta, \mu)$. Among these parameters, μ (location parameter) and σ (scale parameter) are just for parallel and linear transformations of distribution. Parameter α controls the shape of both tails, referred to as the the *index*, characteristic exponent, or tail exponent, explained later in Section 1.1.4. Parameter β, or skewness parameter, controls the asymmetry of distribution. If $\alpha = 2$, then the distribution recovers the Gaussian distribution with mean μ and variance $2\sigma^2$ (β

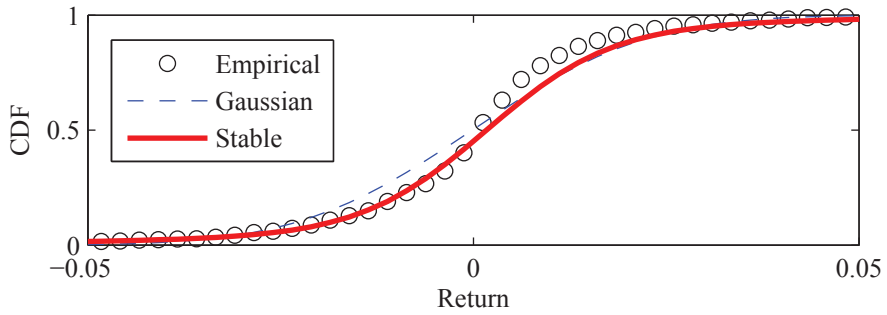
[5]In opposition, a negative-kurtosis distribution is said to be *platykurtosis* or *platykurtic*. A zero-kurtosis distribution is said to be *mesokurtosis* or *mesokurtic*.

disappears in this case). As α decreases from two, the tails of this distribution becomes heavier, meaning that extreme returns occur more frequently. Also, when $\alpha \neq 0$, parameter β appears in the expression. The distribution is symmetric if $\beta = 0$, and gets skewed to the right or left side according to $\beta < 0$ or > 0.

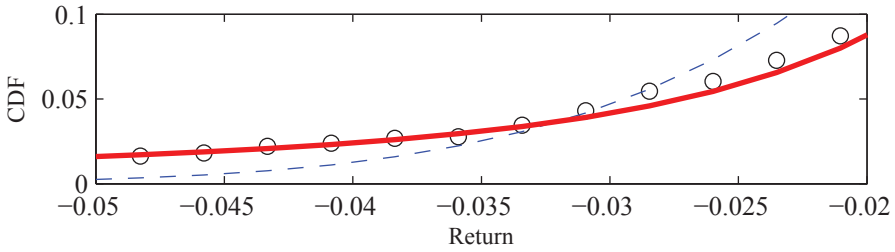
Figure 1.1 demonstrates an example of the α-stable distribution. In this figure, we fit a Gaussian distribution and an α-stable distribution to the daily returns of the Standard and Poor's 500 Index from 2008 to 2012. The first two charts show the cumulative distribution functions (CDF) of these two distributions. In addition, the empirical CDF (defined in eq.(1.23)) is shown as the circles. From these charts, it can be seen that the Gaussian distribution does not match the empirical distribution at all the levels of CDF. The α-stable distribution matches the empirical distribution well at the bottom half, even though it is as bad as the Gaussian at the upper half of the distribution. The third chart of Figure 1.1 shows the probability distribution functions (PDF) of the Gaussian and the α-stable distribution, and it also shows the histogram of the empirical distribution. From this chart, it can be confirmed that the α-stable distribution has heavier tails than the Gaussian distribution, and is closer to the empirical distribution. In this way, the stable distribution can describe heavier tails and skewed distribution, and this is the primary reason for its utilization for returns. It is shown by statistical approaches that the fitting of α-stable distribution to the empirical return data is better than that of Gaussian; for example, see Rachev and Mittnik (2000).

The α-stable distribution has further preferable features. For example, it can be an approximation of an unknown distribution with infinite variances. This is because the average of independent and identically distributed (i.i.d.) random variables with infinite variance tends to an α-stable distribution, which is referred to as the generalized central limit theorem (Shiryaev, 1996). This is similar to the case that the Gaussian distribution can be used as the approximation of unknown distributions because of the central limit theory.

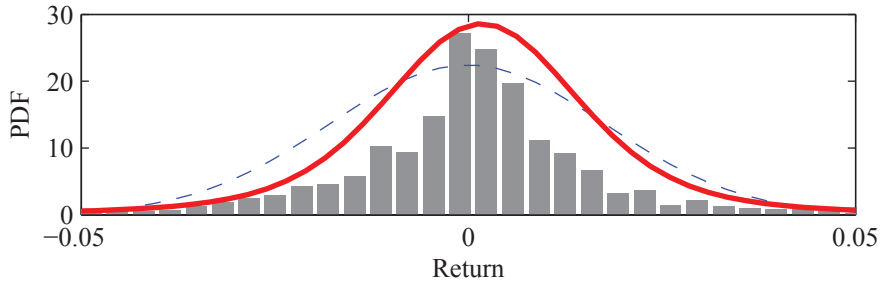
On the other hand, the α-stable distribution has an disadvantage that its variance becomes infinity except the Gaussian case ($\alpha = 2$). In more detail, given that X is a α-stable distribution, then $EX^p < \infty$ if and only if $p \leq \alpha$. Practically, this is undesirable since the variance of empirical distribution is always finite. The classical tempered stable distribution discussed in Section 1.1.5 can avoid this problem.



(a) Cumulative distribution function (CDF)



(b) CDF at the lower tail



(c) Probability distribution function (PDF)

Figure 1.1: Cumulative and probability distribution functions of the α-stable distribution, compared to the empirical and the Gaussian distribution. The data is the daily returns of the Standard and Poor's 500 index from 2008 to 2012. The bars in the PDF chart show the empirical histogram. It can be seen that the lower tail of the α-stable distribution is heavier and closer to the empirical distribution than that of the Gaussian.

1.1.4 Lévy process

The α-stable distribution is not coming from a stand-alone argument, but from the context of the study of stochastic processes. In Section 1.1.2 we gave the conditions of Brownian motion. By relaxing these conditions, we obtain the Lévy process:

Definition 4 (Lévy process) *Suppose a scalar stochastic process X_t defined over the probability space $(\Omega, \mathcal{F}, P)$. If X_t satisfies the following conditions, then X_t is referred to as a Lévy process:*

1. *$X_0 = 0$ a.s.,*
2. *Independence of increments: For any $0 \leq u \leq t < s$, $X_s - X_t$ is independent of X_u,*
3. *Càd-làg sample path:[6] For any $\omega \in \Omega$, the sample path $X_t(\omega)$ is right side continuous and has a left limit, and*
4. *Stationarity of increments:[7] For any $s \geq 0$, $X_{t+s} - X_t$ is independent of t.*

Comparing Definition 2 in Section 1.1.2 and Definition 4 above, we can find that the continuity of sample paths in Brownian motion is replaced with the càd-làg condition of sample paths, and the normality of increments is replaced with the stationarity. Since all continuous sample paths are càd-làg and the normality of Brownian motion increments satisfies stationarity, the Brownian motion is also the Lévy process.

The Lévy process includes a wide variety of stochastic processes other than the Brownian motion. For example, a Poisson process is also a Lévy process. A Poisson process X_t ($t > 0$) with parameter $\lambda > 0$ is defined as the number of jumps which obey the exponential distribution with parameter λ. Note that jumps do not occur simultaneously. Then, $X_{t+s} - X_t$ ($s > 0$) represents the number of jumps between time t and $t+s$. $X_{t+s} - X_t$ ($s > 0$) is dependent on the interval s, but is independent of the data prior to time t, due to the memoryless property of the exponential distribution. Therefore the independence and stationarity of increments are satisfied, so X_t satisfies the

[6] "Càd-làg" originally stands for a French word "continu à droite, limites à gauche." It is now common to use this terminology.

[7] Without stationarity, the process is what is referred to as the *additive process*.

conditions of the Lévy process. Likewise, it can be shown that the compound Poisson process is the Lévy process.

Infinitely divisible If X_t is a Lévy process, then we can divide X_t for each t into the following finite sum, using the independence and the stationarity of increments:

$$X_t = X_{t/n} + (X_{2t/n} - X_{t/n}) + \cdots + (X_t - X_{t(n-1)/n}). \tag{1.8}$$

where n is any integer. This shows that X_t is *infinitely divisible* for each t.

The definition of being infinite divisible is that a random variable X can be represented as the sum of i.i.d random variables Y_n, that is, $X = \Sigma_{1 \leq i \leq n} Y_n$ in distribution, for any given n. Let ϕ denote the characteristic function of X, then this definition is equivalent to that, for any given n, ϕ has the n-th root psi_n, that is, $\phi = {\psi_n}^n$.

Conversely, suppose a infinitely divisible random variable X is given. Then, it can be shown that there exist a Lévy process X_t where $X_1 = X$ in distribution, and such a distribution is unique in distribution (Sato, 1999). Therefore, we can have a one-to-one mapping between the set of Lévy processes and the set of infinitely divisible distributions. For example, a standard Gaussian distribution $X \sim \mathrm{N}(0, 1)$ corresponds to a standard Brownian motion $X_t \sim \mathrm{N}(0, t)$. A Poisson distribution with parameter λ, denoted as $X \sim \mathrm{Po}(\lambda)$, corresponds to a Poisson process with parameter λ, which holds $X_t \sim \mathrm{Po}(\lambda t)$.

Lévy -Khintchine representation An infinitely divisible variable X has the following standard form of characteristic function:

$$\phi_X(z) = \exp\left[i\mu z - \frac{1}{2}\sigma^2 z^2 + \int_{\mathcal{R}} (e^{izx} - 1 - izx \cdot \mathbf{1}_{|x| \leq 1}) \nu(dx) \right], \tag{1.9}$$

where σ and μ are scalar parameters and $\nu(dx)$ is a measure satisfying $\int_{\mathcal{R}} \min(|x|^2, 1)\nu(dx) < \infty$. The set of parameters (σ, ν, μ), referred to as a Lévy triplet or a generating triplet, has one-to-one correspondence to the set of infinitely divisible distributions, and their relation is given by eq.(1.9). For example, a standard Gaussian distribution corresponds to the triplet $(1, \nu \equiv 0, 0)$, and a Poisson distribution with parameter λ corresponds to the triplet $(0, \lambda\delta(x-1), 0)$. This representation is called as the Lévy -Khintchine representation, which was found by Lévy (1934) and Khintchine (1937).

Stability of distribution In Section 1.1.3, we define the α-stable distribution. This name is derived since the distribution is stable. Being stable is defined as follows:

Definition 5 (Stability of distribution) *Suppose a random variable X and independent random variables $X_1, \ldots, X_n$ are generated from the same distribution. The distribution is said to have the property of stability or to be stable if the following relation holds for all n:*

$$a_n(X_1 + \cdots + X_n) + b_n = X, \qquad \textit{in distribution,} \tag{1.10}$$

where a_n and b_n are constants. If b_n=0, then the distribution is said to be strictly stable.

It is shown that a stable distribution is infinitely divisible (Feller, 1971). Therefore, there exists the corresponding Lévy process, which is referred to as a *stable process*. Likewise, the Lévy process corresponding to a strictly stable distribution is referred to as the *strictly stable process*.

Let X_t $(t > 0)$ denote a stable process. It can be shown that X_t and $t^{1/\alpha}X_1 + b_t$ has the same distribution where α is a constant independent of t and b_t is a constant dependent on t. The constant α is referred to as the *index* of the stable process, consistent with that of the α-stable distribution. Especially, if X_t is a strictly stable process, $b_t = 0$ and X_t is $1/\alpha$-*self-similar*.[8] From these properties, it can be shown that a stable distribution has a characteristic function shown in eq.(1.7). For details, see Sato (1999) or Rachev (2000).

1.1.5 Classical tempered stable distribution

As we noted, infinite variance is a problem of the α-stable distribution. It is proposed by Mantegna and Stanley (1994) to avoid infinite variance by truncating a given distribution. However, this prescription breaks the infinite divisibility. Instead, a infinitely divisible distribution referred to as the *classical tempered stable distribution* (CTS) is proposed (Koponen 1995; Rosiński, 2007).[9] This distribution is obtained from the stable distribution

[8] A stochastic process X_t is referred to as a H-self-similar process if and only if, for all $a > 0$, there exists $b > 0$ such that X_{at} and bX_t has the same distribution.

[9] In the literature different names are used for this distribution: Lévy process by Koponen, KoBoL by Boyarchenko and Levendorskiĭ, and CGMY by Carr et al. The name CTS is adapted from Kim et al. (2008)

by changing Lévy measure (referred to as *tempering*), and it has both skewness and heavy-tails. Actually, Carr et al. (2002) and Kim et al. (2008) shows that the CTS distribution has a good fit to asset prices, better than that of the stable distribution. Interestingly, Boyarchenko and Levendorskiĭ (2000) option pricing formula for this distribution.

A CTS distribution is defined as a scalar random variable X whose characteristic function $\phi_X(u)$ is represented as follows:

$$\begin{aligned}\phi_X(u) &= \phi_X(u;\alpha,C_1,C_2,\lambda_+,\lambda_-,m)\\ &= \exp[ium + C_1\Gamma(-\alpha)\{(\lambda_+ - iu)^\alpha - \lambda_+^\alpha\}\\ &\qquad + C_2\Gamma(-\alpha)\{(\lambda_- + iu)^\alpha - \lambda_-^\alpha\}].\end{aligned} \tag{1.11}$$

It has five scalar parameters, and they satisfy $C_1, C_2, \lambda_+, \lambda_- > 0$ and $\alpha \in (0,2)$. Parameter α is again referred to as an index, controlling the heaviness of tails. Especially, when parameters satisfy the following conditions, the random variable X has zero mean and unit variance:

$$\begin{aligned}C = C_1 = C_2 &= [\Gamma(2-\alpha)(\lambda_+^{\alpha-2} + \lambda_-^{\alpha-2})]^{-1},\\ m &= -\Gamma(1-\alpha)(C_1\lambda_+^{\alpha-1} - C_2\lambda_-^{\alpha-1}).\end{aligned} \tag{1.12}$$

In this case, X is referred to as a standard CTS, which has three parameters α, λ_+, and λ_-.

Figure 1.2 shows the CDF of CTS fit to the empirical returns. The data is same with Figure 1.1, and the CDF of Gaussian and α-stable distributions are shown again for comparison. As shown, the CTS has better fitting than the other two distributions.

1.1.6 Normal tempered stable distribution

A *normal tempered stable distribution* (NTS) distribution is an infinitely-divisible distribution introduced and applied in Barndorff-Nielsen and Levendorskiĭ (2001) and Barndorff-Nielsen and Shephard (2001). It is defined as a scalar random variable X whose characteristic function $\phi_X(u)$ is as follows:

$$\begin{aligned}\phi_X(u) &= \phi_X(u;\alpha,\theta,\mu,\beta,\gamma)\\ &= \exp\left[i(\mu-\beta)u - \frac{2\theta^{1-\alpha/2}}{\alpha}\Big(\big(\theta - i\beta u + \gamma^2u^2/2\big)^{\alpha/2} - \theta^{\alpha/2}\Big)\right],\end{aligned} \tag{1.13}$$

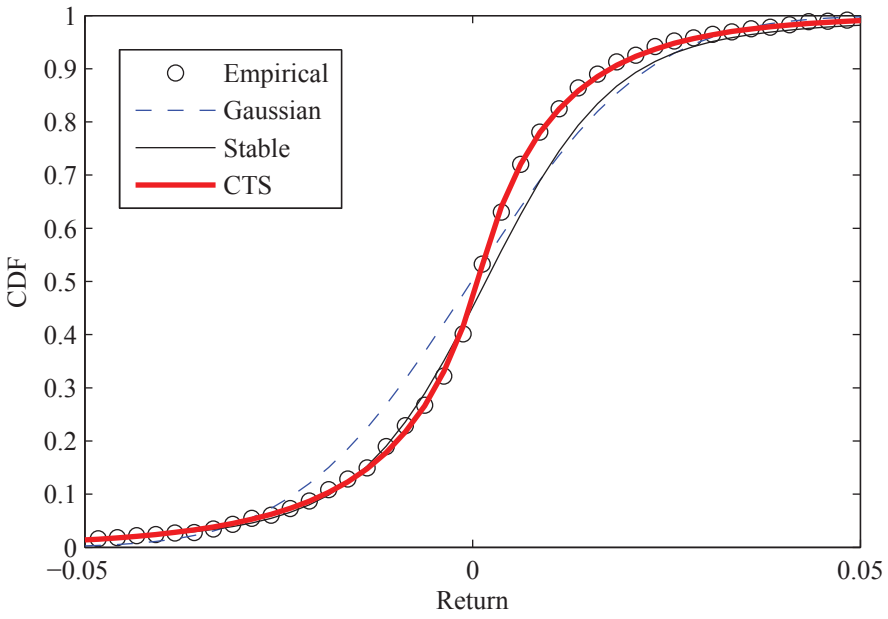


Figure 1.2: Cumulative distribution function of the CTS distribution, compared to the α-stable, the Gaussian, and the empirical distribution. The data is the daily S&P 500 index return from 2008 to 2012. It can be seen that the CTS provides the best fit to the empirical distribution.

where $0 < \alpha < 2$, $\theta > 0$, and $\gamma > 0$. Parameter α controls the heaviness of tails. The average is given by $E[X] = \mu$ and the variance by $\mathrm{var}(X) = \gamma^2 + \beta^2(2-\alpha)/2\theta$. By setting these values to zero and unity respectively, we obtain the standard NTS. The standard NTS has three parameters: α, θ, and β.

The NTS distribution can be also obtained from the Brownian motion by replacing time t with another random variable. Suppose a Brownian motion X_t with drift β and diffusion γ starting at $X_0 = \mu - \beta$, then $X_t = \mu + \beta(t-1) + \gamma B_t$. B_t can be rewritten by a standard Gaussian variable ε as $B_t = \sqrt{t}\varepsilon$. Replacing time t with a Lévy process T_t, we obtain

$$X_t = \mu + \beta(T_t - 1) + \gamma\sqrt{T_t}\varepsilon, \tag{1.14}$$

which is referred to as the subordinated expression. Since time t is always increasing, the Lévy process T_t should be increasing as t increases. Such a Lévy process is referred to as an *increasing Lévy process* or a *subordinator*. Especially, a subordinator T_t is referred to as a *CTS subordinator* if its characteristic function is given as

$$\phi_{T_t}(u) = \exp\left[-\frac{2\theta^{1-\alpha/2}}{\alpha}((\theta - iu)^{\alpha/2} - \theta^{\alpha/2})\right]. \tag{1.15}$$

In this case, and the distribution X_t is then a NTS. One of the advantages of NTS is that the subordinator expression eq.(1.14) can be easily extended to a multivariate case, as explained in Section 1.3.8.

1.2 Risk measurement

1.2.1 Volatility

Volatility plays a central role in finance. It is defined as the standard deviation of asset return, and considered as the extent of uncertainty in future. Volatility is classified as a *dispersion measure*. Given a random variable X representing return, the dispersion measure $D(X)$ satisfies (Rachev, 2008),

Positive shift $D(X + C) \leq D(X)$ for all X and constant $C \geq 0$,

Positive homogeneity $D(0) = 0$ and $D(\lambda X) = \lambda D(X)$ for all X and constant $\lambda > 0$, also referred to as *convexity*, and

Positivity $D(X) \geq 0$ for all X, $D(X) > 0$ for all nonconstant X.

Generally, the central absolute moment of order k, that is,

$$\mathrm{E}[|X - \mathrm{E}\, X|^k] \tag{1.16}$$

satisfies the requirements for dispersion measure. The standard deviation corresponds to $k = 2$, and the case of $k = 1$ is referred to as the mean absolute deviation (MAD). The other example of dispersion measure is the *colog* measure, which is defined as $\mathrm{E}\, X \log(X) - \mathrm{E}\, X\, \mathrm{E} \log X$.

It should be noted that volatility is a measure for uncertainty, not for risk. If the average return of an investment is high enough to compensate its high volatility, then the investment should not be described as risky, even though it can be described as uncertain. More importantly, the notion of risk should incorporate the asymmetry of return. For example, buying a call option has large uncertainty since its maximum profit is infinite, but its downside uncertainty is limited. On the other hand, selling the same call option has large downside uncertainty since its maximum loss is infinite. In this case, apparently, selling is much riskier than buying.

The semi-standard deviation can be used to measure this kind of asymmetric return. This is also a dispersion measure, and defined as

$$\sigma_X^{\pm} = [\mathrm{E}(X - \mathrm{E}\, X)_{\pm}^2]^{1/2}, \tag{1.17}$$

$$(a)_+ = \max(0, a), \quad (a)_- = \min(0, a). \tag{1.18}$$

1.2.2 Value at risk

The *value at risk* (VaR) is a risk measure which incorporates asymmetry of returns.

Definition 6 (Value at risk (VaR)) *Suppose a random variable X representing the return of an investment. A value at risk (VaR) at confidence level α (usually set to 95% or 99%) is defined as the minimum (inferior) loss*[10] *whose probability of realization is less than $1 - \alpha$, that is,*

$$\mathrm{VaR}_\alpha(X) = \inf\{-x \in \mathcal{R} | P(X \leq x) \leq 1 - \alpha\}. \tag{1.19}$$

[10]It is also common to define the VaR in the opposite sign, that is, the maximum return whose probability of realization is less that $1 - \alpha$.

If X is a continuous random variable, then the VaR matches the sign change of the $(1-\alpha)$-quantile of X:

$$\mathrm{VaR}_\alpha(X) = -F_X^{-1}(1-\alpha), \tag{1.20}$$

where $F_X(x)$ is the cumulative distribution function (CDF) of X, as shown in Figure 1.3. Note that VaR sometimes takes a negative value, that is, it breaks the property of positivity, which appears in dispersion measures, shown in subsection 1.2.1. Theoretically, the major properties of VaR are threefold:

Monotonicity $\rho(X) \leq \rho(Y)$ if $X \leq Y$ for almost sure, where Y represents the another return,

Invariance $\rho(X + C) = \rho(X) - C$ for all X and constant $C \in \mathcal{R}$,

Positive homogeneity $\rho(0) = 0$ and $\rho(\lambda X) = \lambda\rho(X)$ for all X and constant $\lambda > 0$.

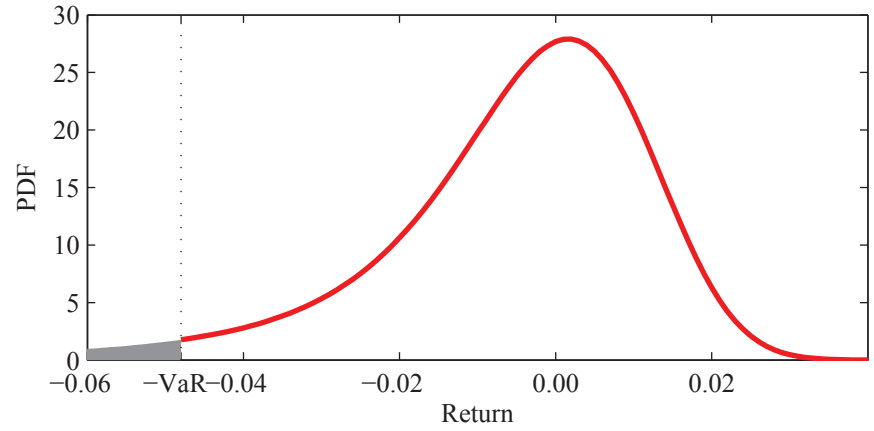
However, the VaR does not satisfy the following property:

Subadditivity $\rho(X + Y) \leq \rho(X) + \rho(Y)$ for all X and Y.

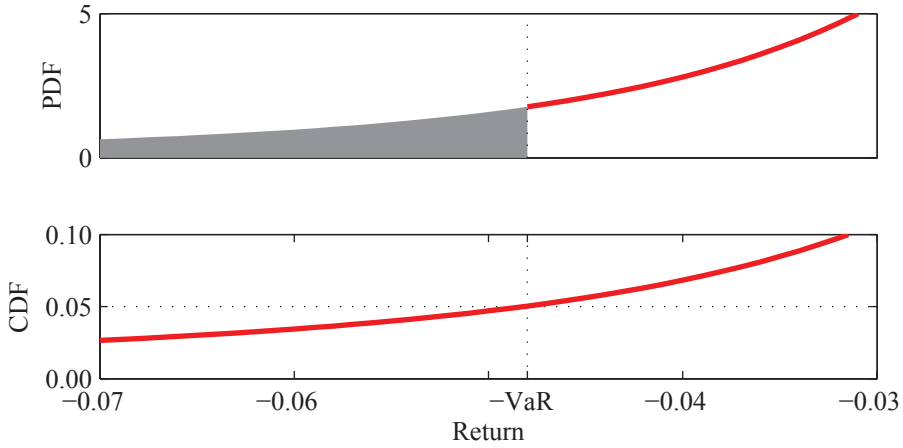
The combination of these four properties is the requirements for what is referred to as a *coherent risk measure*, introduced by Artzner (1999). It should be mentioned that subadditivity of the VaR holds for the finite-variance distribution in the *elliptic distribution family* (Embrechts et al, 2001), which includes major practical distributions such as the Gaussian and the Student-t distributions. The detailed discussion of a coherent risk measure is in Section 1.2.6.

1.2.3 Practical computation of VaR

From its definition, the computation of VaR needs the knowledge about the distribution of X. Statistically, X can be modeled both parametrically and non-parametrically. Given a parametric model, the distribution can be computed both analytically and numerically. Consequently, the computation of VaR can be practically classified into three approaches: the parametric-analytical approach, the parametric-numerical approach, and the non-parametric approach. These approaches correspond to the covariance approach, the Monte-Carlo simulation, and the historical method, respectively.



(a) Entire PDF



(b) Zoom-up around the VaR

Figure 1.3: The VaR at 95% confidence level. The vertical dotted line represents −VaR. The area of the shaded region is 0.05.

Covariance approach Due to its simplicity, the most important approach among these three is the parametric and analytical approach. This approach is referred to as the covariance approach, named after its utilization of covariance, or the approach of RiskMetrics, named after the group of its first publication.

In this method, the joint distribution of assets is usually assumed to have the multivariate Gaussian distribution, which is completely specified by its mean vector and covariance matrix. Since the linear transformation of the Gaussian distribution is also the Gaussian distribution, the return of the portfolio consisted of the assets also has the univariate Gaussian distribution. Let X denote the portfolio return, μ denote its mean, and σ^2 denote its variance, then the VaR of X is given as

$$\mathrm{VaR}_\alpha(X) = -\mu + k_\alpha \sigma, \tag{1.21}$$

$$k_\alpha = -\Phi^{-1}(1-\alpha) = \Phi^{-1}(\alpha) > 0 \tag{1.22}$$

where $\Phi(x)$ is the CDF of the standard Gaussian distribution. Note that $k_{0.95} = 1.64$ for 95% VaR and $k_{0.99} = 2.33$ for 99% VaR, respectively.

What is necessary in this computation is only the mean and variance of X, which leads to the simplicity of this approach. Meanwhile, a major flaw of this approach is that the Gaussian distribution is not appropriate as return distribution, both in marginal and joint. To fix this problem, alternative distributions such as the multivariate Student-t distribution are investigated (Lamantia et al., 2006a; Lamantia et al., 2006b). However, these alternative distributions generally requires computational resources, and their analytical expressions are difficult to be obtained in general cases.

Historical method The historical method does not impose any distributional assumptions, but draws the distribution from the observations in past. The *empirical distribution*, defined as follows, is usually used as the distribution in future:

$$F_{X,\mathrm{emp}}(x) = \frac{1}{N} \sum_{t=1}^{N} \mathbf{1}_{x_t < x}, \tag{1.23}$$

where x_t is the t-th observed value of X in past. The historical method inherits both the advantages and the disadvantages from non-parametric models. For example, the historical method is more robust than a parametric

model in the sense that it can capture the irregular shape of distribution which parametric distributions may miss out, which is an advantage. The other advantage is its computational lightness as long as we use a simple formulation such as eq.(1.23). On the other hand, the followings are the major obstacles of the historical method:

- It assumes that past trend can be applied to future. In reality, it is possible to observe unprecedented events in future, and these events may be extreme.
- It assumes that return distributions are i.i.d. between observations. This assumption is usually rejected since return series has autocorrelation and volatility clustering. This problem can be avoided by transforming the original return series into a stationary series via time series model, which is discussed in section 1.2.4.
- It sometimes faces data insufficiency. For daily observations, for example, one-year comprises only about 250 observations, which is statistically not enough to decide the 1% quantile (99% VaR). Prolonging the observation period may solve this problem, but in turn faces the problem of chronological shift of return distribution itself.

Monte Carlo simulation The Monte Carlo simulation is a numerical solution for parametric approaches. It consists of the following steps:

1. *Select a statistical model*: For instance, suppose a multivariate Gaussian distribution for asset returns.
2. *Estimate parameters of the model*: In the above instance, determine the mean vector and covariance matrix of the multivariate Gaussian distribution based on the observations in past.
3. *Generate return scenarios from the estimated model*: In the above instance, draw a sample value from the obtained multivariate Gaussian model, and compute the sample return of the portfolio. Repeat this drawing and computing many times, for example, 10,000 times. Each time corresponds to one scenario, and afterward 10,000 simulated returns can be obtained. Let $r^{(s)}$, $s = 1, \ldots, 10,000$ denote the return of scenario s.

4. *Construct a portfolio distribution*: Construct the distribution of portfolio return from the 10,000 scenarios using eq.(1.23), and then compute the VaR.

