

ABSTRACT

YANG, SONG. Financial Risk Management: Portfolio Optimization. (Under the direction of Dr. Tao Pang.)

Risk management is a core activity by financial institutions. There are different types of financial risks, e.g. market, credit, operational, model, liquidity, business, etc. Managing these risks to minimize potential losses is essential to ensure viability and good reputations for financial institutions. Therefore, it is necessary to have an accurate model and a proper measurement that describes the risk.

In this thesis, we model the risks with proper measurement, like Value-at-risk (VaR) and Conditional Value-at-Risk (CVaR). The dependence between risks is described by the so-called copula, which can connect marginal distributions with joint distribution. Among many popular copulas, we find a proper copula to describe the correlations between risks and between financial data. Portfolio optimization problems with VaR and CVaR as risk measurement are solved and numerical results indicate that the model can describe the real world risk very well. In addition, we propose another method, called Independent Component Analysis. By linear transformation, we obtain models for independent components with the same optimal solution. The time of solving the new models is highly reduced with the same accuracy.

© Copyright 2011 by Song Yang

All Rights Reserved

Financial Risk Management: Portfolio Optimization

by
Song Yang

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Operations Research

Raleigh, North Carolina

2011

APPROVED BY:

Dr. Peter Bloomfield

Dr. Min Kang

Dr. Negash G. Medhin

Dr. Tao Pang
Chair of Advisory Committee

DEDICATION

I dedicate this dissertation to my mother, Yexin Han, and to my father, Shubin Yang.

BIOGRAPHY

Song yang was born on April 29, 1981, in Harbin, Heilongjiang, China. He received both his B.S. and M.S. degree in Mathematics in 2004 and 2007 from Nankai University. In August 2007, he joined the Ph.D. program in Operations Research of the North Carolina State University. He is now working as a Quantitative Operations Associate at Bank of America.

ACKNOWLEDGEMENTS

I would like to express my deepest appreciation to Dr. Tao Pang for his guidance, encouragement, and support throughout my candidacy at the North Carolina State University. I would also like to thank my committee members, Dr. Negash G. Medhin, Dr. Peter Bloomfield and Dr. Min Kang for their valuable comments and suggestions. My thanks also go to Dr. Fang, Dr. Fathi, Dr. Lexin Li, Ms. Linda Smith, Dr. Zhilin Li and Dr. Jeffrey Thompson. I owe my sincere appreciation to my parents for their love, understanding, encouragement, and support. Acknowledgement and thanks are also due to Kun Huang, Rong Huang, Tao Hong, Qingwei Jin, Lan Li, Le Li, Yue Li, Ning Liu, Ran Liu, Zhe Liu, Hui Wang, Pu Wang, Ling Xiang, Lu Yu, Jinyu Zhang, and many others for their help.

TABLE OF CONTENTS

List of Tables	vii
List of Figures	viii
Chapter 1 Introduction	1
1.1 Background and Literature Review	1
1.2 Contribution of This Research	4
1.3 Outline of the Thesis	6
Chapter 2 Preliminary	7
2.1 Risk Measurement	7
2.2 GST and Stable Distribution	13
2.2.1 Generalized Skewed t Distribution	13
2.2.2 Stable Distribution	15
2.3 Copula Function	19
2.3.1 Copula	19
2.3.2 Parameter Estimation for Copula	21
2.4 Time Series Models	22
2.4.1 Measure of Dependence	22
2.4.2 ARMA Model	23
2.4.3 Heteroscedastic Model	24
2.5 Data Selection	25
Chapter 3 Risk of Stock Against Benchmark	30
3.1 Relative Return Based on The Benchmark	30
3.2 Copula Based Risk Measurements	32
3.3 Backtesting	33
3.4 Numerical Results and Discussion	33
Chapter 4 Portfolio Optimization with CVaR	39
4.1 Risk-Return Portfolio Optimization	39
4.2 Minimization of CVaR	40
4.3 Numerical Results and Discussion	44
4.3.1 Portfolio Optimization Allocation	44
4.3.2 Optimal CVaR with Time Series Model	47
4.3.3 Further Discussion for Stable Distribution	50
Chapter 5 Portfolio Optimization with VaR	54
5.1 Optimization with VaR	54
5.2 Numerical Results	56
5.2.1 Results without Time Dependence	56
5.2.2 Results with Time Dependence	58

Chapter 6 Independent Component Analysis and Its Application in Portfolio Risk Management	63
6.1 Independent Component Analysis Method	64
6.1.1 Introduction	64
6.1.2 ICA Estimation	64
6.2 Applying ICA in Portfolio Optimization Problem	66
6.3 Numerical Results	69
6.3.1 Optimization with CVaR by ICA	70
6.3.2 Optimization with VaR by ICA	75
6.3.3 Running Time for Simulations	79
References	82

LIST OF TABLES

Table 2.1	Statistics for 6 different observed loss data	26
Table 3.1	Backtesting result for PGN and S&P500	35
Table 3.2	Backtesting result for IBM and S&P500	35
Table 3.3	Backtesting result for BAC and S&P500	35
Table 3.4	Backtesting result for F and S&P500	36
Table 3.5	Backtesting result for HD and S&P500	36
Table 3.6	VaR and CVaR of the relative risk for each stock with student's t copula .	37
Table 4.1	Parameter estimations of GST distribution for 5 different observed loss data	45
Table 4.2	Portfolio allocation with $\epsilon = 0.95$ and minimum return requirement $\omega^* = 0.001$ for GST distribution	46
Table 4.3	Optimal portfolio CVaR allocation with $\epsilon = 0.95$ and minimum return requirement $\omega^* = 0.001$ for historical data	46
Table 4.4	Parameter estimations of stable distribution for 5 different observed loss data	50
Table 4.5	Optimal portfolio CVaR allocation with $\epsilon = 0.95$ and minimum return requirement $\omega^* = 0.001$ for stable distribution	50
Table 5.1	Optimal VaR portfolio allocation with $\epsilon = 0.95$ and minimum return requirement $\omega^* = 0.001$ for GST distribution	57
Table 5.2	Optimal VaR portfolio allocation with $\epsilon = 0.95$ and minimum return requirement $\omega^* = 0.001$ for stable distribution	57
Table 6.1	Statistics for 5 Independent Components	69
Table 6.2	Optimal portfolio CVaR allocation with $\epsilon = 0.95$ and minimum return requirement $\omega^* = 0.001$ for GST distribution using ICA	71
Table 6.3	Optimal portfolio CVaR allocation with $\epsilon = 0.95$ and minimum return requirement $\omega^* = 0.001$ for stable distribution using ICA	71
Table 6.4	Optimal portfolio VaR allocation with $\epsilon = 0.95$ and minimum return requirement $\omega^* = 0.001$ for GST distribution using ICA	76
Table 6.5	Optimal portfolio VaR allocation with $\epsilon = 0.95$ and minimum return requirement $\omega^* = 0.001$ for stable distribution using ICA	76
Table 6.6	Average time ellipse for ICA and copula in seconds	81

LIST OF FIGURES

Figure 2.1	Standard GST distribution with $\lambda = 0.5$ and $\nu = 4, 10, 20$	14
Figure 2.2	Standard stable distribution with $\beta = 0, \alpha = 2, 1.7, 1.2$	17
Figure 2.3	Standard stable distribution with $\alpha = 1.8$ and $\beta = -0.9, 0, 0.9$	18
Figure 2.4	Rate of loss time series for 5 stocks and S&P500 index	27
Figure 2.5	ACF and PACF of 5 stocks and S&P500 index rate of loss	29
Figure 4.1	VaR and CVaR	41
Figure 4.2	Efficient 95% CVaR Frontier for GST distribution	47
Figure 4.3	Predicted daily optimal portfolio CVaR _{0.95} for GST distribution and S&P500 index from Jan 1,2008 to Aug 10, 2009	48
Figure 4.4	Predicted daily optimal portfolio CVaR for GST distribution at different levels	49
Figure 4.5	Efficient 95% CVaR Frontier comparison for GST and stable distributions	51
Figure 4.6	Predicted daily optimal portfolio CVaR from Jan 1, 2008 to Aug 10, 2009 for GST and stable distributions at level 0.95	51
Figure 4.7	Predicted daily optimal portfolio CVaR from Jan 1, 2008 to Aug 10, 2009 for GST and stable distributions at level 0.975	52
Figure 4.8	Predicted daily optimal portfolio CVaR from Jan 1, 2008 to Aug 10, 2009 for GST and stable distributions at level 0.99	52
Figure 5.1	Efficient 95% VaR Frontier for both GST and stable distribution	58
Figure 5.2	Predicted optimal daily 95% VaR compared with optimal 95% CVaR for GST distribution	59
Figure 5.3	Predicted optimal daily 95% VaR compared with optimal 95% CVaR for stable distribution	60
Figure 5.4	Predicted optimal daily 95% VaR compared with 95% VaR from optimal CVaR for GST distribution	60
Figure 5.5	Predicted optimal daily 95% VaR compared with 95% VaR from optimal CVaR for stable distribution	61
Figure 5.6	The difference between VaR from optimal CVaR and optimal VaR for GST distribution	61
Figure 5.7	The difference between VaR from optimal CVaR and optimal VaR for stable distribution	62
Figure 6.1	ICA Procedure	64
Figure 6.2	Independent Components series	70
Figure 6.3	Efficient frontier with CVaR _{0.95} for GST distribution using ICA	72
Figure 6.4	Efficient frontier with CVaR _{0.95} for stable distribution using ICA	73
Figure 6.5	Predicted daily optimal portfolio CVaR from Jan 1, 2008 to Aug 10, 2009 for GST distribution with and without ICA at 0.95	74
Figure 6.6	Predicted daily optimal portfolio CVaR from Jan 1, 2008 to Aug 10, 2009 for stable distribution with and without ICA at 0.95	75

Figure 6.7	Efficient frontier with $\text{VaR}_{0.95}$ for GST distribution using ICA	77
Figure 6.8	Efficient frontier with $\text{VaR}_{0.95}$ for stable distribution using ICA	78
Figure 6.9	Predicted daily optimal portfolio VaR from Jan 1, 2008 to Aug 10, 2009 for GST distribution with and without ICA at 0.95	79
Figure 6.10	Predicted daily optimal portfolio VaR from Jan 1, 2008 to Aug 10, 2009 for stable distribution with and without ICA at 0.95	80

Chapter 1

Introduction

Risk management is a core activity by financial institutions. There are different types of financial risks, e.g. market risk, credit risk, operational risk, modeling risk, liquidity risk, business risk, etc. Managing these risks to minimize potential losses is essential to ensure viability and good reputations for financial institutions. Therefore, it is necessary to have an accurate model and a proper measurement that describes the risks.

1.1 Background and Literature Review

People believe that the past behaviors of financial data are rich in information about their future behaviors. Therefore, we can use the past history of financial data to get meaningful predictions of the future. There are two methods to model the financial returns/losses to predict future. One is called historical method and the other is called parametric method. The former is based on the observed data and use these data to obtain the information we need. Parametric method is used under the assumption that the observed data follow some rules or models with unknown parameters. We can use the data to get the estimations of parameters and use the rule or the model we set up to obtain what we need. Usually, the historical method depends on observed data and if the size of the data is not large enough, there is a greater chance to estimate the subject we need inaccurately. Therefore, parametric method is widely used and it can estimate what we need more precisely. For parametric method in risk management, there are mainly two kinds of approaches: unconditional approach and conditional approach.

Unconditional approach is based on the assumption that the financial returns/losses for each time period are independent and identically distributed (i.i.d.) random variables. Gaussian distribution is first introduced since it is easy and has explicit expression. However, empirical

distribution exhibits that the financial data have some properties that cannot be explained by Gaussian distribution. In both Fama (1965,[14]) and Hull and White (1998,[22]), the authors pointed out that empirical financial data are leptokurtic, that is there are too many values near the mean and too many out in the extreme tails. Therefore, Gaussian assumption is questioned and some alternative families of distributions are suggested. The family of generalized skewed t (GST) distribution, which is a generalization of student's t distribution allowing skewness, is introduced in Theodossiou (1998, [40]). In that paper, the author used GST distribution to fit the empirical distributions of 6 financial data. By statistical test, he concluded that GST distribution fits these data very well. However, for some financial data, the tail of GST distribution may not be heavy enough to describe their properties. Another candidate of distribution family is stable distribution. Stable distribution is a generalization of Gaussian distribution, which allows skewness and heavy tails and has many intriguing mathematical properties. Khindanova (2001,[24]) pointed out that stable distribution seems to be the most appropriate distribution to fit financial data. It is because that stable distribution can describe both heavy tails and asymmetry, which are common characteristics of financial data. Although stable distribution seems to be the best one for financial data, it has its own drawbacks. It does not have explicit expression for the probability density function and cumulative distribution function, we have to calculate them numerically, which is time consuming.

On the other hand, the assumption of i.i.d. for the unconditional approach is questioned. In real world, financial series are not independent with each other, therefore, conditional approach is introduced. Conditional approach admits the financial return/loss series depending on the past information. Traditionally, serial dependence is modeled by autoregressive moving average (ARMA) structure, which gives a stationary series. However, the ARMA models assume that the variance is constant and they are not able to capture the volatility clustering, i.e. large (small) changes tend to be followed by large (small) changes. Since time-varying volatility clustering is a common characteristic of financial returns/losses, researchers try to find solutions to this property. The most popular one is Autoregressive Conditional Heteroscedasticity (ARCH) model proposed by Engle (1982,[13]) and followed by Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model introduced by Bollerslev (1986,[7]). Both of the models can also combine with an ARMA specification for the mean equation, referred as ARMA-ARCH and ARMA-GARCH model. Nowadays, GARCH model is widely used. In standard GARCH model, it is assumed that the innovation distribution is Gaussian. However, for many financial return/loss series, Gaussian is not adequate for leptokurtosis. Therefore, several non-Gaussian distributions are used, for example GST distribution (Bali and Theodossiou 2007 [4] and Hansen 1994 [19]), stable paretian distribution (Liu and Brorsen 1995 [27] and Panorska, Mittnik and Rachev 1995 [36] etc). We will introduce these time series models in details in the next chapter.

Another issue in risk management is how to measure the risk of financial returns/losses. In Gaussian models, variance is used to measure the risk. As time goes on, people realized that variance is not good enough to describe the risk and some alternative measurements are then introduced. One of the most famous risk measurements is Value-at-Risk (VaR), which is the absolute value of the lower/upper percentile of the financial return/loss distributions. VaR measures the highest possible loss at a given confidence level. It is widely used to measure the risk and even becomes part of the industrial regulatory mechanisms (Basel II Accord). However, it is pointed out by Artzner, Delbaen, Eber and Heath (1998, [2]) that VaR is not a coherent measurement, and it is not convex. A risk measure is coherent if it satisfies certain axioms of monotonicity, sub-additivity, positive homogeneity and translation invariance, which will be introduced in details in chapter 2. An alternative coherent measurement of risk is introduced in Artzner, Delbaen, Eber and Heath (1998, [2]), which is called Conditional Value-at-Risk (CVaR), defined by the expected loss exceeding VaR. As a coherent measurement, CVaR has perfect properties and can make a lot of research work easier. Nowadays, both VaR and CVaR are widely used to measure the risk in both research and real world.

After we know how to measure risk, portfolio optimization problem is naturally proposed. Portfolio optimization method can be traced back to Markowitz (1952,[28]), where the author introduced mean-variance framework. In his works, Markowitz tried to find the optimal allocation for each asset of the investment, by minimizing the variance (regarded as the risk measurement) under a minimum return requirement condition. This is based on the assumption that individual assets follow Gaussian distributions and the correlation between assets is given by a correlation matrix. We can take the place of variance with some other risk measurements like VaR and CVaR. As for VaR is not a coherent measurement, there are lots of local minimizers. It is not easy to find the optimal allocation of the portfolio to minimize VaR. Gaivoronski and Pflug (2005,[16]) provided an approach to approximate VaR by a smoothed measurement called SVaR, which filters out local irregularities. However, it is quit complex to obtain the smoothed approximation. On the other hand, CVaR is coherent and convex, we can find unique global minimizer. Therefore, CVaR is widely used in portfolio optimization. Rockafeller and Uryasev (2000,[37]) discussed an easy way to calculate the portfolio optimization problem to minimize CVaR, which can transform the problem into a classic linear programming problem. Thereafter, a lot of work extended their work, for example Andersson, Mausser, Rosen and Uryasev (2001 [1]), Charpentier and Oulidi (2009, [9]), Glasserman, Heidelberger and Shahabuddin (2002, [18]) and Krokmal and Uryasev (2002, [25]). This approach does provide us a convenient way, but their work all use the so-called historical approach, in which they only use historical data to estimate the portfolio allocation. If we do not have enough

historical data, we may under-estimate the risk seriously. Another issue they did not address is the serial dependence of the historical data. It will be better if we use time series models, like ARMA and GARCH to model their serial dependence.