The Monte Carlo simulation has several advantages over the other two approaches. Compared to the covariance approach, it has more flexibility in selection of statistical model. Compared to the historical method, it requires less data for proper estimation of distribution, and it allows unprecedented extreme events as long as the model indicates. Especially, as for the derivative securities, the Monte Carlo simulation is practically the only method for risk evaluation, since they have strong non-linearity of risk and unprecedented risk events play an important role in their pricing. Meanwhile, the Monte Carlo simulation has the disadvantages of its heavy computational burden and possible estimation errors. The latter problem can be mitigated by increasing the number of scenarios.

1.2.4 Time series model

Return distribution is not generally i.i.d., but has the following stylized facts, reportedly (for example, McNeil et al, 2005, pp.117):

- Series of returns show weak autocorrelation,
- Series of squared returns shows strong autocorrelation,
- Volatility varies over time, which is referred to as *volatility clustering*,
- Return series is heavy-tailed, and
- Extreme returns tends to continue during a specific period.

These facts indicate the necessity to incorporate a *time series model* which describes chronological shift of return distribution. As for serial correlation, the *autoregressive moving average* (ARMA) model is a standard model in order to incorporate serial correlation. Let X_t $(t = 1, 2, \dots)$ represents a time series data, and then the ARMA model of order (p, q), denoted as $ARMA(p, q)$, is defined as follows:

$$X_t = c + \sum_{i=1}^{p} \phi_i X_{t-i} + \sum_{j=1}^{q} \psi_j u_{t-j} + u_t. \tag{1.24}$$

where c, ϕ_i $(i = 1, \ldots, p)$, and ψ_j $(j = 1, \ldots, q)$ are constants. Variable u_t is a random variable referred to as a *white noise*, a *innovation term* or a *residual term*. The distribution of u_t is usually assumed to be i.i.d. with zero mean.

Volatility clustering is also referred to as the *heteroscedasticity*. The simplest model to describe this feature is the *generalized autoregressive conditional heteroscedasticity* (GARCH) model, which is defined as follows. Suppose the random variable u_t can be factorized as

$$u_t = \sigma_t \varepsilon_t. \tag{1.25}$$

In this expression, σ_t is assumed to be a non-random variable (filtered by the information up to time t), and ε_t is assumed to be an i.i.d. random variable. Usually, ε_t is assumed to have the standardized Gaussian or the Student-t distribution, and referred to as the residual term (so the meaning of "residual terms" depends on the context; it may be u_t or ε_t). The GARCH model of order (p, q) is defined as follows:

$$\sigma_t^2 = K + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2 + \sum_{j=1}^{q} \alpha_j u_{t-j}^2, \tag{1.26}$$

where K, α_j and β_i are constant parameters. Note that σ_t is a constant after the information up to time $(t-1)$ is revealed.[11]

The combination of these two models is referred to as the ARMA-GARCH model. If the model is appropriate, the original source of randomness X_t is now converted to a stationary and standardized series of ε_t, which makes it easy to analyze the return series. Given the order (p, q), the parameters are decided by the ML estimation. The order is generally determined by a correlogram or information criteria (IC).

Once the ARMA-GARCH model is determined, it is possible to predict the return distribution in future according to the model. This prediction is generally done by a Monte Carlo simulation using eqs.(1.24), (1.25), and (1.26). If ε_t is assumed to be a simple distribution such as the Gaussian, X_t can be computed analytically.

It is possible that the residuals ε_t based on the actual data do not obey the Gaussian distribution. One solution for this problem is to use non-Gaussian

[11] Meanwhile, volatility clustering is also modeled by the stochastic volatility (SV) model, in which the process $\{\sigma_t\}_t$ is an adapted stochastic process, that is, the value of σ_t is unknown until time t.

distribution of ε_t for prediction purpose, and the other is to use the empirical distribution of ε_t in past. The former solution is usually accompanied with the Monte Carlo simulation due to analytical intractability of non-Gaussian distribution. The latter solution is proposed by Barone-Adesi et al. (1999) and referred to as the filtered historical simulation (FHS).

Time series model has various topics. To see the details of these problems, see major text books such as Hamilton (1994). Among them, we introduce two major problems related to risk analysis. The first is the goodness of fitting (g-o-f) of the ARMA-GARCH to the actual data. Usually, the g-o-f is measured as for the in-sample data primarily, and then measured as for the out-of-sample data by methods like cross validation. In risk analysis, however, modeling is usually for the prediction of future returns, and therefore it is important to see the g-o-f based on the out-of-sample data in the direction of future. Therefore, it is common to adopt a method which is referred to as backtesting, which is explained in Section 1.2.5.

The second problem is how to introduce a time series model of multiple variables. We have at least two approaches for this problem. One is to introduce a multivariate time series model such as the vector autoregressive model (VAR), and the other is to use a joint distribution of residual series ε_t between different time series.

1.2.5 Backtesting

We need to check if the VaR reflects risk in future. For this purpose, we can use the past data. This process is referred to as *backtesting*. It is consisted of the following steps:

1. Decide a time period for backtesting. Let T denote the number of days included to the period.

2. For each day t on the period, compute the VaR using the data up to the day t. Let $\mathrm{VaR}_{\alpha,t+1}$ denote the VaR. Note that the subscript is $(t+1)$ since this VaR is the estimate of the loss on the following day, $(t+1)$.

3. Observe the return on the day $t+1$, compute the loss L_{t+1}, and compare the loss L_{t+1} with$\mathrm{VaR}_{\alpha,t+1}$. If $L_{t+1} > \mathrm{VaR}_{\alpha,t+1}$, then this is the case of an *exceedance*.

4. Count the number of exceedances, and compare the number with its expected number, $(1-\alpha)T$.

If there are too many exceedances, then it indicates that the VaR is not correctly computed, and consequently it signifies that the return model is not appropriate. More precisely, the number of exceedances can be modeled by the following way: Let I_t denote if an exceedance occurs ($I_t = 1$) or not ($I_t = 0$) at the observation at time t. By the definition of VaR, it is expected that $P(I_t = 1) = 1-\alpha$ and $P(I_t = 0) = \alpha$. This means I_t obeys the Bernoulli distribution with parameter $(1-\alpha)$. Let X denote the number of exceedances during T observations, then $X = \sum_{1 \le t \le T} I_t$. If series $\{I_t\}_t$ is i.i.d., then the distribution of X is the binomial distribution $B(T, 1-\alpha)$, that is,

$$P(X = x) = \frac{T!}{x!(T-x)!}(1-\alpha)^x \alpha^{T-x}. \tag{1.27}$$

The example of $P(X \le x)$ is shown in Table 1.1, which is excerpted from the official publication by the Basel Committee on Banking Supervision (1996). This is the case of $T = 250$ observations and 99% VaR, $(1-\alpha) = 0.01$. For example, we can see 95.88% at the last column in the Table, and the corresponding number at the second column is 5. This means that the probability of five or less exceedances is 95.88%, or equivalently, the probability of 6 or more exceedances is 4.12%. Therefore, if we have six or more exceedances with 5% significance level (note that this significance level can be different from the confidence level of the VaR), then we can reject the null hypothesis that the VaR captures 99% loss, that is, we can reject that the VaR is valid.

For regulation purposes, the Committee defines the case of five to nine exceptions (exceedances) as the "yellow zone" and that of ten or more as the "red zone." Financial institution with these numbers of exceedances is punished by the additional requirement for its capital, and the extent of punishment increases as the number of exceptions increases.

The flaw of the previous model is that it neglects the effect of dependency between exceedances, that is, the series $\{I_t\}_t$ is assumed to be identical and independent. If a return model were precisely specified, that is, if a true return model were available, then the time series effect of return series would be completely included into the return model itself. In this case, the exceedance series $\{I_t\}_t$ would have no autocorrelation and be completely modeled by the series of Bernoulli trials. In reality, however, such a complete specification

Table 1.1: The cumulative probability $P(X \leq x)$ (fourth column) according to x, the number of exceptions (exceedances; second column) against the 99% VaR. The first column is the "zones" indicating the level of exceptions, and the third column is the level of additional capital requirement. This table is excerpted from the official publication by the the Basel Committee on Banking Supervision (1996).

Zone	Number of exceptions	Increase in scaling factor	Cumulative probability
Green Zone	0	0.00	8.11 %
	1	0.00	28.58 %
	2	0.00	54.32 %
	3	0.00	75.81 %
	4	0.00	89.22 %
Yellow Zone	5	0.40	95.88 %
	6	0.50	98.63 %
	7	0.65	99.60 %
	8	0.75	99.89 %
	9	0.85	99.97 %
Red Zone	10 or more	1.00	99.99 %

of return is never reached, neither a completely precise computation of VaR is. Consequently, consecutive exceedances of VaR sometimes occur. Recognition of this dependency is practically important because consecutive large losses lead to a precipitous decline in asset value. Even if the same number of large losses is given, it is practically more difficult to recover from the large consecutive losses than from the large but interspersed losses.

The test proposed by Christoffersen (1998) is one of the statistical tests to evaluate the serial correlation of VaR, consisted of three steps. The first step is the likelihood ratio (LR) version of eq.(1.27), which is initially proposed by Kupiec (1995). Let n_1 denote the number of exceedances and n_0 that of non-exceedances. In other words, $n_i = n(I_t = i)$, $i = 0, 1$ where $n(A)$ is the number of set A. Suppose $\pi = P(I_t = 1)$, then the likelihood function is $L_1(\pi) = \pi^{n_1}(1-\pi)^{n_0}$. Especially, if $\pi = 1-\alpha$, then $L_1(1-\alpha) = (1-\alpha)^{n_1}\alpha^{n_0}$. According to the LR test, the LR between $L_1(1-\alpha)$ and $L_1(\pi)$ asymptotically obeys the χ^2-distribution of parameter 1,

$$\mathrm{LR}_u = -2\log\frac{L_1(1-\alpha)}{\sup_\pi L_1(\pi)} = -2\log\frac{L_1(1-\alpha)}{L_1(\hat{\pi})} \sim \chi^2(1), \tag{1.28}$$

where $\hat{\pi} = n_1/n$, $n = n_0 + n_1$. This is an approximation version of the test based on eq.(1.27). The second step is the test of independence. Let us assume that the series $\{I_t\}_t$ is a two-state, first-order Markov chain whose transition matrix is given as

$$\Pi = \begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix} = \begin{pmatrix} 1-p_{01} & p_{01} \\ 1-p_{11} & p_{11} \end{pmatrix}, \tag{1.29}$$

where $p_{ij} = P(I_t = j | I_{t-1} = i)$. Note that $p_{00} + p_{01} = p_{10} + p_{11} = 1$. Then the approximate likelihood is

$$L_2(\Pi) = (1-p_{01})^{n_{00}} p_{01}^{n_{01}} (1-p_{11})^{n_{10}} p_{11}^{n_{11}}, \tag{1.30}$$

where $n_{ij} = n\{(I_{t-1}, I_t) = (i, j)\}$. In this notation, independence of I_t from I_{t-1} means $p_{01} = p_{11} = \pi$, and the likelihood function is

$$L_2(\pi) = (1-\pi)^{n_{10}+n_{00}} \pi^{n_{01}+n_{11}}. \tag{1.31}$$

Since the difference of degree of freedom is 1, the limiting distribution of LR between $L_2(\pi)$ and $L_2(\Pi)$ obeys the $\chi^2(1)$-distribution,

$$\mathrm{LR}_i = -2 \log \frac{L_2(\pi)}{\sup_\Pi L_2(\Pi)} = -2 \log \frac{L_2(\pi)}{L_2(\hat{\Pi})} \sim \chi^2(1), \tag{1.32}$$

where

$$\hat{\Pi} = \begin{pmatrix} n_{00}/(n_{00}+n_{01}) & n_{01}/(n_{01}+n_{01}) \\ n_{10}/(n_{10}+n_{11}) & n_{11}/(n_{10}+n_{11}) \end{pmatrix}. \tag{1.33}$$

This test can be used to test if the series $\{I_t\}_t$ is independent or not. Finally, the third step is the joint test of $\pi = 1-\alpha$ and $p_{01} = p_{11} = \pi$. The LR is

$$\mathrm{LR}_c = -2 \log \frac{L_1(1-\alpha)}{\sup_\Pi L_2(\Pi)}. \tag{1.34}$$

The following approximation is shown by Christoffersen:

$$\mathrm{LR}_c \approx \mathrm{LR}_u + \mathrm{LR}_i \sim \chi^2(2). \tag{1.35}$$

Equation (1.35) can be used in order to test if the estimated VaR correctly reflects tail risk.

There are other tests to evaluate VaR exceedances. The conditional autoregressive value at risk (CAViaR), proposed by Engle ang Manganelli (2004), is one of them. Berkowitz et al. (2009) comprehensively review various tests for this purpose using the actual trading desk data.

1.2.6 Average value at risk

The VaR has a flaw that it does not take the extent of loss into account. This can be problematic in actual situations. For example, suppose two ventures whose expected VaRs are the same but maximum losses are different. In this case, the one with the lower maximum loss will be preferred to the other by risk averse investors, given other conditions are the same. In other words, the theory of VaR does not explain the extent of loss conditional to the case that the large loss comes into reality.

This line of discussion is theoretically detailed as follows (Rachev, 2008). As we discussed in section 1.2.2, the VaR does not satisfy the subadditivity in general. This fact leads to the following situation: Suppose two ventures whose returns are represented by random variables X and Y. According to the discussion of the second order stochastic dominance (SSD), the following inequalities can stand together:

$$\mathrm{VaR}_\alpha(X) \leq \mathrm{VaR}_\alpha(Y), \text{ and} \tag{1.36}$$

$$\mathrm{E}[-X|-X < \mathrm{VaR}_\alpha(X)] \geq \mathrm{E}[-Y|-Y < \mathrm{VaR}_\alpha(X)]. \tag{1.37}$$

The first inequality shows that X is less risky than Y according to the measure of VaR. On the other hand, the second inequality shows that the extent of loss in extreme cases is larger in X than in Y, which intuitively means that X is riskier than Y. Even though it is a rare case, this is a contradiction in the definition of risk measure. If the subadditivity holds, then the SSD assures that these inequalities are exclusive each other. This shows the importance of *coherent risk measure* (Artzner et al., 1999), which includes the subadditivity property in addition to the other three properties in section 1.2.2.

The *average value at risk* (AVaR) is the risk measure to solve such an incoherency problem of VaR. Its definition is as follows:

Definition 7 (Average value at risk (AVaR))

$$\mathrm{AVaR}_{1-\alpha}(X) = \frac{1}{1-\alpha}\int_{1-\alpha}^{1} \mathrm{VaR}_{1-p}(X)dp. \tag{1.38}$$

The AVaR is the average of VaR conditional to the case that the loss over VaR is observed. The synonyms of AVaR are the conditional value at risk (CVaR) or the expected shortfall (ES). Also, if X is continuous distribution,

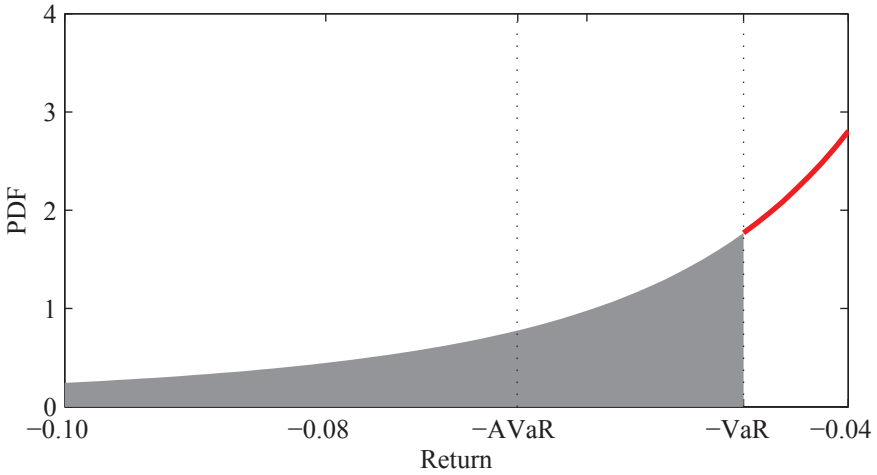


Figure 1.4: The AVaR at 95% confidence level. The two vertical lines correspond to the AVaR and the VaR, respectively. The AVaR is the average loss conditional that the loss is larger than the VaR.

it is shown that

$$\mathrm{AVaR}_{1-\alpha}(X) = -\,\mathrm{E}[X|X < -\mathrm{VaR}_{1-\alpha}(X)], \tag{1.39}$$

which is referred to as the expected tail loss (ETL), while this equation does not hold if X has a discrete distribution or mass points. Figure 1.4 demonstrates the AVaR. Since it is the average of VaR conditional to the case that the loss over VaR is observed, the value of AVaR is larger than that of the VaR.

The practical computation of AVaR is similar to the case of VaR. The Monte Carlo simulation is available to compute the AVaR. The covariance approach is available if we assume the return distribution X is the Gaussian distribution. Given $\mathrm{E}\,X = \mu$ and $\mathrm{var}X = \sigma^2$, the AVaR is

$$\mathrm{AVaR}_{\alpha}(X) = -\mu + C_\alpha \sigma, \tag{1.40}$$

$$C_\alpha = \frac{\exp(-(\Phi^{-1}(\alpha))^2/2)}{(1-\alpha)\sqrt{(2\pi)}} > 0. \tag{1.41}$$

This has the same form with the case VaR in eqs.(1.21) and (1.22) except the magnitude of the constant depending on α. Especially, C_α takes $C_{0.95} = 2.06$ and $C_{0.99} = 2.67$ for the 95% and 99% AVaR, respectively.

As for the historical method, the data insufficiency becomes a severe problem compared to the case of VaR, since the AVaR uses the shape of the distribution below VaR. For example, given 250 daily observations, which corresponds to one year, the expected number of exceedances is only two or three, which is apparently insufficient. Therefore, it is usual to use modifications of the historical method, such as weighting of extreme scenarios or bootstrapping.

Similar to the mean-variance portfolio introduced by Markowitz (1952), it is possible to obtain the optimal portfolio based on the AVaR. The optimization problem is shown to be convex and easy to be computed (Bertsimas et al., 2004).

The backtesting of AVaR is as follows. For example, let Y_t denote the loss over AVaR, that is, $Y_t = -X_t - \mathrm{AVaR}_{\alpha,t}$. If we assume the distribution is continuous and eq.(1.39), eq.(1.39) is equivalent to

$$\mathrm{E}[Y_t | L_t > \mathrm{VaR}_{\alpha,t}] = 0. \tag{1.42}$$

Let $\hat{Y_t}$ denote the observed value of Y_t. This value can be used to estimate the difference between the realized loss and the AVaR. If $\hat{Y_t} > 0$, then it means that the realized loss is larger than the expected by the model, implying the model does not capture the tail risk properly.

However, it is usually difficult to conduct a statistical test to validate the AVaR for several reasons. First, if we use time series model, the distribution of Y_t depends on time t. In addition, even if we know the distribution of Y_t, it is possible that data is not enough to conduct a statistical test. In the case of 250 observations, the probability of no observation of 99% VaR exceedance is $0.99^{250} = 8.1\%$. It is quite possible that we do not observe any exceedances, and we cannot consider the extent of large loss in this case. The detailed discussion about these problems is introduced by Rachev et al. (2008).

1.2.7 Backtesting revisit

The purpose of backtesting is to test if a model can correctly predict the out-of-sample data. For example, the 99% VaR backtesting tests a return model by seeing if the lower 1% quantile of return is correctly predicted by the return model. The backtesting of AVaR should be, if available, a test of the return model using the PDF below the lower 1% quantile.

If we drop the constraint of "below 1% quantile," several statistical tests are available. Let $F_t(x_t)$ denote the CDF of X_t, and suppose a random variable $Z_t = F_t(X_t)$. Then Z_t is expected to be a uniform random variable over $(0, 1)$ being independent of different t, that is,

$$Z_t \sim U(0, 1), \quad \text{i.i.d.} \tag{1.43}$$

In realistic situation, the true distribution $F_t(x_t)$ is unknown, and replaced with the model distribution $\hat{F}_t(x_t)$. Then we obtain the following:

$$\hat{Z}_t = \hat{F}_t(\hat{X}_t). \tag{1.44}$$

It is not assured that this series satisfies the uniformity condition eq.(1.43). Therefore, the uniformity of $\hat{Z}_t$ over (0,1) can be tested to see if $\hat{F}_t(x_t)$ is a proper estimation of $F_t(x_t)$.

This approach is referred to as the multiple α model by Campbell (2005), since it is equivalent to utilize all the levels of VaR confidence. The advantage of the above approach is that it is more sensitive to the return model misspecification. In addition, by introducing all the data available, the test can be improved for small numbers of exceedances (Berkowitz, 2001).

On the other hand there is a criticism that this approach implicitly broadens the scope of risk, since the loss less than the VaR has conceptually nothing to do with the risk. For example, the current regulatory guidelines dictate that the risk should be measured by 99% VaR, and the test for the entire return distribution is not required. In addition to such a conceptual criticism, this approach has a practical problem that we cannot use the extreme value theory (EVT) model, since the EVT focuses on just large losses and have nothing to do with usual returns. For detailed discussions, see Haas (2001) and Campbell (2005).

1.3 Copula model

The copula function can be used to describe the joint distribution. Its definition is as follows:

Definition 8 (Copula function) *Suppose a d-dimensional continuous random vector* $\mathbf{X} = (X_1, X_2, \ldots, X_d)^{\mathrm{T}}$ *whose joint cumulative distribution is* $F(x_1, \ldots, x_d)$ *and marginal cumulative distribution is* $F_i(x_i)$, $i = 1, \ldots, d$. *Let* $u_i = F_i(x_i)$, *then, according to the Sklar's theorem of representation*

(Sklar, 1959; Nelsen, 1999, Theorem 2.10.9), there exists a unique function $C : (0,1)^d \to (0,1)$ *satisfying*

$$C = C(u_1, \ldots, u_d) = F(x_1, \ldots, x_d). \tag{1.45}$$

Then the function $C = C(u_1, \ldots, u_n)$ *is referred to as a copula function.*

In other words, a copula function is the cumulative distribution function of the random vector $U = (U_1, \ldots, U_d)^{\mathrm{T}}$, where $U_i = F_i(X_i)$ is the uniformly-distributed random variable over (0,1). The joint distribution function $F = F(x_1, \ldots, x_d)$ can be then separated into the marginal distributions $\{F_i\}_{i=1,\ldots,d}$ and the dependency structure C. C is considered to represent dependency structure of d variables; for example, rank correlation such as the Kendall's tau[12] can be computed only from the copula function, being independent of the marginal distribution (Nelsen, 1999, pp. 127; Embrechts et al., 2001). This separation is the advantage of copula models.

The major examples of copulas are: The Gaussian copula, the Student-t copula, and the Archimedean copula family. Given any continuous joint distributions, it is possible to construct a copula function from the distribution. For example, the Student-t copula is constructed from the multivariate Student-t distribution. It should be noted that a marginal distribution can be different from the joint distribution on which the copula is based. For example, it is possible to combine the univariate Gaussian distribution as the marginal distributions and the Student-t copula as the copula distribution.

1.3.1 Gaussian copula

The Gaussian copula is one of the easiest copula, which is also referred to as the normal copula. Its copula function satisfies

$$\frac{\partial^d}{\partial u_1 \cdots \partial u_d} C(u_1, \ldots, u_d) = \frac{1}{|\Sigma|^{1/2}} \exp\left(-\frac{\omega^{\mathrm{T}}(\Sigma^{-1} - I)\omega}{2}\right), \tag{1.46}$$

where

$$\omega = (\omega_1, \ldots, \omega_d)^{\mathrm{T}} \tag{1.47}$$

$$= (\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d))^{\mathrm{T}}. \tag{1.48}$$

[12] Also referred to as the Kendall's rank correlation.

and $\Phi^{-1}(u)$ is the inverse CDF of the standard Gaussian distribution. Parameter Σ, a d-by-d matrix, is referred to as a correlation matrix, since it is positive definite and its diagonal elements are unity. The d-th-order derivative at the left hand side of eq.(1.46) is referred to as the probability distribution function (PDF) of the copula, denoted as $c(u_1, \ldots, u_d)$.

The Gaussian copula is constructed from a multivariate Gaussian distribution. Under the notation of Definition 8, suppose $\mathbf{X} = (X_1, \ldots, X_d) \sim \mathrm{N}(\mathbf{0}, \Sigma)$ and $U_i = \Phi(X_i)$. Then the copula function is given by eq.(1.45), and its derivative by $x_1, \ldots, x_d$ yields

$$c(u_1, \ldots, u_d) \prod_{i=1}^{d} f_{X_i}(x_i) = f_X(x_1, \ldots, x_d), \tag{1.49}$$

where $f_{X_i}(x_i)$ is the marginal PDF of X_i and $f_{\mathbf{X}}$ is the joint PDF of $\mathbf{X}$. In this case, since $\mathbf{X} \sim \mathrm{N}(\mathbf{0}, \Sigma)$,

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}\mathbf{x}^{\mathrm{T}}\Sigma^{-1}\mathbf{x}\right), \tag{1.50}$$

where $\mathbf{x} = (x_1, \ldots, x_d)^{\mathrm{T}}$. In addition, since the diagonal elements of Σ are unity, $X_i \sim \mathrm{N}(0, 1)$, and $f_{X_i}(x_i) = \exp(-x_i^2/2)/\sqrt{2\pi}$. Combining these equations, we obtain eq.(1.46).

The simple PDF formula given by eq.(1.46) enables easy parameter estimation in high dimensions. Owing to this computational tractability, the Gaussian copula is widely used in finance; for example, Li (2000) used it in his famous paper evaluating the price of CDO.

1.3.2 Tail dependency

A criticism against the Gaussian copula is that it cannot capture what is referred to as the tail dependency. The following is an explanation by the other literatures (Coles et al., 1999; Demarta and McNeil, 2005). Let us define the coefficient of upper tail dependence of X_1 and X_2 as

$$\lambda_U = \lim_{u \to 1^-} P(X_2 > F_2^{-1}(u) | X_1 > F_1^{-1}(u)). \tag{1.51}$$

where $F_1(x_1)$ and $F_2(x_2)$ are the CDFs of X_1 and X_2. Likewise, the coefficient of lower tail dependence is defined as

$$\lambda_L = \lim_{u \to 0^+} P(X_2 < F_2^{-1}(u) | X_1 < F_1^{-1}(u)). \tag{1.52}$$

Let us introduce the two-dimensional copula $C(u_1, u_2)$, then these coefficients are shown to be dependent only on the copula function, as

$$\lambda_U = \lim_{u \to 1^-} \frac{1 - 2u + C(u, u)}{1 - u}, \tag{1.53}$$

$$\lambda_L = \lim_{u \to 0^+} \frac{C(u, u)}{u}. \tag{1.54}$$

These coefficients satisfy $\lambda_U, \lambda_L \in [0, 1]$. If $\lambda_L = 0$ (or $\lambda_U = 0$), then X_1 and X_2 are independent of each other at the limit of $u \to 0^+$ ($u \to 1^-$). On the other hand, if $\lambda_L = 1$ (or $\lambda_U = 1$), then X_1 and X_2 are completely dependent on each other at the limit of $u \to 0^+$ ($u \to 1^-$).

Suppose C is a Gaussian copula with parameter ρ (which is the off-diagonal element of Σ). Except the case of complete dependency ($\rho = 1$), the coefficients of upper and lower tail dependence[13] are

$$\lambda_U = \lambda_L = 2 \lim_{x \to \infty} \left(1 - \Phi(x \sqrt{(1 - \rho)/(1 + \rho)} \right) = 0. \tag{1.55}$$

This implicates that components of a multivariate Gaussian distribution are virtually independent each other in tails, regardless of their correlation matrix. In a practical context, this means that the Gaussian copula cannot describe the tail dependent events such as the simultaneous defaults of bonds or market crash of stocks. Since such events are possible in reality, it is better to find another copula which has positive values of these tail dependence coefficients.

1.3.3 Student-t copula

The Student-t copula is the copula based on the multivariate Student-t distribution. Demarta and McNeil (2005) provide its comprehensive explanation. The copula is characterized by a correlation matrix (positive definite and unit diagonal elements) Σ and degree of freedom $\nu > 0$. Its PDF is represented as

$$c(u_1, \ldots, u_d; \Sigma, \nu) = \frac{\Gamma(\frac{\nu+d}{2})[\Gamma(\frac{\nu}{2})]^d (1 + \frac{1}{\nu} \omega^{\mathrm{T}} \Sigma^{-1} \omega)^{-(\nu+d)/2}}{\sqrt{|\Sigma|} \Gamma(\frac{\nu}{2})[\Gamma(\frac{\nu+1}{2})]^d \prod_{i=1}^{d} (1 + \frac{\omega_i^2}{\nu})^{-(\nu+1)/2}}, \tag{1.56}$$

[13] The upper and lower coefficients of tail dependence coincides with each other if the copula distribution belongs to the elliptically symmetric distribution.

where $\Gamma(x)$ is the Gamma function,

$$\omega = (\omega_1, \ldots, \omega_d)^{\mathrm{T}} = (t_\nu^{-1}(u_1), \ldots, t_\nu^{-1}(u_d))^{\mathrm{T}}, \tag{1.57}$$

and $t_\nu^{-1}(x)$ is the inverse CDF of the univariate Student-t distribution with degree of freedom ν.

The tail dependence coefficient of a two-dimensional Student-t copula is given as

$$\lambda_L = \lambda_U = 2t_{\nu+1}\left(-\sqrt{\frac{(1-\rho)(\nu+1)}{1+\rho}}\right), \tag{1.58}$$

where ρ is the off-diagonal element of Σ. These coefficients take finite values, which is preferable feature for financial assets with tail dependency.

The parameters of Student-t copula is usually estimated by the maximum likelihood (ML) method. There are two parameters to be estimated: Σ and ν. Given the value of ν, the optimal $\Sigma = \Sigma_\nu$ based on the ML satisfies

$$\Sigma_\nu = \frac{\nu+n}{N}\frac{\sum_{1\le t\le N}\omega^t{\omega^t}^{\mathrm{T}}}{\nu + \sum_{1\le t\le N}{\omega^t}^{\mathrm{T}}\Sigma_\nu^{-1}\omega^t}, \tag{1.59}$$

where ω^t is t-th sample of ω given ν.

Suppose two random variables are connected with a Student-t copula whose correlation matrix is ρ (off-diagonal element). Then, the Kendall's tau of these two variables is

$$\tau = \frac{2}{\pi}\arcsin\rho. \tag{1.60}$$

This equation can be used to estimate ρ.

1.3.4 Estimation of copula parameters

Suppose a parametric copula such as the Gaussian copula is given. In order to fit the model parameters to the samples $(u_1^t, \ldots, u_d^t)$, $t = 1, \ldots, N$, it is the most standard to use the maximizing likelihood (ML) method.

However, the samples $(u_1^t, \ldots, u_d^t)$ cannot be observed directly, usually. Instead, they are computed from the observations of $\mathbf{X}^t$, denoted as $\hat{\mathbf{X}}^t = (\hat{x_1}^t, \ldots, \hat{x_d}^t)$, and the marginal distribution $F_i(x_i)$,$i = 1, \ldots, d$, as $u_i^t = F_i^{-1}(\hat{x_i}^t)$. If $F_i(x_i)$ is the true marginal distribution, then u_i^t is the true value

of the sample. However, we usually do not know $F_i(x_i)$, and assume a model distribution $\hat{F}_i(x_i)$ instead. Therefore, what we obtain is not $u_i^t = F_i(\hat{x_i}^t)$ but $\hat{u_i}^t = \hat{F}_i(\hat{x_i}^t)$, which depends both on the observed data and the model assumption. As a result, we cannot separate the problems of parameter estimation into the marginal and the copula, but need to consider them simultaneously, abandoning the advantage of copula.

The inference functions for marginals (IFM), introduced by Joe (1997), partly recovers the advantage of copula model. In the IFM, marginal parameters are first estimated by the ML estimation, and then copula parameters are estimated by the ML, the marginal likelihoods being plugged into the copula likelihood. This approach is close to the true ML, but has an disadvantage that it depends on the selection of parametric marginal distribution.

An alternative approach is called as the *pseudo-likelihood* method by Demarta and McNeil (2005) or the *canonical maximum likelihood* (CML) method by Romano (2002). It is extensively investigated by Genest et al. (1995). When we estimate parameters, the pseudo-likelihood method approximates the marginal cumulative distribution (CDF) as

$$\hat{F}_i(x_i) = \frac{1}{N+1} \sum_{t=1}^{N} \mathbf{1}_{\hat{x_i^t} < x_i}. \tag{1.61}$$

This is similar to the empirical distribution defined in eq.(1.23), but the normalized factor is replaced with $1/(N+1)$, in order to assure $\hat{F}_j(x_j) < 1$.

In the case of the Gaussian copula, the ML estimation can be easily computed. However, in the case of Student-t copula, the computational burden of parameter estimation rapidly increases as the dimension d increases. Generally, in high dimensions, the ML usually has the problem of computational intractability.

An alternative parameter estimation for the case of Student-t copula is to use eq.(1.60) instead. We can compute the Kendall's tau $\hat{\tau}$ directly from the observation.[14] Then, it is possible to estimate the value of ρ from $\hat{\tau}$ using eq.(1.60). If the dimension is larger than 2, it is possible to estimate each

[14]In two dimensions, it is computed as follows: First, draw a scatter plot of two variables. Suppose there are N points. For each point, we count up the number of points located at the up-and-right side of the point, and count down the number of points located at the down-and-left side of the point (ignoring points located at the up-and-left and down-and-right side). Sum up these numbers and divide by $N(N-1)/2$, then we obtain the Kendall's tau.

element of Σ from the pairwise taus, and then we can obtain the estimation of the correlation matrix $\hat{\Sigma}$. After that, we can compute ν using the ML, which is easy to solve since it is one-dimensional optimization. This idea is similar to the momentum method of parameter estimation in the sense that we just match some descriptive statistics.

1.3.5 Empirical copula

Another important copula is an *empirical copula*, usually defined by

$$C_{emp}(u_1, \ldots, u_d) = \frac{1}{N} \sum_{t=1}^{N} 1_{u_1 < u_1^t, \ldots, u_n < u_d^t}, \tag{1.62}$$

where $(u_1^t, \ldots, u_d^t)$ $(t = 1, \ldots, N)$ stands for the t-th sample. Given a model copula and the empirical copula, we can compute the "distance" between them, in order to measure how well the model copula fit the empirical one. An example is the copula version of the Cramér-von Mises statistics,

$$S = \int_{[0,1]^N} N\{C_{\text{emp}}(u_1, \ldots, u_d) - C(u_1, \ldots, u_d)\}^2 dC_{\text{emp}}(u_1, \ldots, u_d) \tag{1.63}$$

$$= \sum_{t=1}^{N} \{C_{\text{emp}}(\hat{u_1}^t, \ldots, \hat{u_d}^t) - C(\hat{u_1}^t, \ldots, \hat{u_d}^t)\}^2, \tag{1.64}$$

where C is the model copula. It is also possible to compute a p-value of the null hypothesis that the model copula matches the empirical copula. However, the computation is generally numerically burdensome. For details of discussion about copula fitness, see Genest et al. (2009) and Berg (2009).

1.3.6 Archimedean copula

The Archimedean copula is a major copula class. We do not refer this kind of copula in this dissertation, but quickly review this copula in the case of two dimensions, since it is frequently used in practice,

Let us consider a scalar function $\phi = \phi_\theta(u)$ satisfying $\phi(1) = 0$ and parameterized by $\theta \in \Theta$. Assume that $\phi(u)$ is strictly decreasing and convex, that is, $\phi'(u) < 0$, and $\phi''(u) \geq 0$.[15] Then, we define a copula $C(u_1, u_2)$ as

$$C(u_1, u_2) = \phi^{[-1]}(\phi(u_1) + \phi(u_2)), \tag{1.65}$$

[15]If the number of dimensions is d, these conditions are replaced with, over the support

where $\phi^{[-1]}(v)$ is the pseudo-inverse function of $\phi(u)$, which matches $\phi^{-1}(v)$ over $v \in [0, \phi(1)]$ but takes 0 otherwise. The function $\phi(u)$ is referred to as the *Archimedean generator*. The follwing are the examples of the Archimedean generators:

- Clayton copula, $\phi_\theta(u) = \frac{1}{\theta}(u^{-\theta} - 1)$,
- Ali-Milhail-Haq copula, $\phi_\theta(u) = \log((1 - \theta(1 - u))/u)$,
- Gumbel copula, $\phi_\theta(u) = (-\log u)^\theta$,
- Frank copula, $\phi_\theta(u) = -log\left(\frac{\exp(-\theta u)-1}{\exp(-\theta)-1}\right)$, and
- Joe copula,$\phi_\theta(u) = 1 - (1 - \exp(-u))^{1/\theta}$.

The advantage of the Archimedean copulas is that it has only one scalar parameter, which leads computational simplicity. Also, Some of these models have tail dependency; for example, the Clayton copula has lower tail dependency $\lambda_L = 2^{-1/\theta}$. The disadvantage is that it is too simple to denote a complicated dependency. Especially, an Archemedean copula treats each dimension equally, ignoring the different dependency among pairs of dimensions.

1.3.7 Skewed Student-t distribution

The multivariate Student-t distribution is classified to the class of *normal mixture distribution*, since a d-dimensional Student-t random variable $\mathbf{X}$ with correlation Σ and degree of freedom ν has the following representation:

$$\mathbf{X} = \sqrt{W}\mathbf{Z}, \tag{1.66}$$

where W and $\mathbf{Z}$ are independent random variables such that $\nu/W \sim \chi^2_\nu$ and $\mathbf{Z} \sim N(\mathbf{0}, \Sigma)$. This expression simplify the generation of random samples obeying the multivariate Student-t distribution and its copula. A larger class of multivariate normal mixture is the mean-variance mixtures, obtained by

$$\mathbf{X} = \mu + \gamma g(W) + \sqrt{W}\mathbf{Z}, \tag{1.67}$$

of $\phi(u)$, 1) $\phi(u)$ is $(d-2)$-times differentiable, 2) $(-1)^k\phi^{(k-2)}(u) \geq 0$ for $k = 0, 1, \ldots, d-2$, and 3) $(-1)^{d-2}\phi^{(d-2)}(u) \geq 0$ is non-increasing and convex. These conditions are referred to as *d-monotone*, whose definition is given by McNeil and Nešlehova (2009).

for some $g : [0, \infty) \to [0, \infty)$ and d-dimensional parameters μ and γ (Demarta and McNeil, 2005). A preferable feature of this class is that it has skewness due to the term $\gamma g(W)$, while it is not elliptically symmetric for the same reason. It is also preferable that generation of random samples is easy based on eq.(1.67).

Suppose the case that $g(W) = W$ and W has an inverse gamma distribution $IG(\nu/2, \nu/2)$. Then, a random variable

$$\mathbf{X} = \mu + \gamma W + \mathbf{Z}\sqrt{W}, \tag{1.68}$$

is said to have a *multivariate skewed-t distribution*, which is denoted as $t_d(\nu, \mu, \Sigma, \gamma)$ (Demarta and McNeil, 2005). Its PDF function has the following form:

$$f_{t_d}(\mathbf{x}; \nu, \mu, \Sigma, \gamma) = c \times \frac{K_{(\nu+d)/2}(\sqrt{(\nu + a)b}) \exp((\mathbf{x} - \mu)^{\mathrm{T}}\Sigma^{-1}\gamma)}{(\sqrt{(\nu + a)b})^{-(\nu+d)/2}(1 + a/\nu)^{(\nu+d)/d}}, \tag{1.69}$$

$$a = a(\mathbf{x}; \mu, \Sigma) = (\mathbf{x} - \mu)^{\mathrm{T}}\Sigma^{-1}(\mathbf{x} - \mu), \tag{1.70}$$

$$b = \gamma^{\mathrm{T}}\Sigma^{-1}\gamma, \tag{1.71}$$

$$c = \frac{2^{[2-(\nu+d)]/2}}{\Gamma(\nu/2)(\pi\nu)^{d/2}|\Sigma|^{1/2}}, \tag{1.72}$$

where K_λ denotes the modified Bessel function of the third kind. According to this formula, it is possible to estimate parameters by maximizing the likelihood. The distribution recovers the Student-t distribution when $\mu = 0$ and $\gamma \to 0$. Its expected value and covariance are

$$\mathrm{E}\,\mathbf{X} = \mu + \frac{\nu}{\nu - 2}\gamma, \tag{1.73}$$

$$\mathrm{cov}\mathbf{X} = \frac{\nu}{\nu - 2}\Sigma + \frac{2\nu^2}{(\nu - 2)^2(\nu - 4)}\gamma\gamma^{\mathrm{T}}. \tag{1.74}$$

The copula dependency of the multivariate Student-t distribution is referred to as the *skewed t copula*. Especially, a copula of $t_d(\nu, 0, \Sigma, \gamma)$ distribution is denoted as $C^t_{\nu,\Sigma,\gamma}$. In order to convert the skewed-t distribution to the skewed-t copula, the following formula about the marginal distribution of skewed-t distribution is useful:

$$X_i \sim t_d(\nu, 0, 1, \gamma_i), \quad i = 1, \ldots, n. \tag{1.75}$$

1.3.8 Multivariate normal tempered stable distribution

In the subordinated expression of the NTS distribution eq.(1.14), suppose that μ, β and γ are d-dimensional vectors and ε is a d-dimensional multivariate standardized Gaussian distribution $\varepsilon \sim N(\mathbf{0}, \Sigma)$, where Σ is a correlation matrix. Then, we obtain a d-dimensional random variable $\mathbf{X}$:

$$\mathbf{X} = \mu + \beta(T_t - 1) + \gamma\sqrt{T_t}\varepsilon, \tag{1.76}$$

The distribution of $\mathbf{X}$ is referred to as the *multivariate NTS* (MNTS) distribution.

The MNTS model has five scalar or vector parameters: α, θ, μ, β, and γ. The number of parameter variables is then $3d + 2$. This seems too much to estimate. To estimate these parameters, we first decide the values of parameters as for the CTS subordinator in eq.(1.15), that is, α and θ. This can be done by considering the average of d variables or by referring market index. In addition, if we assume that $\mathrm{E}\, X_i = 0$ and $\mathrm{var} X_i = 1$, we can set $\mu = \mathbf{0}$ and $\beta = \sqrt{1 - \gamma^2(2-\alpha)/(2\theta)}$. Then the parameter estimation becomes feasible for higher dimensions. Kim et al. (2012) apply the MNTS to the Dow Jones Industrial Average and obtain a good fitting.

1.4 Independent component analysis

1.4.1 Factor analysis in finance

The factor model is generally defined as

$$\mathbf{X} = A\mathbf{Y} + \varepsilon. \tag{1.77}$$

Variable $\mathbf{X}$ is a d-dimensional random vector representing observable variables, and $\mathbf{Y}$ is a q-dimensional random vector representing unobservable or latent variables. Each component of $\mathbf{Y}$ is referred to as a factor. A is a d-by-q constant matrix which is referred to as factor loadings. ε is a d-dimensional random vector representing noise or factors not included in $\mathbf{Y}$. For simplicity, the averages of $\mathbf{X}$, $\mathbf{Y}$ and ε are assumed to be zero. Also, we usually assume $\mathbf{Y}$ is standardized and non-correlated as $\mathrm{cov}\mathbf{Y} = I_q$, since the covariance structure can be moved to A or ε. As a result, we obtain the covariance of $\mathbf{X}$ as

$$\mathrm{cov}\mathbf{X} = C_{\mathbf{X}} = AA^{\mathrm{T}} + Q, \tag{1.78}$$

where $Q = \mathrm{cov}\varepsilon$. Since $C_{\mathbf{X}}$ can be obtained by the sample covariance matrix, it is an important problem to determine A and Q given $C_{\mathbf{X}}$. There are a lot of works as for this problem, and this field of research is referred to as *factor analysis*. For details, see Harman (1967), for example.

In the field of finance, the most famous factor model will be the capital asset pricing model (CAPM), which was proposed in 1960s independently by several studies (Sharpe, 1964; Lintner, 1965; Mossin, 1966). According to the model, the return of asset r follows

$$r = r_f + \beta(r_M - r_f) + \varepsilon, \tag{1.79}$$

where r_f is the return of risk-free asset, r_M is the return of the market, and ε is the return specific to the asset. It is usually assumed as $\mathrm{E}\,\varepsilon = 0$ since it can be derived from the CAPM theory.[16] The parenthesis term $(r_M - r_f)$ is usually referred to as a *risk premium*. β is a constant specific to the asset, and is considered to represent the sensitivity to the risk premium. Given the volatility of the market as σ_M, the volatility of the asset σ satisfies

$$\sigma^2 = \beta^2\sigma_M^2 + \mathrm{var}\varepsilon. \tag{1.80}$$

Equations (1.79) and (1.80) are the $q = 1$ case of eqs. (1.77) and (1.78).

Later in 1976, Ross proposed the arbitrage pricing theory (APT), in which the asset return r is assumed to be

$$r = a + \beta_1 f_1 + \beta_2 f_2 + \cdots + \beta_q f_q + \varepsilon. \tag{1.81}$$

In this equation, random variables f_j, $j = 1, \ldots, q$, are considered to represent fundamental factors such as the risk premium or economic indicators, and the coefficients β_j are the factor loadings for these variables. Ignoring the constant term a, this equation corresponds to the general case of eq.(1.77), and it is more flexible compared to the CAPM. On the other hand, the APT does not give how to decide factors f_j. So its implication and implementation are not as clear as in the case of the CAPM.

In early 1990s, Fama and French proposed their three factor model as for stock returns (Fama and French, 1992, 1993). As the name indicates, it is a three-factor model as

$$r = \beta_M r_M + \beta_{\mathrm{HML}}\mathrm{HML} + \beta_{\mathrm{SMB}}\mathrm{SMB} + \varepsilon. \tag{1.82}$$

[16] If it is not assumed to be zero, then the term *alpha* is commonly used to show the non-zero average of return specific to the asset.

Here, HML is the difference of the book-to-price (BP) ratios between stocks of large BP and those of small BP in the market. The word HML stands for "High Minus Low." Meanwhile, SMB stands for "Small Minus Big," which is defined as the difference of the returns between stocks of small capitalization and those of large capitalization. These two plus the market return (r_M) are the three factors of the Fama-French model, and the coefficients β and the noise term ε are specific to the asset.

The primary advantage of these factor models is the reduction of the number of parameters. For example, suppose the return of d assets. If we model them by the Gaussian distribution, the number of parameters is $d(d+1)/2$ for the covariance matrix and d for the average return. On the other hand, if we model them by the CAPM and the Gaussian distribution of r_M and ε, then the number of parameters is only $2d+3$ (d for β, d for σ, two for the average and variance of r_M, and one for r_f).

The secondary advantage is clearness in interpretation. For example, in the case of CAPM, the factor is simply interpreted as the market trend, and the model can be interpreted as the separation of returns into the common and the specific part. On the other hand, the APT model does not seem clear, since the model itself does not give what are the factors. Even though, it is clear in the sense that it assumes some hidden factors which have effect on returns.

1.4.2 Principal component analysis

These factor models assume the existence of latent variables affecting on returns. On the other hand, the principal component analysis (PCA) does not assume the existence of such variables, and find the most influential component just from the data.

Suppose a d-dimensional random vector $\mathbf{X} = (X_1, \ldots, X_d)$. For simplicity, assume $\mathrm{E}\, X = 0$. The first principal component (PC) P_1 is defined as $P_1 = \mathbf{w_1} \cdot \mathbf{X}$ where

$$\mathbf{w}_1 = \mathrm{argmax}_{||\mathbf{w}||=1} \,\mathrm{var}(\mathbf{w} \cdot \mathbf{X}), \tag{1.83}$$

that is, the first PC is the linear transformation of X whose variance is maximized. The second PC P_2 is defined

$$\mathbf{w}_2 = \mathrm{argmax}_{||\mathbf{w}||=1, \mathbf{w}_1 \perp \mathbf{w}} \,\mathrm{var}(\mathbf{w} \cdot \mathbf{X}). \tag{1.84}$$

This is similar to (1.83), but the direction of linear transformation is restricted to the direction orthogonal to $\mathbf{w}_1$. Likewise, the third and the following PCs are defined as the linear transformation of X whose variance is maximized under the condition that the transformation is orthogonal to the preceding PCs.

It is known that the result of these maximizing steps is given by the eigenvalues of the covariance matrix. If the covariance matrix $C_{\mathbf{X}}$ is diagonalizable, we can find a d-by-d unitary matrix U and characteristic values λ_i, $i = 1, \ldots, d$, satisfying

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_1 > 0, \tag{1.85}$$

$$D = \mathrm{diag}(\lambda_1, \ldots, \lambda_d), \tag{1.86}$$

$$C_{\mathbf{X}} = U^{\mathrm{T}} D U. \tag{1.87}$$

On the other hand, $C_{\mathbf{X}} = \mathrm{cov}\mathbf{X}\mathbf{X}^{\mathrm{T}}$. Combining this and eq.(1.87), we obtain $\mathrm{cov}\mathbf{X}\mathbf{X}^{\mathrm{T}} = U^{\mathrm{T}} D U$. Multiplying U and U^{T} from the left and the right of this equation, respectively, we obtain

$$\mathrm{cov}\mathbf{P}\mathbf{P}^{\mathrm{T}} = D, \tag{1.88}$$

where $\mathbf{P} = (P_1, \ldots, P_d)^{\mathrm{T}} = U\mathbf{X}$. P_i, $i = 1, \ldots, d$, are the principal components obtained by eqs.(1.83) and (1.84), and the i-th row of U corresponds to $\mathbf{w}_i$. Conversely, we can write as

$$\mathbf{X} = U^{\mathrm{T}} P = P_1 \mathbf{w}_1 + \cdots + P_d \mathbf{w}_d. \tag{1.89}$$

The characteristic value λ_i represents the variance explained by P_i, that is, $\mathrm{var}P_i = \lambda_i$. The explanatory power of the i-th PC is defined as $\lambda_i / \Sigma_{1 \leq j \leq d} \lambda_j$. By picking up the first q PCs and ignoring the last $(d - q)$ PCs in eq.(1.89), we obtain the model

$$\mathbf{X} = A\mathbf{P}_T + \varepsilon. \tag{1.90}$$

In this equation, the first q PCs are summarized as $\mathbf{P}_T = (P_1, \ldots, P_q)^{\mathrm{T}}$, and $A = [\mathbf{w}_1, \ldots, \mathbf{w}_q]$ is a d-by-q constant matrix. The other $(d - q)$ PCs are combined into the error term $\varepsilon = P_{q+1}\mathbf{w}_{q+1} + \cdots + P_d \mathbf{w}_d$. Note that $\mathrm{var}\varepsilon = \lambda_{q+1} + \cdots + \lambda_d$.

Equation (1.90) has the same form with eq.(1.77). The difference of these two equations is that the former, PCA, does not assume any latent variables, while the latter, factor analysis, assumes some latent variables.

This is actually a large difference, since PCA does not mean the existence of any fundamental factors, but just summarize the data. However, the purpose of these models is the same, that is, to reduce the dimension of variables and to find the important factors or components affecting on the result. Recently, the independent component analysis is proposed as a novel method to find the important linear component similar to eqs.(1.77) and (1.90).

1.4.3 Independent component analysis

Suppose a d-dimensional random vector $\mathbf{X} = (X_1, \ldots, X_d)$ representing the observable variables, and assume $\mathrm{E}\,\mathbf{X} = 0$. The independent component analysis (ICA) is the following model of $\mathbf{X}$:

$$\mathbf{X} = A\mathbf{S}, \tag{1.91}$$

where $\mathbf{S} = (S_1, \ldots, S_q)^{\mathrm{T}}$ is a q-dimensional random variable and A is a d-by-q constant matrix. S_j, $j = 1, \ldots, q$, is referred to as the *independent component* (IC).

Equation (1.91) has the same form with the PCA, eq.(1.89). In fact, both the PCA and the ICA are the linear transformation technique, and the difference is just how to determine the matrix A. In the PCA, A is decided so that the PCs has *no correlation* with each other. In the ICA, A is decided so that ICs are *independent* of each other, that is,

$$f_{\mathbf{S}}(s) = f_{S1}(s_1) \cdots f_{Sq}(s_q), \tag{1.92}$$

where $f_{\mathbf{S}}$ and f_{sj}, $j = 1, \ldots, q$, are the joint and marginal PDF of $s = (s_1, \ldots, s_q)$ and s_j, respectively.

Independence is a much stronger property than no correlation, and has a great advantage. For example, if the variables are independent, the modeling of multiple variables is reduced to that of univariate random variables. In addition, if independent, it is easier to distinguish the sources of variables, which can be used as a tool to find factors in factor analysis.

On the other hand, the ICA cannot determine the variance of each component, while the PCA can. This is because A is restricted to a unitary matrix in the PCA but not in the ICA. Due to its ambiguity, the variances of ICs are conventionally set to unity:

$$\mathrm{cov}\mathbf{S} = I_q. \tag{1.93}$$

Consequently, the ICA cannot evaluate the importance of each component, while the PCA can pick up the most important one according to its variance. This is a disadvantage of the ICA.

There are many applications of the ICA. The most famous example is what is referred to as the *blind source separation*, which is to find a separated sound sources from the recorded sound waves. Since the recorded waves are the mixture of different sources and noises, it is important to reconstruct the original sources.

In finance, the earliest application of the ICA will be the stock market analysis by Back and Weigend (1997). Mălăroiu et al. (2000) applied the ICA to the foreign exchange market and obtained good forecasting results. More recently, Kumiega et al. (2011) found that the IC of the S&P sectorial indexes can be interpreted as the proxy of the energy, financial and the other sectors.

For details of further properties and applications of the ICA, see Hyväreinen et al. (2001) and Hyvärinen (2013).

1.4.4 Theoretical background of ICA

In this subsection, we explain how to obtain A in eq.(1.91) based on Hyväreinen et al. (2001). For simplicity, we assume $d = q$ hereafter.

Entropy The *entropy* is the fundamental idea of information theory. It is also referred to as the *Shannon entropy*. Given a discrete random variable X, its entropy H is defined as

$$H = -\sum_i P(X = a_i) \log P(X = a_i), \tag{1.94}$$

where a_i is a possible value of X. The value H is considered to measure the uncertainty of the random variable X. For a continuous random variable X, we define the *differential entropy*

$$H(X) = -\int f_X(\xi) \log f_X(\xi) d\xi, \tag{1.95}$$

where $f_X(\xi)$ is the PDF of X. Equation (1.95) can be extended to a random vector. Note that a transformation of variables changes the value of entropy. Especially, for a linear transformation $Y = MX$,

$$H(Y) = H(X) + \log |\det M|. \tag{1.96}$$

On the other hand, note that $H(X) = H(Y)$ for a parallel transformation $Y = X + A$.

Negentropy The smallest differential entropy is negative infinity, considering the localized distribution such as a delta distribution (the term $\log f_X(\xi)$ approaches to negative infinity as ξ approaches to the localized point).[17] We consider the problem what distribution gives the largest differential entropy. According to eq.(1.96), we can increase the entropy as much as we want by rescaling, so we impose the condition that the variance (or the diagonal elements of the covariance matrix) of X is set to unity. Under this constraint, it can be proved by functional derivative that the Gaussian distribution maximizes the differential entropy compared to the other distributions. The maximum entropy is $(1 + \log 2\pi)/2$ if it is one-dimensional. If the variance is σ^2, the maximum entropy is $(1 + \log 2\pi\sigma^2)/2$, according to eq.(1.96).

Therefore, the following quantity is assured to be non-negative:

$$J(X) = H(X_G) - H(X) \geq 0, \tag{1.97}$$

where H_G is the Gaussian distribution whose variance (or covariance) is set to var$X$ (cov$X$). Equation holds in the case that X is also a Gaussian distribution. This J is defined as the *negentropy*, which measures the non-Gaussianity, that is, how different the distribution is from the Gaussian distribution. It can be shown from eqs.(1.96) and (1.97) that the negentropy is invariant as for a linear transformation, that is,

$$J(MX) = J(X). \tag{1.98}$$

Mutual information Let $\mathbf{X} = (X_1, \ldots, X_q)^{\mathrm{T}}$ denote a q-dimensional random vector. We can obtain the joint entropy $H(\mathbf{X})$ and the marginal entropies $H(X_j)$, $j = 1, \ldots, q$. Then it is natural to define the *mutual information* $I(X_1, \ldots, X_q)$ as

$$I(X_1, \ldots, X_q) = \sum_{j=1}^{q} H(X_j) - H(\mathbf{X}). \tag{1.99}$$

It is shown that the mutual information always takes a non-negative value, and takes zero if and only if X_j, $j = 1, \ldots, q$ are independent of each other.

[17] In the case of discrete random variable, H is always non-negative since $0 < P(X = a_i) \leq 1$ and $(-\log P(X = a_i))$ is always non-negative.

Therefore, in order to obtain independent variables by a change of variables, it is a natural idea to minimize the mutual information. The ICA is based on this idea.

Suppose a change of variable from $\mathbf{X}$ to $\mathbf{Y} = M\mathbf{X}$. From eqs.(1.96), the mutual information is transformed to

$$I(Y_1, \ldots, Y_q) = \sum_{j=1}^{q} H(Y_j) - H(\mathbf{X}) - \log|\det M|. \tag{1.100}$$

Using the negentropies $J(Y_j)$, $j = 1, \ldots, q$, we obtain

$$I(Y_1, \ldots, Y_q) = C - \sum_j J(Y_j), \tag{1.101}$$

$$C = \sum_j H(Y_{j,G}) - H(\mathbf{X}) - \log|\det M|. \tag{1.102}$$

The value of $H(\mathbf{X})$ is fixed since the distribution of $\mathbf{X}$ is given. Consider an additional constraint that $\mathrm{cov}\mathbf{Y} = I_q$, that is, $\mathbf{Y}$ is standardized and non-correlated. Under this constraint, $H(Y_{j,G})$ takes a constant $(1 + \log 2\pi)/2)$since $Y_{j,G}$ is a standard Gaussian distribution. In addition, $\det M$ is a constant since $\det M = \sqrt{\det \mathrm{cov}\mathbf{Y} / \det \mathrm{cov}\mathbf{X}}$ =constant. Therefore, C is a constant under the constraint.

According to (1.101), minimizing of the mutual information is equivalent to maximizing the negentropy. This means that the ICA can also be formalized as the maximization of negentropy. The constraint $\mathrm{cov}\mathbf{Y} = I_q$ is allowed by eq.(1.93). As for maximizing the negentropy, Hyvärinen (2000) proposed a quick algorithm with the fixed point method, which is referred to as the *fastICA*. The algorithm is as quick as the usual PCA.