In portfolio optimization problem, what we really need is the joint distribution of the financial data, rather than the individual marginal distributions. In both academic and practical world, financial data are not usually perfectly independent, they usually have connections between each other. It is fact that there are correlations between different stock returns and between different types of risks. Therefore, describing their correlations is also an important topic in financial risk management. It is essential to obtain the joint distributions of the issues we are looking into. In the past, some simple but improper assumptions of joint distribution were assumed, for example, multivariate normal distribution (Markowitz 1952 [28] etc). Under multivariate normal distribution, dependence is described by the correlation matrix and it is easy to obtain the risk measurement we need. However, these simple joint distribution assumptions do not work well for financial data. Therefore, another approach is used to describe the dependence between financial data: copula function approach. Copula provides a way to connect the marginal distributions with the joint distribution. In addition, it has been proved that there exists at least one copula for the joint distributions with known marginal distributions. As a result, if we can obtain the marginal distributions such as GST distribution and stable distribution as we discussed above for financial data, we can get the joint distribution with proper copula functions. This approach is discussed in many papers, for example Dobrić and Schmid (2005, [12]) and He and Gong (2009, [21]). In [12], the author suggested a chi-square test of fit for parametric families of bivariate copulas. The test was applied to investigate the dependence structure of daily German asset returns. It turns out that the Gaussian copula is inappropriate to describe the dependencies in the data. A student's t copula with low degrees of freedom performs better. In [21], the authors discussed different type of risks individually and described the correlation between risk types by student's t copula. However, in theory, if we do not know the real joint and marginal distributions and we do not really know which kind of copula is proper.

1.2 Contribution of This Research

The first part of my research contributes to the copula selection. It can provide an approach to find which copula works well to obtain the dependence of stock returns and the benchmark returns. In this part, we focus on relative returns, which is defined as the difference of stock returns and benchmark returns. We use time series ARMA-GARCH model with the innovation of stable paretian distribution to describe both the returns, respectively. We use different copulas

to describe the correlations between the returns. The copulas we use are Gaussian, Student's t , Clayton, Gumbel and Frank. Monte Carlo simulation is used to estimate the VaR for the relative returns for each copula. An approach called backtesting is proposed. By this approach, we can tell which copula function can describe the correlation better. From numerical results, we can conclude that if our objective is risk measurement VaR, both normal and student's t copula work better than the other copulas. In addition, among the two, student's t copula is relatively a little bit better.

As discussed above, the paper of Rockafellar and Uryasev (2000, [37]) provides an easy way to solve the portfolio optimization problem taking CVaR as the risk measurement. However, it only used historical method and did not consider the serial dependence. The second part of this thesis solves this issue. In this part, we also consider the portfolio optimization problem with CVaR. Instead of historical method, we use parametric method. We extend the Rockafellar and Uryasev's idea and find the efficient frontier for the portfolio, when the marginal distribution of each asset to be GST or stable paretian distribution and their correlation is described by student's t copula. Furthermore, we extend their idea to the time series ARMA-GARCH model with innovation distribution of GST or stable. We find that the series of CVaR of the optimal portfolio allocation we obtain describe the real world's market risk efficiently. This indicates that the model we set up is quite good.

Compared to CVaR, VaR is used more widely. Therefore, we are considering how to minimize the VaR of the portfolio. Since VaR is not a convex risk measurement, it does not have unique global minimizer. On the contrary, VaR has many local minimizers. This makes it difficult to find the global minimizer of the risk of portfolio. In the paper of Benati and Rizzi (2007, [5]), the authors proposed a mixed integer linear programming method to solve this problem, but they only considered the so-called historical method. In this thesis, we use the parametric method with student's t copula to describe the correlations between the assets, time series model, as well as mixed integer linear programming method to find the minimum risk for the portfolios at each time period. We can obtain the similar results as to minimum CVaR, which also indicate that the model we set up for minimizing VaR is quite good.

In all the above discussions, we use copula to describe the correlations between the assets. The advantage of this method is that when we do simulation, we do not need to know the joint distribution of the assets. We only need the marginal distribution and the copula. However, there are disadvantages too. One of them is that it may cost more time to generate the simulated data. Therefore, in the last part of this thesis, we are trying to find a method that can save time of simulation to solve the portfolio optimization problems using the model we set up above.

The easiest way to simulate multivariate data is when the data we are trying to simulate are independent of each other. If so, we can simulate the data independently. Therefore, we propose to transform our original data into independent variables, and then transform our models into models with independent variables. Inspired by this idea, we consider to use a statistical method called Independent Component Analysis (ICA), which can transform the mixed data into statistically independent sources, by linear transformation. Using this ICA method, we can transform the original models with dependent variables into the models with independent variable. As a result, this procedure may save us a lot of time.

1.3 Outline of the Thesis

The rest of this thesis is arranged as follows. In Chapter 2, we introduce the preliminary knowledge we need in the research, including the definition of VaR and CVaR, GST and stable distributions, copula functions, the basic time series models and the data we use in the research.

In Chapter 3, we consider the risk of relative return. We present, in details, how to set up models for stock and benchmark returns, how to use Monte Carlo method to estimate the VaR and how to find the proper copula functions by backtesting approach.

Portfolio optimization problem with CVaR as the risk measurement is discussed in Chapter 4. We will explain how to use the idea of Rockafellar and Uryasev ([37]), and how to apply the ARMA-GARCH time series model in this approach.

In Chapter 5, we will solve portfolio optimization problem with VaR as the risk measurement, using both GST and stable distributions as the marginal distributions.

Independent Component analysis method is introduced in Chapter 6. Using this method, we try to solve portfolio optimization problem with both CVaR and VaR as the risk measurement. The results are compared with those using copula method.

Chapter 2

Preliminary

Some preliminary theory and results are introduced in this chapter. In Section 2.1, we introduce two of the most popular risk measurements VaR and CVaR. GST and stable distribution are discussed in Section 2.2. In Section 2.3, copula function is defined, which can describe the dependent structure between marginal distributions. In Section 2.4, we introduce some heteroscedastic models, which allow volatility clustering. Finally in Section 2.5, the data we use in the research is given and their properties are discussed.

2.1 Risk Measurement

The problem of how to measure risk is an old one in statistics, economics and finance. Variance is one of the first risk measurements proposed to define risk (see Markowitz 1952 [28]). However, as time goes on, people realize that variance is not good enough to describe risk, especially when the return/loss distributions are non-Gaussian. Therefore, some other risk measurements such as VaR and CVaR come into our view.

Definition 2.1.1. *Let R be a continuous random variable, which can represent the rate of return of an portfolio, with the cumulative distribution function $F_R(x)$. $L = -R$ is also a random variable regarded as the rate of loss of this asset, with the cumulative distribution function $F_L(x)$. Then, the Value-at-Risk (VaR) at level ϵ for the asset is defined as the absolute value of the lower/upper $1 - \epsilon$ percentile of the return/loss, i.e.*

$$\begin{aligned} \text{VaR}_\epsilon(R) &\triangleq -\max\{x \in \mathbb{R} : P(R \leq x) \leq 1 - \epsilon\} \\ &= -\max\{x \in \mathbb{R} : F_R(x) \leq 1 - \epsilon\} \end{aligned} \tag{2.1}$$

and

$$\begin{aligned} \text{VaR}_\epsilon(L) &\triangleq \min \{x \in \mathbb{R} : P(L \geq x) \leq 1 - \epsilon\} \\ &= \min \{x \in \mathbb{R} : F_L(x) \geq \epsilon\}, \end{aligned} \tag{2.2}$$

where $0 < \epsilon < 1$ and is usually taken to be close to 1, e.g., 0.95, 0.975 or 0.99.

VaR is widely accepted by financial institutes because it is easy to understand. It measures the highest possible loss at a given confidence level. However, in general, VaR is not a coherent and convex measurement (see Artzner, Delbaen, Eber and Heath 1999 [2]), except for the assumption of elliptically distributed return/loss. Here a coherent measurement is defined as follows:

Definition 2.1.2. *A risk measurement ξ of a linear space \mathbb{S} is said to be a coherent measurement if it satisfies the following properties:*

(i) Monotonicity: *If $Z_1, Z_2 \in \mathbb{S}$ and $Z_1 \leq Z_2$, then $\xi(Z_1) \geq \xi(Z_2)$. That is, if portfolio Z_2 has better values than portfolio Z_1 under all scenarios, the risk of Z_2 should be less than the risk of Z_1 .*

(ii) Sub-additivity: *If $Z_1, Z_2 \in \mathbb{S}$, then $\xi(Z_1 + Z_2) \leq \xi(Z_1) + \xi(Z_2)$. That is the risk of two portfolios together cannot get any worse than adding the two individual risks together.*

(iii) Positive homogeneity: *If $\alpha > 0$ and $Z \in \mathbb{S}$, then $\xi(\alpha Z) = \alpha \xi(Z)$. That is, if you double your portfolio, you double your risk.*

(iv) Translation invariance: *If $a \in \mathbb{R}$ and $Z \in \mathbb{S}$, then $\xi(Z + a) = \xi(Z) - a$. The value a is just adding cash to your portfolio Z , which acts like an insurance: the risk of $Z + a$ is less than the risk of Z , and the difference is exactly the added cash a .*

Generally, VaR violates the sub-additivity property, so it is not coherent. Therefore, Conditional Value-at-Risk (CVaR) is introduced in Artzner, Delbaen, Eber and Heath 1999 ([2]), which is a coherent measurement.

Definition 2.1.3. *Conditional Value-at-Risk (CVaR) at level ϵ for an asset is defined as the expected loss given that the loss exceeds VaR, i.e.*

$$\text{CVaR}_\epsilon(R) \triangleq -\mathbf{E}[R | R \leq -\text{VaR}_\epsilon(R)] \tag{2.3}$$

and

$$CVaR_\epsilon(L) \triangleq \mathbf{E}[L|L \geq VaR_\epsilon(L)]. \quad (2.4)$$

There are several approaches to estimate VaR and $CVaR$ numerically.

Historical Approach: Given the observed historical data of return R_1, R_2, \dots, R_n or data of loss L_1, L_2, \dots, L_n , the empirical distribution of the return $F_R(x)$ or loss $F_L(x)$ can be obtained by

$$F_R(x) = \frac{1}{n} \sum_{i=1}^n I_{\{R_i \leq x\}}$$

and

$$F_L(x) = \frac{1}{n} \sum_{i=1}^n I_{\{L_i \leq x\}},$$

where $I_{\{\cdot\}}$ is the indicator function. Then, the historical VaR and $CVaR$ with confident level of ϵ can be estimated as:

$$\begin{aligned} \widehat{VaR}_\epsilon(R) &= -F_R^{-1}(1 - \epsilon) = -R_{[ni]}, \quad p \in \left(\frac{i-1}{n}, \frac{i}{n} \right], \\ \widehat{VaR}_\epsilon(L) &= F_L^{-1}(\epsilon) = L_{[ni]}, \quad \epsilon \in \left(\frac{i-1}{n}, \frac{i}{n} \right], \\ \widehat{CVaR}_\epsilon(R) &= -\frac{1}{[Np]} \sum_{i=1}^{[Np]} R_{[i]}, \\ \widehat{CVaR}_\epsilon(L) &= \frac{1}{[Np]} \sum_{i=[N\epsilon]}^N L_{[i]}, \end{aligned}$$

where $p = 1 - \epsilon$, $R_{[1]}, \dots, R_{[n]}$ and $L_{[1]}, \dots, L_{[n]}$ is the ascending order statistics and $[\cdot]$ is the ceiling of a number.

Gaussian Approach: This approach is based on the assumption that the observed data follow a normal distribution $N(\mu, \sigma^2)$, with unknown parameters μ and $\sigma > 0$, that is $R_1, R_2, \dots, R_n \sim$

$N(\mu_R, \sigma_R^2)$ and $L_1, L_2, \dots, L_n \sim N(\mu_L, \sigma_L^2)$. The parameters are estimated by

$$\begin{aligned}\widehat{\mu}_R &= \frac{1}{n} \sum_{i=1}^n R_i, \\ \widehat{\mu}_L &= \frac{1}{n} \sum_{i=1}^n L_i, \\ \widehat{\sigma}_R^2 &= \frac{1}{n-1} \sum_{i=1}^n (R_i - \widehat{\mu}_R)^2, \\ \widehat{\sigma}_L^2 &= \frac{1}{n-1} \sum_{i=1}^n (L_i - \widehat{\mu}_L)^2.\end{aligned}$$

Then, the estimated VaR and $CVaR$ is

$$\begin{aligned}\widehat{VaR}_\epsilon(R) &= -\mu_R - \Phi^{-1}(p)\sigma_R, \\ \widehat{VaR}_\epsilon(L) &= \mu_L + \Phi^{-1}(\epsilon)\sigma_L, \\ \widehat{CVaR}_\epsilon(R) &= -\sigma_R \mathbf{E}[Z|Z \leq \Phi^{-1}(p)] - \mu_R, \\ \widehat{CVaR}_\epsilon(L) &= \sigma_R \mathbf{E}[Z|Z \geq \Phi^{-1}(\epsilon)] + \mu_L,\end{aligned}$$

where $p = 1 - \epsilon$, $Z \sim N(0, 1)$ is a standard Normal random variable and Φ is its cumulative distribution function.

Extreme Value Theory Approach: For VaR and $CVaR$, we only concentrate on the extreme events. We do not have to know the overall distribution and what we need is only the tail's behavior. Extreme value theories provide good ways to describe the extreme event's distribution. One of the theories is called Peak Over Threshold (POT), considering the distribution of exceedances over a certain threshold (for details see Gilli and Kellezi 2006 [17] and McNeil and Frey 2000 [29]). Let X be a random variable with cumulative distribution function F , we are interested in the conditional distribution that X is above a threshold u given that $X > u$. Let F_u be its cumulative distribution function, which is called conditional excess distribution function. We will have that

$$F_u(y) = P(X - u \leq y | X > u), \quad 0 \leq y \leq x_F - u,$$

where $y = x - u$ are the excesses and x_F is the right endpoint of F . Therefore, we can rewrite $F_u(y)$ as

$$F_u(y) = \frac{F(u+y) - F(u)}{1 - F(u)} = \frac{F(x) - F(u)}{1 - F(u)}. \quad (2.5)$$

From Extreme Value Theory, the following theorem gives a powerful result about the conditional excess distribution function.

Theorem 2.1.4. (*Pickands (1975), Balkema and de Haan (1974)*) For a large class of underlying distribution function F , the conditional excess distribution function $F_u(y)$, for large u , is well approximated by:

$$F_u(y) \approx G_{\xi, \sigma}(y), \quad u \rightarrow \infty,$$

where

$$G_{\xi, \sigma}(y) = \begin{cases} 1 - \left(1 + \frac{\xi}{\sigma}y\right)^{-\frac{1}{\xi}} & \text{if } \xi \neq 0 \\ 1 - e^{-\frac{y}{\sigma}} & \text{if } \xi = 0, \end{cases}$$

for $y \in [0, x_F - u]$, if $\xi \geq 0$ and $y \in [0, -\frac{\sigma}{\xi}]$, if $\xi < 0$. $G_{\xi, \sigma}$ is the so called Generalized Pareto Distribution (GPD).

The estimation of the parameters $\hat{\xi}$ and $\hat{\sigma}$ in $G_{\xi, \sigma}$ can be obtained by maximum likelihood method. In addition, from (2.5), we can obtain

$$F(x) = (1 - F(u))F_u(y) + F(u).$$

Replacing F_u by the GPD and $F(u)$ by the estimation $\frac{(n-N_u)}{n}$, where n is the total number of observations and N_u is the number of observations above the threshold u , we obtain the estimated $F(x)$ as

$$\hat{F}(x) = 1 - \frac{N_u}{n} \left(1 + \frac{\hat{\xi}}{\hat{\sigma}}(x - u)\right)^{-\frac{1}{\hat{\xi}}}, \quad \hat{\xi} \neq 0. \quad (2.6)$$

Therefore we have

$$\hat{F}(\widehat{VaR}) = 1 - \frac{N_u}{n} \left(1 + \frac{\hat{\xi}}{\hat{\sigma}}(\widehat{VaR} - u)\right)^{-\frac{1}{\hat{\xi}}} = \epsilon \quad (2.7)$$

Solve for \widehat{VaR}_ϵ from (2.7), we get

$$\widehat{VaR}_\epsilon = u + \frac{\hat{\sigma}}{\hat{\xi}} \left(\left(\frac{n}{N_u}p\right)^{-\hat{\xi}} - 1 \right),$$

where $p = 1 - \epsilon$.