Note that it is not assured that the mutual information can be minimized to zero. Even after minimization, we still have a positive mutual information, that is, dependency structure. Note also that the maximization of negentropy is effective only for non-Gaussian distributions. For a Gaussian distribution, the negentropy is zero, so we cannot consider the maximization problem. These facts shows the fundamental importance of non-Gaussianity considering independent component analysis; for details, see Hyvärinen et al. (2001).

Chapter 2

Mean-ETL Portfolio Selection under Maximum Weight and Turnover Constraints based on Fundamental Security Factors

In this chapter, we model stock returns using fundamental data, minimizing AVaR, and multi-period portfolio selection with weight and turnover constraints. Equity returns are decomposed into returns explained by fundamental factors and non-fundamental factors. While the former are found to be independent, the latter factors are found to be highly dependent among various stocks. Based on this fact, we construct return forecast models using the ARMA-GARCH models with different innovation distributions. In addition, we compare the objective functions of portfolio optimization, not only the mean-variance approach but also the mean-AVaR approach and so on. As results, we find that (1) the ARMA-GARCH model with classical tempered stable distribution provides better prediction than the normal and Student-t distribution and (2) the AVaR is a better risk measure than the variance. We also see how portfolio performance changes under weight and turnover constraints, and suggest it is effective to reduce the trading universe to large capitalization stocks.

2.1 Introduction

Since Markowitz introduced mean-variance portfolio selection in his celebrated paper in 1952, several ways have been suggested to enhance the framework and/or inputs employed.

First is improved modeling of mean and variance. For example, the sample mean and variance have been proven to be poor choices in the high-dimensional settings, especially when the available sample size is close to or even smaller than the number of instruments in the portfolio. More fundamentally, off-sample distribution in future is generally different from in-sample distribution in past.

Second is dealing with tail risk. Usually, the normal distribution of returns is assumed, and variance (or covariance matrix in multidimensional cases) is used as a risk measure. If the normal distribution is assumed, all risk can be captured by variance. However, empirical evidence suggests that return distributions are not normal.

Finally, incorporating constraints and multiple investment periods into the portfolio selection model have been proposed. Practically, all investors confront weight constraint, turnover constraint and multiperiod investment. For example, most investors are not allowed to make short positions due to investment policy or regulatory constraints. Also, maximum weight for a single asset is usually introduced in order to diversify the risk. These weight constraints reduce feasible set of portfolio selection. In addition, since the investment environment itself has uncertainty in future, multiperiod investment should be considered. Multiperiod investment also confronts the problem of transaction cost and turnover constraint.

In this chapter, we take these enhancements into account, suggesting a practical example of portfolio construction. First, in order to predict future returns, we introduce a return forecast model based on monthly fundamental financial data and ARMA-GARCH model with non-normal innovation terms. In this model, we first do regression of returns over 21 fundamental factors such as earnings and momentum indices, divide returns into two parts: Returns explained by these 21 factors, which is defined as "factor returns," and returns explained as residuals, defined as "non-factor returns." Second, since factor returns are approximately independent, we apply ARMA-GARCH model to factor returns in order to obtain out-of-sample forecast. Third, since non-factor returns are dependent among stocks, we extract principal components (PC) of non-factor returns, and do regressions of non-factor re-

turns by these PCs. Fourth, we apply ARMA-GARCH model into these PCs and residuals to get forecast of out-of-sample values. This method of forecasting is shown to result in much better performance than simple sample return and covariance by constructing actual portfolios.

Second, when applying ARMA-GARCH, we use heavy-tailed innovations such as the Student-t distribution and classical tempered stable (CTS) distribution (Kim et al., 2008; Rosiński, 2007) in addition to normal distribution, since it is supposed that return distribution is not normal. Furthermore, we do not only estimate parameters of these distributions, but also generate large number of scenarios of out-of-sample returns for Monte Carlo simulation, which enables various optimizations of risk measures without any presumptions on the joint distribution of returns. In addition, we use both variance and average value at risk, AVaR as risk measure to be minimized for portfolio selection. This is a generalization of Markowitz framework consistent with the expected utility maximization, and it is a convex optimization problem which can be solved efficiently (Bertsimas et al., 2004; Rachev, 2008). Finally, we see that minimizing AVaR with CTS innovations gives better result than other cases.

Third, we check how our portfolio performance changes under weight and turnover constraints. First, we see the performance under no short selling is allowed. Next, we add the maximum weight constraint of 4%, which does not have big effect. Finally, we consider turnover constraint whose maximum turnover is 8% monthly. We adopt the strategy to reduce the number of stock universe and trade large capitalization securities. It is shown that the strategy gives better performance.

The rest of this chapter is structured as follows. First, we describe the data we use, introduce the forecasting method of returns we adopt, and discuss optimization approaches. Next, we investigate the efficiency of various forecasts and approaches by constructing actual portfolio, and discuss how to construct portfolio within turnover constraint.

2.2 Data

We use the monthly data from December 1979 to November 2009 provided McKinley Capital. The original data is provided by Wharton Research Data Service, WRDS. We extract fundamental financial data such as earnings per share and forecast of these fundamental data as fundamental factors. We also

use momentum indices such as USER, which is introduced and explained by Guerard et al. (2009), as fundamental factors. We finally obtain 21 factors for each stock for each month.

The database provides returns of about 15,000 stocks. However, about half of these stocks do not have their data more than 60 months, and only 310 stocks exist for the entire period, or 360 months. In order to apply statistical method discussed later, we have to exclude some stocks which seldom appear.

Our exclusion is as follows: first, we set the number of regression period length T. Next, for each month t_0, we select stocks which have successive monthly data from month $(t_0 - T + 1)$ to month t_0. We call the set of these stocks as the universe U_{t_0} of month t_0. In other words, U_{t_0} consists of stocks which have more than T successive entries in the past. We adopt $T = 201$ until November 1999 and 240 from December 1999 due to data constraint. Since we focus on not factors themselves but their changes, we add one to our desired period length, 200 and 240. Then, the number of stocks in $N_{t_0} = n(U_{t_0})$ distributes mostly between 300 and 400 with average 347.

Let r_t^i denote return and $\tilde{F}_{t,k}^i$ factor values where t denotes month, $k = 1, 2, ..., K$ is the index of factor $(K = 21)$ and $i \in U_{t_0}$ is the index of individual stock. Hereafter, we focus on changes of factors $F_{t,k}^i = \tilde{F}_{t,k}^i - \tilde{F}_{t-1,k}^i$ where $1 \leq t \leq 359$, or from January 1980 to November 2009. Note that t_0 moves in the range $200 \leq t_0 \leq 359$, or from August 1996 to November 2009.

2.3 Forecasting model

The estimation of mean return and covariance matrix is very important. The method of our model consists of following three parts: (1) Constructing factor model, (2) applying time series model and estimating parameters, and (3) simulating scenarios of one-step-ahead returns via simulating factors and noise terms. In this section, for notation simplicity, let t $(1 \leq t \leq T)$ denote the in-sample period, and $t = T + 1$ denote out-of-sample month to be forecasted.

2.3.1 Factor model

First, in order to extract the predictable trends out of the returns, we do regression of the monthly returns of each company with respect on the afore-

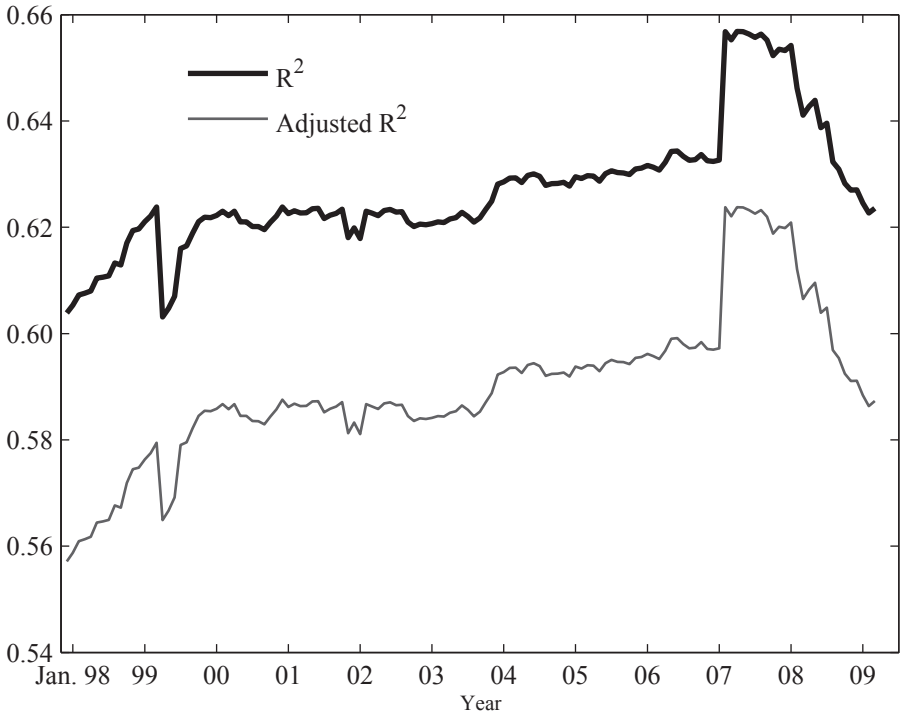


Figure 2.1: Explanatory power of 21 fundamental factors, according to R^2 and adjusted R^2 of eq. (2.1). About 60% of returns can be explained by these factors.

mentioned $K = 21$ fundamental factors:

$$r_t^i = \alpha_0^i + \sum_{k=1}^{K} \beta_k^i F_{k,t}^i + \varepsilon_t^i, \tag{2.1}$$

where $1 \le t \le T$, $1 \le i \le N$, and ε_t^i represents noise terms. The coefficient parameters α_0^i and β_k^i are estimated via least square method.

Figure 2.1 shows both R^2 and adjusted R^2 are around 0.6 in averages when we do regression of returns r_t^i over factor changes $F_{t,k}^i$ for each stock i in each universe U_{t_0} based on eq. (2.1). Note that month t_0 is from December 1997 to November 2009 and that $T = 200$ or 240 according to month t_0.

We define $F_t^i := \sum_k \beta_k F_{k,t}^i$ as a factor return of stock i at time t. We

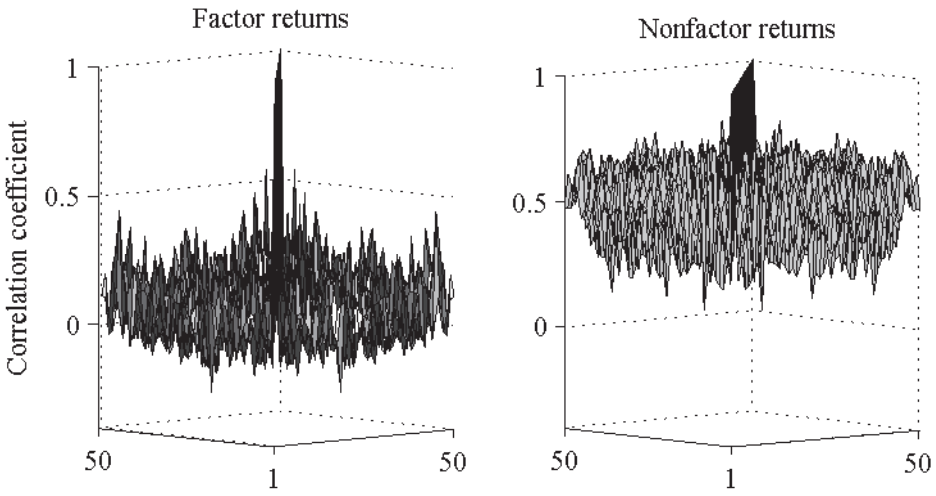


Figure 2.2: Sample correlation of factor and non-factor returns. While correlation between factor returns is weak, that between nonfactor returns is considerable, which may imply hidden factors in nonfactor returns.

also define ε_t^i as a non-factor return. Figure 2.2 shows the sample correlation coefficient of factor and nonfactor returns among 50 stocks randomly selected. It is shown that factor returns F_t^i are close to independent while non-factor returns ε_t^i have dependency.

The dependency of non-factor returns ε_t^i is not expected since they stand for random noises. This dependency is partly because there exist hidden factors not chosen in the regression model in eq. (2.1). In order to model this dependence parsimoniously, we extract principal components $H_{m,t}$ $(1 < m < N)$ of the non-factor returns ε_t^i $(1 \leq i \leq N)$. Then, we adopt the top M components of $H_{m,t}$ $(M \approx 30)$, and regress ε_t^i with regard to $H_{m,t}$:

$$\varepsilon_t^i = \alpha_\varepsilon^i + \sum_{m=1}^{M} \gamma_m^i H_{m,t} + \eta_t^i, \tag{2.2}$$

where η_t^i are independent. Note that this regression is done for each stock i $(1 \leq i \leq N)$.

Figure 2.3 shows the relation between the explanatory power of eq. (2.2) and the number of principal components M. It is shown that the explanatory power increase much faster with respect to the increase of M. In our model,

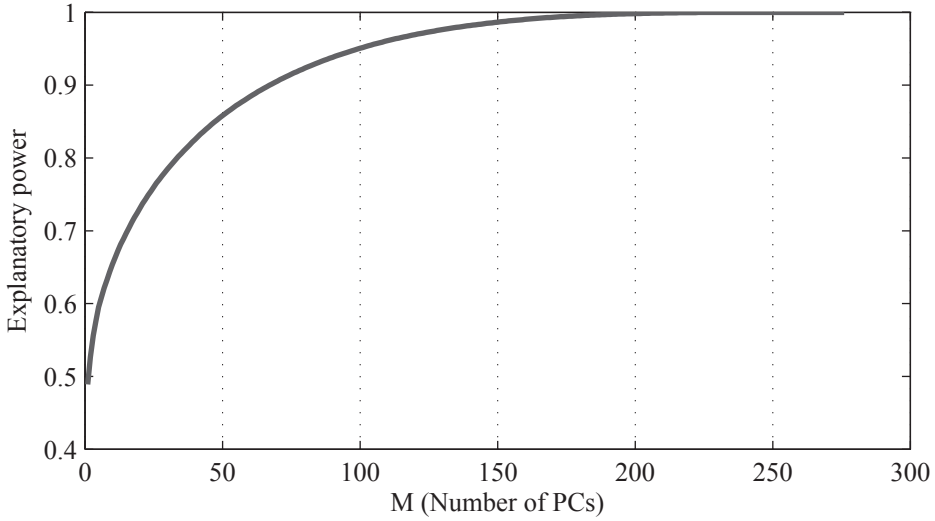


Figure 2.3: Explanatory power with respect to the number of principal components. The explanatory power increase much faster with respect to the increase of M. In our model, we choose the number of components such that the explanatory power is greater than 80%, approximately $M \approx 30$ factors.

we choose the number of components such that the explanatory power is greater than 80%, approximately $M \approx 30$ factors.

Note that it is possible to consider that principal components $H_{m,t}$ correspond to macroeconomic factors, which do not appear in our model explicitly. Our $K = 21$ factors aforementioned are corporate specific factors such as earning per share, which is why our factor returns F_t^i show weak correlation. If we adopt this idea, we can use the knowledge of macro environment to forecast $H_{m,t}$. In this chapter, however, we do not consider this problem.

2.3.2 Time series model

Factor returns F_t^i are returns explained by factors. Since these returns have several stylized facts such as volatility clustering, it is natural to apply ARMA(1,1)-GARCH(1,1) model to these returns.

In addition, we apply ARMA-GARCH model into $H_{m,t}$ and η_t^i. Note that we apply the model into each principal component m and each stock i.

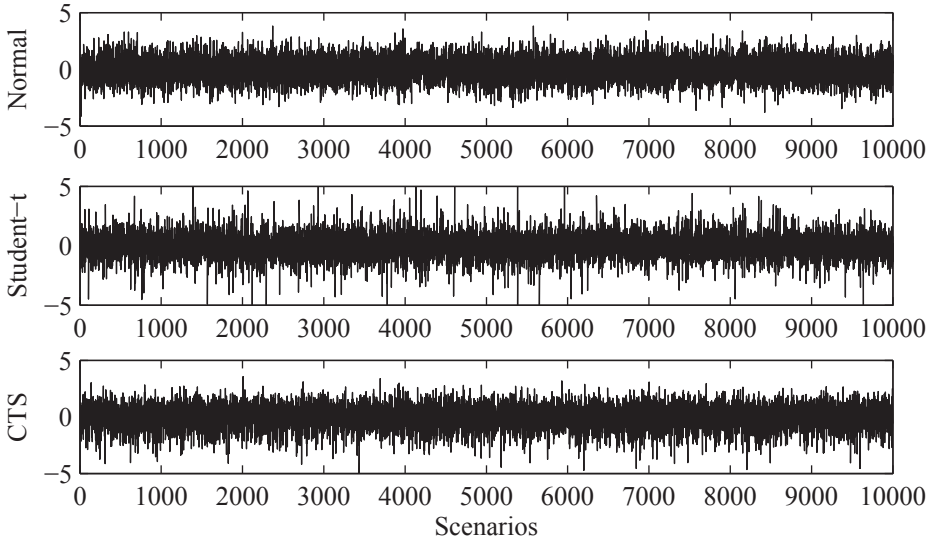


Figure 2.4: Simulated scenarios of the first hidden factors by normal (top), Student-t (middle) and CTS (bottom) distributions. It is seen that normal distribution underestimates the downside risk, while Student-t and CTS distributions capture the downside risk. However, Student-t distribution has a disadvantage that it is symmetric, which is not typical in return distribution. CTS distribution has both tail risk and asymmetric downside risk.

To express the fat-tailedness and skewness of the return distribution, Student-t and classical tempered stable (CTS) distribution are introduced for innovation terms in addition to normal (Gaussian) distribution. Figure 2.4 shows the simulated scenarios (values) of $H_{m=1,t}$. It is seen that normal distribution underestimates the downside risk, while Student-t and CTS distributions capture the downside risk. However, Student-t distribution has a disadvantage that it is symmetric, which is not typical in return distribution. CTS distribution has both tail risk and asymmetric downside risk.

2.3.3 Simulation

At this point, we obtain the return distribution of time $T + 1$ by combining factor model and ARMA-GARCH models. However, this combination is not analytically tractable, generally. In addition, the portfolio VaR or AVaR is

generally not obtainable or too complicated in an analytical form. Therefore, Monte Carlo simulation is used to generate a large number of scenarios, based on which the portfolio VaR and AVaR are easily computed via historical simulation method. The method is as follows:

1. Generate S scenarios of F^i_{T+1}, η^i_{T+1} $(1 \leq i \leq N)$ and $H_{m,T+1}(1 \leq m \leq M)$.

2. Make S scenarios of ε^i_{T+1} by substituting η^i_{T+1} and $H_{m,T+1}$ into eq. (2.2),
$$\varepsilon^i_{T+1} = \alpha^i_\varepsilon + \sum_{m=1}^{M} \gamma^i_m H_{m,T+1} + \eta^i_{T+1}. \tag{2.3}$$

3. Make S scenarios of r^i_{T+1} by substituting ε^i_{T+1} and F^i_{T+1} into eq. (2.1),
$$r^i_{T+1} = \alpha^i_0 + F^i_{T+1} + \varepsilon^i_{T+1}. \tag{2.4}$$

We use $S = 10,000$. Figure 2.5 shows the historical correlations and simulated correlations among 50 stocks randomly selected. There is a slight reduction of our simulated covariance compared with the sample covariance due to possibly: (1) loss of the rest 20% variance in the residuals; (2) overlook of the dependence of fundamental factors. However, to simplify our model, we do not consider these at the present, which could be improved by allowing more explanatory power or applying copula model or PCA on the combined factors as well.

2.4 Optimization approaches

In mean-variance approach, the first step of optimization is obtaining an efficient frontier, on which variance of portfolio is minimized for a given level of return. Variance here is used as a measure of risk. Under the assumption of normal distribution of returns, variance is the only and the enough measure of risk. Variance has also a practical advantage that it can be easily obtained by a single covariance matrix without knowledge of return distribution.

However, if the distribution is other than normal, variance is just one of statistics which describes risk structure. The disadvantage of variance is that it cannot capture the tail risk. Candidates of well-known risk measures other than variance are VaR and ETL, or usually equivalently, AVaR. These

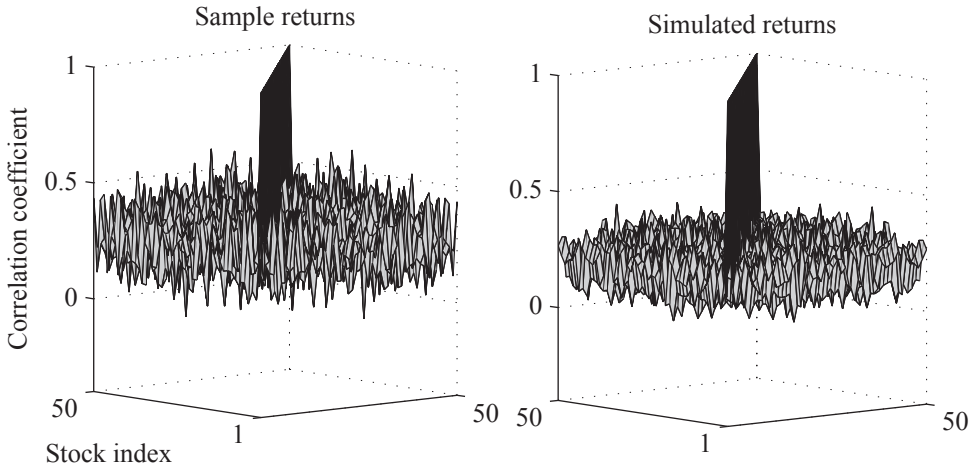


Figure 2.5: Correlation of sample returns and correlation of simulated next month's returns. There is a slight reduction of our simulated covariance compared with the sample covariance due to possibly: (1) loss of the rest 20% variance in the residuals; (2) overlook of the dependence of fundamental factors.

measures are expected to reflect tail-risk better than variance. Minimizing these risk measures for a given level of return gives an alternative of efficient frontier in mean-variance approach. Especially, the substitution of AVaR into variance is called mean-ETL.

In addition, to obtain a specific portfolio, the desired return level r should be specified. Within the return range of the efficient frontier, r is a free parameter. Therefore, we have to give some criteria in order to decide r. One idea is to maximize $u = \mu/\sigma$, the return for a unit volatility where μ is the average return and σ is the volatility. We call u a risk utility. Under the existence of risk-free asset, however, r can be decided as it maximizes Sharpe ratio, $u = \mu_e/\sigma$, where μ_e is the average excess return of portfolio over risk-free asset. One natural extension of this idea to our situation is to substitute VaR or AVaR into the volatility in the expression of Sharpe ratio. Especially, $u = \mu_e/\mathrm{AVaR}$, the substitution by AVaR, is called STARR, stable tail-adjusted return ratio. We also call these ratios risk utilities.

In summary, two variables, risk measure s and risk utility u, specify our portfolio selection approach. In formulae, the approach has two steps. The first step is to minimize s,

$$w(r) = \arg\min_{w \in W(r)} s(w) \tag{2.5}$$

where $W(r) = \{w; w \in W_C \wedge \mu \cdot w = r\}$, W_C is a set of portfolios satisfying constraints and μ is a vector representing the expected return of stocks in the universe. The second step is to maximize u,

$$w = w(\arg\max_{r} u(w(r))). \tag{2.6}$$

Note that the most standard setup of modern portfolio theory corresponds to use standard deviation as a risk measure and Sharpe ratio as a risk utility. Under the expected utility hypothesis, our approach is the natural generalization of the standard mean-variance approach, which initially introduced by Markowitz.

2.5 Comparison of forecasts and approaches

Next, we compare portfolio selection approaches discussed by constructing actual portfolio. In addition, we also compare return forecasts obtained

in the previous section by making actual portfolio. Then, we can select the best combination of the forecast and approach by comparison. There are four available forecasts: Historical average and covariance (referred as "historical" hereafter), forecast based on ARMA-GARCH-Normal (AGNormal) model, forecast based on ARMA-GARCH-T (AGT) model, and forecast based on ARMA-GARCH-CTS (AGCTS) model. We examine two risk measures: Variance and AVaR at 95% significance level. We examine three risk utilities: Sharpe ratio (Sharpe), return/volatility ratio (Ret/Vol) and STARR (STARR). For each combination, we computed various statistics such as historical volatility or beta. We adopted U.S. TB 3M rate from Bloomberg (USGG3M Index) as risk-free rate and Russell 3000 Growth Index with dividends as market index. All assets are invested into stocks and investing to risk-free asset is prohibited. We use numerical search function `fmincon` of MATLAB for portfolio optimization. At this point, we do not impose constraints other than no-short constraint, since we want to see the characteristics of forecast and approach.

The result is shown in Table 2.1. For reference, statistics of the risk-free rate (TB 3M, TB3M) and the market index (Russell 3000 Growth with dividends, R3KGD) are shown. In addition, statistics of S&P 500 (SPX) is shown for comparison purpose. Note that the cumulative return of R3KGD is lower than that of TB3M, while the monthly average of R3KGD is higher than that of TB3M. This is due to a high volatility and large drawbacks of R3KGD.

Table 2.1: Comparison of forecasts and approaches

Forecast	Risk measure	Risk utility	Average return	Volatility	Skew	Kurtosis	Upside months	Downside months
TB3M			0.26	N/A	N/A	N/A	144	0
R3KGD			0.31	5.70	-0.67	3.65	79	65
SPX			0.21	4.76	-0.69	3.92	84	60
Historical	Variance	Sharpe	0.44	3.52	-0.38	3.64	81	63
Historical	Variance	Ret/Vol	0.47	3.44	-0.41	3.48	84	60
Historical	Variance	STARR	0.47	3.44	-0.41	3.48	84	60
AGNormal	Variance	Sharpe	1.30	6.64	0.96	7.90	81	63
AGNormal	Variance	Ret/Vol	1.29	6.61	0.98	7.99	82	62
AGNormal	Variance	STARR	1.22	6.93	0.47	8.41	82	62
AGNormal	AVaR	Sharpe	1.34	6.70	0.81	7.55	78	66
AGNormal	AVaR	Ret/Vol	1.33	6.64	0.85	7.72	80	64
AGNormal	AVaR	STARR	1.32	7.02	0.49	8.12	79	65
AGT	Variance	Sharpe	1.12	6.73	0.82	9.25	82	62
AGT	Variance	Ret/Vol	1.04	6.71	0.84	9.38	81	63
AGT	Variance	STARR	1.10	6.84	0.64	9.39	83	61
AGT	AVaR	Sharpe	1.13	6.88	0.97	9.75	81	63
AGT	AVaR	Ret/Vol	1.08	6.86	1.00	9.90	80	64
AGT	AVaR	STARR	1.26	7.22	0.89	9.16	82	62
AGCTS	Variance	Sharpe	1.29	6.60	1.05	8.24	80	64
AGCTS	Variance	Ret/Vol	1.25	6.59	1.06	8.31	80	64
AGCTS	Variance	STARR	1.32	6.88	0.69	8.16	80	64
AGCTS	AVaR	Sharpe	1.29	6.73	0.95	7.09	80	64
AGCTS	AVaR	Ret/Vol	1.27	6.70	0.97	7.20	81	63
AGCTS	AVaR	STARR	1.22	6.91	0.69	7.28	79	65

Table 2.1: Comparison of forecasts and approaches (continued)

Forecast	Risk measure	Risk utility	Beta	Alpha	p(H)	Specific error	Sharpe ratio	Treynor ratio
TB3M			0.00	0.00	1.00	0.00	0.0000	N/A
R3KGD			1.00	0.00	1.00	0.00	0.0091	0.05
SPX			0.78	-0.09	0.83	1.63	-0.0095	-0.06
Historical	Variance	Sharpe	0.36	0.17	0.16	2.83	0.0536	0.52
Historical	Variance	Ret/Vol	0.33	0.20	0.13	2.86	0.0625	0.64
Historical	Variance	STARR	0.33	0.20	0.13	2.86	0.0625	0.64
AGNormal	Variance	Sharpe	0.71	1.01	0.02	5.25	0.1579	1.47
AGNormal	Variance	Ret/Vol	0.71	1.00	0.02	5.22	0.1566	1.45
AGNormal	Variance	STARR	0.76	0.92	0.03	5.40	0.1390	1.26
AGNormal	AVaR	Sharpe	0.73	1.05	0.01	5.27	0.1620	1.49
AGNormal	AVaR	Ret/Vol	0.72	1.04	0.01	5.23	0.1622	1.50
AGNormal	AVaR	STARR	0.76	1.02	0.02	5.52	0.1513	1.40
AGT	Variance	Sharpe	0.76	0.82	0.04	5.13	0.1279	1.12
AGT	Variance	Ret/Vol	0.76	0.75	0.06	5.12	0.1175	1.03
AGT	Variance	STARR	0.78	0.80	0.05	5.22	0.1238	1.09
AGT	AVaR	Sharpe	0.77	0.83	0.05	5.31	0.1267	1.14
AGT	AVaR	Ret/Vol	0.76	0.79	0.06	5.30	0.1210	1.09
AGT	AVaR	STARR	0.79	0.97	0.03	5.65	0.1398	1.28
AGCTS	Variance	Sharpe	0.71	1.00	0.02	5.20	0.1566	1.44
AGCTS	Variance	Ret/Vol	0.71	0.96	0.02	5.19	0.1510	1.39
AGCTS	Variance	STARR	0.75	1.02	0.02	5.36	0.1545	1.40
AGCTS	AVaR	Sharpe	0.70	0.99	0.02	5.41	0.1533	1.47
AGCTS	AVaR	Ret/Vol	0.70	0.98	0.02	5.37	0.1517	1.45
AGCTS	AVaR	STARR	0.74	0.93	0.03	5.46	0.1403	1.30

Notes: Figures are shown in percentage in Average return, volatility, maximum drawdown, alpha, specific error and Treynor ratio. The market index is R3KGD, Russell 3000 Growth with Dividends. TB3M denotes US Treasury bill three month, and SPX denotes Standard and Poor's 500. H denotes the null hypothesis that $\alpha = 0$.

The fourth line (historical data, variance as risk measure and Sharpe ratio as risk utility, denoted as "Historical-Variance-Sharpe" hereafter) corresponds to the standard setup. This combination is better than the market index in many aspects: Higher in mean and cumulative return, lower volatility, and lower maximum loss. In Table 2.1 all portfolio selections based on historical data give better results than investment on the market index.

Comparing ARMA-GARCH forecasts (AGNormal, AGT, and AGCTS) and historical, portfolios based on ARMA-GARCH forecasts give much higher average and cumulative returns than those based on historical data. As for volatility and maximum loss, ARMA-GARCH forecasts give higher values than historical data. Using regression, most of ARMA-GARCH forecasts give significant positive alpha with confidence level 95%, while historical data give no significant positive alpha. Sharpe and Treynor ratios are also higher in ARMA-GARCH forecasts than in historical data. In summary, ARMA-GARCH forecasts are better than historical data in many aspects other than volatility and maximum loss.

Comparing variance and AVaR as risk measure, AVaR gives higher realized returns than variance in many cases using ARMA-GARCH forecasts. For example, comparing AGNormal-Variance-Sharpe and AGNormal-AVaR-Sharpe, the average and cumulative returns of the latter are higher than those of the former by 0.4% and 21%, respectively. Among nine cases of such comparisons (three cases of ARMA-GARCH forecasts and three cases of risk utility), seven cases report higher returns in AVaR risk measure than in variance risk measure. There are two exceptions: the AGCTS-AVaR-Sharpe case reports lower cumulative return by 8% compared to AGCTS-Variance-Sharpe, and the case of AGCTS-AVaR-STARR reports lower average and cumulative returns by 0.10% and 60% respectively compared to AGCTS-Variance-STARR. However, in total, we consider that AVaR is better as risk measure than variance in general.

Comparing ARMA-GARCH forecasts, AGT case has less average return, cumulative return, alpha, Sharpe ratio, and Treynor ratio than AGNormal and AGCTS, generally. On the other hand, volatilities and maximum loss of AGT tend to be higher than AGNormal and AGCTS. In addition, in some combinations using AGT data, p-values of null hypothesis that α=0 is higher than the significance level (5%). For these reasons, we reject AGT forecast.

Comparing AGNormal and AGCTS, they resemble each other. As for returns, AGNormal reports higher values than AGCTS. However, as for maximum loss, AGCTS reports lower values than AGNormal in all cases. Since

this is a favorable property for portfolio selection in order to avoid market collapse such as 2001 or 2008, we adopt AGCTS forecast model.

At this point, we have selected AGCTS as forecast and AVaR as risk measure, equivalent to mean-ETL. Our last selection of approaches is risk utility: Sharpe, Ret/Vol, and STARR. Among these three, Sharpe seems to be the best in the sense that it reports highest average and cumulative returns, lowest drawback. However, in this chapter, we adopt STARR ratio, since it is more consistent with using AVaR as risk measure.

2.6 Portfolio selection with turnover constraint

In our portfolio selection, we have two constraints: turnover condition (less than 8% monthly or 100% annually) and weight range condition (between zero and 4%). Weight condition does not affect so much on portfolio return. Figure 2.6 shows the development of cumulative returns with weight maximum of 4% and without weight maximum. Hereafter, we use AGCTS-AVaR-STARR in order to obtain an optimized portfolio, following the discussion in the previous subsection.

On the other hand, turnover constraint causes an important change on our portfolio selection. Since it restricts changes between successive months, portfolio selection becomes path dependent. Especially, the selection of the initial month is important. If we select stocks which can contribute portfolio return in the initial month but cannot at all after the following months, it becomes difficult to obtain a good result. In order to avoid this situation, we have to know future returns not only one month ahead but also many months ahead. However, it is practically impossible.

Another idea is to restrict stocks universe into those which appear most frequently in top n% in portfolio selection without turnover constraints. We call n as the selection level. As a result, the possibility is avoided that the portfolio is trapped into meaningless stocks. The problem here is how to find such a universe of stocks. One idea is to use the universe of stocks which appear frequently in portfolio selection without turnover constraint in past. The members of the universe are replaced slightly every month according to the portfolio selection of the latest months. Namely, this is a rolling selection based on historical data.

We use $n = 5$ and 10 as selection levels, in which case the universe is consisted of 99 and 127 stocks, respectively. As for restriction of the

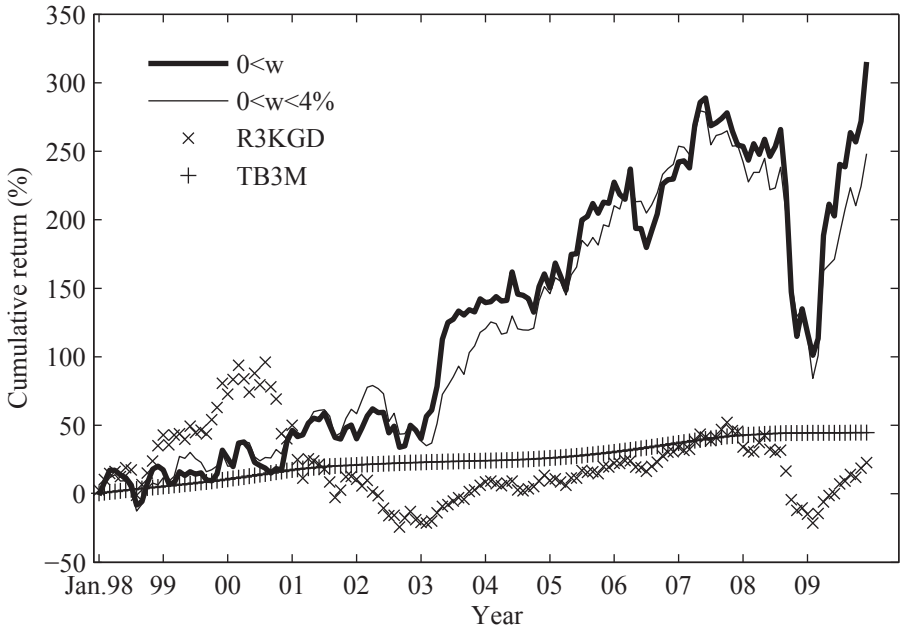


Figure 2.6: Cumulative returns with weight constraint, without turnover constraint. The portfolio with the maximum weight constraint ($0 < w < 4\%$, solid slight line) is slightly lower than that without the constraint ($w > 0$, solid bold line), but its performance is significantly much better than that of market indices. Note that we use AGCTS-AVaR-STARR in order to obtain an optimized portfolio, following the discussion in the previous subsection.

universe, we adopt the stocks based on no turnover results from December 1996 to December 2009, instead of rolling selection. Strictly speaking, this selection uses information of future portfolio selection, which is infeasible in practical situation.

We also set the initial month of portfolio selection as December 1996 based on the data of November 1996. This purpose is to avoid the effect of initial selection. Since no turnover constraint is imposed on the initial selection, it is possible that the return of the initial selection is relatively high compared to other months.

2.7 Large market capitalization securities

Here, we consider tracking of stock index by holding large market capitalization securities representing the market. This is also expected to reduce turnover.

We first select three stocks: IBM (IBM, CUSIP:459200101), Johnson and Johnson (JNJ, 478160104), Exxon Mobil (XOM, 30231G102, from January 2000). Before 2000, we use GE (GE, 369604103) instead of Exxon, since our data lacks the data before the merger of Exxon and Mobil in 1999. We call these three stocks LCS, large capitalization securities.

Next, we decide weights of these stocks according to the following rule:

1. If USER (see Guerard et al. [2009]) value of a stock becomes larger than 70, its weight is increased to 4%.

2. If USER value of a stock becomes smaller than 40, its weight is decreased to 0%.

Therefore, the maximum weight of these three stocks is up to 12%. Here, the variable we refer to decide the weight of LCS is USER value only. The weight of LCS in the stock index is not referred.

The selection of LCS is basically heuristic, but has some reasons. First, these stocks are selected because they are steady members of top market capitalization stocks. The average rank of GE from 1980 to 2009 is 2.1, IBM 5.5, Johnson and Johnson 11.5. The average of Exxon and Mobil is 2.4 from 2000 to 2009. Second, these companies are representative in industries they belong to.

2.8 Result

In summary, our portfolio selection has three steps:

1. First, we decide the universe of stocks according to the result with weight and without turnover constraints.

2. Second, we decide the weight of LCS according to their USER.

3. Third, we optimize the portfolio with weight and turnover constraints. Here, the sum of portfolio weights is decided by the weights of LCS, which means its sum is not always one.

Figure 2.7 shows the cumulative return for $n = 5$ and 10 cases and Table 2.2 shows statistics of these cases. It is shown that turnover constraints decrease the average and cumulative return, but they are still much higher than the market index, Russell 3000 Growth with dividends.

It is also shown that the $n = 10$ case provides less return compared to the $n = 5$ case. In addition, positive alpha of $n = 10$ case is not significant at 95% confidence interval. This reason is unknown. One possible explanation is that the necessary turnover to achieve the best portfolio may become larger when the number of stocks in the universe is large. In this case, for a fixed turnover level, the gap between the best and feasible portfolios becomes larger, causing decrease of average return.

Finally, we select $n = 5$ case. In this case, we obtain 0.95% average return and 213% cumulative return. The historical volatility is 5.67%, which is lower than that of the market index, 5.70%. Its beta is 0.70 and alpha is 0.57 in 95% confidence level.

2.9 Conclusion

Our portfolio selection is divided into two parts: Forecast part and portfolio selection part. In the forecast part, we first do regression of stock returns with respect to fundamental factors of each company, define factor and non-factor returns, and find that the main dependence comes from nonfactor returns. Then, principal component analysis is used to extract the hidden factors out of the non-factor returns and to approximate the covariance. The factor and nonfactor returns are modeled by ARMA-GARCH, which gives scenarios of future returns.

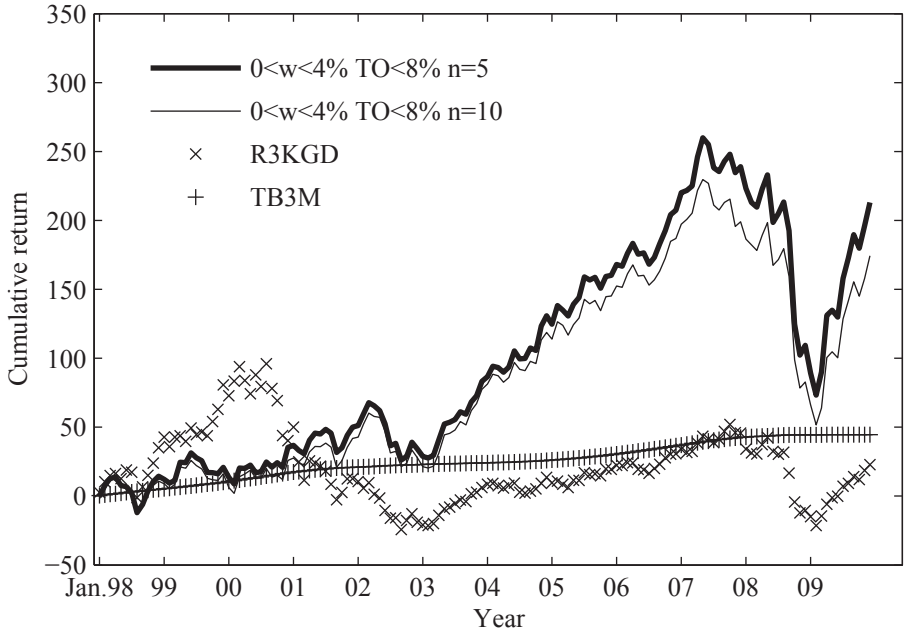


Figure 2.7: Cumulative returns with weight and turnover constraint. It is shown that turnover constraints decrease the average and cumulative return, but they are still much higher than the market index, Russell 3000 Growth with dividends.

Table 2.2: Comparison of selection levels

#	Constraints			Average return	Volatility	Skew
1		TB3M		0.26	NaN	NaN
2		R3KGD		0.31	5.70	-0.67
3	0<w<4%	TO<8%	n=5	0.95	5.52	-0.39
4	0<w<4%	TO<8%	n=10	0.86	5.67	-0.28

#	Kurtosis	Upside months	Downside months	Maximum drawdown	Cumulative return	Beta
1	NaN	144	0	0.00	45	0.00
2	3.65	79	65	-17.93	23	1.00
3	5.94	84	60	-23.21	213	0.70
4	5.72	83	61	-23.32	174	0.71

#	Alpha	p(H)	Specific error	Sharpe ratio	Treynor ratio	Information ratio
1	0.00	1.00	0.00	0.0000	NaN	NaN
2	0.00	1.00	0.00	0.0091	0.05	NaN
3	0.66	0.02	3.82	0.1259	0.99	0.15
4	0.57	0.06	3.98	0.1075	0.86	0.13

Notes: w is weight of each stock, TO is turnover of each month, and n is selection level. The AGCTS-AVaR-STARR approach is selected.

In portfolio selection part, we first construct portfolios combining two risk measures, three risk utilities and four forecasts, and find that using ARMA-GARCH-CTS forecast model and AVaR risk measure gives the relatively best result. Next, we discuss restricting of the universe of stocks and introducing of large capitalization securities. Finally, we obtain a portfolio which is better than Russell 3000 Growth Index with dividends, the market index.

Chapter 3

Time Series and Copula Dependency Analysis for Eurozone Sovereign Bond Returns

In this chapter, we analyze the distribution of returns on seven major Eurozone sovereign bonds (France, Germany, Greece, Ireland, Italy, Portugal, and Spain) and their co-movement. To obtain a good forecast for the return distribution, it is necessary to forecast future market volatility. We empirically investigate five ARMA-GARCH models for forecasting market volatility based on different assumptions about the innovations: Gaussian, Student-t, classical tempered stable, normal tempered stable, and α-stable. For each of the five models, we apply the relative copula dependence structure. Finally, we assess the forecasting performance of these models. Daily returns from 2001 to 2011 for the Barclays Capital Index are used, with backtesting based on returns from 2006 to 2011. To investigate the extent to which the models studied can provide a forward-looking measure to detect the exacerbating of the financial crisis of Greece, we analyze the evolution of the tail parameter over time.

3.1 Introduction

Recent turmoil in the European sovereign bond market demonstrated that these debt obligations exhibit the same type of tail risk as other investment instruments such as equities. It has been the failure to properly model tail risk that market observers identify as one of the reasons for the recent global financial crisis. More specifically, it is alleged that risk models prior to 2007 failed to properly assess the risks associated with large adverse price behavior. The forecasting of price behavior of financial instruments is an essential activity in the implementation of risk management and portfolio allocation.

In modelling price behavior, there exist two distinguishing characteristics which have been observed. First is the effect of returns in the near past on returns in near future. Empirically, this observation can be captured using an autoregressive and moving average (ARMA) model. The other is volatility clustering, which is the tendency that volatility remains high in near future if it is high in near past and remains low if it is low. The autoregressive conditional heteroscedasticity (ARCH) model and its generalization (GARCH) are examples of econometric models for capturing volatility clustering. The combination of these models, referred to as ARMA-GARCH models, is one of the simplest models which can capture price behavior that has been observed in markets.

The problem remains, however, that the ARMA-GARCH model cannot completely capture tail risk, although it works better in capturing tail risk compared to many other models. This problem is partly attributed to the failure of the Gaussian assumption to properly account for random variables in financial time series. Kim et al. (2011) have demonstrated how applying heavier-tailed innovations improve the situation and for that reason we incorporate their methods in this chapter.

The other aspect of the risk problem is attributed to systemic risk and has been dealt with by some researchers by papers focusing on macroeconomic factor shocks that tend to affect all institutions/countries in an economy. A macro shock causes an increase in correlated default losses, with detrimental effects on intermediaries and thus financial stability. Aikman et al. (2009) propose a model to assess the impact of macroeconomic and financial shocks on the banking system. Giesecke and Kim (2011) define systemic risk as the conditional probability of failure of a large number of financial institutions, based on a dynamic hazard rate model with macroeconomic covariates. A related study using a large number of macroeconomic and financial covari-

ates by Schwaab et al. (2011) derives measures and early signal indicators of distress using a mixed dynamic factor model approach. Longstaff et al. (2011) show that the credit spreads for Mexico, Turkey, and Korea share a strong common relation to U.S. stock market volatility as measured by the volatility index (VIX) index. Such a common dependence could induce significant correlations among sovereign credit spreads.

There are studies in the literature recognizing that the cross sectional nature of systemic risk largely results from multivariate tail dependence. Applying principal components analysis and predictive causality tests, Billio et. al. (2010) capture dependence between intermediaries through. Hartmann et al. (2005) use multivariate extreme value theory to estimate the systemic risk in the international banking system using equity data. Similarly, De Jonghe (2010) presents estimates of tail betas for European financial firms as their systemic risk measure.

In line with the idea that non-Gaussianity is essential for marginal modeling and that increasing dependency between common factors is related to tail dependency of returns particularly during financial crises, in this chapter we investigate which model offers a more reliable risk assessment for the Eurozone sovereign market. In particular, we try a set of marginal distributions with different degrees of heaviness in the tail (Gaussian, Student-t, classical tempered stable, normal tempered stable, and α-stable) and different dependence structure (Gaussian copula, multivariate Student-t copula, skewed-t copula, and multivariate normal tempered stable copula). We use ARMA-GARCH model plus heavy-tailed residual distribution the as marginal distribution, showing finding that our model is better than that which uses the normal distribution. Also, we show that the Gaussian copula fails to incorporate dependence structure, and heavy-tailed dependence structure provides better results. In addition, we discuss the parameters indicating heavy-tailedness, and then try to forecast the Greek sovereign debt crisis by combining these parameters. The reason why we focus on Greece is that its rating was downgraded faster than any of the other Eurozone countries.[1]

The rest of this chapter is structured as follows. In Section 3.2, we introduce the data we use. Section 3.3 describes our methodology. The fitting results of the model are presented in Section 3.4. Based on the fitting re-

[1]For example, Standard and Poor's downgraded Greece (Hellenic Republic) to BB+ in April 2010. Greece is the first Eurozone sovereign to receive this non-investment grade rating after 2007.

sult, we consider the heavy-tail property of returns and try to apply them to predict future risk. Our conclusions are provided in the final section.

3.2 Data

Performance indices of bonds measuring total return are constructed and disseminated by major banks. In this chapter, we use selected bond indices published by Barclays Capital Index (previously published by Lehman Brothers).[2] The total return includes both price changes and interest payments. Because they have a long and continuous history, the Barclays Capital major sovereign bond indices are followed by institutional investors and often used by clients as benchmarks. For this reason, we use this bond index for each of our seven countries.

We use both daily and weekly (returns between weekends) frequencies in our analysis. In order to avoid the effects of non-synchronous trading and day-of-week effect, some researchers adopt weekly returns. Because of recent events in the sovereign bond market, however, significant price changes within a week call for the use of daily data in addition to weekly data in order to capture these rapid price movements.

For our analysis, we select seven countries in the Eurozone – France, Germany, Greece, Ireland, Italy, Portugal, and Spain – for the 11-year period from January 2001 to December 2011. The daily data consist of 2,818 observations and the weekly data 573 observations. Each observation is a seven-dimensional vector representing returns of seven countries. We denote observation at time t as $r_t = (r_t^1, \ldots, r_t^7)^T$ where $t = 1, \ldots, 2,818$ in daily or $t = 1, \ldots, 573$ in weekly.

In order to reduce the number of dimension and therefore reduce the amount of computation, we have selected seven countries. The selected countries include both major economic powers in the Eurozone (France, Germany) and peripheral countries in crisis (Greece, Ireland, Italy, Portugal, and Spain) and represent the characteristics of the entire Eurozone during the financial crisis. During the period of analysis, the ratings of France and Germany were not downgraded, while those of the other peripheral countries were downgraded. France was downgraded after the analysis period. The amount outstanding of debt securities of these countries is more than 80% of that

[2] At `http://www.barcap.com/indices`.

of the 17 countries consisting of the entire Eurozone.[3] The correlation coefficient between the weighted daily returns of the entire Eurozone and that of the selected seven countries (in which the weights are proportional to the amount of debt outstanding) is 0.98.

3.3 Methodology

In this chapter, we employ the following methodology:

> Step 1: Fit the ARMA-GARCH model to the return series and then compute the residuals.
>
> Step 2:Fit marginal distribution models to the residual series.
>
> Step 3:Fit a joint copulae to the cumulative distribution function (CDF) of the residuals.

For the backtesting period, the three steps are repeated.

3.3.1 Calculating the residuals from the ARMA-GARCH model

In Step 1 the ARMA-GARCH model is used to compute the residuals because this model captures two stylized facts about financial time series: autoregressive nature of returns and volatility clustering. In particular, we use the ARMA(1,1)-GARCH(1,1) model given by

$$\begin{aligned} r_{t+1}^i &= \mu^i + \phi^i r_t + \psi^i u_t^i + u_{t+1}^i, \\ u_t^i &= \sigma_t^i \varepsilon_t^i, \\ (\sigma_{t+1}^i)^2 &= K^i + \beta^i (\sigma_t^i)^2 + \alpha^i (u_t^i)^2, \end{aligned} \tag{3.1}$$

where ε_t^i is n independent and identically distributed (i.i.d.) random standardized (zero mean and unit variance) variables called a residual. The distribution of residuals is assumed to be the standard Gaussian or the standard Student-t. We apply this model to index returns r_t^i for each country i independently. Correlation between country index returns is not considered here and will be introduced later.

[3] As of July 2012, computed from statistics available from the European Central Bank.

3.3.2 Fitting the marginal distribution to the residuals

As noted in the introduction, Kim et al. (2011) have demonstrated that heavier-tailed innovations improve the ARMA-GARCH fitting. For this reason in Step 2 we fit the following five marginal distributions to the residuals: standard Gaussian, standard Student-t, standard classical tempered stable (CTS), standard normal tempered stable (NTS), and α-stable. With the exception of the standard Gaussian marginal distribution, these distributions have parameters. The estimation of these parameters is done using the maximum likelihood estimation (MLE) method. That is, the estimated parameters are the values that maximize the joint probability of the time series of the residuals obtained from the ARMA-GARCH model. For the CTS, NTS, and α-stable distributions, the residuals are obtained from the ARMA-GARCH fitting using the Student-t distribution as explained in Section 3.3.1.

The Student-t distribution has degree-of-freedom parameter $\nu_i > 0$. As for the standard CTS, the standard NTS, and the α-stable distributions, the most important parameter is $\alpha^i \in (0, 2)$ among all parameters. This is because parameters ν^i and α^i indicate how the distribution differs from the Gaussian. In the case of ν^i, smaller values mean heavier tails than the Gaussian, and the distribution tends to the Gaussian when $\nu \to \infty$. In the case of α^i, smaller values mean heavier tails than the Gaussian, and $\alpha \to 2$ corresponds to the Gaussian.

A goodness-of-fit test must be performed in order to check if these model distributions and fitted parameters are close to the empirical distribution. We employ the Kolmogorov-Smirnov (KS) test and the Anderson-Darling (AD) test for this purpose.[4] The null hypothesis is that the empirical distribution matches the model distribution. The p-value is defined as the probability that the null hypothesis holds. Therefore, if the p-value is less than ϵ $(0 < \epsilon < 1)$, then the null hypothesis is rejected with $1 - \epsilon$ confidence level, suggesting that the empirical distribution is different from the model distribution tested. While both of these tests can be used as a goodness-of-fit test, the AD statistic has an advantage in that it is more sensitive to the tail part of the distribution, a desirable feature in testing heavy-tail distributions.

[4]We use the limiting distribution of these test functions, that is, the distribution for the case $n \to \infty$ where n is the number of observations.

3.3.3 Applying the joint copula to the CDF of the residuals

The degree of capital market integration between sovereign bond markets is an important topic. For this purpose, we describe the relation by analyzing the residuals. We use residuals instead of returns because residuals are independent and identically distributed while returns are not due to volatility clustering and mean reversion. Let $F = F(\varepsilon^1, \ldots, \varepsilon^7)$ be the joint cumulative distribution function of the residuals where the superscript denotes the i-th sovereign bond market ($i = 1, \ldots, 7$). Then, a copula function is defined as follows:

$$\begin{aligned} C(u^1, \ldots, u^7) &= F(\varepsilon^1, \ldots, \varepsilon^7), \\ u^i &= F_i(\varepsilon^i), \end{aligned} \tag{3.2}$$

where $F_i(\varepsilon^i)$ is the marginal CDF of residual ε^i (therefore, variable u^i has a uniform distribution over $(0, 1)$). The copula function can separate the joint distribution F into the copula function C and the marginal part F_i .

As for copula function C, in order to capture tail dependency, we use four different functions: multivariate Gaussian, multivariate Student-t, skewed-t, and multivariate NTS (MNTS). Fitting to these copulae is based on the maximum likelihood method. As for the marginal part F_i, the empirical distribution is the best description of empirical tail-heaviness. Therefore, we first use the empirical marginal distribution in the copula model. However, to see how the marginal mismatch between the theoretical and marginal distributions affect the joint mismatch between them, we also use theoretical marginal distributions in the copula model for comparison purposes. Since they are a natural combination, we use the Gaussian marginal distribution for the Gaussian copula, the Student-t for the multivariate Student-t, the CTS for the skewed-t, and the NTS for the MNTS.

In order to measure the distance from the theoretical copula model and the empirical joint distribution, we use the Cramér-von Mises statistic.[5] This statistic takes the value of zero if the theoretical copula completely matches the empirical one, and gets larger positive values as the difference between the empirical and the model distribution increases.

[5]It is also possible to obtain p-values from the statistic; see Genest et al. (2009)

3.3.4 Backtesting

The three-step methodology described above provides an analysis for a single period. To check the efficiency of a given model, backtesting of each model is required. We conduct backtesting for each observation during the six years from 2006 to 2011. For each observation in this period, we use five-year data prior to the backtested observation day, and then follow the three steps described above. That is, we backtest based on the recent six years using five-year rolling data. By doing so, we obtain a time series of p-values of the null hypothesis for each country and for each model that are used to test whether the empirical distribution matches the given model. Also, we compute a time series of Cramér-von Mises statistics for the multivariate Gaussian and multivariate Student-t copulae.[6]

In our backtesting, we investigate how each model predicts the risk measure endorsed by bank regulators, value at risk (VaR). VaR, which is a kind of out-of-sample interval forecast, is defined as

$$\mathrm{VaR}_\epsilon(R) = -\mathrm{Inf}_x\{x|\mathrm{Prob}(R \leq x) \geq \epsilon\}, \tag{3.3}$$

where R is a random variable representing return and ϵ is a given confidence level. For example, for a 99% VaR, violation against VaR should be realized for only 1% of the total observations.

Therefore, in order to test if VaR violations are within the expected numbers given the confidence level, the number of realized violations is important. In addition, there is a statistical test introduced by Christoffersen (1998) that can be employed. By using the likelihood ratio test, Christoffersen's method tests the null hypothesis that the probability of violations against the given interval forecast (for example, 99% VaR) matches the probability the forecast assumes (for example, 1%). If the test fails, the probability of the event is higher or lower than the given value. An advantage of Christoffersen's method is that it can incorporate the tendency of consecutive occurrence of VaR violations (i.e., VaR violations tends to continue once they occur). This tendency has been observed in markets during periods of financial turmoil.

VaR has, however, a flaw: it does not consider the heavy-tail effect. Even if two distributions have the same VaR, the heavier-tailed distribution is riskier since it may result in a larger loss compared to the less heavier-tailed

[6]We omit the cases of the skewed-t and MNTS copulae, because it is shown in Section 3.4 that the fitting of these copulae is worse than the multivariate Student-t copula.

one. In order to measure this heavy-tail effect, the average VaR (AVaR[7]) can be used. AVaR is expressed mathematically by

$$\mathrm{AVaR}_\epsilon(R) = \frac{1}{\epsilon}\int_0^\epsilon \mathrm{VaR}_p(R)dp. \tag{3.4}$$

The following equation holds if the distribution is continuous:

$$\mathrm{AVaR}(R) = E[-R|-R > \mathrm{VaR}(R)] = 0. \tag{3.5}$$

3.4 Empirical results

In this section, we discuss the results from fitting the ARMA-GARCH model to each country's bond return series. The goodness-of-fit test results for each country's marginal distribution models are shown in Tables 3.1 and 3.2. More specifically, the table reports the p-values based on the KS and AD statistics. If a model has a p-value below ϵ, it means that the model is rejected with $(1-\epsilon)$ confidence. A high p-value does not mean that the model is correct but means that the model is not rejected.

In the daily analysis, the p-values based on the KS statistic are below 5% for the following combinations: Gaussian-Greece, Gaussian-Ireland, Gaussian-Italy, Gaussian-Portugal, CTS-Greece, NTS-Greece, α-stable-Ireland, and α-stable-Italy. The p-values based on the daily data and the AD statistics give similar results; that is, Gaussian-Spain also gives $p < 0.05$ in addition to the combinations in the KS. As a result, the Student-t distribution is the only distribution which is not rejected for any of the seven countries included in this study. For the weekly return analysis, the CTS distribution is not rejected for any of the countries for both the KS and AD statistics.