It is known that the mean excess function for the GPD with parameter $\xi < 1$ is

$$e(z) = \mathbf{E}[Y - z | Y > z] = \frac{\sigma + \xi z}{1 - \xi}, \quad \sigma + \xi z > 0. \quad (2.8)$$

Therefore, let $Y = X - u$ and $z = \widehat{VaR}_\epsilon - u$, we have

$$\begin{aligned}
\widehat{CVaR}_\epsilon &= \mathbf{E} \left[X | X > \widehat{VaR}_\epsilon \right] \\
&= \widehat{VaR}_\epsilon + \mathbf{E} \left[X - \widehat{VaR}_\epsilon | X > \widehat{VaR}_\epsilon \right] \\
&= \widehat{VaR}_\epsilon + \mathbf{E} \left[Y - (\widehat{VaR}_\epsilon - u) | Y > \widehat{VaR}_\epsilon - u \right] \\
&= \widehat{VaR}_\epsilon + \frac{\widehat{\sigma} + \widehat{\xi}(\widehat{VaR}_\epsilon - u)}{1 - \widehat{\xi}} \\
&= \frac{\widehat{VaR}_\epsilon}{1 - \widehat{\xi}} + \frac{\widehat{\sigma} - \widehat{\xi}u}{1 - \widehat{\xi}}.
\end{aligned}$$

Parametric Approach: Parametric approach is based on the assumption that the returns/losses follow a family of distributions with unknown parameters. The parameters can be estimated by maximum likelihood method from the observed data. In order to estimate VaR and $CVaR$ numerically, we can use Monte Carlo method. If we know the distribution of R/L , we can obtain N realizations from the family of distribution with the estimated parameters. Suppose the N realizations of the rate of return R are r_1, r_2, \dots, r_N and the N realizations of the rate of loss L are l_1, l_2, \dots, l_N , then, the estimations of VaR and CVaR is obtained as:

$$\begin{aligned}
\widehat{VaR}_\epsilon(R) &= -r_{[Np]}, \\
\widehat{VaR}_\epsilon(L) &= l_{[N\epsilon]}, \\
\widehat{CVaR}_\epsilon(R) &= -\frac{1}{Np} \sum_{i=1}^{Np} r_{[i]}, \\
\widehat{CVaR}_\epsilon(L) &= \frac{1}{Np} \sum_{i=N\epsilon}^N l_{[i]},
\end{aligned}$$

where $r_{[1]}, \dots, r_{[N]}$ and $l_{[1]}, \dots, l_{[N]}$ is the ascending order statistics for the realization of R and L , respectively and $p = 1 - \epsilon$.

All the above four methods have their own advantages and disadvantages. For historical method, we only estimate the risk by the observed data. It is easy to calculate, but should not be accurate if we do not have enough extreme observations. Gaussian method is also relatively easy, but the assumption of normal distribution is questioned. For extreme value theory approach, we do not need to know the overall distribution, which is good. However, how to choose the threshold u is a big issue. Parametric approach can give accurate estimation of the risk if we fit the data by proper distributions, but Monte Carlo simulation is time consuming. In our research, in order to estimate the risk measurements accurately, we use the last approach.

2.2 GST and Stable Distribution

Finding a proper distribution to describe the financial data is an essential topic in risk management. Gaussian distribution is firstly taken into account because of its wide use and easy expression. However, empirical results show that financial data are usually asymmetric and fat-tailed, which Gaussian is lack of. This motivates people to look for some other families of distributions that can describe both asymmetry and fat-tail. Here, we introduce two families of distributions: generalized skewed t (GST) distribution and stable distribution, both of which can describe the skewness and kurtosis characteristics of financial data.

2.2.1 Generalized Skewed t Distribution

Definition 2.2.1. *Generalized skewed t (GST) distribution is the generalization of student's t distribution, which allows skewness. The probability density function for the standard GST distribution is defined as follows:*

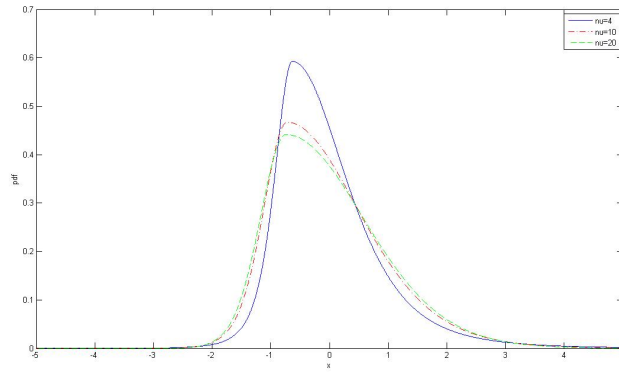
$$f(x; \nu, \lambda) = \begin{cases} bc(1 + \frac{1}{\nu-2}(\frac{bx+a}{1-\lambda})^2)^{-\frac{\nu+1}{2}} & x < -\frac{a}{b} \\ bc(1 + \frac{1}{\nu-2}(\frac{bx+a}{1+\lambda})^2)^{-\frac{\nu+1}{2}} & x \geq -\frac{a}{b}, \end{cases}$$

where $2 < \nu < \infty$, $-1 < \lambda < 1$ and

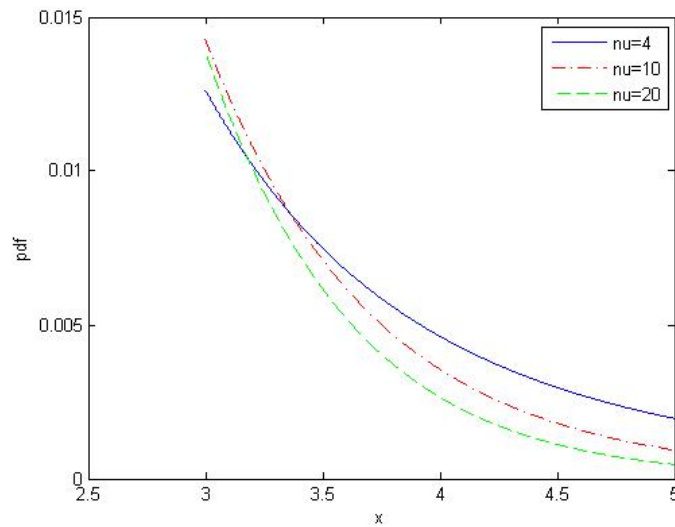
$$\begin{aligned} a &= 4\lambda c \frac{\nu-2}{\nu-1}, \\ b &= \sqrt{1 + 3\nu^2 - a^2}, \\ c &= \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\pi(\nu-2)}\Gamma(\frac{\nu}{2})}. \end{aligned}$$

It can be proved that this is a proper density function with mean 0 and variance 1 (see [19]). The parameter ν controls the tail thickness and λ controls the skewness. When $\nu \rightarrow \infty$, it is reduced to skewed normal distribution. When $\lambda = 0$, it is reduced to student's t distribution. The density function has a single mode at $-\frac{a}{b}$, which is of opposite sign of λ . Therefore, if $\lambda > 0$, the mode of the density is to the left of zero and the variable is skewed to the right, and vice-versa when $\lambda < 0$. Figure 2.1 gives the probability density function of GST with different parameters. We can see that the smaller ν is, the fatter the tail is.

If a random variable Z follows a standard GST distribution with parameter ν and λ , we write it as $Z \sim GST(\nu, \lambda)$. A non-standardized GST random variable X with mean μ and variance σ^2 is written as $X \sim GST(\mu, \sigma, \nu, \lambda)$ and $\frac{X-\mu}{\sigma} \sim GST(\nu, \lambda)$. Its probability density function can be derived from the standard GST probability density function.



(a) GST distribution



(b) GST right tail distribution

Figure 2.1: Standard GST distribution with $\lambda = 0.5$ and $\nu = 4, 10, 20$

2.2.2 Stable Distribution

Definition 2.2.2. A random variable X is said to follow stable distribution if for any $a > 0$ and $b > 0$, there exist constants $c > 0$ and $d \in \mathbb{R}$ such that

$$aX_1 + bX_2 \stackrel{d}{=} cX + d,$$

where X_1 and X_2 are independent copies of X and $\stackrel{d}{=}$ denotes the equality in distribution.

In general, stable distributions do not have closed form expressions for probability density function (PDF) and cumulative distribution function (CDF). A stable random variable X is commonly described by its characteristic function (CF), which is defined by

$$\begin{aligned} \Phi_X(t; \alpha, \beta, \sigma, \mu) &= \mathbf{E}[\exp(iXt)] \\ &= \exp(-\sigma^\alpha |t|^\alpha (1 - i\beta \operatorname{sgn}(t) \tan(\frac{\pi\alpha}{2})) + i\mu t), \quad \text{if } \alpha \neq 1, \end{aligned}$$

$$\begin{aligned} \Phi_X(t; \alpha, \beta, \sigma, \mu) &= \mathbf{E}[\exp(iXt)] \\ &= \exp(-\sigma |t| (1 + i\beta \frac{2}{\pi} \operatorname{sgn}(t) \ln(t)) + i\mu t), \quad \text{if } \alpha = 1, \end{aligned}$$

where

$$\operatorname{sgn}(t) = \begin{cases} 1 & \text{if } t > 0 \\ 0 & \text{if } t = 0 \\ -1 & \text{if } t < 0, \end{cases}$$

$0 < \alpha \leq 2$ is the index of stability, $-1 \leq \beta \leq 1$ is the skewness parameter, $\sigma \geq 0$ is the scale parameter, and $\mu \in \mathbb{R}$ is the location parameter. To indicate the dependence of a stable random variable X on its parameters, we write $X \sim S(\alpha, \beta, \sigma, \mu)$. Define that

$$Z = \frac{X - \mu}{\sigma}, \tag{2.9}$$

then, $Z \sim S(\alpha, \beta, 1, 0)$ is called a standard stable random variable with PDF $f(x; \alpha, \beta, 1, 0)$. Therefore, the PDF for X , $f(x; \alpha, \beta, \sigma, \mu)$, can be expressed by the standardized stable PDF such that

$$f(x; \alpha, \beta, \sigma, \mu) = \frac{1}{\sigma} f\left(\frac{x - \mu}{\sigma}; \alpha, \beta, 1, 0\right). \tag{2.10}$$

In addition, for $\sigma = 1$, $\mu = 0$ and $\alpha \neq 1$, the CF becomes

$$\Phi_X(t) = \exp(-|t|^\alpha + i\beta t |t|^{\alpha-1} \tan(\frac{\pi\alpha}{2})) \tag{2.11}$$

In empirical studies, the modeling of financial data is done typically by stable distributions with $1 < \alpha < 2$, which is called stable paretian distribution. Stable distributions are unimodal and the smaller α is, the stronger the leptokurtic feature the distribution has (the peak of the density becomes higher and the tails are heavier). Therefore, the index of stability α can be interpreted as a measure of kurtosis. When $\alpha > 1$, the location parameter μ measures the mean of the distribution. If the skewness parameter $\beta = 0$, the distribution of X is symmetric, if $\beta > 0$, the distribution is skewed to the right and if $\beta < 0$, the distribution is skewed to the left. Larger magnitude of β indicates stronger skewness. Furthermore, the p^{th} absolute moment $\mathbf{E}[|X|^p]$ is finite only when $\alpha > p$ or $\alpha = 2$.

Since we can not get closed form of PDF and CDF for stable distributions (except that $\alpha = 2$), we calculate them numerically. The approach to approximate the PDF is applying Fast Fourier Transform (FFT) to the CF, see [31], [33]. To briefly summarize the FFT-based approximation, recall that the PDF can be written in terms of CF as

$$f(x; \alpha, \beta, \sigma, \mu) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ixt} \Phi_X(t; \alpha, \beta, \sigma, \mu) dt. \quad (2.12)$$

The integral in (2.12) can be calculated for N equally-spaced points with distance h , such that $x_k = (k - 1 - N/2)h$, $k = 1, \dots, N$. Setting $t = 2\pi\omega$, (2.12) implies

$$f((k - 1 - \frac{N}{2})h) = \int_{-\infty}^{\infty} e^{-i2\pi\omega(k-1-\frac{N}{2})h} \Phi_X(2\pi\omega) d\omega.$$

The integration can be approximated by

$$\begin{aligned} f((k - 1 - \frac{N}{2})h) &\approx s \sum_{n=1}^N \Phi_X(2\pi s(n - 1 - \frac{N}{2})) e^{-i2\pi(n-1-\frac{N}{2})(k-1-\frac{N}{2})hs} \\ &= s(-1)^{k-1-\frac{N}{2}} \\ &\quad \times \sum_{n=1}^N (-1)^{n-1} \Phi_X(2\pi s(n - 1 - \frac{N}{2})) e^{-i2\pi(n-1)(k-1)/N}, \end{aligned} \quad (2.13)$$

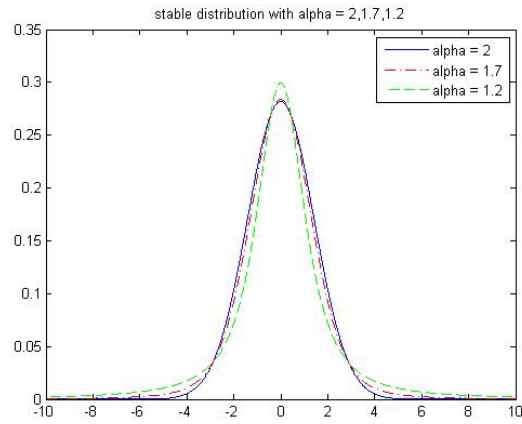
where $s = (hN)^{-1}$. The summation in (2.13) can be efficiently computed by applying FFT to the sequence

$$(-1)^{n-1} \Phi_X(2\pi s(n - 1 - \frac{N}{2})), \quad n = 1, \dots, N.$$

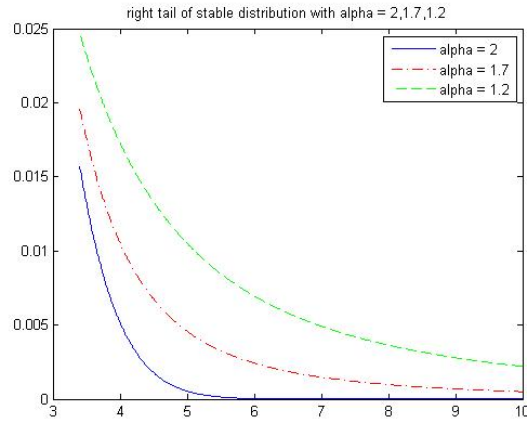
Normalizing the k^{th} element of this sequence by $s(-1)^{k-1-\frac{N}{2}}$, we obtain the approximate PDF value for each grid point. For any x , we can use linear interpolation to obtain its PDF. By substituting (2.11) into (2.13), standardized PDF values can be calculated and, via (2.9) and

(2.10), PDF can be obtained with any desired parameter combinations.

When $\alpha = 2$ and $\beta = 0$, $S(2, 0, 1, 0)$ is reduced to Gaussian distribution. When $\alpha < 2$, $S(\alpha, 0, 1, 0)$ have fatter tail than Gaussian distribution. We can see this from Figure 2.2, the smaller α is, the fatter the tail is. We can tell how β influences the skewness from Figure 2.3.



(a) stable pdf



(b) stable right tail pdf

Figure 2.2: Standard stable distribution with $\beta = 0$, $\alpha = 2, 1.7, 1.2$

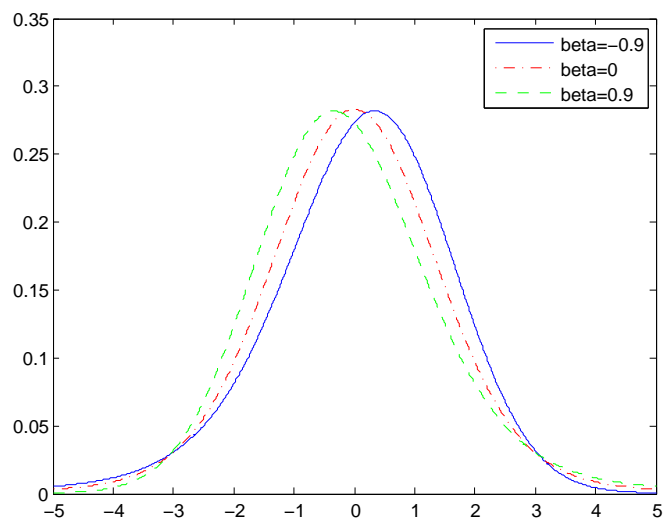


Figure 2.3: Standard stable distribution with $\alpha = 1.8$ and $\beta = -0.9, 0, 0.9$

2.3 Copula Function

2.3.1 Copula

To describe the financial data properly, the marginal distribution for each individual asset is not enough. We also need the joint distribution of the assets. In early time, people just assumed that marginal distributions of assets are normal and the joint distribution is multi-variate normal, which is easy. However, as discussed in the previous section, the marginal distribution may not be Gaussian, but some other fat-tailed distributions. In order to get the joint distribution, we need a tool to connect the joint distribution of the assets with the marginal distributions of each asset. Copula provides us such a convenient tool. The definition of copula is given as follows.

Definition 2.3.1. *A d -dimensional copula is the joint distribution of random variables U_1, U_2, \dots, U_d , each of which is marginally uniformly distributed as $U(0, 1)$, i.e.*

$$C(u_1, u_2, \dots, u_d) = P(U_1 \leq u_1, U_2 \leq u_2, \dots, U_d \leq u_d).$$

Sklar's theorem states the importance of copula as follows:

Theorem 2.3.2. (Sklar's Theorem) *For any random variable X_1, X_2, \dots, X_d with joint cumulative distribution function (CDF)*

$$F(x_1, x_2, \dots, x_d) = P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_d \leq x_d)$$

and marginal CDFs

$$F_j(x) = P(X_j \leq x), \quad j = 1, 2, \dots, d,$$

there exists a copula C such that

$$F(x_1, x_2, \dots, x_d) = C(F_1(x_1), F_2(x_2), \dots, F_d(x_d)). \quad (2.14)$$

In addition,

$$C(u_1, u_2, \dots, u_d) = F(F_1^{-1}(u_1), F_2^{-1}(u_2), \dots, F_d^{-1}(u_d)),$$

where F_j^{-1} is the generalized inverse of F_j , $j = 1, 2, \dots, d$. If each X_j is a continuous random variable, C is unique.

Sklar's theorem allows us to formulate the joint distribution by separating the marginal distributions $F_j(x)$ and the dependence structure, which is expressed by copula C .

There are a lot of copulas. In this thesis, we introduce two parametric families of copulas: the elliptical copulas and the archimedean copulas. Elliptical copula includes Gaussian and student's t copula and archimedean copula contains clayton, gumbel and frank copula.

Multivariate Gaussian Copula The multivariate Gaussian copula is the copula of the multivariate normal distribution. It can be written as

$$C_R^{Ga}(u_1, \dots, u_d) = \phi_R^d(\phi^{-1}(u_1), \dots, \phi^{-1}(u_d)),$$

where ϕ^{-1} is the inverse of the standard univariate normal distribution function ϕ . ϕ_R^d is the d -dimensional multi-normal distribution function, with R to be the correlation matrix. Gaussian copula is a very popular elliptical copula, but it cannot fit to possess tail dependence. Gaussian copula has neither upper nor lower tail dependence.

Multivariate Student's t Copula The multivariate student's t copula is the copula of the multivariate student's t distribution, which is given by

$$C_{\rho, R}^t(u_1, \dots, u_d) = t_{\rho, R}^d(t_{\rho}^{-1}(u_1), \dots, t_{\rho}^{-1}(u_d)),$$

where $t_{\rho, R}^d$ is the standardized d -dimensional multivariate student's t distribution with correlation matrix R and degree of freedom ρ and t_{ρ}^{-1} is the inverse of the univariate cumulative distribution function of student's t with ρ as degree of freedom.

The student's t copula is generally superior to the Gaussian one on the tail dependence, which allows joint fat tails and an increased probability of joint extreme events compared with the Gaussian copula. Student's t copula has an additional parameter, namely the degree of freedom ρ . Increasing the value of ρ decreases the tendency to exhibit extreme co-movements.

Multivariate Clayton Copula The clayton copula, which is an asymmetric copula, exhibits greater dependence in the negative tail than in the positive. The multivariate clayton copula can be written as

$$C_{\gamma}^{Cl}(u_1, \dots, u_d) = \left[\sum_{i=1}^d u_i^{-\gamma} - d + 1 \right]^{-\frac{1}{\gamma}},$$

where $\gamma > 0$ is a parameter controlling the dependence. Perfect dependence is obtained if $\gamma \rightarrow \infty$, while $\gamma = 0$ implies independence.

Multivariate Gumbel Copula The gumbel copula is also an asymmetric copula, but it exhibits greater dependence in the positive tail than in the negative. The multivariate gumbel

copula is given by

$$C_\gamma^{Cu}(u_1, \dots, u_d) = \exp \left[- \left[\sum_{i=1}^d (-\ln(u_i))^\gamma \right]^{\frac{1}{\gamma}} \right],$$

where $\gamma > 0$ is a parameter controlling the dependence. Perfect dependence is obtained if $\gamma \rightarrow \infty$, while $\gamma = 1$ implies independence.

Multivariate Frank Copula The Frank copula is a symmetric copula. The multivariate Frank copula is given by

$$C_\gamma^{Fr}(u_1, \dots, u_d) = -\frac{1}{\gamma} \ln \left[1 + \frac{\prod_{i=1}^d (e^{-\gamma u_i} - 1)}{(e^{-\gamma} - 1)^{d-1}} \right]$$

where $\alpha > 0$ is a parameter controlling the dependence.

2.3.2 Parameter Estimation for Copula

There are two approaches to estimate the parameters for copulas. The most direct estimation method is to simultaneously estimate all the parameters in both marginal distributions and copula using full maximum likelihood method. However, this approach will give big computational burden. Therefore, we introduce the second approach, named 2-step maximum likelihood method, where the parameters for marginal distributions are estimated in the first step and the dependence parameters in copula is estimated in the second step after the estimated marginal distributions have been substituted into it. The estimation in the first step can be obtained by maximum likelihood method. The estimation in the second step is as follows. Suppose the parameter of the copula is θ , from (2.14), the probability density function for the joint distribution of X_1, X_2, \dots, X_d can be derived as:

$$\begin{aligned} f(x_1, x_2, \dots, x_d; \theta) &= \frac{\partial^d F(x_1, x_2, \dots, x_d; \theta)}{\partial x_1 \partial x_2 \cdots \partial x_d} \\ &= \frac{\partial^d C(F_1(x_1), F_2(x_2), \dots, F_d(x_d); \theta)}{\partial x_1 \partial x_2 \cdots \partial x_d} \\ &= \frac{\partial^d C(F_1, F_2, \dots, F_d; \theta)}{\partial F_1 \partial F_2 \cdots \partial F_d} \frac{\partial F_1(x_1)}{x_1} \cdots \frac{\partial F_1(x_1)}{x_1} \\ &= c(F_1(x_1), F_2(x_2), \dots, F_d(x_d); \theta) f_1(x_1) f_2(x_2) \cdots f_d(x_d), \end{aligned} \tag{2.15}$$

where $c(u_1, u_2, \dots, u_d; \theta) = \frac{\partial^d C(u_1, u_2, \dots, u_d; \theta)}{\partial u_1 \partial u_2 \dots \partial u_d}$, and f_j is the marginal probability density function for the j^{th} variable, which is known in the first step, $j = 1, 2, \dots, d$. If we have n observations of the multivariate random vector $X = (X_1, X_2, \dots, X_d)$, which is $x_i = (x_{i1}, x_{i2}, \dots, x_{id})$, $i = 1, 2, \dots, n$. Then the log-likelihood function can be obtained from (2.15)

$$\begin{aligned}
L(\theta) &= \ln\left(\prod_{i=1}^n f(x_{i1}, x_{i2}, \dots, x_{id}; \theta)\right) \\
&= \ln\left(\prod_{i=1}^n c(F_1(x_{i1}), F_2(x_{i2}), \dots, F_d(x_{id}); \theta) f_1(x_{i1}) f_2(x_{i2}) \cdots f_d(x_{id})\right) \\
&= \sum_{i=1}^n \sum_{j=1}^d \ln(f_j(x_{ij})) + \sum_{i=1}^n \ln(c(F_1(x_{i1}), F_2(x_{i2}), \dots, F_d(x_{id}); \theta)).
\end{aligned} \tag{2.16}$$

Then, the estimator of parameters θ for a copula is the maximizer of the log-likelihood function $L(\theta)$ via θ .

2.4 Time Series Models

Let $\{r_t\}$ be a stochastic process via $t \in \{1, 2, \dots\}$. Usually, if r_t represents financial data, empirical study shows that they are not i.i.d. and the current values of r_t may be influenced by the past values. If the present values can be plausibly modeled in terms of only the past values, we have the enticing prospect that forecasting will be possible.

2.4.1 Measure of Dependence

The dependence of the series r_t can be measured by a function called autocorrelation function (ACF), that is defined as follows.

Definition 2.4.1. *The **autocovariance function** of the series r_t is defined as the second moment product*

$$\gamma_r(s, t) = \mathbf{E}[(r_s - \mu_s)(r_t - \mu_t)],$$

for all $s, t = 1, 2, \dots$, where $\mu_t = \mathbf{E}[r_t]$. The **autocorrelation function (ACF)** is defined as

$$\rho_r(s, t) = \frac{\gamma_r(s, t)}{\sqrt{\gamma_r(s, s)\gamma_r(t, t)}}.$$

The series we are interested in are something called **stationary times series**. Usually, for stationary here, we mean weakly stationary defined below.

Definition 2.4.2. A *weakly stationary* time series, r_t , is a finite variance process such that

- (i) the mean value function μ_t is constant and does not depend on time t , and
- (ii) the autocovariance function, $\gamma_r(s, t)$, depends on s and t only through their difference $|s - t|$.

Therefore, the autocovariance function and ACF for a stationary time series r_t with the time difference h will be written as

$$\gamma_r(h) = \mathbf{E}[(r_{t+h} - \mu)(r_t - \mu)]$$

and

$$\rho_r(h) = \frac{\gamma_r(t+h, t)}{\sqrt{\gamma_r(t+h, t+h)\gamma_r(t, t)}} = \frac{\gamma_r(h)}{\gamma_r(0)},$$

where $\mu_t = \mu$ is a constant. To obtain the sample autocovariance function and ACF for a stationary time series r_t , suppose we have n sample of r_t , r_1, r_2, \dots, r_n . Let

$$\bar{r} = \frac{1}{n} \sum_{i=1}^n r_i.$$

Then, the sample autocovariance function is defined as

$$\hat{\gamma}_r(h) = \frac{1}{n} \sum_{i=1}^{n-h} (r_{i+h} - \bar{r})(r_i - \bar{r}),$$

and the sample ACF is defined as

$$\hat{\rho}_r(h) = \frac{\hat{\gamma}_r(h)}{\hat{\gamma}_r(0)}.$$

There are several models for stationary time series with self dependence structure, the most famous one is autoregressive moving average (ARMA) model.

2.4.2 ARMA Model

The autoregressive and moving average (ARMA) models are used in time series analysis to describe stationary time series. These models represent time series that are generated by passing white noise through a recursive and through a nonrecursive linear filter, consecutively.

The ARMA model with the order p and q is defined as:

$$r_t - \sum_{i=1}^p \phi_i r_{t-i} = \phi_0 + \epsilon_t + \sum_{i=1}^q \psi_i \epsilon_{t-i},$$

where r_t is the time series of data and ϵ_t is the white noise. The model consists of two parts, a p -order autoregressive (AR) part and a q -order moving average (MA) part. ϕ_i , $i = 1, \dots, p$ are the coefficients of the AR part, ψ_i , $i = 1, \dots, q$ are the coefficients of the MA part and ϕ_0 is constant. Therefore, we call it ARMA(p, q) model. ARMA model introduces an important parametric family of stationary series, under the condition that the roots of $1 - \sum_{i=1}^p \phi_i z^i = 0$ all lie outside the unit circle. To fit the ARMA model to the sample data, we need to decide the order p and q . To obtain the order p and q from the sample of r_t , we need another function called partial autocorrelation function (PACF).

Definition 2.4.3. Let r_h^{h-1} denote the regression of r_h on $\{r_{h-1}, r_{h-2}, \dots, r_1\}$, which we write

$$r_h^{h-1} = \beta_1 r_{h-1} + \beta_2 r_{h-2} + \dots + \beta_{h-1} r_1.$$

In addition, let r_0^{h-1} denote the regression of r_0 on $\{r_1, r_2, \dots, r_{h-1}\}$, which we write

$$r_0^{h-1} = \beta_1 r_1 + \beta_2 r_2 + \dots + \beta_{h-1} r_{h-1}.$$

Then the **partial autocorrelation function (PACF)** of a stationary process r_t denoted ϕ_{hh} , for $h = 1, 2, \dots$, is

$$\phi_{11} = \text{corr}(r_1, r_0) = \rho(1)$$

and

$$\phi_{hh} = \text{corr}(r_h - r_h^{h-1}, r_0 - r_0^{h-1}), \quad h \geq 2.$$

The order of ARMA can be obtained from both ACF and PACF we defined above. In theory, for MA(q) model, the ACF cuts off after q lags and for AR(p) model, the PACF cuts off after p lags. With this property, we can check the sample ACF and PACF, they can help us to decide the order of ARMA.

ARMA models describe time series with self dependence structure with constant variance. However, they do not allow volatility clustering, or heteroscedasticity. But volatility clustering is a common characteristic of financial data. Therefore, ARMA model alone can not describe the financial series properly, we need some heteroscedastic models, which is GARCH model. It will be introduced in the next section.

2.4.3 Heteroscedastic Model

Autoregressive Conditional Heteroscedasticity (ARCH) model is first introduced by Engle (1982,[13]) to describe heteroscedasticity. It considers the variance of the current error term or innovation to be a function of the actual sizes of the previous time periods' error terms. Often, the variance

is related to the squares of the previous innovations, but we also admit some other functions like absolute value, logarithm etc. The ARCH model is formulated as

$$r_t = \sigma_t z_t,$$

$$\sigma_t^2 = a_0 + \sum_{i=1}^m a_i r_{t-i}^2,$$

where σ_t is the standard deviation of r_t , a_0 is a constant, $a_i > 0$, $i = 1, \dots, m$ are the coefficients and must satisfy $\sum_{i=1}^m a_i < 1$. z_t are i.i.d. and they follow a distribution with mean 0 and variance 1. We call this model as ARCH(m) model. Traditionally, z_t is standard normal distribution, but we can use some other distributions.

Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model is an extended model for ARCH introduced by Bollerslev (1986,[7]). GARCH allows the variance of the current error term depends also on the previous error variance term. GARCH(m,s) model is written as

$$r_t = \sigma_t z_t,$$

$$\sigma_t^2 = a_0 + \sum_{i=1}^m a_i r_{t-i}^2 + \sum_{i=1}^s b_i \sigma_{t-i}^2,$$

where z_t , a_0 and a_i , $i = 1, \dots, m$ are the same as ARCH model, and $b_i > 0$, $i = 1, \dots, s$ are also the coefficients. Moreover, a_i , $i = 1, \dots, m$ and b_i , $i = 1, \dots, s$ must satisfy $\sum_{i=1}^{\max(m,s)} (a_i + b_i) < 1$ (for $i > m$, $a_i = 0$, for $i > s$, $b_i = 0$).

Usually, GARCH(m,s) model can be modeled together with ARMA(p,q) model, we call it ARMA(p,q)-GARCH(m,s) model, which is widely used for financial data. In order to reduce the computational burden, we use $m = 1$, $s = 1$ for GARCH model. Empirical studies support that GARCH(1,1) model works well for financial data (see Bali and Theodossiou [4], Hansen 1994 [19], Liu and Brorsen 1995 [27] and Panorska, Mittnik and Rachev 1995 [36] etc). Given the financial data, we can estimate the parameters for the model using maximum likelihood method.