In Table 3.3, we show the Cramér-von Mises statistics of four different models of joint copulae. The multivariate Student-t model is closest to the empirical distribution. Except for the case between the Gaussian marginal and the empirical marginal distributions under the Gaussian copula, there is no large difference between the empirical and the model cases in the marginal distributions. This shows that the model distributions can capture well the tails in the marginal distribution. The major difference in the Gaussian case may be due to the poor fitting of the marginal Gaussian distribution reported in Tables 3.1 and 3.2.

[7]AVaR coincides with expected shortfall or conditional VaR (CVaR) if return distribution is continuous. See Rachev et al. (2008), p. 210.

Table 3.1: The p-values for each marginal distribution and each country, using KS statistic.

	France	Germany	Greece	Ireland	Italy	Portugal	Spain
Residuals from daily data							
Gaussian	0.40	0.21	0.00	0.00	0.03	0.00	0.06
Student-t	0.79	0.60	0.26	0.82	0.80	0.54	0.99
CTS	0.97	0.94	0.01	0.21	0.47	0.11	0.93
NTS	0.91	0.99	0.00	0.10	0.34	0.14	0.72
Stable	0.73	0.83	0.38	0.00	0.03	0.55	0.84
Residuals from weekly data							
Gaussian	0.08	0.02	0.06	0.12	0.23	0.42	0.06
Student-t	0.08	0.02	0.25	0.83	0.50	0.78	0.21
CTS	0.19	0.09	0.34	0.63	0.45	0.56	0.79
NTS	0.08	0.02	0.27	0.85	0.37	0.51	0.35
Stable	0.21	0.09	0.11	0.52	0.00	0.80	0.11

Table 3.2: The p-values for each marginal distribution and each country, using AD statistic.

	France	Germany	Greece	Ireland	Italy	Portugal	Spain
Residuals from daily data							
Gaussian	0.17	0.11	0.00	0.00	0.00	0.00	0.00
Student-t	0.90	0.80	0.05	0.55	0.38	0.33	0.89
CTS	0.99	0.98	0.02	0.48	0.31	0.15	0.94
NTS	0.98	0.99	0.00	0.24	0.17	0.12	0.73
Stable	0.70	0.86	0.10	0.00	0.00	0.37	0.84
Residuals from weekly data							
Gaussian	0.29	0.15	0.03	0.12	0.17	0.13	0.17
Student-t	0.30	0.16	0.14	0.57	0.47	0.44	0.38
CTS	0.50	0.33	0.15	0.72	0.61	0.40	0.67
NTS	0.30	0.16	0.10	0.64	0.51	0.31	0.46
Stable	0.47	0.34	0.11	0.42	0.02	0.47	0.35

Table 3.3: Cramér-von Mises statistics for each joint copula.

Marginal dist.	Copula distribution Gaussian	Multivariatet	Skewed-t	MNTS
		Residuals from daily data		
Empirical	7.92	4.53	9.96	10.46
Model	Gaussian	t	CTS	NTS
	11.75	4.19	9.87	10.48
		Residuals from weekly data		
Empirical	1.53	0.94	2.07	2.09
Model	Gaussian	t	CTS	NTS
	1.99	1.30	2.05	2.11

In Tables 3.4 and 3.5, the number of backtesting days on which the model distribution is rejected is reported, as for each model distribution and country, using both the KS and AD statistics and 95% confidence level. From these tables, it can be seen that the Gaussian and α-stable distributions are rejected in most countries and backtesting days. With daily residuals, the Student-t and the CTS distributions have the least number of rejections. With weekly residuals, the CTS distribution has the least number of rejections, followed by the Student-t and the NTS distributions. From these results, a fair conclusion is that CTS distribution has the best fit, while the fitting of Student-t and NTS distributions is comparable to the CTS distributions.

In Figure 3.1, the time series of the Cramér-von Mises statistics for the Gaussain and Student-t copulae for the seven countries is presented. These two copulae are selected since based on the results reported on Table 3.3, they generally give a better fit. The multivariate Student-t model is generally less than the multivariate Gaussian distribution.

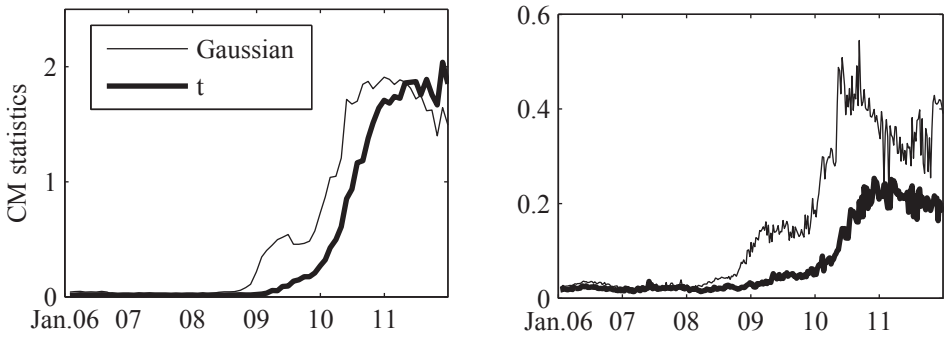
In Tables 3.6 and 3.7, we show the forecasting results based on each model. The tables show (1) the number of violations of VaR, (2) p-values for the conditional Christoffersen test assuming the tendency of consecutive VaR violations, (3) p-values for the unconditional Christoffersen test ignoring the consecutive tendency of VaR violations, and (4) the average loss minus AVaR in the case of VaR violations. For the Christoffersen tests, for example, p-values less than 5% suggest that the probability of VaR violation is not 1% with 95% confidence level.

Table 3.4: Number of rejected days in backtesting of each marginal model, using KS statistic.

	France	Germany	Greece	Ireland	Italy	Portugal	Spain
				Residuals from daily data			
Gaussian	4	4	452	523	43	350	69
Student-t	1	2	5	0	0	0	1
CTS	0	2	6	0	0	0	0
NTS	0	2	27	0	0	0	0
Stable	45	37	97	24	23	37	66
				Residuals from weekly data			
Gaussian	6	3	35	66	2	19	3
Student-t	2	4	6	15	1	9	3
CTS	0	0	0	0	0	1	0
NTS	2	4	6	13	0	9	3
Stable	1	1	7	5	3	4	1

Table 3.5: Number of rejected days in backtesting of each marginal model, using AD statistic.

	France	Germany	Greece	Ireland	Italy	Portugal	Spain
				Residuals from daily data			
Gaussian	3	1	513	544	199	546	108
Student-t	2	2	21	0	0	0	1
CTS	1	2	21	0	0	0	0
NTS	1	2	40	0	0	0	0
Stable	48	40	95	23	24	33	65
				Residuals from weekly data			
Gaussian	0	0	28	38	0	4	0
Student-t	0	0	1	1	0	0	0
CTS	0	0	0	0	0	1	0
NTS	0	0	0	1	0	1	0
Stable	0	0	2	0	0	0	0



(a) Daily

(b) Weekly

Figure 3.1: Time series of Cramér-von Mises statistics. In most observations, the multivariate Student-t model gives lower statistic values than the multivariate Gaussian distribution, which means that the multivariate Student-t model has a better fit than the multivariate Gaussian. Note that daily results are computed at each month-end in order to reduce the computational load.

Table 3.6: VaR violations, Christoffersen test, and average loss over AVaR: Daily result.

	France	Germany	Greece	Ireland	Italy	Portugal	Spain
	Violations of 99% VaR (for 1,536 daily observations, 2006-2011)						
Gaussian	22	16	41	34	30	35	26
Student-t	17	15	38	32	25	26	21
CTS	17	15	34	32	21	25	21
NTS	17	15	33	32	21	25	22
Stable	20	16	38	35	25	30	25
	Conditional Christoffersen test						
Gaussian	0.07	0.55	0.00	0.00	0.00	0.00	0.01
Student-t	0.46	0.58	0.00	0.00	0.01	0.00	0.12
CTS	0.46	0.58	0.00	0.00	0.12	0.00	0.12
NTS	0.46	0.58	0.00	0.00	0.12	0.00	0.07
Stable	0.18	0.55	0.00	0.00	0.02	0.00	0.01
	Unconditional Christoffersen test						
Gaussian	0.11	0.87	0.00	0.00	0.00	0.00	0.01
Student-t	0.68	0.93	0.00	0.00	0.02	0.01	0.17
CTS	0.68	0.93	0.00	0.00	0.17	0.02	0.17
NTS	0.68	0.93	0.00	0.00	0.17	0.02	0.11
Stable	0.26	0.87	0.00	0.00	0.02	0.00	0.02
	Average loss over AVaR in case of violation against AVaR						
Gaussian	0.01	0.05	0.54	0.24	0.16	0.45	0.10
Student-t	0.00	0.02	0.34	0.11	0.15	0.51	0.09
CTS	-0.02	0.00	0.34	0.11	0.16	0.52	0.08
NTS	-0.02	0.00	0.37	0.08	0.15	0.50	0.05
Stable	-0.02	0.01	-0.01	-0.09	0.09	0.28	0.01

Table 3.7: VaR violations, Christoffersen test, and average loss over AVaR: Weekly result.

	France	Germany	Greece	Ireland	Italy	Portugal	Spain
	Violations of 99% VaR (for 311 weekly observations, 2006-2011)						
Gaussian	4	4	13	10	10	13	6
Student-t	5	4	10	10	10	11	6
CTS	5	3	8	9	8	7	5
NTS	5	4	10	12	10	10	7
Stable	5	4	10	10	9	9	5
	Conditional Christoffersen test						
Gaussian	0.56	0.56	0.00	0.00	0.00	0.00	0.12
Student-t	0.28	0.56	0.00	0.00	0.00	0.00	0.12
CTS	0.28	0.80	0.02	0.00	0.01	0.05	0.28
NTS	0.28	0.56	0.00	0.00	0.00	0.00	0.28
Stable	0.28	0.56	0.00	0.00	0.00	0.00	0.28
	Unconditional Christoffersen test						
Gaussian	0.63	0.63	0.00	0.00	0.00	0.00	0.14
Student-t	0.32	0.63	0.00	0.00	0.00	0.00	0.14
CTS	0.32	0.95	0.02	0.01	0.02	0.06	0.32
NTS	0.32	0.63	0.01	0.00	0.01	0.00	0.32
Stable	0.32	0.63	0.00	0.00	0.01	0.01	0.32
	Average loss over AVaR in case of violation against AVaR						
Gaussian	-0.03	-0.04	1.36	0.80	0.30	0.71	0.36
Student-t	0.00	-0.06	1.33	0.47	0.08	0.48	0.20
CTS	-0.04	-0.10	1.39	0.49	0.06	0.79	0.25
NTS	-0.02	-0.09	1.30	0.39	0.04	0.57	0.24
Stable	-0.03	-0.07	-2.14	-0.21	-0.11	-0.53	0.13

The number of violations always exceeds the expected value – 15 for daily data and three for weekly data. Comparing models, however, the least number of violations occurs for the CTS model, while the largest is for the Gaussian model.

Based on Christoffersen's test with 95% confidence, for the daily analysis the null hypothesis of 1% VaR violation is rejected for five countries for the Gaussian and α-stable distributions both in the conditional and unconditional cases. Likewise, in the daily analysis, the null hypothesis is rejected for four countries in the case of the Student-t distribution and for three countries in the cases of the CTS and NTS distributions. Turning to the weekly analysis, the null hypothesis is rejected for four countries for all distributions except the CTS distribution, while it is rejected for three countries for the CTS distribution. Therefore, our findings suggest that VaR forecasts based on the CTS model generate less errors than the other models.

There is no big difference in the average loss minus AVaR among models. which should be zero according to equation (3.5). This means that the expected loss in the case of VaR violations does not differ based on the model used. However, Gaussian distribution has a tendency to have a positive deviation from zero, indicating that this model has a tendency to predict lower losses, which is an undesirable feature in return prediction.

Overall, based on the empirical evidence described above, the CTS distribution seems to be the best-fit model, while the Gaussian distribution does not have sufficient empirical support to use it as a marginal model in modeling sovereign bond returns. Although not the best-fit models, Student-t, NTS, and α-stable models have more empirical support for their use than the Gaussian model.

3.5 Tail heaviness and its application to risk prediction

3.5.1 Tail parameters

The evidence presented in Section 3.4 suggests that the residual distribution is not Gaussian. In order to measure how different the actual distribution is from the Gaussian distribution, we will look at the estimated parameters to assess the deviation from that distribution.

There exist two parameters measuring non-Gaussianity: (1) the degree

Table 3.8: Tail parameters for the entire period.

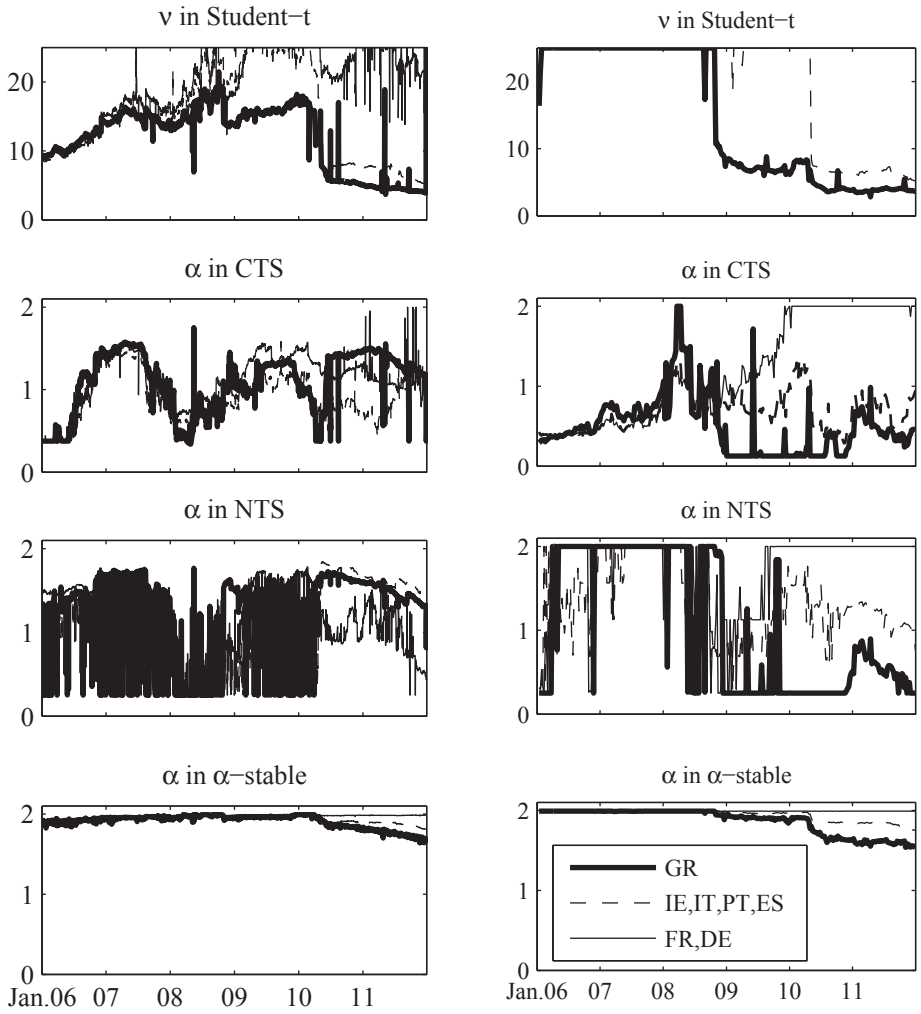
	(param.)	France	Germany	Greece	Ireland	Italy	Portugal	Spain
				Residuals from daily data				
Student-t	nu	15.22	13.89	4.94	5.96	7.46	6.84	7.82
CTS	alpha	1.03	1.20	1.30	0.84	1.46	0.35	1.30
NTS	alpha	0.52	1.42	1.44	1.39	1.67	1.67	1.66
Stable	alpha	1.96	1.95	1.82	1.85	1.89	1.85	1.89
				Residuals from weekly data				
Student-t	nu	>200	>200	4.24	6.05	8.10	5.84	8.45
CTS	alpha	1.29	1.16	0.20	<0.13	<0.13	0.38	<0.13
NTS	alpha	2.00	2.00	<0.25	<0.25	<0.25	1.04	1.79
Stable	alpha	1.99	1.99	1.80	1.84	1.93	1.86	1.92

Note: Inequality (> 200, < 0.13 and < 0.25) means that computation is stopped due to hitting boundary.

of freedom ν in the Student-t distribution, and (2) parameter α in the CTS, NTS, and α-stable distributions. These parameters for the entire period are shown in Table 3.8, and the time series is shown in Figure 3.2. In Figure 3.2, we create three groups of the seven countries in the analysis based on a country's credit rating: (1) Greece, (2) Ireland, Italy, Portugal, and Spain, and (3) France and Germany. The first group is the country whose rating was downgraded most rapidly, the second group consists of the countries whose ratings were downgraded during the analysis period, and the third group includes the countries which were not downgraded during the analysis period. The vertical scale of ν in the Student-t is cut at $\nu = 25$. The weekly time series of ν for France and Germany always exceeds 25. The vertical scale of α in the CTS, NTS, and α-stable is in the range of α, (0,2).

If degree of freedom ν decreases, the tail of the Student-t distribution becomes heavier. From Table 3.8, peripheral countries such as Greece have a tendency to have heavier tails. From Figure 3.2 it can be seen that the heaviness of the tail of the peripheral countries measured in ν increases after the middle of 2010, the time when the Greek sovereign debt problem emerged.

The parameter α for the CTS, NTS, and α-stable models takes the value between 0 and 2, with α equal to 2 corresponding to the Gaussian case. From the table, α is less than 2 for all countries. In Figure 3.2, α of the



(a) Daily

(b) Weekly

Figure 3.2: Time series of tail-heaviness parameters. The distributions have heavier tails as parameters ν and α decrease. In the Student-t and α-stable models, these parameters in peripheral countries such as Greece decrease after the Greek crisis. In the CTS model, the parameter α in Greece is much less than 2 in almost all observations, showing that the model has much heavier tails than the Gaussian.

peripheral countries for the CTS model is always much less than 2, and α of these countries for the α-stable model decreases from 2 subsequent to the commencement of the Greek crisis.

3.5.2 VaR spread

As can be seen in the case of Greece, the heaviness of the tail has a tendency to increase when a crisis starts. Therefore, there is the potential to use this parameter as a signal of forthcoming turmoil in a market.[8]

In order to relate the heaviness of a tail to risk, we compare the VaR based on both Gaussian and non-Gaussian distributions. If the market is under duress, heaviness of its tail is likely to increase as more attention will be paid toward tail risk, and the difference between the VaR based on Gaussian and the VaR based on non-Gaussian becomes larger. We refer to this difference as VaR spread.

In Figure 3.3, we show the realized return and daily VaR spread for Greek debt around May 2010 when Greek government bond prices dropped and then recovered sharply.[9] It can be seen that the VaR spread exhibited a tendency to increase since the beginning of April 2010, reflecting increased anxiety regarding the tail risk for Greek government bonds. In fact, Greek bond returns began oscillating in larger amplitude since then.

In the figure, we show two VaR spreads: Gaussian-Student-t and Gaussian-CTS. For the latter, the spread is larger than the former. This is because the heaviness of the return distribution is better captured by the CTS distribution.

[8] A commonly used measure to proxy for market tension is volatility. However, it is sometimes a lagging indicator which goes behind the market turmoil. The advantage of VaR spread is that it captures the risk which has not appeared but is expected to appear in the near future.

[9] Reflecting bad news such as unreported debt, downgrading, strikes and discord of international rescue packages, the Greek bond market had been under considerable tension since the end of 2009. The total return index of Greek bonds dropped by 14% for five consecutive business days from April 30 to May 7, 2010 and then recovered by more than 26% on May 10, 2010.

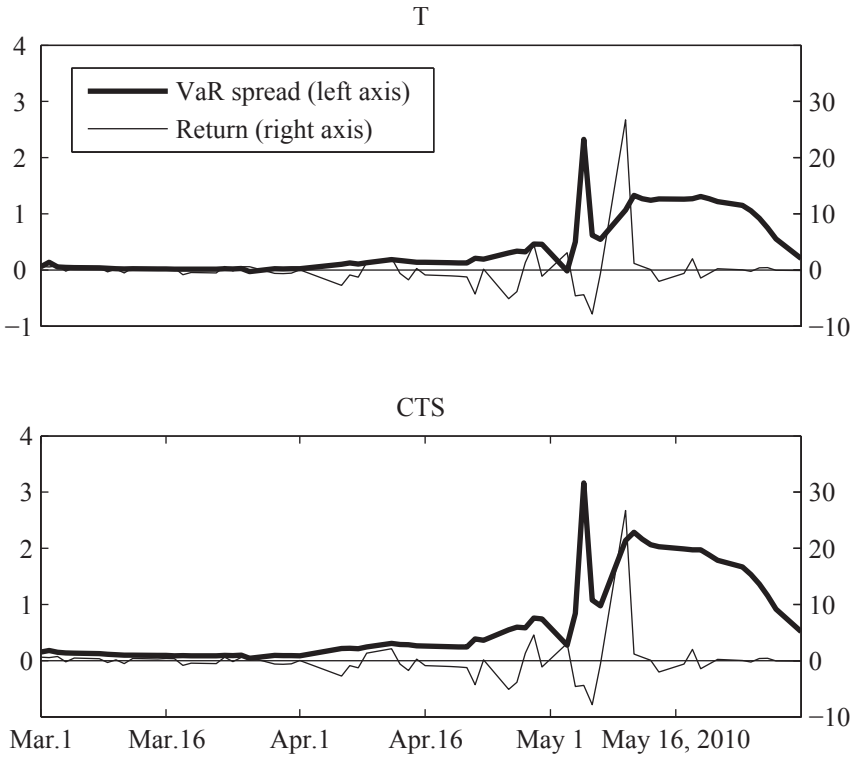


Figure 3.3: Daily VaR spread of Greece from March to May 2010. Units are percentages in both VaR spreads and returns. VaR spread is defined as the difference between the VaR based on the Gaussian model and the VaR based on the non-Gaussian model. The daily VaR spread exhibited a tendency to increase since the beginning of April 2010, reflecting increased anxiety regarding the tail risk for Greek government bonds.

3.6 Conclusion

In this chapter, we use alternative statistical models in an attempt to identify the best performing model to describe sovereign bond returns for seven governments in the Eurozone. These returns exhibit autocorrelation effect and volatility clustering effect. We apply the ARMA(1,1)-GARCH(1,1) model to returns for each of the seven countries, and apply five marginal distribution models to ARMA-GARCH generated residuals. We find that Gaussian distribution cannot describe the empirical distribution well, suggesting that heaviness of tails in return distributions cannot be ignored in model development. Our empirical analysis suggests that the CTS distribution is the distribution that most closely matches the empirical. As for the joint distribution, the multivariate Student-t distribution gives the best fit to the empirical one. The heaviness of the return distribution changes over time, reflecting future risk. We compute VaR based on different heavy-tailed models, and compute their spread over Gaussian VaR. These spreads signals forthcoming risk in the case of the financial crisis in Greece around May 2010.

Chapter 4

ICA-GARCH Analysis of Joint Default Probability of Eurozone Sovereigns

In this chapter, we provide an empirical application of the independent component analysis (ICA) based on the Eurozone sovereign CDS. First, we fit the data to the various ICA models, and compare their goodness-of-fitting based on the log-likelihood, the Akaike information criterion and the Bayesian information criterion. The data is the sovereign CDS premiums of 11 Eurozone countries (Austria, Belgium, Finland, France, Germany, Greece, Ireland, Italy, Netherlands, Portugal, and Spain) from 2007 to 2013. The result shows that the ICA model is as good as the direct modeling of joint distribution via copula functions. Especially, the ICA model with the GARCH postprocess model is as better as the ARMA-GARCH model with the copula postprocess model. Based on this result, we introduce a structural sovereign default model with CDS premium, and compute the joint default probability of sovereigns based on the ICA models and copula models. Accordingly, the ICA model reports higher joint probability of defaults compared to the copula models. Consequently the instability of Eurozone sovereigns shown by ICA has increased before 2010, in which year the Eurozone crisis occurred, and is generally larger than that shown by copula model.

4.1 Introduction

Modeling of a multivariate random variable is an important topic for risk analysis. Movements of a specific asset is affected by that of another asset. Especially, in the period of turmoil, we can observe the increased dependency of different assets – not only among the same asset class, but across the asset classes, markets, or countries. In order to avoid a precipitous decline in asset value, it is important to capture the dependency structure of variables.

For this purpose, a joint distribution of all variables mathematically provides all the information of the dependency. For example, the multivariate Gaussian distribution can express all the pairwise correlations. If the number of variables is limited to two or three, it is efficient and exact to use the Gaussian joint distribution, as long as the model is correct. However, as the number of variable increases, the number of pairwise correlations increases rapidly (quadratically), creating the problem of parameter estimation. In addition, in the modeling of asset returns, such a number of parameters is usually unnecessary, since assets are characterized by fewer number of parameters, such as industry or capitalization.

For these reasons, it becomes an important problem to find an appropriate representation of a high-dimensional random vector. Such a representation should be in the way that we can easily understand and see the nature of the variable. Especially, if the representation is given by a linear transformation, then it becomes quite easy to interpret and compute the variable. Conventionally, factor analysis and principal component analysis (PCA) are used for this purpose.

Independent component analysis is a novel method to obtain the representation of a random vector. It is a linear transformation similar to PCA. However, different from PCA, it discovers the most independent variables of the original variables based on the information theory. It is a powerful property for modeling and interpretation. The research about ICA and its applications has been rapidly developing recently (Hyvärinen et al., 2001).

The first application of ICA to finance dates back to 1997, as long as we investigate (Back and Weigend, 1997). The authors of the paper applied ICA to the daily returns, finding that these returns are separated into frequent and non-frequent components. Mălăroiu et al. (2000) applied ICA to a preprocess of AR model. They showed the forecast of ten foreign exchange rates can be predicted more accurately using ICA compared to the case without ICA. More recently, Kumiega et al. (2011) applied ICA to the S&P sectorial

indexes consisted of 10 time series, and they showed that the independent components can be interpreted as the proxy of the energy index, financial sector index, and the index of other sectors.

For the modeling of return series, it is well known that the series has volatility clustering. In order to model it, recent works combine ICA with stochastic volatility or GARCH models. Especially, Garcia-Ferrer et al. (2008), Xu and Wirjanto (2009) and Kumiega et al. (2011) investigate the combination of ICA and GARCH

Default correlation is an important field of research into which dependency analysis should be applied. Especially, since the financial and sovereign crisis from 2008, the analysis of joint default of banks or sovereigns has been an important topic; for example, Ang and Longstaff (2013) and Caporin et al. (2013). In this chapter, we are trying to explain the joint default probability based on the CDS data and ICA. We adopt the structural model of sovereign defaults based on the CDS premium proposed by Lucas et al. (2011). Combining the CDS model and their definition of defaults, we can compute the joint distribution of defaults.

This chapter is constructed as follows. Section 4.2 explains the data. Section 4.3 explains the ICA models and show the fitting results based on the data. Section 4.4 explains both how to construct a joint default model from the CDS data and how to define the statistics indicating the level of joint default. The results are also shown in the section. Section 4.5 is devoted to the conclusion.

4.2 Data

The CDS premium used in this research is the maturity of five-years and denominated by USD. The data is sourced from the CMA and downloaded from the Bloomberg terminal. Its period is from April 30, 2007 to March 8, 2013. We use daily data and convert them into daily log returns. We use 11 Eurozone sovereigns: Austria (AT), Belgium (BE), Finland (FI), France (FR), Germany (DE), Greece (GE), Ireland (IE), Italy (IT), Netherlands (NE), Portugal (PT), and Spain (ES).

During the data period, the collapse of Lehman Brothers and the Eurozone sovereign crisis occurred. It is expected that the market conditions are not uniform during the period. Therefore, we divide the entire period into three subperiods:

Table 4.1: Data periods and countries for each period. All the eleven countries are included by Period 2: Austria Belgium, Finland, France, Germany, Greece, Ireland, Italy, Netherland, Portugal, and Spain. Some of them are excluded from the other periods.

#	Period	Days	# of countries	Excluded countries
1	Apr. 30, 2007 to Jan. 28, 2009	458	9	Finland and Ireland
2	Jan. 29, 2009 to Sept. 9, 2011	682	11	None
3	Sept. 12, 2011 to Mar. 8, 2013	390	10	Greece
4	Apr. 30, 2007 to Mar. 8, 2013	1530	8	Finland, Greece, Ireland

1. From April 30, 2007 to January 28, 2009,

2. From January 29, 2009 to September 9, 2011, and

3. From September 2011 to March 8, 2013.

We decided these subperiods mainly for the data availability. For example, we identified Period 1 since the data of Finland and Ireland is not available during the period. However, we also give an interpretation for this decision. Period 1 corresponds to the period prior to the Lehman Brothers collapse, and the level of the bid-ask spread was much lower than that of later periods (For example, see Calice et al., 2012). So it is natural to separate Period 1 from the others.

On the other hand, Period 3 is decided since the continuity of Greek CDS premium is lost during the period. In the third quarter of 2011, the Greek premium increased to the level that conventional quotation is impossible and 'upfront basis' quotation of premium is required (CMA, 2011), which is the fundamental reason we set the end of Period 2 in September 2011.