2.5 Data Selection

Throughout this thesis, we pick up five members of S&P500 index, which are from five totally different industrials, together with S&P500 index. These five members are Progress Energy Inc.

(PGN), International Business Machines Corp. (IBM), Bank of America Corp. (BAC), Ford Motor Co. (F) and Home Depot Inc. (HD). These five stocks represent the five major industrials constitute into S&P500 index. We observe the price of these five stocks and the S&P500 index from the time period Jan 1, 2000 to Aug 10, 2009, which gives us 2415 observations. for the index and each stock, let R_t be the daily financial rate of return of the stock or index at day t and P_t be the stock or index at day t . Assume that the rate of return is continuous, then, we have

$$P_t = P_{t-1}e^{R_t}.$$

Therefore, daily rate of return data are obtained by taking the first differences of the logarithm of series of stock prices

$$R_t = \ln(P_t) - \ln(P_{t-1}).$$

Since the daily rate of return is too small, we consider the percentage, i.e.

$$R_t = 100 [\ln(P_t) - \ln(P_{t-1})]. \quad (2.17)$$

The daily rate of loss is then $L_t = -R_t$. Table 2.1 shows the statistics of these six financial rate of loss data.

Table 2.1: Statistics for 6 different observed loss data

stock	industrial	mean	variance	skewness	kurtosis
PGN	Utilities	-0.0049	0.4185	0.0127	9.9113
IBM	Information Technology	-0.000414	0.7121	-0.0140	9.3080
BAC	Financials	0.0192	2.7190	2.7023	65.2862
F	Auto-Cars/Light Trucks	0.0344	2.1649	1.0004	27.3520
HD	Retail-Building Products	0.0158	1.1333	0.9304	20.7729
S&P500	Index	0.0066	0.3772	0.0926	10.6326

From the Table 2.1, for each stock and the index, we can see that the skewness of the data is not zero, meaning that the data are asymmetric. In addition, the kurtosis of the data is big, which suggests that the data are fat-tailed. Therefore, Gaussian distribution is not a good choice to model the data. We use some other distributions, such as GST and stable distribution, which can describe both the skewness and kurtosis characteristics of the data. The series plot of the 5 stocks and S&P500 index losses are shown in Figure 2.4. From Figure

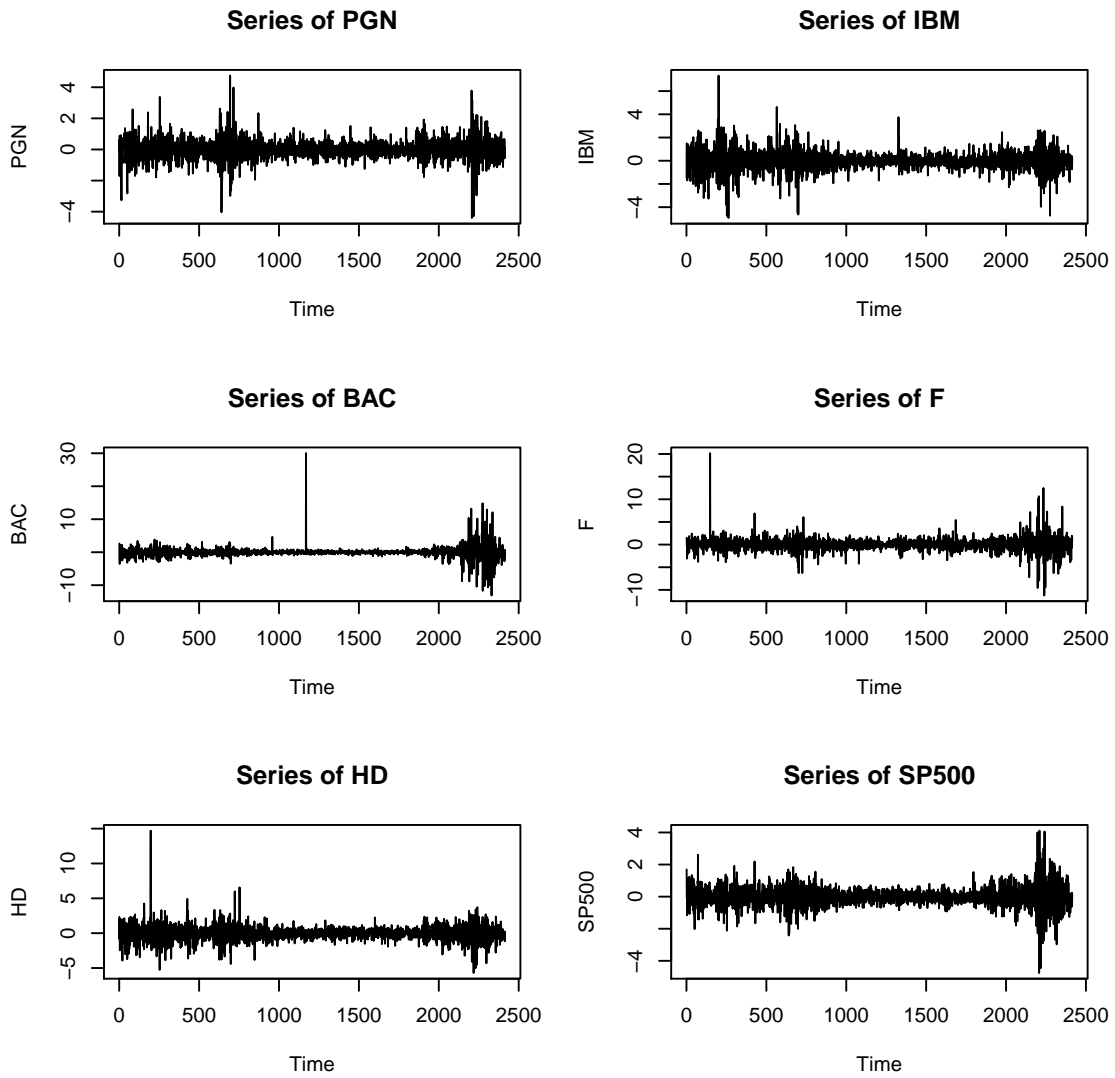


Figure 2.4: Rate of loss time series for 5 stocks and S&P500 index

2.4, we can see that for each of the series, large (small) changes tend to be followed by large (small) changes. Therefore, ARMA model alone can not describe the financial series properly, we need the heteroscedastic model. Here we use GARCH model introduced in previous section. For the order of ARMA, we check the ACF and PACF for the series. We show the ACFs and PACFs of the 5 stocks' and S&P500 index's loss of returns in Figure 2.5. From Figure 2.5, we can see that for most of the series, the ACF and PACF are cut off at order zero. Therefore, to make the model simple and easier, we take $p = q = 0$. For GARCH we take the simplest one $m = s = 1$.

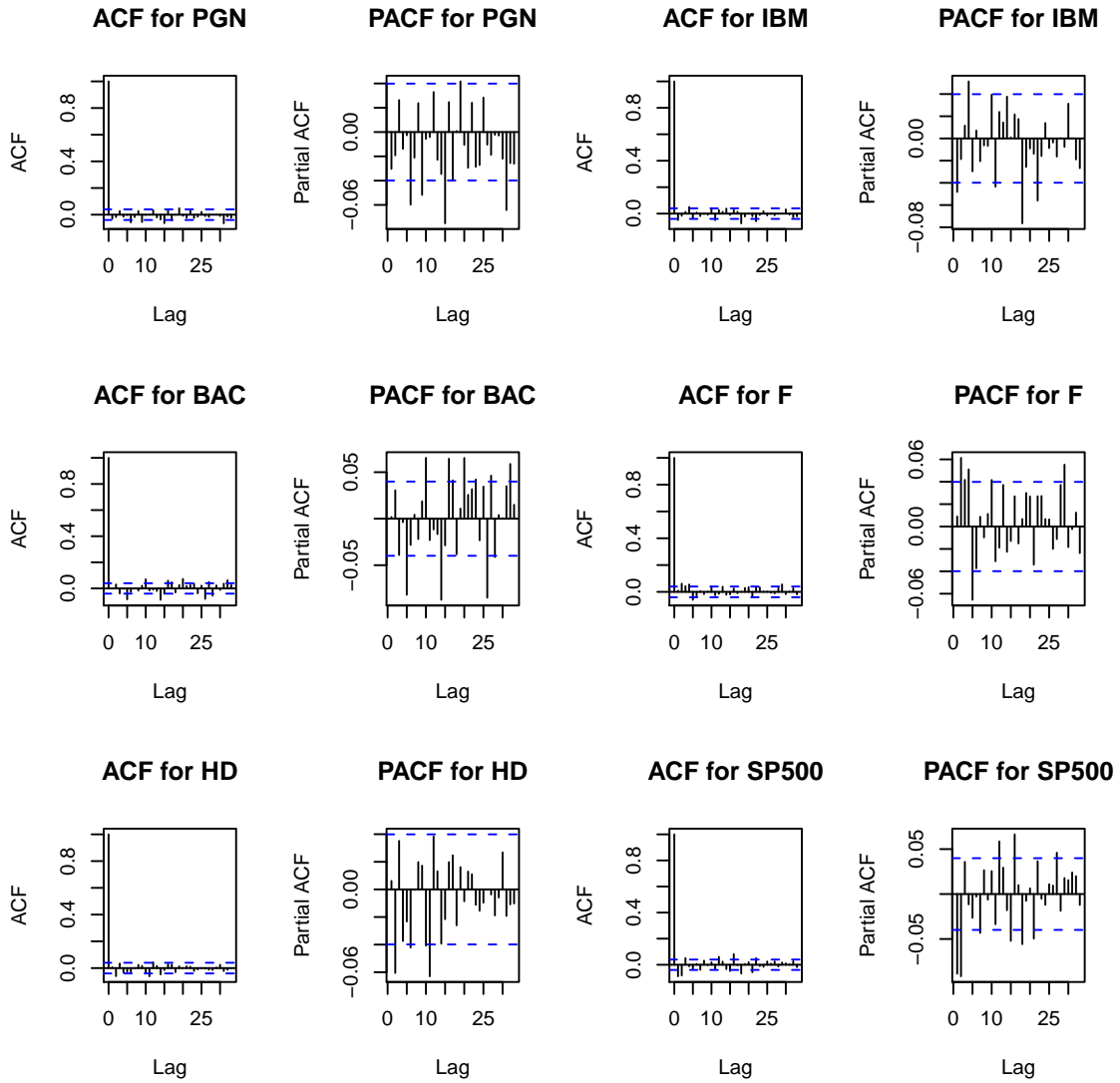


Figure 2.5: ACF and PACF of 5 stocks and S&P500 index rate of loss

Chapter 3

Risk of Stock Against Benchmark

Performance of a stock can be tracked by the behavior of the stock price. However, stock prices can not fully represent its performance. Frequently, people tend to compare it with overall market performance, which can serve as benchmark. Benchmark is a base against which the performance of an individual security or group of securities is measured. A benchmark is usually an index of securities of the same or similar class. Stocks are usually compared against stocks; bonds against bonds, etc. For example, S&P500 and DJIA are the most widely used benchmarks for stocks. In this chapter, we use S&P500 as the benchmark. We describe model both individual stock return and benchmark return with time series model and describe their correlations by copula. By Monte Carlo simulation method, we are able to estimate the VaR and CVaR for the relative return, which is defined by the difference of the two returns. By an approach called backtesting, we figure out the proper copula for dependence structure of stock and benchmark returns is student's t copula.

The rest of this chapter is arranged as follows. In Section 3.1, we set up the basic time series models for stock and benchmark returns, respectively. In Section 3.2, copula is introduced to describe the tail dependence between the two returns. Backtesting method is discussed in Section 3.3. Numerical results and discussion are given in Section 3.4.

3.1 Relative Return Based on The Benchmark

Performance of the stock is represent by the relative return of individual stock return against the return of the Benchmark. We use S&P500 as the benchmark for stocks. The returns are defines as in (2.17):

$$R_t = 100 [\ln(P_t) - \ln(P_{t-1})],$$

where P_t is the individual stock price or S&P500 index at time t . As we discussed in section 2.5, financial data show not only asymmetry and fat-tail (see Table 2.1), but also time-varying volatility clustering (see Figure 2.4). Thus, we can use ARMA-GARCH to the market risk. To avoid computational burden, we use the simplest ARMA(0,0)-GARCH(1,1) model. To describe the asymmetry and fat-tail property, we use stable distribution as the innovation in GARCH model. In addition, since stable distributions do not have the second absolute expectation (except that $\alpha = 2$), we use absolute function instead of the square function in GARCH model. The model is set up as follows.

Individual stock return is modeled as

$$R_t = a_0 + \epsilon_t = \mu_t + \sigma_t z_t, \quad (3.1)$$

$$|\sigma_t| = b_0 + b_1 |\epsilon_{t-1}| + b_2 |\sigma_{t-1}|, \quad (3.2)$$

where R_t is the series of individual stock return at time t , μ_t and σ_t are, respectively, the conditional mean and conditional standard deviation of R_t . $\epsilon_t = \sigma_t z_t$, where z_t are i.i.d. standardized stable paretian random variables, $z_t \sim S(\alpha_1, \beta_1, 1, 0)$, $1 < \alpha_1 < 2$.

The benchmark S&P500 return is modeled as

$$R_{mt} = c_0 + \epsilon_{mt} = \mu_{mt} + \sigma_{mt} z_{mt}, \quad (3.3)$$

$$|\sigma_{mt}| = d_0 + d_1 |\epsilon_{mt-1}| + d_2 |\sigma_{mt-1}|, \quad (3.4)$$

where R_{mt} is the series of S&P500 index return at time t , μ_{mt} and σ_{mt} are, respectively, the conditional mean and conditional standard deviation of R_{mt} . $\epsilon_{mt} = \sigma_{mt} z_{mt}$, where z_{mt} are also i.i.d. standardized stable paretian distributions random variables, $z_{mt} \sim S(\alpha_2, \beta_2, 1, 0)$, $1 < \alpha_2 < 2$.

To estimate the parameters in (3.1), (3.2) (3.3) and (3.4), we use Maximum Likelihood Estimation (MLE) method, where the PDFs of z_t and z_{mt} are approximated by FFT method discussed in 2.2.2.

After we set up models for the stock and benchmark returns, we can obtain the relative return as the difference of the above two returns, i.e.

$$H_t = R_t - R_{mt}. \quad (3.5)$$

H_t can describe the performance of the stock at time t , compared with the overall market performance, which is described by the benchmark S&P500. If $H_t > 0$, the stock outperforms the overall market and vice versa.

3.2 Copula Based Risk Measurements

Because S&P500 is a weighted index of the prices of 500 large-cap common stocks actively traded in the United States, we can see that the individual stock return and the benchmark return are not independent with each other, i.e., z_t and z_{mt} in (3.1) and (3.3) are dependent. The dependent structure can be described by copula introduced in Section 2.3. Our goal is to measure the risk of the relative return, which is the difference of the stock return and benchmark return. Let the relative return of the stock (H) be defined as in (3.5), where R_t and R_{mt} are dependent with each other. To measure the risk, we use VaR and CVaR introduced in 2.1. Then, we can use Monte Carlo simulation to simulate the stock and benchmark returns with different copulas, obtaining the simulated sample total returns h_1, \dots, h_N , by the difference of the two simulated returns, which are obtained from (3.1)-(3.4). Finally, we will get the estimated VaR and CVaR for the relative return by point estimation:

$$\widehat{VaR}_\epsilon(H) = -h_{[Np]} \quad (3.6)$$

and

$$\widehat{CVaR}_\epsilon(H) = -\frac{1}{Np} \sum_{i=1}^{Np} h_{[i]}, \quad (3.7)$$

respectively. where $h_{[1]}, \dots, h_{[N]}$ is the ascending order statistics for the realization of H and $p = 1 - \epsilon$.

The detailed algorithm is shown below:

- Step 1 Get the observed data R_i and R_{mi} , $i = 1, \dots, t$ of the individual stock return and the index return.
- Step 2 Use MLE to estimate the parameters for the GARCH model and the stable distribution in (3.1)-(3.4). Forecast μ_{t+1} , σ_{t+1} and μ_{mt+1} , σ_{mt+1} by (3.1)-(3.4).
- Step 3 Calculate the observed z_i and z_{mi} from (3.1)-(3.4), using the parameter we get in step 2.
- Step 4 Fit the copula we used with the observed z_t and z_{mt} to get the parameters for the copula.

- Step 5 Simulate N realizations of the coupled z_{t+1} and z_{mt+1} with the copula we used, get $\widehat{z}_{t+1,i}$ and $\widehat{z}_{mt+1,i}$. Obtain the estimated N of stock return $\widehat{r}_{t+1,i}$ and benchmark return $\widehat{r}_{mt+1,i}$ from (3.1)-(3.4), $i = 1, \dots, N$.
- Step 6 Calculate the simulated relative return $h_{t+1,i} = \widehat{r}_{t+1,i} - \widehat{r}_{mt+1,i}$, $i = 1, \dots, N$.
- Step 7 Calculate estimation of VaR and CVaR for H using the simulated data $h_{t+1,i}$, $i = 1, \dots, N$, by (3.6) and (3.7).

3.3 Backtesting

Hypothesis testing will be discussed in this section. Let ϵ be the confidence level for VaR calculation. Suppose we use the most recent k historical data to forecast the current VaR. For example, if $k = 502$, we use the most recent two years' historical data to predict the current VaR. If we have totally n data of historical returns, define the indicator for VaR violations as follows:

$$I_s = \begin{cases} 1, & \text{if } H_{k+s} < -VaR_{k+s}; \\ 0, & \text{otherwise;} \end{cases}$$

where $s = 1, \dots, n - k$. Note that according to [15] and [20], I_{s1} and I_{s2} are independent for $s1 \neq s2$. As a result,

$$X_\epsilon = \sum_{s=1}^{n-k} I_s \sim B(n - k, 1 - \epsilon),$$

i.e. the total number of violations is binomially distributed with parameter $n - k$ and $1 - \epsilon$. Let the null hypothesis be that VaR is accurate, and the null distribution is binomial with parameter $n - k$ and $1 - \epsilon$. From the binomial law, the expected (mean) number of violations is $(1 - \epsilon)(n - k)$. The p -value in the hypothesis testing is the probability of falsely rejecting the null hypothesis. A small p -value can be interpreted as weak support for the null hypothesis, while a large p -value signals a strong support for the null hypothesis. Here we take 0.05 as the level, i.e. when the p -value is less than 0.05 we reject the hypothesis that VaR is accurate. This test is called backtesting.

3.4 Numerical Results and Discussion

We choose the data of the stock market from Jan 1, 2000 to Aug 10, 2009, five individual stocks (PGN, IBM, BAC, F and HD) for stock return and stock index S&P500 for benchmark return. The total number of the observed return data is $n = 2414$. We use the most recent 2 year historical data to estimate the parameters of (3.1)-(3.4) and copulas, then we use these parameters to predict the VaR for the next time, that is we take $k = 502$. Therefore, the expected

number of violations $\mathbf{E}[X_\epsilon]$ is $(1 - \epsilon)(n - k) = 1912(1 - \epsilon)$. We take $\epsilon = 0.95, 0.975$ and 0.99 . Following the algorithm in Section 3.2, we conduct the backtesting procedure in 3.3. We give the backtesting results in Table 3.1 to 3.5 for the five different copulas and five stocks.

Table 3.1: Backtesting result for PGN and S&P500

Copula	$\epsilon = 0.95, \mathbf{E}[X_\epsilon] = 95.56$		$\epsilon = 0.975, \mathbf{E}[X_\epsilon] = 47.78$		$\epsilon = 0.99, \mathbf{E}[X_\epsilon] = 19.12$	
	$X_{.95}$	p-value	$X_{.975}$	p-value	$X_{.99}$	p-value
Gaussian	66	0.0001	38	0.083	17	0.3672
Student's t	69	0.0024	33	0.0144	16	0.2817
Clayton	57	0.0000	32	0.0093	9	0.0080
Gumbel	60	0.0000	30	0.0036	11	0.0321
Frank	64	0.0002	32	0.0093	7	0.0013

Table 3.2: Backtesting result for IBM and S&P500

Copula	$\epsilon = 0.95, \mathbf{E}[X_\epsilon] = 95.56$		$\epsilon = 0.975, \mathbf{E}[X_\epsilon] = 47.78$		$\epsilon = 0.99, \mathbf{E}[X_\epsilon] = 19.12$	
	$X_{.95}$	p-value	$X_{.975}$	p-value	$X_{.99}$	p-value
Gaussian	82	0.0822	47	0.4917	24	0.1110
Student's t	83	0.1002	47	0.4917	21	0.2832
Clayton	60	0.0000	29	0.0021	17	0.3672
Gumbel	70	0.0031	37	0.0614	17	0.3672
Frank	76	0.0198	35	0.0212	15	0.2058

Table 3.3: Backtesting result for BAC and S&P500

Copula	$\epsilon = 0.95, \mathbf{E}[X_\epsilon] = 95.56$		$\epsilon = 0.975, \mathbf{E}[X_\epsilon] = 47.78$		$\epsilon = 0.99, \mathbf{E}[X_\epsilon] = 19.12$	
	$X_{.95}$	p-value	$X_{.975}$	p-value	$X_{.99}$	p-value
Gaussian	90	0.3002	49	0.3933	22	0.2141
Student's t	93	0.4187	47	0.4917	16	0.2817
Clayton	70	0.0031	37	0.0614	17	0.3672
Gumbel	81	0.0668	40	0.1415	15	0.2058
Frank	83	0.1002	40	0.1415	13	0.0928

Table 3.4: Backtesting result for F and S&P500

Copula	$\epsilon = 0.95, \mathbf{E}[X_\epsilon] = 95.56$		$\epsilon = 0.975, \mathbf{E}[X_\epsilon] = 47.78$		$\epsilon = 0.99, \mathbf{E}[X_\epsilon] = 19.12$	
	$X_{.95}$	p-value	$X_{.975}$	p-value	$X_{.99}$	p-value
Gaussian	88	0.2302	46	0.4333	26	0.0507
Student's t	91	0.3382	44	0.3208	21	0.2832
Clayton	82	0.0823	44	0.3208	26	0.0507
Gumbel	76	0.0198	39	0.1096	20	0.3629
Frank	86	0.1701	41	0.1788	20	0.3629

Table 3.5: Backtesting result for HD and S&P500

Copula	$\epsilon = 0.95, \mathbf{E}[X_\epsilon] = 95.56$		$\epsilon = 0.975, \mathbf{E}[X_\epsilon] = 47.78$		$\epsilon = 0.99, \mathbf{E}[X_\epsilon] = 19.12$	
	$X_{.95}$	p-value	$X_{.975}$	p-value	$X_{.99}$	p-value
Gaussian	98	0.3750	53	0.1998	26	0.0507
Student's t	94	0.4602	53	0.1998	22	0.2141
Clayton	70	0.0031	42	0.2215	17	0.3672
Gumbel	75	0.0150	39	0.1096	18	0.4581
Frank	85	0.1441	41	0.1788	17	0.3672

In Table 3.1-3.5, the p -value greater than 0.05 is highlighted in yellow, which means we can not reject the hypothesis that VaR is accurate. The more the p -value closes to 0.5, the better the VaR is.

- Observation 1. Different stock may have different tail dependence structures with benchmark and may have different copulas to fit it. There are totally 15 tests for each copula. For Gaussian copula, $\frac{14}{15}$ of the backtesting procedures can not reject the null hypothesis that VaR is accurate. For student's t copula, it is $\frac{13}{15}$. It is $\frac{10}{15}$, $\frac{9}{15}$, $\frac{8}{15}$ for frank, gumbel and clayton, respectively. Therefore, Gaussian and student's t copula works best among the 5 copulas.
- Observation 2. To compare Gaussian and student's t copula, we also highlight the value $1912(1 - \epsilon)$ of X_ϵ in green, which is closest to the theoretical value it should be. We can see that student's t copula performs a little big better than normal copula.

Next we use the last 502 observed data to predict the VaR and CVaR for the next day (Aug 11, 2009) using student's t copula. Table 3.6 gives the result we forecast for all five stocks, as well as the index of stability α for each z_t of the stock. From the table we can tell that

Table 3.6: VaR and CVaR of the relative risk for each stock with student's t copula

	α	$VaR_{.95}$	$CVaR_{0.95}$	$VaR_{.975}$	$CVaR_{0.975}$	$VaR_{.99}$	$CVaR_{0.99}$
PGN	1.9392	0.7518	1.0724	0.9465	1.3073	1.2209	1.6826
IBM	1.9576	0.6331	0.8860	0.7986	1.0642	1.0231	1.3314
BAC	1.6337	1.7702	3.6057	2.2845	5.2156	3.4528	10.9506
F	1.9392	2.3199	3.5603	2.8533	4.5638	3.6754	6.6025
HD	1.8984	0.7866	1.0152	0.9486	1.1171	1.1551	1.2664

- Observation 3. For risk comparison, IBM have the smallest risk of relative return. PGN and HD have the second and third smallest risk. But we cannot tell which one is better, since their VaR and CVaR are almost at the same level. On the other hand, F and BAC are more risky than the other three. The stability parameter of BAC is really small. That means the tail of BAC is very heavy and this makes it more risky. Even though the VaR of BAC is smaller than that of F, as for BAC has heavy tail, the CVaR for BAC is remarkably larger than the CVaR for F. Therefore, we can say that BAC is the most risky stock. We can see that when difference between the index of stability is not so big, there maybe some other factors such as skewness can affect the risk.

But when the index of stability is much more smaller than others, like BAC, the risk is much bigger. Therefore, the index of stability is the main factor for the risk.

Chapter 4

Portfolio Optimization with CVaR

Portfolio optimization method can be traced up to Markowitz (1952), where he introduced mean-variance framework. In his work, he tries to manage the portfolio by minimizing the variance (regarded as a risk) under the constraint that the return is no less than some value. This idea is still very popular in recent risk management, except that we use various different measurements for the risk, such as VaR and CVaR. Moreover, according to Rockafellar and Uryasev (2002 [38]), minimizing CVaR of the portfolio can be reformulated and solved by linear programming. In this chapter, we introduce their idea and extend it to portfolio optimization problem with CVaR as risk measurement, together with time series model. From the numerical results, we find out that the model we set up describes the practical economic market very well.

The rest of this chapter is organized as follows. In Section 4.1, we introduce the risk-return portfolio optimization framework, which is an extension of Markowitz (1952, [28]). Rockafellar and Uryasev's idea of how to solve the CVaR minimization problem is introduced in Section 4.2. Numerical results and discussions are given in Section 4.3.

4.1 Risk-Return Portfolio Optimization

Let $L = (L_1, L_2, \dots, L_n)^T$ be a set of n assets' loss. Suppose the allocation of these n assets in portfolio is $x = (x_1, x_2, \dots, x_n)^T$, such that $x_i \geq 0$ (no short sales permitted) and $\sum_{i=1}^n x_i = 1$ (budget constraint). Then, the total loss of the portfolio is

$$\tilde{L} = x^T L = \sum_{i=1}^n x_i L_i.$$

As L is a multi-variate random variable, \tilde{L} is a univariate random variable. Hence, we can define VaR and CVaR for the total loss (\tilde{L}) by (2.2) and (2.4) as $VaR_\epsilon(\tilde{L}, x)$ and $CVaR_\epsilon(\tilde{L}, x)$, where x is the parameter of risk allocation vector. Without confusing, we just write $VaR_\epsilon(x)$ and $CVaR_\epsilon(x)$ in short. Then, we can generalize the classical concept of mean-variance portfolio optimization problem for the case of the risk measure we defined above. We can formulate the following optimization problem:

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & \mathfrak{R}(\tilde{L}, x) = \mathfrak{R}(x^T L) \\ \text{s.t.} \quad & x^T \omega \geq \omega^* \\ & \sum_{i=1}^n x_i = 1 \\ & x_i \geq 0, \quad i = 1, \dots, n \end{aligned} \tag{4.1}$$

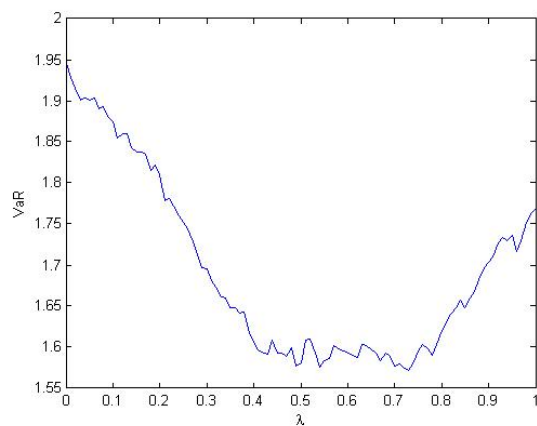
where \mathfrak{R} represents the risk measurement $VaR_\epsilon(x)$ or $CVaR_\epsilon(x)$, w.r.t. fixed x , $\omega = (\omega_1, \omega_2, \dots, \omega_n)^T$ is the vector of the expected return for the assets and ω^* is the minimum return requirement. As discussed in Section 2.1, VaR is not a coherent and convex function and it is not easy to find its global minimizer. While $CVaR$ is both coherent and convex and it is possible to find the optimal allocations. This can be seen in Figure 4.1. In Figure 4.1, we plot the portfolio VaR and CVaR for two asset: BAC and F with the historical data we selected and the allocation $x = (\lambda, 1 - \lambda)^T$. From Figure 4.1, we can see that VaR has many local minimizers and CVaR is convex with only one minimizer. Therefore, in this chapter, we use $CVaR$ as our objective risk measurement. Our optimization problem is

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & CVaR_\epsilon(x) \\ \text{s.t.} \quad & x^T \omega \geq \omega^* \\ & \sum_{i=1}^n x_i = 1 \\ & x_i \geq 0, \quad i = 1, \dots, n \end{aligned} \tag{4.2}$$

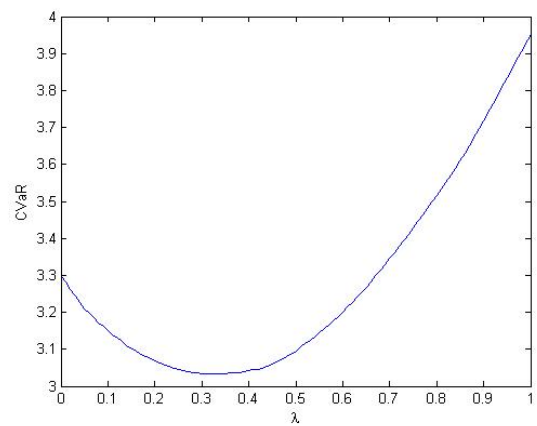
According to Rockafellar and Uryasev (2002), the minimization of $CVaR$ can be simplified to a linear programming, which is very easy to solve. In the next section, we will sketch their idea.

4.2 Minimization of CVaR

Let $f(x, L)$ be the loss function associated with the risk allocation vector x , where $x \in \mathbb{X}$ and \mathbb{X} is a certain subset of \mathbb{R}^n and the random vector $L \in \mathbb{R}^n$. For each x , the loss function $f(x, L)$ is a random variable with a distribution in \mathbb{R} . If the probability density function for L is $p(l)$,



(a) VaR of the portfolio of two assets



(b) CVaR of the portfolio of two assets

Figure 4.1: VaR and CVaR

$l \in \mathbb{R}^n$, the probability of the loss function $f(x, L)$ not exceeding a threshold ξ is given by

$$\Psi(x, \xi) \triangleq Pr(f(x, L) \leq \xi) = \int_{f(x, l) \leq \xi} p(l) dl.$$

As a function of ξ , for fix x , $\Psi(x, \xi)$ is the cumulative distribution function for the loss $f(x, L)$ associate with x . Then, the VaR_ϵ and $CVaR_\epsilon$ for the loss $f(x, L)$ associate with x is

$$VaR_\epsilon(x) = \min\{\xi \in \mathbb{R} : \Psi(x, \xi) \geq \epsilon\}$$

and

$$CVaR_\epsilon(x) = (1 - \epsilon)^{-1} \int_{f(x, l) \geq VaR_\epsilon} f(x, l) p(l) dl.$$

We define a function F_ϵ on $\mathbb{X} \times \mathbb{R}$ by

$$F_\epsilon(x, \xi) \triangleq \xi + (1 - \epsilon)^{-1} \int_{l \in \mathbb{R}^n} [f(x, l) - \xi]^+ p(l) dl,$$

where $[t]^+ = \max\{t, 0\}$. $CVaR_\epsilon(x)$ and $VaR_\epsilon(x)$ can be characterized in terms of the function F_ϵ on $\mathbb{X} \times \mathbb{R}$ defined above. The crucial features of F_ϵ , under the assumption above, are as follows (Rockafellar and Uryasev 2002).