For these reasons, we define the three periods above, and some of the countries are excluded from 11 countries. However, as a result, the length of data is shortened. In order to mitigate this disadvantage, we also define Period 4 from the beginning and to the end of the data. During this period, three countries (Finland, Greece and Ireland) are excluded due to the data unavailability. The summary of the periods and the countries is shown in Table 4.1.

4.3 CDS return model based on ICA

4.3.1 ICA model

Our major interest is to apply the ICA model and ICA-GARCH model into the CDS. Let $\mathbf{r}_t = (r_{1t}, \ldots, r_{dt})$, $t = 1, \ldots, T$ denote the log-return of d sovereign CDS premiums on day t. The independent component analysis (ICA) converts $\mathbf{r}_t$ into $\mathbf{s}_t$ as

$$\mathbf{r}_t = \mu + A\mathbf{s}_t, \tag{4.1}$$

$$\mathbf{s}_t = (s_{1t}, \ldots, s_{qt})^{\mathrm{T}}\text{: independent components,} \tag{4.2}$$

$$\mathrm{var}_t \mathbf{s}_{tj} = 1, \quad j = 1, \ldots, q \tag{4.3}$$

$$\mu\text{: constant vector, and} \tag{4.4}$$

$$A\text{: } d \times q\text{-constant matrix.} \tag{4.5}$$

Parameter μ is set to the sample mean $\hat{\mu} = \Sigma_t r_t / T$. Matrix A is estimated by the fastICA algorithm. Hereafter, for simplicity, we assume $d = q$.

The series of $\{s_{jt}\}_{1 \le t \le T}$, is referred to as a *independent component* (IC). The sign of an IC is arbitrary, so we set it as it shows negative skewness. Due to the definition of the ICA, we expect that time series $\{s_{jt}\}_t$, $j = 1, \ldots, d$, are independent and standardized, that is,

$$\mathrm{cov}_t \mathbf{s} = I_q, \tag{4.6}$$

$$f_{\mathbf{s}t}(\mathbf{s}_t) = f_{1t}(s_{1t}) f_{2t}(s_{2t}) \cdots f_{qt}(s_{qt}), \tag{4.7}$$

where $f_{\mathbf{s}t}(\mathbf{s}_t)$ and $f_{jt}(s_{jt})$ are the PDF of $\mathbf{s}_t$ and s_{jt}, respectively.

We still have to model the marginal distribution of s_{jt}. In our analysis, we define Model 1 by assuming that they have a Gaussian distribution. Also, Model 2 assumes that s_{jt} has a standardized Student-t distribution with degree-of-freedom (dof) ν_j.[1]

In Models 1 and 2, the PDFs of $\mathbf{S}_t$ and S_{jt} are independent of t. Let $f_{\mathbf{s}}(\mathbf{S}_t)$ and $f_j(S_{jt})$ denote their PDFs, respectively. Due to eq.(4.1), we can

[1] A Student-t distribution multiplied by factor $\sqrt{(\nu-2)/\nu}$ in order to set its variance unity.

define the PDF of $\mathbf{r}_t$, $f_{\mathbf{r}}(\mathbf{r}_t)$ from $f_{\mathbf{s}}(\mathbf{s}_t)$ and $f_j(s_{jt})$, as

$$f_{\mathbf{r}}(\mathbf{r}_t) = |\frac{\partial \mathbf{s_t}}{\partial \mathbf{r_t}}| f_{\mathbf{s}}(\mathbf{s}_t) \tag{4.8}$$

$$= \frac{1}{|\det A|} f_1(s_{1t}) \cdots f_d(s_{dt}). \tag{4.9}$$

By taking the logarithm of eq.(4.9) and summing them as for $t = 1, \ldots, T$, we obtain the log-likelihood of ICA model.

4.3.2 ICA-GARCH model

Since ICs are a linear transformation of returns, ICs are also expected to inherit the properties of serial correlation and volatility clustering, which returns usually have. In order to model volatility clustering, we apply GARCH(1,1) model to s_{jt}:

$$s_{jt} = \sigma_{jt}\varepsilon_{jt}, \tag{4.10}$$

$$\sigma^2_{j,t+1} = \omega_j + a_j s^2_{jt} + b_j \sigma^2_{jt}. \tag{4.11}$$

ε_{jt} is a random variable whose variance is unity. We use the Gaussian distribution in Model 3 and the standardized Student-t distribution in model 4.

In addition, since returns have skewness, ICs are also expected to have skewness. In order to express this property, we can use EGARCH(1,1) model (Nelson, 1991) as Model 5 and GJR-GARCH(1,1) model (Glosten et al., 1993) as Model 6. These models replaces eq.(4.11) with

$$\log \sigma^2_{jt} = \omega_j b_j \log \sigma^2_{jt} + a_j \varepsilon_{jt} + c_j(|\varepsilon_{jt} - \mathrm{E}\,|\varepsilon_{jt}|), \tag{4.12}$$

in EGARCH and

$$\sigma^2_{jt} = \omega_j + (a_j - 1_{s_{jt}<0} c_j) s^2_{jt} + b_j \sigma^2_{jt}. \tag{4.13}$$

in GJR-GARCH.

Let us consider the likelihood of these models. In these models, the PDF of $s_{j,t+1}$ is conditional to the data up to time t. Let $f_{j,t+1|t}(s_{j,t+1})$ denote the conditional PDF. Therefore the joint PDF $f_j(s_{j1}, s_{j2}, \ldots, s_{jT})$ is represented as

$$f_j(s_{j1}, s_{j2}, \ldots, s_{jT}) = f_{j1}(s_{j1}) f_{j2|1}(s_{j2}) \cdots f_{jT|T-1}(s_{jT}). \tag{4.14}$$

where $f_{jt|t-1}(s_{jt})$ is the PDF of s_{jt} conditional to the information up to time $t-1$. The parameters in eqs.(4.11), (4.12), and (4.13) are obtained by maximizing (4.14).

The PDF of $\mathbf{r}_{t+1}$ is also conditional to the past data up to time t. The distribution of $\mathbf{r}_{t+1}$ is

$$f_{\mathbf{r},t+1|t}(\mathbf{r}_{t+1}) = \frac{1}{|\det A|} f_{\mathbf{s},t+1|t}(\mathbf{s}_{t+1}). \tag{4.15}$$

Therefore, the PDF of all $\mathbf{r}_t$ is

$$f(\mathbf{r}_1, \ldots, \mathbf{r}_T) = \prod_{t=1}^{T} f_{\mathbf{r}t|t-1}(\mathbf{r}_t) \tag{4.16}$$

$$= \frac{1}{|\det A|^T} \prod_{t=1}^{T} f_{\mathbf{s}t|t-1}(\mathbf{s}_t) \tag{4.17}$$

$$= \frac{1}{|\det A|^T} \prod_{t=1}^{T} \prod_{j=1}^{q} f_{jt|t-1}(s_{jt}) \tag{4.18}$$

$$= \frac{1}{|\det A|^T} \prod_{j=1}^{q} f_j(s_{j1}, s_{j2}, \ldots, s_{jT}), \tag{4.19}$$

where $f_{\mathbf{r}1|0}(\mathbf{r}_1) = f_{\mathbf{r}1}(\mathbf{r}_1)$ and $f_{j1|0}(s_{j1}) = f_{j1}(s_{j1})$. From eqs.(4.14) and (4.19), we can define the model PDF and likelihood.

4.3.3 AR-ICA-GARCH model

It is known that return series usually has a weak autocorrelation. In order to express this effect, we also try to use AR(1) or ARMA(1,1) models. Let us consider to replace eq.(4.1) with

$$r_{i,t+1} = \mu_i + \phi_i r_{it} + u_{i,t+1}, \quad i = 1, \ldots, q, \tag{4.20}$$

and define $(u_{i1}, \ldots, u_{iq})^{\mathrm{T}} = A\mathbf{S}$, then we obtain an AR(1)-ICA model. We define Model 7 as the AR-ICA-GARCH model and Model 8 as the AR-ICA-EGARCH model. Likewise, let us consider to replace eq.(4.1) with

$$r_{i,t+1} = \mu_i + \phi_i r_{it} + \psi_i u_{it} + u_{i,t+1}, \quad i = 1, \ldots, q, \tag{4.21}$$

then we obtain an ARMA(1,1)-ICA model. We define Models 9 and 10 as the ARMA-ICA-GARCH and the ARMA-ICA-EGARCH models, respectively. The computation of likelihood is similar to eq.(4.19).

4.3.4 ARMA-GARCH and copula model

We use the ARMA-GARCH and copula model as a reference case. It is frequently used in order to capture return autocorrelation and volatility clustering. First, we use the ARMA-GARCH model for each marginal distribution:

$$\begin{aligned} r^i_{t+1} &= \mu^i + \phi^i r_t + \psi^i u^i_t + u^i_{t+1}, \\ u^i_t &= \sigma^i_t \varepsilon^i_t, \\ (\sigma^i_{t+1})^2 &= K^i + \beta^i (\sigma^i_t)^2 + \alpha^i (u^i_t)^2, \end{aligned} \tag{4.22}$$

where ε^i_t is n independent and identically distributed (i.i.d.) random standardized (zero mean and unit variance) variables called a residual. Next, in order to capture the joint distribution, we assume that the residual vector $(\varepsilon^i_t, \ldots, \varepsilon^i_t)^{\mathrm{T}}$ obeys copula dependency. Models 11 and 12 use the Gaussian distribution for ε^i_t, while Models 12 and 13 use the standardized Student-t distribution. Models 11 and 13 use the Gaussian copula for the dependency structure of ε, while Models 12 and 14 use the Student-t copula.

4.3.5 Empirical results of fitting

In summary, we consider 14 models shown in Table 4.2. The log-likelihood for each model and each period and its ranking among the 14 models are shown in Table 4.3. Also, the values of and the rankings based on Akaike and Bayesian information criteria (AIC and BIC) are shown in Tables 4.4 and 4.5.

According to these tables, it can be seen that Model 14 (ARMA-GARCH-T and T copula model) is almost always the best in almost all periods and measures (likelihood, AIC or BIC). The second best models seems to be Models 4 or 5 (ICA-GARCH or ICA-EGARCH models) according to the likelihood and the AIC, and seems to be Model 13 (ARMA-GARCH-T and Gaussian copula model) according to the BIC.

In order to see the differences of these best models (Models 4, 5, 13 and 14) in detail, we then do a rolling analysis. For every five day, we pick up the 252 consecutive days prior to the day, and fit the models to the data of these 252 days. The result is shown in Figures 4.1 and 4.2. Figure 4.1 is as for 11 countries and Figure 4.2 is as for 8 countries. Note that the former includes Greece but the latter does not. From these figures, the order of goodness of these four models depends on time. For example, based on Figure 4.1,

Table 4.2: The list of models we consider. For details, see the main text.

#	Model description
1	ICA-Gaussian
2	ICA-T
3	ICA-GARCH(1,1)-Gaussian
4	ICA-GARCH(1,1)-T
5	ICA-EGARCH(1,1)-T
6	ICA-GJR(1,1)-T
7	AR(1)-ICA-GARCH(1,1)-T
8	AR(1)-ICA-EGARCH(1,1)-T
9	ARMA(1,1)-ICA-GARCH(1,1)-T
10	ARMA(1,1)-ICA-EGARCH(1,1)-T
11	ARMA(1,1)-GARCH(1,1)-Gaussian with Gaussian copula
12	ARMA(1,1)-GARCH(1,1)-Gaussian with T copula
13	ARMA(1,1)-GARCH(1,1)-T with Gaussian copula
14	ARMA(1,1)-GARCH(1,1)-T with T copula

Table 4.3: The list of log-likelihood for each model and period. The numbers in parentheses are the ranking of goodness-of-fitting, in which 1 shows the best fit and 14 the worst. According to these rankings, Model 14 is the best, and Models 4-5 are the second best.

Model	Period								Rank
#	1		2		3		4		ave.
1	4,113	(14)	16,145	(14)	9,631	(14)	17,457	(14)	14.00
2	6,160	(10)	17,020	(10)	9,966	(10)	23,991	(10)	10.00
3	4,762	(13)	16,830	(12)	9,805	(12)	22,056	(12)	12.25
4	6,451	(6)	17,323	(4)	10,072	(4)	24,964	(4)	4.50
5	6,526	(3)	17,337	(2)	10,080	(1)	25,004	(2)	2.00
6	6,464	(5)	17,334	(3)	10,075	(2)	24,970	(3)	3.25
7	6,413	(7)	17,176	(7)	10,038	(8)	24,737	(7)	7.25
8	6,466	(4)	17,191	(6)	10,044	(7)	24,758	(6)	5.75
9	6,335	(9)	17,084	(9)	10,046	(6)	24,527	(9)	8.25
10	6,385	(8)	17,095	(8)	10,055	(5)	24,529	(8)	7.25
11	4,774	(12)	16,721	(13)	9,770	(13)	21,384	(13)	12.75
12	4,925	(11)	16,875	(11)	9,828	(11)	22,408	(11)	11.00
13	7,193	(2)	17,268	(5)	10,006	(9)	24,927	(5)	5.25
14	7,528	(1)	17,434	(1)	10,074	(3)	25,813	(1)	1.50

Table 4.4: The list of AIC for each model and period. The numbers in parentheses are the ranking of goodness-of-fitting, in which 1 shows the best fit and 14 the worst. According to these rankings, Model 14 is the best, and Models 4-5 are the second best.

Model	Period								Rank
#	1		2		3		4		ave.
1	-8,064	(14)	-32,049	(14)	-19,062	(14)	-34,787	(14)	14.00
2	-11,996	(10)	-33,557	(10)	-19,533	(10)	-47,726	(10)	10.00
3	-9,326	(13)	-33,374	(12)	-19,370	(12)	-43,953	(12)	12.25
4	-12,686	(6)	-34,338	(4)	-19,884	(2)	-49,753	(3)	3.75
5	-12,818	(3)	-34,345	(2)	-19,880	(3)	-49,816	(2)	2.50
6	-12,694	(5)	-34,339	(3)	-19,870	(4)	-49,748	(4)	4.00
7	-12,609	(7)	-34,043	(7)	-19,816	(7)	-49,298	(7)	7.00
8	-12,698	(4)	-34,052	(6)	-19,808	(8)	-49,324	(6)	6.00
9	-12,453	(9)	-33,861	(8)	-19,831	(5)	-48,877	(8)	7.50
10	-12,536	(8)	-33,860	(9)	-19,830	(6)	-48,866	(9)	8.00
11	-9,386	(12)	-33,222	(13)	-19,350	(13)	-42,633	(13)	12.75
12	-9,686	(11)	-33,528	(11)	-19,464	(11)	-44,678	(11)	11.00
13	-14,206	(2)	-34,294	(5)	-19,801	(9)	-49,702	(5)	5.25
14	-14,874	(1)	-34,624	(1)	-19,936	(1)	-51,471	(1)	1.00

Table 4.5: The list of BIC for each model and period. The numbers in parentheses are the ranking of goodness-of-fitting, in which 1 shows the best fit and 14 the worst. According to these rankings, Model 14 is the best, and Model 13 is the second best.

Model #	Period 1		2		3		4		Rank ave.
1	-7,730	(14)	-31,501	(14)	-18,666	(14)	-34,445	(14)	14.00
2	-11,328	(10)	-32,462	(13)	-18,740	(13)	-47,043	(10)	11.50
3	-8,918	(13)	-32,727	(11)	-18,894	(12)	-43,526	(12)	12.00
4	-12,240	(4)	-33,642	(3)	-19,368	(3)	-49,283	(4)	3.50
5	-12,335	(3)	-33,598	(4)	-19,325	(4)	-49,305	(2)	3.25
6	-12,212	(6)	-33,592	(5)	-19,315	(6)	-49,236	(5)	5.50
7	-12,163	(7)	-33,346	(6)	-19,300	(7)	-48,829	(6)	6.50
8	-12,215	(5)	-33,305	(7)	-19,253	(9)	-48,812	(7)	7.00
9	-12,007	(9)	-33,164	(8)	-19,316	(5)	-48,408	(8)	7.50
10	-12,053	(8)	-33,113	(9)	-19,274	(8)	-48,354	(9)	8.50
11	-9,052	(12)	-32,725	(12)	-18,973	(11)	-42,270	(13)	12.00
12	-9,347	(11)	-33,026	(10)	-19,083	(10)	-44,310	(11)	10.50
13	-13,835	(2)	-33,746	(2)	-19,385	(2)	-49,296	(3)	2.25
14	-14,498	(1)	-34,072	(1)	-19,516	(1)	-51,061	(1)	1.00

the ICA-based models are better than the copula-based models both around March 2010 and July 2011, while the copula-based models are better around September 2010. Figure 4.2 shows the interesting feature that the fitness of ICA models is much worse than that of copula models before 2009. The reason is not clear, but it is possible from this figure that the market structure is completely changed at around 2009. This timing of change matches the beginning of the European sovereign crisis.

4.4 Joint default model

4.4.1 Default probability of single country

There are two sources of the default probability (DP) of sovereign entities. One is that based on credit ratings, and the other is that implied in the CDS. A rating is a discrete variable issuedby credit agencies. For example, Standard and Poor's adopts 17 levels[2] of credit worthiness, ranging from AAA to D. These rating are provided by major credit agencies. These agencies decide the rating basing on a wide range of information available, both quantitative and qualitative. There are preceding empirical studies which converts these ratings into the default probability. For example, these agencies themselves publish the transition probability to default based on their definition of ratings; for example, see the report by Standard and Poor's (Standard and Poor's, 2013).

On the other hand, a CDS premium is a virtually continuous variable, compared with a rating.[3] Its value is decided by market participants, who are supposed to use all the public information available up to the trading time. A CDS premium can be theoretically connected to the default probability. Especially, ignoring the risk premium or counter party risk and assuming continuous payments of premium, the relation becomes simple (Hull and White, 2000), according to which the default probability (DP) p implied in a CDS preimum c is

$$p = c \times \frac{1 + r_f}{1 - R}, \tag{4.23}$$

where r_f is the risk-free rate and R the recovery rate. For example, the

[2]Ignoring positive/negative outlooks.

[3]Strictly, a CDS premium is also a discrete value since it is a multiplication of price tick.

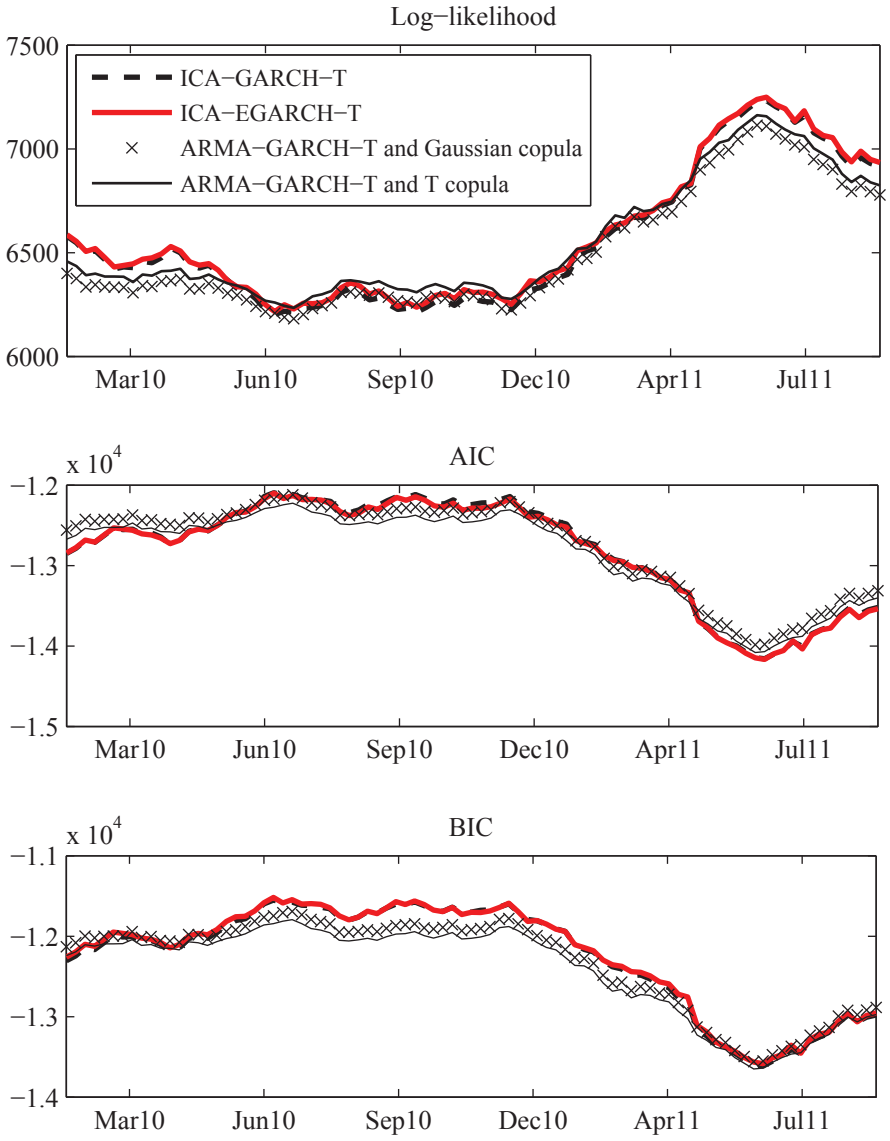


Figure 4.1: Time series of log-likelihood, AIC and BIC during Period 2. The order of goodness of these four models depends on time. For example, the ICA-based models are better than the copula-based models both around March 2010 and July 2011, while the copula-based models are better around September 2010.

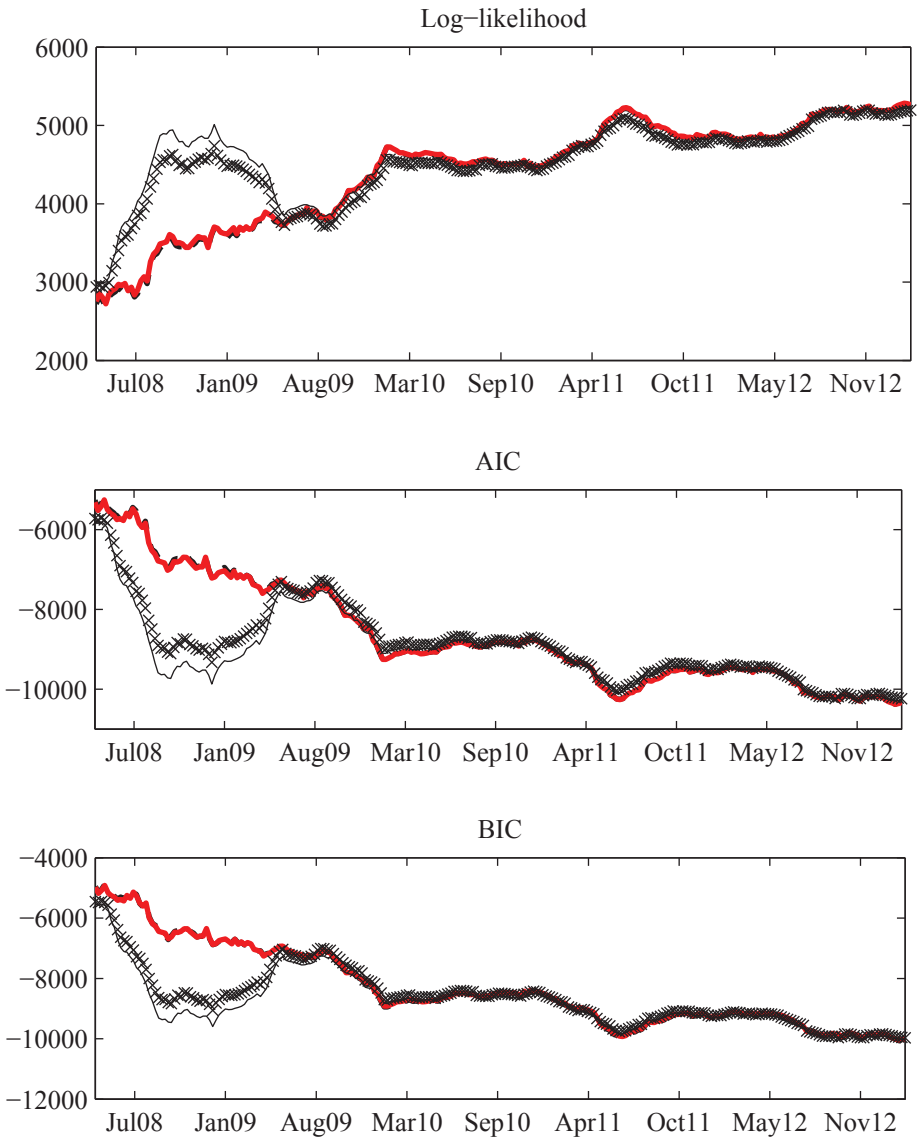


Figure 4.2: Time series of log-likelihood, AIC and BIC during Period 4. The fitness of ICA models is much worse than that of copula models before 2009. The reason is not clear, but it is possible from this figure that the market structure is completely changed at around 2009. This timing of change matches the beginning of the European sovereign crisis.

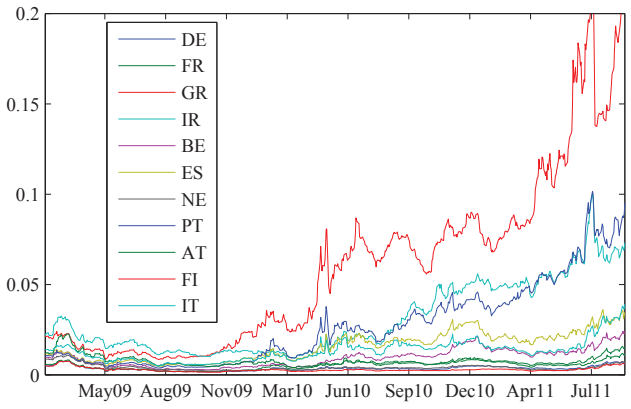


Figure 4.3: Semiannual default probability of each country during Period 2. Based on CDS premium of five year maturity, six month LIBOR, and 40% recovery rate.

example of semiannual default probability based on eq.(4.23) is shown in Figure 4.3. In this figure, the semiannual default probabilities of 11 countries during Period 2 are shown, based on the five-year maturity CDS premium provided by CMA, the six-month USD LIBOR as the risk-free rate, and 40% as the recovery rate. Equation (4.23) assumes a continuous constant premium applied in all maturities. There are more sophisticated default probability models incorporating maturity structure or risk premium, but we simply uses the relation 4.23 since our purpose is to check the effect of ICA.

These two sources have both advantages and disadvantages. For example, changes of credit ratings are not so frequent. Greek has experienced frequent changes of its rating due to recent turmoil, but the number of changes is around ten since 2000. On the other hand, a CDS premium is updated daily, hourly, or even at every seconds in the market. From the viewpoint of up-to-date information, a CDS premium is better since it reflects the newest news. On the other hand, a CDS premium is affected by not only the situation of the reference entity itself but the situation of the market. As a result, it contains a lot of erroneous information.

4.4.2 Connecting default probability with CDS

A default of an corporate is often modeled as the event that the asset value of the corporate falls below the debt amount at the time of measurement. In another words, a default is defined as the asset deficit at the measurement time. A default model based on this idea is usually referred to as a *structural model* of default. For example, Merton model is a structural model importing the geometric Brownian motion as the asset value process (Merton, 1974). Black Cox model is same with the Merton model except that the measurement time is not limited to discrete times but all time t (Black and Cox, 1976). Based on this idea, it is natural to link defaults of a set of corporates via a vector stochastic process representing their asset values (McNeil, Frey and Embrechts, 2005).

In the case of sovereign default, it is not common to extend the idea of asset deficit. The alternative idea to use the cost and benefit of default, and define a default as the case the benefit surpasses the cost (Calvo, 1988). Especially, by regarding the CDS premium as the difference between the benefit and the cost, a default model similar to the structural model is obtained; for details, see Lucas et al. (2011). The only difference is the direction of inequality; since larger CDS premium is closer to default, it is natural to define a default event as the CDS premium goes over some threshold value.

Let us define a default model based on the CDS premium pricing. Suppose we have the market information up to time t, and let $p_{t,t+\tau}$ denote the DP of a country by time $t+\tau$. Let $A_{t,t+\tau}$ denote the set of default events between time t and $t+\tau$. By definition,

$$p_{t,t+\tau} = P_t(A_{t,t+\tau}), \tag{4.24}$$

where P_t means the probability filtered by the information up to time t.

Let C_t denote the CDS premium at time t. As we noted, the fundamental idea of connecting CDS with default is to define a default event as the case that CDS spread goes over some threshold value. In equation,

$$A_{t,t+\tau} = \{C_{t+\tau} > C^0_{t,t+\tau}\}, \tag{4.25}$$

where $C^0_{t,t+\tau}$ is the threshold.