Theorem 4.2.1. *As a function of ξ , $F_\epsilon(x, \xi)$ is convex and continuously differentiable. The $CVaR_\epsilon(x)$ of the loss $f(x, L)$ associate with any $x \in \mathbb{X}$ can be determined from the formula*

$$CVaR_\epsilon(x) = \min_{\xi \in \mathbb{R}} F_\epsilon(x, \xi).$$

Let

$$A_\epsilon(x) = \arg \min_{\xi \in \mathbb{R}} F_\epsilon(x, \xi)$$

be the set of values of ξ for which the minimum is attained. So $A_\epsilon(x)$ is a nonempty, closed, bounded interval, and VaR_ϵ of the loss $f(x, L)$ is given by

$$VaR_\epsilon(x) = \text{left endpoint of } A_\epsilon(x).$$

If the set $A_\epsilon(x)$ is reduced to a single point, then $VaR_\epsilon(x) = \arg \min_{\xi \in \mathbb{R}} F_\epsilon(x, \xi)$. It is also proved in (Rockafellar and Uryasev 2002) that

Theorem 4.2.2. *Minimizing the $CVaR_\epsilon(x)$ of the loss $f(x, L)$ associated with x over all $x \in \mathbb{X}$ is equivalent to minimizing $F_\epsilon(x, \xi)$ over all $(x, \xi) \in \mathbb{X} \times \mathbb{R}$, i.e.*

$$\min_{x \in \mathbb{X}} CVaR_\epsilon(x) = \min_{(x, \xi) \in \mathbb{X} \times \mathbb{R}} F_\epsilon(x, \xi), \quad (4.3)$$

where a pair (x^*, ξ^*) achieves the right hand side minimum if and only if x^* achieves the left hand side minimum and $\xi^* \in A_\epsilon(x^*)$. In particular, when $A_\epsilon(x^*)$ reduces to a single point, we have $\xi^* = VaR_\epsilon(x^*)$.

Furthermore, when $f(x, L)$ is convex w.r.t. x , $F_\epsilon(x, \xi)$ is convex w.r.t. (x, ξ) and $CVaR_\epsilon(x)$ is convex w.r.t. x . Therefore, if we take $f(x, L) = x^T L$, which is the portfolio of the assets' loss, then both $F_\epsilon(x, \xi)$ and $CVaR_\epsilon(x)$ are convex. Moreover, if the multi-variate distribution of L is continuous, the set $A_\epsilon(x)$ reduced to a single point. Hence from (4.3), our portfolio optimization problem (4.2) can be reformulated as

$$\begin{aligned} \min_{(x, \xi) \in \mathbb{R}^n \times \mathbb{R}} \quad & \xi + (1 - \epsilon)^{-1} \mathbf{E}[x^T L - \xi]^+ \\ \text{s.t.} \quad & x^T \omega \geq \omega^* \\ & \sum_{i=1}^n x_i = 1 \\ & x_i \geq 0, \quad i = 1, \dots, n, \end{aligned} \tag{4.4}$$

where (4.4) is derived by

$$F_\epsilon(x, \xi) = \xi + (1 - \epsilon)^{-1} \int_{l \in \mathbb{R}^n} [x^T l - \xi]^+ p(l) dl = \xi + (1 - \epsilon)^{-1} \mathbf{E}[x^T L - \xi]^+.$$

Then, we can discretize L use Monte Carlo method to simulated N realization of it, $l^j = (l_1^j, l_2^j, \dots, l_n^j)$, $j = 1, 2, \dots, N$ and thus the expectation can be estimated by

$$\mathbf{E}[x^T L - \xi]^+ = \frac{1}{N} \sum_{j=1}^N [x^T l^j - \xi]^+$$

and the portfolio optimization problem (4.4) becomes:

$$\begin{aligned} \min_{(x, \xi) \in \mathbb{R}^n \times \mathbb{R}} \quad & \xi + (1 - \epsilon)^{-1} \frac{1}{N} \sum_{j=1}^N [x^T l^j - \xi]^+ \\ \text{s.t.} \quad & x^T \omega \geq \omega^* \\ & \sum_{i=1}^n x_i = 1 \\ & x_i \geq 0, \quad i = 1, \dots, n. \end{aligned} \tag{4.5}$$

Let $z_j = [x^T l^j - \xi]^+$, we can rewrite (4.5) as:

$$\begin{aligned}
\min_{(x, \xi) \in \mathbb{R}^n \times \mathbb{R}} \quad & \xi + (1 - \epsilon)^{-1} \frac{1}{N} \sum_{j=1}^N z_j & (4.6) \\
s.t. \quad & x^T \omega \geq \omega^* \\
& x^T l^j - \xi \leq z_j, \quad j = 1, \dots, N \\
& \sum_{i=1}^n x_i = 1 \\
& x_i, z_j \geq 0, \quad i = 1, \dots, n, \quad j = 1, \dots, N,
\end{aligned}$$

which is a classical linear programming problem, which is relatively easy to solve.

In Rockfellar (2002,[38]), they used the historical method to estimate the optimal CVaR. In order to estimate the optimal CVaR more precisely, we extend their idea by parametric method, using GST and stable distribution to fit the historical data. We can estimate CVaR by Monte Carlo method. To describe the volatility clustering property inside each of the data, we use the GARCH(1,1) model and the rolling window of 2 years to predict the optimal CVaR at any time. The numerical results are discussed in the next section.

4.3 Numerical Results and Discussion

4.3.1 Portfolio Optimization Allocation

We choose five members of S&P500 index, which are from five totally different industrials. These five members are PGN, IBM, BAC, F and HD.

Now we consider an investor who will invest on stocks chosen from these five stocks. We describe their daily percentage loss using the first differences of the logarithm of series of their stock prices as in (4.7),

$$L_t = -100 [\ln(P_t) - \ln(P_{t-1})], \quad (4.7)$$

where $t = 1, \dots, \dots 2414$ corresponding to the period from Jan 1,2000 to Aug 10, 2009. The statistics for the observed loss data for each member is in Table 2.1. We can see that all this five loss data are fat-tailed and asymmetric. Therefore, GST distribution is a good one to describe their marginal distributions. In addition, the results in Chapter 3 and in Dobrić and Schmid (2005,[12]) indicate that student's t copula with low degree of freedom is a good copula for financial data. Therefore, we use it to describe the dependence structure for the loss data. We use these observed loss data to estimate the parameters for the GST distribution, which is shown in Table 4.1.

Table 4.1: Parameter estimations of GST distribution for 5 different observed loss data

stock	μ	σ	ν	λ
PGN	0.0005	0.6600	3.5379	0.0617
IBM	0.0019	0.9439	2.8967	0.0114
BAC	0.0193	67.5801	2.0001	0.0303
F	0.0217	1.4899	2.9711	-0.0619
HD	0.0124	1.1146	3.2322	-0.0282

We can see from Table 4.1 that the variance of *BAC* data is very big, which means that *BAC* is not very steady. In addition, the parameter ν of *BAC* is very close to 2, which means that its distribution is fatter than the others, i.e. it is more risky. We take the mean return of the history data as the expected return of each asset, i.e negative mean of the loss data. So from Table 2.1, $\omega = (0.0049, 0.000414, -0.0192, -0.034, -0.0158)^T$. We can see that only *PGN* and *IBM* have positive mean return during the time period. Thus, the minimum return requirement ω^* should be less than 0.0049. We first take $\omega^* = 0.001$ to see how diversification of investment can reduce the risk. We solve the risk-return portfolio optimization problem

$$\begin{aligned}
 \min_{(x,\xi) \in \mathbb{R}^n \times \mathbb{R}} \quad & \xi + (1 - \epsilon)^{-1} \frac{1}{N} \sum_{j=1}^N z_j & (4.8) \\
 \text{s.t.} \quad & x^T \omega \geq 0.001 \\
 & x^T l^j - \xi \leq z_j, \quad j = 1, \dots, N \\
 & \sum_{i=1}^n x_i = 1 \\
 & x_i, z_j \geq 0, \quad i = 1, \dots, n, \quad j = 1, \dots, N
 \end{aligned}$$

where $\epsilon = 0.95$ and l^j are the loss simulations we get from the joint distribution, which is obtained by the student's *t* copula and marginally GST distribution with parameters in Table 4.1. Since only *PGN* and *IBM* have positive expected returns, we try to solve (4.8) for the investment combination of 2 assets (*PGN* and *IBM*), 3 assets (*PGN*, *IBM*, and *BAC*), 4 assets (*PGN*, *IBM*, *BAC* and *F*) and 5 assets (*PGN*, *IBM*, *BAC*, *F* and *HD*) respectively. The result of optimal portfolio allocation is showed in Table 4.2.

We also list the parameter estimation $\hat{\rho}$ for the student's *t* copula and omit the parameter estimation for *R*. We can see from Table 4.2 that firstly the degree of freedom for student's *t* copula are all round 5, which are consist with Dobrić (2005,[12]) referring the low degrees

Table 4.2: Portfolio allocation with $\epsilon = 0.95$ and minimum return requirement $\omega^* = 0.001$ for GST distribution

	$\hat{\rho}$	opt-CVaR	VaR	Portfolio allocation				
				PGN	IBM	BAC	F	HD
2 assets	4.7677	1.4763	0.9488	0.7989	0.2011			
3 assets	4.4818	1.3889	0.8848	0.6694	0.2282	0.1024		
4 assets	4.8391	1.3736	0.8747	0.6820	0.2475	0.0009	0.0696	
5 assets	4.6501	1.3670	0.8734	0.6701	0.2004	0.0014	0.0000	0.1282

of freedom. Secondly, the diversification of the investment does reduce the risk, from 1.4763 of CVaR for 2 assets to 1.3670 for 5 assets, but it is not reduced dramatically. The reason is that since we have the minimum return requirement $\omega^* = 0.001$, and only PGN and IBM gives a positive mean return and all the other 3 assets have negative returns. As a result, the allocation of the assets for each case are investing almost all the money on PGN and IBM. The rest of the asset have almost no investment. Thirdly, in each of the portfolio, the allocation for BAC is relatively very small. That is because that BAC is more risky than the other stocks (the index of tail for its GST distribution is 2.0001, which means its tail is much more fatter than the others). In order to compare with the historical method used in Rockfellar's paper ([37]), we show the results of portfolio optimization with 2,3,4,5 assets in Table 4.3, using the historical method. Comparing the results in Table 4.2 and 4.3, we can see that the optimal CVaRs from historical method are smaller than those from parametric method in all scenarios. We can see that although we have 2414 historical data to estimate CVaR, the historical method still under-estimates the risk. This shows that parametric method does superior to historical method, in the sense of accuracy.

Table 4.3: Optimal portfolio CVaR allocation with $\epsilon = 0.95$ and minimum return requirement $\omega^* = 0.001$ for historical data

	CVaR	VaR	Portfolio allocation				
			PGN	IBM	BAC	F	HD
2 assets	1.1726	0.7961	0.6327	0.3673			
3 assets	1.1639	0.8099	0.6145	0.3530	0.0325		
4 assets	1.1595	0.7785	0.5776	0.3565	0.0310	0.0349	
5 assets	1.1357	0.7515	0.5310	0.3447	0.0006	0.0257	0.0980

Now let's enforce the minimum return requirement and change the value of ω^* from 0.001 to 0.0045 for the 5 assets' portfolio. This gives us an efficient frontier which is shown in Figure 4.2. We can see that the efficient frontier is quite similar to that in the mean-variance framework. The portfolio allocation of the assets at each point is the tangent line at this point.

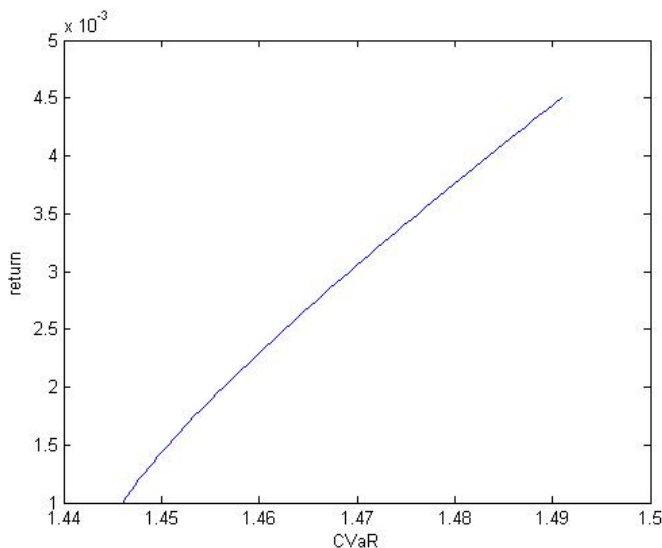


Figure 4.2: Efficient 95% CVaR Frontier for GST distribution

4.3.2 Optimal CVaR with Time Series Model

We know that short time financial data shows some volatility clustering, so it is better to use time series GARCH model to describe them. As discussed in section 2.5, we use the simplest GARCH(1,1) model to fit the loss data with GST as innovation distribution for each asset loss as follows

$$L_t = \mu + \sigma_t z_t,$$

$$\sigma_t^2 = a_0 + \sum_{i=1}^m a_i (L_{t-i} - \mu)^2 + \sum_{i=1}^s b_i \sigma_{t-i}^2,$$

where $z_t \sim GST(\nu, \lambda)$ for each asset. The innovation distributions for each asset are not independent, they are correlated. Therefore, we use student's t copula to fit their correlations to get the joint distribution of z_t s. To estimate the parameters in GARCH model and copula,

we use the most recent 2 year historical data to estimate them. Then, use those parameters to simulate the coupled loss data by Monte Carlo method, and use those data to predict the current optimal portfolio CVaR for the loss, by solving the model we set up in section 4.2. Here in Figure 4.3, we show the predicted optimal daily portfolio CVaR for the loss, by solving the model we set up in section 4.2. Here in Figure 4.3, we show the predicted optimal daily portfolio CVaR for time between Jan 1, 2008 and Aug 10, 2009, as well as the S&P500 index for that time period in the same graph.

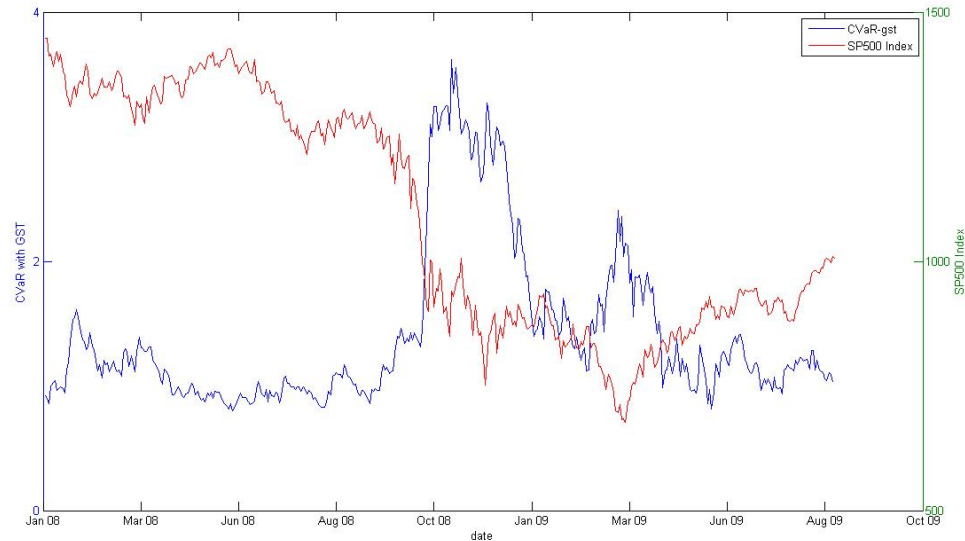


Figure 4.3: Predicted daily optimal portfolio $CVaR_{0.95}$ for GST distribution and S&P500 index from Jan 1,2008 to Aug 10, 2009

We believe that the stock index can represent the economic status in the market. That is, if the economics goes bad, the index decreases and vice versa. We can see from Figure 4.3 that, when the S&P500 index decreases, that means the economics goes bad, the corresponding optimal daily portfolio CVaR increases and vice versa. In addition, there is a big increase of optimal daily portfolio CVaR around time 187 - 200, which is around Sep 26, 2008 - Oct 15, 2008, meaning that the risk at that period is high. While at almost the same time period, the S&P500 index have a serious drop. It is because that during this period, the US economics is suffering the crisis, triggered by the bankruptcy of the Lehman Brothers on Sep 15, 2008. While after that, on October 3, 2008, the US government introduce the bailout of the U.S. financial system, which authorizes the United States Secretary of the Treasury to spend up to US\$700

billion to purchase distressed assets, especially mortgage-backed securities, and make capital injections into banks. This action increases the confidence of the investor and makes the daily CVaR decrease rapidly. This is shown in the figure at the time period around 201-250, where daily CVaR decreased quickly. After that, the risk regarded as daily CVaR oscillates with the index fluctuates by the pattern we described above. In average, at the time when the index is high the daily optimal portfolio daily CVaR is low, i.e. the risk is low, vice versa. We can say the the optimal daily portfolio CVaR we get is quite consistent with the economic environment. This indicates that our model works well.