Combining eqs.(4.24) and (4.25), we obtain

$$p_{t,t+\tau} = P_t(C_{t+\tau} > C^0_{t,t+\tau}). \tag{4.26}$$

Let $F_{t,t+\tau}(C)$ denote the cumulative distribution function (CDF) of $C_{t+\tau}$, that is, $F_{t,t+\tau}(C) = P_t(C_{t+\tau} < C)$. Then, eq.(4.26) is rewritten as $p_{t,t+\tau} = 1 - F_{t,t+\tau}(c^0_{t,t+\tau})$, or equivalently,

$$c^0_{t,t+\tau} = F^{-1}_{t,t+\tau}(1 - p_{t,t+\tau}). \tag{4.27}$$

4.4.3 Joint default probability

The discussion above can be easily extended to multiple entities. Suppose we have only two entities (entity 1 and entity 2), and identify a variable by subscripting, such as, $F_{1,t,t+\tau}(C_i)$ for the CDF of entity 1. Let us consider the joint DP of the entities 1 and 2. The event sets of entity 1 default is defined by $A_{1,t,t+\tau}$, and that as for entity 2 is $A_{2,t,t+\tau}$. Therefore, the set of events where both entity 1 and 2 default during $t \sim t+\tau$ is $A_{1,t,t+\tau} \cap A_{2,t,t+\tau}$. From eq.(4.25),

$$\begin{aligned}
&\{\text{Entity 1 and 2 default simultaneously}\} \\
&\quad = A_{1,t,t+\tau} \cap A_{2,t,t+\tau} && (4.28)\\
&\quad = \{C_{1,t+\tau} > C^0_{1,t,t+\tau} \text{ and } C_{2,t+\tau} > C^0_{2,t,t+\tau}\}. && (4.29)
\end{aligned}$$

Therefore,

$$\begin{aligned}
&P_t(\text{Entity 1 and 2 default simultaneously}) \\
&\quad = P_t(C_{1,t+\tau} > C^0_{1,t,t+\tau} \text{ and } C_{2,t+\tau} > C^0_{2,t,t+\tau}) && (4.30)\\
&\quad = 1 - F_{12,t,t+\tau}(C^0_{1,t+\tau}, C^0_{2,t+\tau}), && (4.31)
\end{aligned}$$

where $F_{12,t,t+\tau}(C_1, C_2)$ is the joint CDF of $(C_{1,t+\tau}, C_{2,t+\tau})$ filtered by time t. According to eq.(4.31), in order to obtain the joint default probability, we need the joint distribution of CDS. The threshold values $C_{1,t+\tau}$ and $C_{2,t+\tau}$ are given by eq.(4.27). Likewise, it can be easily shown that the probability that entity 1 defaults but entity 2 does not is $P_t(C_{1,t+\tau} > C^0_{1,t,t+\tau} \text{ and } C_{2,t+\tau} < C^0_{2,t,t+\tau})$, and so on.

It is clear how to extend this discussion for larger numbers of entities. The probability of defaults of q entities during $t \sim t+\tau$ is given as the joint probability that the CDSs of these entities go over their corresponding thresholds at time $t+\tau$, and these thresholds can be given by the marginal distribution function of CDS and the default probability. If there exist $(d-q)$ another entities which do not default, then the probability is the probability

that q CDS go over their thresholds and another $(d-q)$ go below their thresholds.

Based on these discussions, it is enough to know the joint distribution of CDS and the marginal probability of default, in order to obtain the joint probability of defaults. In this chapter, we use the return model described in Section 4.3 in order to obtain the joint distribution.

4.4.4 Joint default statistics

Based on the previous discussion, we can define the joint default probability. However, for intuitive understanding, we need to invent an indicator showing dependency. For simplicity, let $I_i = 1$ denote default event $C_{i,t+\tau} > C^0_{i,t,t+\tau}$ and $I_i = 0$ denote non-default $C_{i,t+\tau} < C^0_{i,t,t+\tau}$

First, we consider the probability of one or more defaults $D_{\geq 1}$, the probability of only one default D_1, and the probability of two or more defaults $D_{\geq 2}$. Apparently, $D_{\geq 1} = D_1 + D_{\geq 2}$.

Then we can define the conditional probability of simultaneous defaults (CPSD). It is defined as the probability of two or more defaults conditional to one or more defaults. Apparently,

$$\text{CPSD} = \frac{D_{\geq 2}}{D_{\geq 1}} = 1 - \frac{D_1}{D_{\geq 1}} \tag{4.32}$$

This indicates the risk of simultaneous default.

Similar to CPSD, we define the stability index (SI). This is originally defined by Huang (1992). Segoviano and Goodhart (2009) name it as the *banking stability index* since they discuss the stability of banking sectors. It is defined as

$$\text{SI} = \frac{P(I_1 = 1) + \cdots + P(I_d = 1)}{1 - P(I_1 = 0, \ldots, I_d = 0)}. \tag{4.33}$$

SI takes a value in the range from 1 (all defaults are independent) to d (all defaults are completely correlated).

4.4.5 Joint default results

In order to see the joint default statistics in detail, we do backtesting similar to the rolling analysis in Section 4.3.5. For each backtesting day, we pick up the 252 consecutive days prior to the day, fit the models to the data of these

252 days, and compute the joint default probability in six months from the day. The backtesting days are every five days. We pick up Model 4 (ICA-GARCH-T) and 14 (ARMA-GARCH-T and T copula). We show the results based on the data of Period 4 in order to see longer time series, but have the similar results based on Period 2.

Figure 4.4 shows the time series of $D_{\geq 1}$, D_1, and $D_{\geq 2}$. It can be seen that the ICA model reports lower probability of any defaults, $D_{\geq 1}$, compared to the copula model. On the other hand, it can be seen that the probability of multiple defaults $D_{\geq 2}$ based on ICA is about the same with that based on copula. The difference is considered to come from the probability of only one default, D_1. The copula model reports larger probability of single default than the ICA model.

Figure 4.5 shows the time series of the CPSD and the SI. It can be seen that these values based on ICA are generally larger than those based on copula, especially in the early period of analysis. This is because that the ICA model reports lower single default probability D_1. As a result, the conditional probability of eq.(4.32) gets larger. The increased probability of joint defaults pushes the SI up, making the system unstable.

It is interesting that these values based on ICA increased earlier than those based on copula before 2009. This means that the ICA model was more sensitive to the increased joint risk of sovereigns occurring in 2010. Also, these values based on ICA have decreased since 2011, while those based on copula have not. This means that, based on ICA, the stress of Eurozone sovereign has been relaxed since 2011. This is an interesting results since the consecutive rescue programs of Eurozone countries such as the European Financial Stability Facility (EFSF) did have effect on the market, while it did not decreased the risk of the system dramatically by a single shot.

4.5 Conclusion

In this chapter, we provide an empirical application of the independent component analysis (ICA) based on the Eurozone sovereign CDS. According to the result, the ICA-based models are as good as the copula-based models.

Based on this result, we introduce a structural sovereign default model with CDS premium, and compute the joint default probability of sovereigns based on the ICA models and copula models. Accordingly, the ICA model reports higher conditional probability of joint defaults compared to the cop-

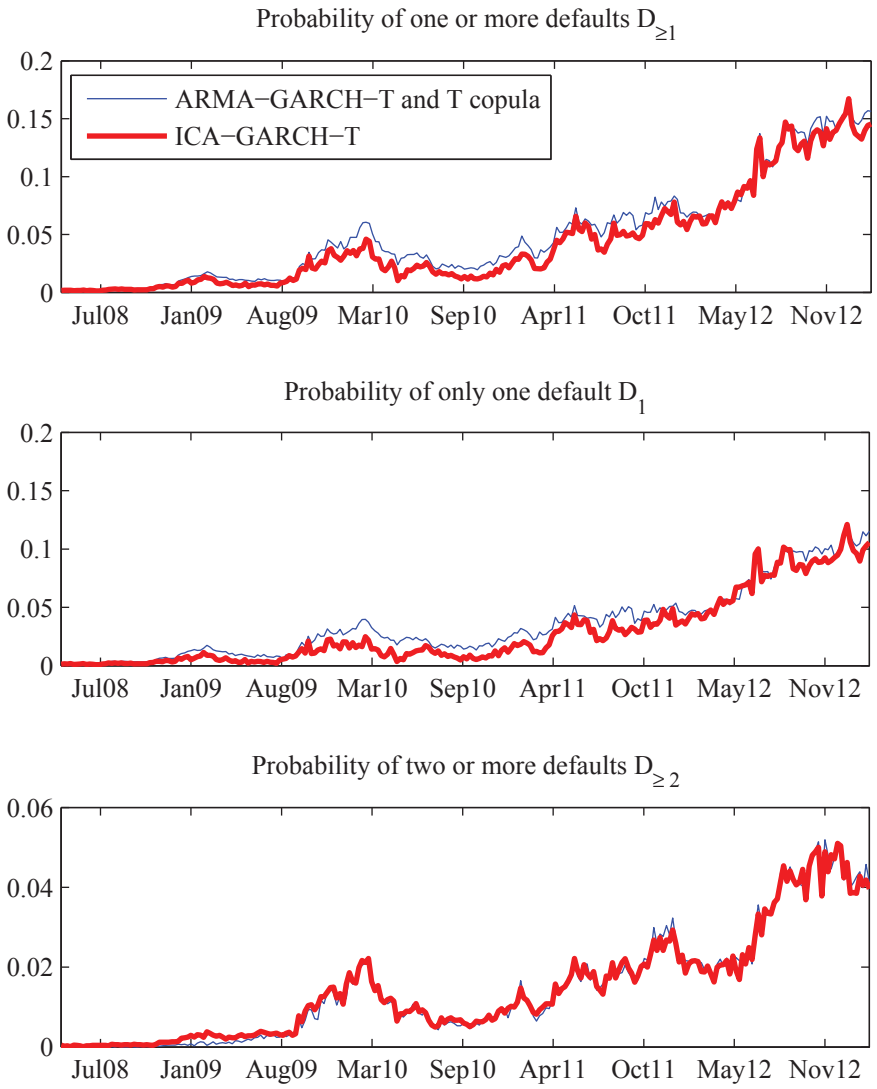


Figure 4.4: Time series of $D_{\geq 1}$, D_1, and $D_{\geq 2}$ during Period 4. It can be seen that the ICA model reports lower probability of $D_{\geq 1}$, compared to the copula model. On the other hand, it can be seen that the probability of multiple defaults $D_{\geq 2}$ based on ICA is about the same with that based on copula.

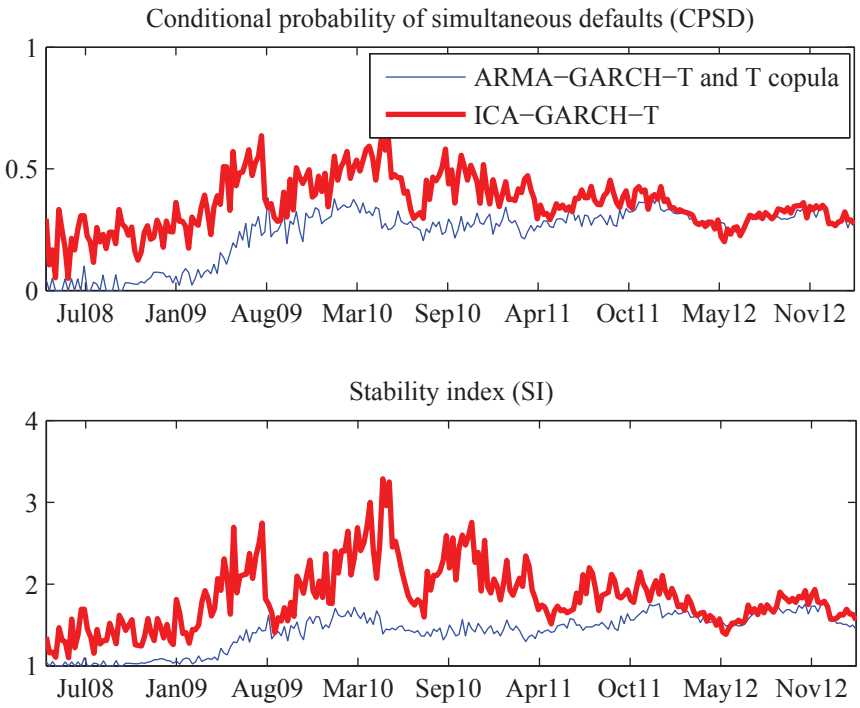


Figure 4.5: Time series of the conditional probability of simultaneous default (CPSD) and the stability index (SI) during Period 4. It can be seen that these values based on ICA are generally larger than those based on copula, especially in the early period of analysis. This is because that the ICA model reports lower single default probability D_1. The increased probability of joint defaults means that the system is unstable.

ula models. Consequently the instability of Eurozone sovereigns shown by ICA is larger than that shown by copula.

In this research, we combined two models; ICA model and joint default model. The advantage of ICA is that it is easy to interpret and compute. For further researches, it should be considered to relate ICs to external economic indicator. Also, it is interesting to apply ICA into higher dimensions, in which the maximum likelihood estimation of Student-t copula parameters becomes computationally intractable. Analysis of joint default in Eurozone sovereigns is an ongoing hot topic at the time we write this chapter. Our research is just an example how to compute the joint risk of Eurozone, and it should be necessary to carefully evaluate it using a wide variety of models and indicators.

Chapter 5

Conclusion

In this dissertation, we discussed the practical applications of risk analysis based on advanced probabilistic models.

In Chapter 1, we summarized the methodology from the three important viewpoints: univariate modeling, multivariate modeling, and definition of risk. As for univariate modeling, we introduced the α-stable distribution and its modifications such as the classical tempered distribution, and we saw that these distributions are connected by the concept of Lévy process. As for multivariate modeling, we introduced two concepts. One is the copula function, which is widely used in pricing of multiple assets. The other is the factor analysis, the principal component analysis and the independent component analysis, which are the basic tool to summarize the multivariate data by single of few number of variables. As for definition of risk, we introduced the concept of Value at Risk, and discussed related topics.

In Chapter 2, we presented a practical example of how to construct a portfolio based on the return model and risk measure. The result was tested by the method of backtesting. Consequently, it was found that (1) the ARMA-GARCH model with classical tempered stable (CTS) distribution provides a superior prediction than the normal and Student-t distribution and (2) AVaR provides a better risk measure than variance. It was also suggested that the number of universe has effect on the portfolio return, and that it is effective to reduce stock universe to large capitalization stocks.

In Chapter 3, we analyzed the distribution of returns on seven major Eurozone sovereign bonds (France, Germany, Greece, Ireland, Italy, Portugal, and Spain) and their co-movement. We investigated the ARMA-GARCH models based on different assumptions about the innovations: Gaussian,

Student-t, CTS, normal tempered stable (NTS), and α-stable. For each of the five models, we applied four copula functions, and assessed the forecasting performance of combinations of these models. In addition, to find a forward-looking measure to detect the exacerbating of the financial crisis of Greece, we analyzed the evolution of the tail parameter over time.

In Chapter 4, we discussed the goodness of fitting of the independent component analysis (ICA) using the sovereign CDS premiums of 11 Eurozone countries (Austria, Belgium, Finland, France, Germany, Greece, Ireland, Italy, Netherlands, Portugal, and Spain). Based on the log-likelihood, the Akaike information criterion and the Bayesian information criterion, we first showed the fitness of the ICA is as good as more complicated models such as the ARMA-GARCH and Student-t copula model. Also, we introduced the joint default probability model based on the CDS prediction model and the marginal default probability, and discussed that the ICA-based joint default indicates the European sovereign crisis earlier than that based on the ARMA-GARCH and copula.

As we discussed at the beginning of Chapter 1, analyzing risk in finance is an important topic for all participants of finance and economy. I hope that the results shown in this dissertation can help improving risk analysis in future.

Bibliography

[1] Aikman, D., Alessandri, P., Eklund, B., Gai, P., Kapadia, S., Martin, E., Mora, N., Sterne, G., Willison, M., 2009, "Funding liquidity risk in a quantitative model of systemic stability," Bank of England Working Paper 372.

[2] Ang, A., and A. A. Longstaff, 2013, "Systemic sovereign credit risk: Lessons from the U.S. and Europe," Journal of Monetary Economics, 60 (5), 493-510.

[3] Artzner, P., F. Delbaen, J. M. Eber, D. Heath, 1999, "Coherent measures of risk," Mathematical Finance, 9(3), 203-228.

[4] Bachelier, L., 1900, "Théorie de la spéculation," Annales Scientifiques de l'École Normale Supérieure, 3 (17), 21-86 (in French).

[5] Back, A., D., A. S. Weigend, 1997, "A first application of independent component analysis to extracting structure from stock returns," Int. J. Neural Systems, 8, 473-484.

[6] Barndorff-Nielsen, O. E., Levendorskiĭ, S. Z., 2001, "Feller processes of normal inverse Gaussian type," Quantitative Finance 1 (3), 318-331.

[7] Barndorff-Nielsen, O. E., Shephard, N., 2001, "Normal modified stable process," Theory of Probability and Mathematical Statistics 65, 1-19.

[8] Barone-Adesi, G., R. F. Engle, L. Mancini, 2008, "A GARCH option pricing model with filtered historical simulation," Review of Financial Studies, 21 (3), 1223-1258.

[9] Basel Committee of Banking Supervision, 1996, "Supervisory framework for the use of 'backtesting' in conjunction with the internal models

approach to market risk capital requirements" Bank for International Settlements, January 1996.

[10] Basel Committee of Banking Supervision, 2006, "Basel II: International convergence of capital measurement and capital standards: A revised framework - comprehensive version," Bank for International Settlements, June 2006.

[11] Berg, D., 2009, "Copula goodness-of-fit testing: an overview and power comparison," The European Journal of Finance, 15 (7-8), 675-701.

[12] Berkowitz, J., 2001, "Testing density forecasts, with applications to risk management," Journal of Business and Economic Statistics, 19 (4), 465-474.

[13] Berkowitz, J., P. Christoffersen, D. Pelletier, 2009, "Evaluating value-at-risk models with desk-level data," Management Science, 57 (12), 2213-2227.

[14] Bertsimas, D., G. J. Lauprete, A. Samarov, 2004, "Shortfall as a risk measure: properties, optimization and applications," Journal of Economic Dynamics and Control, 28 (7), 1353-1381.

[15] Billio, M., Getmansky, M., Lo, A. W., Pelizzon, L., 2010, "Econometric measures of systemic risk in the finance and insurance sectors," NBER Working Paper 16223.

[16] Black, F., J. C. Cox, 1976, "Valuing corporate securities: Some effects of bond indenture provisions," Journal of Finance, 31 (2), 351-367.

[17] Boyarchenko, S. I. , Levendorskiĭ, S. Z., 2000, "Option pricing for truncated Lévy processes," International Journal of Theoretical and Applied Finance 3 (3), 549-552.

[18] Calice, G., J. Chen, J. Williams, 2012, "Liquidity spillovers in sovereign bond and CDS markets: An analysis of the Eurozone sovereign debt crisis," Journal of Economic Behavior & Organization, 85, 122-143.

[19] Calvo, G. A, 1988, "Servicing the public debt: The role of expectations," The American Economic Review, 78 (4), 647-661.

[20] Campbell, S. D., 2005, "A review of backtesting and backtesting procedures," Divisions of Research and Statistics and Monetary Affairs, Federal Reserve Board, 2005-21.

[21] Caporin, M., L. Pelizzon, F. Ravazzolo, R. Rigobon, 2013, "Measuring sovereign contagion in Europe," NBER working paper, 18741.

[22] Carr, P., Geman, H., Madan, D., Yor, M., 2002, "The Fine Structure of Asset Returns: An Empirical Investigation," Journal of Business 75 (2), 305-332.

[23] Christoffersen, P. F., 1998, "Evaluating interval forecasts," International Economic Review 39 (4), 841-862.

[24] CMA, 2011, "CMA Global Sovereign Debt Credit Risk Report: 3rd Quarter 2011."

[25] Coles, S., J. Heffernan, J. Tawn, 1999 "Dependence measures for extreme value analyses," Extremes, 2.4, 339-365.

[26] De Jonghe, O., 2010, "Back to the basics in banking? A micro-analysis of banking system stability," Journal of Financial Intermediation 19, 387-417.

[27] Demarta, S., A. J. McNeil, 2005, "The t copula and related copulas," International Statistical Review, 73 (1), 111-129.

[28] Embrechts, P., A. McNeil, D. Straumann, 2001, "Correlation dependency in risk management: properties and pitfalls," In "Risk management: value at risk and beyond," Eds. M. Dempster, H. Moffatt, Cambridge University Press, 2002, 176-223.

[29] Engle, R. F., S. Manganelli, 2004, "CAViaR: Conditional autoregressive value at risk by regresion quantiles," Journal of Business and Economic Statistics, 22 (4), 367-381.

[30] Fama, E., 1963, "Mandelbrot and the stable Paretian hypothesis," Journal of Business, 36, 420-429.

[31] Fama, E. F., K. R. French, 1992, "The cross - section of expected stock returns," The Journal of Finance, 47 (2), 427-465.

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

[33] Feller, W., 1971, "An introduction to probability theory and its application," 2nd ed., New York, Wiley.

[34] García-Ferrer, A., E. González-Prieto, D. Peña, 2008, "A multivariate generalized independent factor GARCH model with an application to financial stock returns," Statistics and econometrics working papers, Universidad Carlos III, Departamento de Estadística y Econometría.

[35] Genest, C., K. Ghoudi, and L-P. Rivest, 1995, "A semiparametric estimation procedure of dependence parameters in multivariate families of distributions," Biometrika, 82 (3) 543-552.

[36] Genest, C., Rémillard, B., Beaudoin, D., 2009, "Goodness-of-fit tests for copulas: A review and a power study," Insurance: Mathematics and Economics 44, 199-213.

[37] Giesecke, K., Kim, B., 2011, "Systemic Risk: What defaults are telling us," Managament Science 57 (8), 1387-1405.

[38] Glosten, L. R., R. Jagannathan, D. Runkle, 1993, "On the relation between the expected value and the volatility of the nominal excess return on stocks," Journal of Finance, 48 (5), 1779-1801.

[39] Guerard, J. B., Jr., ed., 2009, "Handbook of Portfolio Construction: Contemporary Applications of Markowitz Techniques," 1st ed., Springer.

[40] Haas, M., 2001, "New methods in backtesting," Financial Engineering Research Center, Working Paper.

[41] Hamilton, J. D., 1994, "Time series analysis (Vol. 2)," Princeton, Princeton university press.

[42] Harman, H.H., 1967, "Modern Factor Analysis," University of Chicago Press, 2nd ed.

[43] Hartmann, P., Straetmans, S., De Vries, C., 2005, "Banking system stability: A cross-Atlantic perspective," NBER Working Paper 11698.

[44] Huang, X., 1992, "Statistics of Bivariate Extreme Values," Tinbergen Institute Research Series, Ph.D. Thesis, 22, Erasmus University, Rotterdam, Netherlands.

[45] Huber, P. J., 1981, "Robust statistics," New York, Wiley.

[46] Hull, J. C., A. White, 2000, "Valuing credit default swaps I: No counterparty default risk," The Journal of Derivatives, 8, 29-40

[47] Hyväreinen, A., 2013, "Independent component analysis: recent advances," Phil. Trans. R. Soc. A., 371, 20110534.

[48] Hyväreinen, A., J. Karhunen, E. Oja, 2001, "Independent Component Analysis" Wiley, New York.

[49] Hyvärinen, E. Oja, 2000, "Independent component analysis: algorithms and applications," Neural Networks, 13, 411-430.

[50] Itô (ed.) et al., 2012, "Kakuritsuron Handbook" (Probability theory handbook), Tokyo, Maruzen (in Japanese).

[51] Joe, H., 1997, "Multivariate models and dependence concepts," London, Chapman & Hall/CRC.

[52] Khintchine, A. Ya., 1937, "A new derivation of one formula by Paul Lévy ," Moscow State University Bulletin, 1 (1), 1-5 (in Russian).

[53] Kim, Y. S., S. T. Rachev, M. L. Bianchi, F. J. Fabozzi, 2008, "Financial market models with Levy processes and time-varying volatility," Journal of Banking and Finance, 32 (7), 1363-1378.

[54] Kim, Y. S., Rachev, S. T., Bianchi, M. L., Mitov, I., Fabozzi, F. J., 2011, "Time series analysis for financial market meltdowns," Journal of Banking and Finance 35, 1879-1891.

[55] Kim, Y. S., Giacometti, R., Rachev, S. T., Fabozzi, F. J., Mignacca, D., 2012, "Measuring financial risk and portfolio optimization with a non-Gaussian multivariate model," Annals of Operations Research, 201 (1), 325-343.

[56] Koponen, I., 1995, "Analytic approach to the problem of convergence of truncated Lévy flights towards the Gaussian stochastic process," Physical Review E, 52 (1), 1197-1199.

[57] Kumiega, A., T. Neururer, B. V. Vliet, 2011, "Independent component analysis for realized volatility: Analysis of the stock market crash of 2008," The Quarterly Review of Economics and Finance, 51, 292-302.

[58] Kupiec, P., 1995, "Techniques for verifying the accuracy of risk measurement models," Journal of Derivatives, 3 (2), 73-84.

[59] Lamantia, F., S. Ortobelli, S. T. Rachev, 2006, "An empirical comparison among VaR models and time rules with elliptical and stable distributed returns," Investment Management and Financial Innovations, 3 (3), 8-29.

[60] Lamantia, F., S. Ortobelli, S. T. Rachev, 2006, "VaR, CVaR and time rules with elliptical and asymmetric stable distributed returns," Investment Management and Financial Innovations, 4, 19-39.

[61] Lévy, P. 1934, "Sur les intégrales dont les éléments sont des variables aléatoires indépendantes," Annali della Scuola Normale Superiore di Pisa, Classe di Scienze 2e série, 3-4, 337-366 (in French).

[62] Li, D. X., 2000, "On default correlation: a copula function approach," Journal of Fixed Income, 9 (4), 43-54.

[63] Lintner, J., 1965, "The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets," Review of Economics and Statistics, 47 (1), 13-37.

[64] Longstaff, F. A., Pan, J., Pedersen, L. H., Singleton, K. J., 2011, "How sovereign is sovereign credit risk?" American Economic Journal: Macroeconomics, 3 (2), 75-103.

[65] Lucas, A., B. Schwaab, X. Zhang, "Conditional probabilities for Euro area sovereign default risk," Tinbergen Institute Discussion Paper, 11-176/2/DSF29.

[66] Mălăroiu, S., K., Kiviluoto, E., Oja, 2000, "Time series prediction with independent component analysis," Proc. Int. Conf. on Advanced Investment Technology, Gold Coast, Australia, 2000.

[67] Mandelbrot, B. B., 1962, "The variation of certain speculative prices," IBM Research Report, NC-87 (March).

[68] Mandelbrot, B. B., 1963, "New methods in statistical economics," Journal of Political Economy, 71, 421-440.

[69] Mantegna, R. N., H. E. Stanley, 1994, "Stochastic process with ultraslow convergence to a Gaussian: The truncated Lévy flight," Physical Review Letters, 73, 2946-2949.

[70] Markowitz, H. M., 1952, "Portfolio selection," The Journal of Finance, 7 (1), 77-91.

[71] McNeil, A. J., R. Frey, P. Embrechts, 2005, "Quantitative risk management: concepts, techniques, tools," Princeton, Princeton University Press.

[72] McNeil, A. J., J. Nešlehova, 2009, "Multivariate Archimedean Copulas, d-Monotone Functions and l_1-Norm Symmetric Distributions," The Annals of Statistics, 37 (5B), 3059-3097.

[73] Merton, R. C., 1974, "On the Pricing of Corporate Debt: The Risk Structure of Interest Rates," Journal of Finance, 29 (2), 449-470.

[74] Mossin, J., 1966, "Equilibrium in a Capital Asset Market," Econometrica, 34 (4), 768-783.

[75] Nelsen, R. B., 1999, "An introduction to copulas," Springer Verlag, 1999.

[76] Nelson, D. B., 1991, "Conditional Heteroskedasticity in Asset Returns: A New Approach," Econometrica, 59, 347-370.

[77] Rachev, S., Mittnik, S., 2000, "Stable Paretian Models in Finance," Wiley.

[78] Rachev, S. T., S. V. Stoyanov, F. J. Fabozzi, 2008, Advanced Stochastic Models, Risk Assessment, and Portfolio Optimization: The Ideal Risk, Uncertainty, and Performance Measures, 1st ed.: New York, Wiley.

[79] Romano, C., 2002, "Calibrating and simulating copula functions: an application to the Italian stock market," Risk Management Function, Capitalia, Viale U. Tupini 180.

[80] Rosiński, J., 2008, "Tempering stable processes," Stochastic Processes and Their Applications, 117, 677-707.

[81] Ross, S., 1976, "The arbitrage theory of capital asset pricing," Journal of Economic Theory, 13 (3), 341-360.

[82] Sato, K., 1999, "Lévy Processes and Infinitely Divisible Distributions," Cambridge University Press.

[83] Schwaab, B., Koopman, S. J., Lucas, A., 2011, "Systemic risk diagnostics: Coincident indicators and early warning signals," European Central Bank Working Paper 1327.

[84] Segoviano, M. A., C. Goodhart, 2009, "Banking stability measures," International Monetary Fund working paper, 09/4.

[85] Sharpe, W. F., 1964, "Capital asset prices: A theory of market equilibrium under conditions of risk," Journal of Finance, 19 (3), 425-442.

[86] Shiryaev, A. N., 1996, "Probability," New York, Springer.

[87] Sklar, A., 1959, "Fonctions de repartition à n dimensions et leurs marges," Publications de l'Institut de Statistique de l'Université de Paris, 8, 229-231 (in French).

[88] Standard and Poor's, 2013, "Sovereign defaults and rating transition data, 2012 update."

[89] Tsuchida, N., X. Zhou, S. T. Rachev, 2012, "Mean-ETL portfolio selection under maximum weight and turnover constraints based on fundamental security factors," Journal of Investing, 21, 1 (Spring 2012), 14-24.

[90] Xu, D., T. S. Wirjanto, 2009, "Computation of portfolio VaRs with GARCH models using independent component analysis," Working paper, University of Waterloo.