In addition, if we take $\epsilon = 0.975$ and 0.99 , we will obtain the optimal CVaR corresponding to difference level of ϵ , which is shown in Figure 4.4. From the Figure 4.4, we can see that the pattern of series for 97.5% and 99% are the same as that for 95%. The only difference is that at each time, the optimal CVaR of 99% is greater than that of 97.5%, which is again greater than that of 95%.

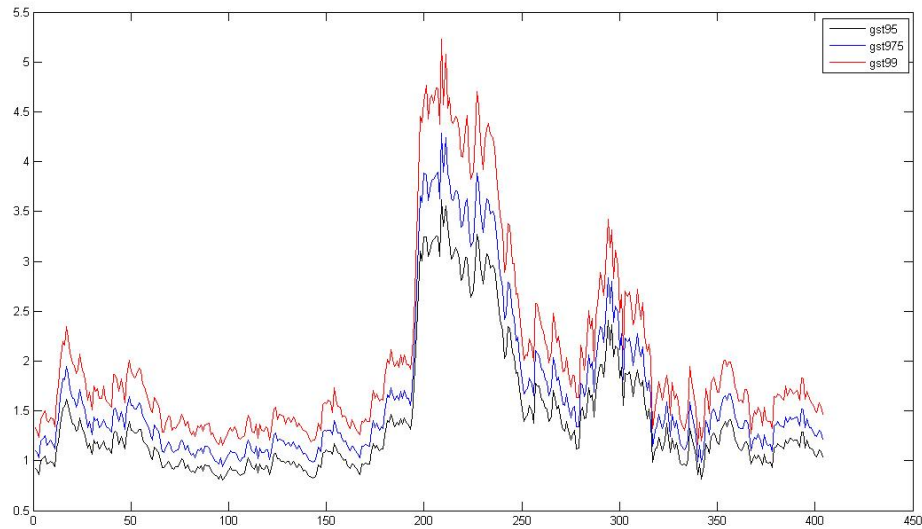


Figure 4.4: Predicted daily optimal portfolio CVaR for GST distribution at different levels

4.3.3 Further Discussion for Stable Distribution

From the above discussion, GST distribution does do a good job. However, there are some potential risks that the tail of GST distribution is not fat enough to describe financial data. This may under-estimate the risk. Therefore, we consider another better candidate distribution - stable distribution, which has fatter tail than GST and can describe the skewness too. However, although stable distribution has better description for tail thickness, it does not have a close form expression of the probability density function. We have to calculate it numerically, which will cost more time.

Next, we use stable distribution as our fitting distribution and use student's t copula for dependence structure, we follow the above framework and the numerical results are given in Table 4.4, Table 4.5, Figure 4.5 to 4.8.

Table 4.4: Parameter estimations of stable distribution for 5 different observed loss data

stock	α	β	σ	μ
PGN	1.6833	0.1518	0.3510	0.0053
IBM	1.5467	0.0409	0.4246	0.0073
BAC	1.1828	0.0582	0.4043	0.0663
F	1.6147	-0.2489	0.6855	-0.0192
HD	1.5951	-0.0554	0.5525	-0.0102

Table 4.5: Optimal portfolio CVaR allocation with $\epsilon = 0.95$ and minimum return requirement $\omega^* = 0.001$ for stable distribution

	$\hat{\rho}$	opt-CVaR	VaR	Portfolio allocation				
				PGN	IBM	BAC	F	HD
2 assets	4.1180	1.7744	0.9162	0.8676	0.1324			
3 assets	3.3971	1.7531	0.9073	0.8324	0.1627	0.0049		
4 assets	3.8897	1.7265	0.9279	0.7775	0.1318	0.0000	0.0907	
5 assets	3.7749	1.7168	0.9163	0.7246	0.1056	0.0000	0.0714	0.0984

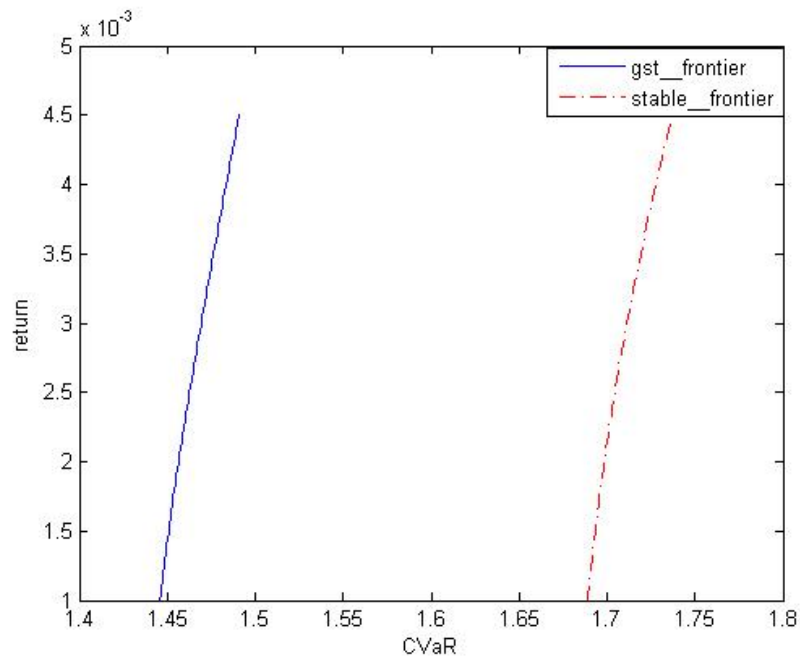


Figure 4.5: Efficient 95% CVaR Frontier comparison for GST and stable distributions

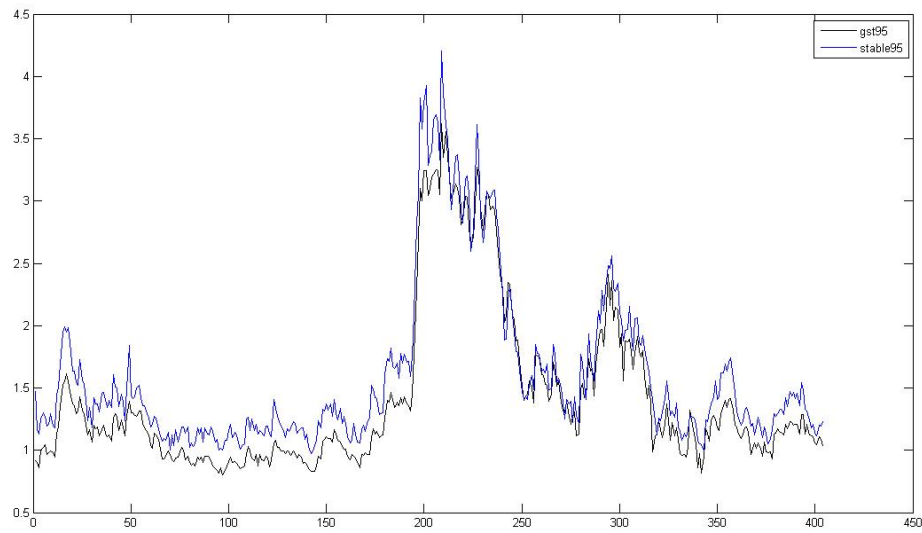


Figure 4.6: Predicted daily optimal portfolio CVaR from Jan 1, 2008 to Aug 10, 2009 for GST and stable distributions at level 0.95

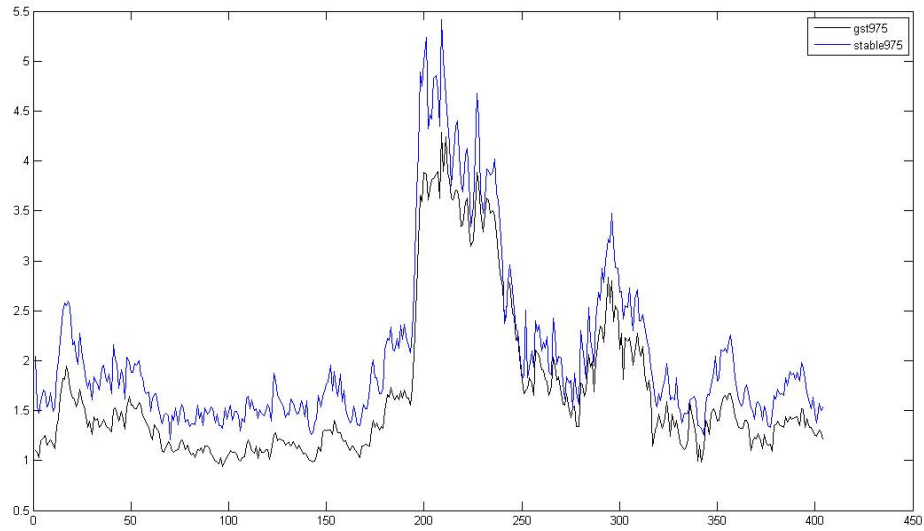


Figure 4.7: Predicted daily optimal portfolio CVaR from Jan 1, 2008 to Aug 10, 2009 for GST and stable distributions at level 0.975

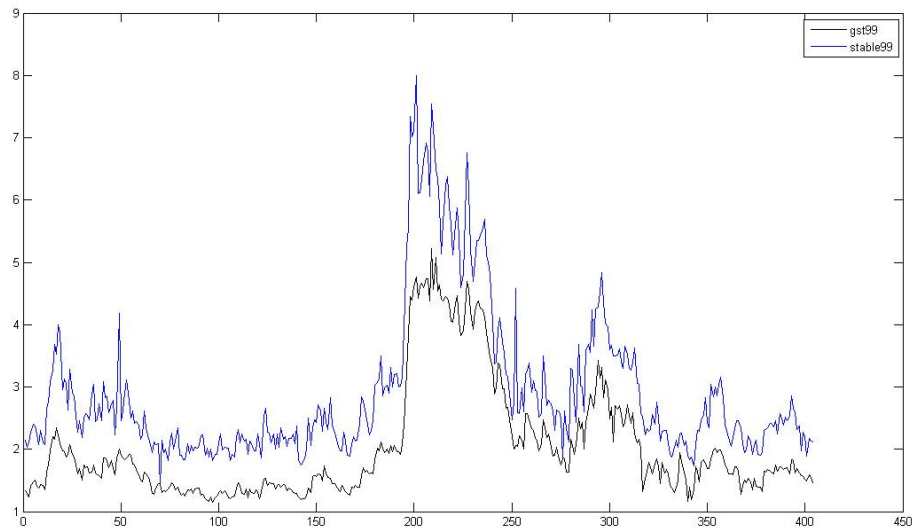


Figure 4.8: Predicted daily optimal portfolio CVaR from Jan 1, 2008 to Aug 10, 2009 for GST and stable distributions at level 0.99

In Table 4.4, we give the parameter estimation for the marginal stable distributions, using the historical data. Table 4.5 gives the optimal CVaR under different number of asset for portfolio. We can see the same phenomena as that of GST, that the diversification can reduce the risk of the portfolio. However, the optimal risks under stable distribution are bigger than that of GST. From Figure 4.5, which draw efficient frontiers under both stable and GST distributions, we can see that the efficient frontier of stable distribution is almost parallel to that of GST and is bigger than GST. Figure 4.6 to 4.8, compares the daily optimal CVaR obtained from our time series models under different levels of threshold. These figures show that under the time series model, the daily optimal CVaR of stable almost have the same pattern as that of GST at each level, which implies that the time series model with stable distribution is also consistent with the real market, but it may predict bigger risks compared to GST distribution.

Chapter 5

Portfolio Optimization with VaR

In chapter 4, we consider portfolio optimization problem with CVaR as the risk measurement. The reason why we use CVaR rather than VaR is that VaR is not a coherent or convex measurement, but CVaR is. However, as we said in Chapter 1, VaR is widely used in practical world and it is used in regulatory mechanisms. Therefore, the research of the optimal problem with VaR as risk measurement is essential. Actually, it can be done in some complex way. Gaivoronski and Pflug(2005, [16]) provided a numerical approach to solve this problem. In that paper, the authors introduced an approach to approximate VaR by a smoothed measurement called SVaR, which filters out local irregularities. However, the way to find this smoothed SVaR is quit complex. In Benati(2007,[5]), a mixed integer linear programming method is discussed for the portfolio optimization problem with VaR as a risk measurement. In that paper, the author proposed a mixed integer linear programming model for the problem, which can be efficiently solved by linear integer optimization software (like Cplex, SAS-OR package). However, their idea is the historical approach, which may lead to under-estimate the risk. In this chapter, I extend their idea to the parametric approach, as well as with the time series model.

The rest of this chapter is arranged as follows. Section 5.1 introduces how we model the portfolio optimization problem by mixed integer linear problem. Numerical results and discussions are shown in Section 5.2.

5.1 Optimization with VaR

Let $R = (R_1, R_2, \dots, R_n)^T$ be the vector of n assets' return. Suppose the allocation of these n assets in portfolio is $x = (x_1, x_2, \dots, x_n)^T$. Therefore, the total return of the portfolio is

$$\tilde{R} = x^T R = \sum_{i=1}^n x_i R_i.$$

Let the Value-at-Risk for \tilde{R} at lever ϵ is defined as (2.1), we write it as $VaR_\epsilon(\tilde{R})$. If we replace $VaR_\epsilon(\tilde{R})$ as the risk measurement in (4.2) instead of CVaR, we will have the portfolio optimization problem with VaR

$$\begin{aligned} \min_{x \in \tilde{\mathbb{R}}^n} \quad & VaR_\epsilon(TR) & (5.1) \\ \text{s.t.} \quad & x^T \omega \geq \omega^* \\ & TR = \sum_{i=1}^n x_i R_i \\ & \sum_{i=1}^n x_i = 1 \\ & x_i \geq 0, \quad i = 1, \dots, n. \end{aligned}$$

We can consider the model (5.1) in another way. The decision maker fixed two parameters, the probability ϵ and the return rate requirement ω^* and one positive variable η^ϵ . He will not accept any investment whose VaR with level ϵ is greater than η^ϵ , i.e. no investment in which $Pr[\tilde{R} \leq -\eta^\epsilon] \geq 1 - \epsilon$ is taken into consideration. In addition, no investment is accepted if the expected rate of return is less than ω^* . At the same time, he needs to minimize the risk variable η^ϵ . Therefore, we can set up the model as (Problem P1)

$$\min_{x, \eta^\epsilon} \quad \eta^\epsilon \quad (5.2)$$

$$\text{s.t.} \quad Pr[\tilde{R} \leq -\eta^\epsilon] \leq 1 - \epsilon \quad (5.3)$$

$$\tilde{R} = \sum_{i=1}^n x_i R_i \quad (5.4)$$

$$x^T \omega \geq \omega^* \quad (5.5)$$

$$\sum_{i=1}^n x_i = 1 \quad (5.6)$$

$$x_i \geq 0, \quad i = 1, \dots, n, \quad (5.7)$$

where constraint (5.6) requires that one unit of wealth must be allocated on different assets and constraint (5.7) prevents short-selling. The decision maker is willing to accept only portfolios for which probability of portfolio under $-\eta^\epsilon$ is less than or equal to the threshold $1 - \epsilon$, this constraint is described in (5.3) and is equivalent to say that $VaR_\epsilon(\tilde{R}) \leq \eta^\epsilon$.

Since R_i and \tilde{R} are all random variables, if each R_i have N realizations r_i^j , $j = 1, \dots, N$ and we have that $Pr[\bigcap_{i=1}^n (R_i = r_i^j)] = p_j$, $j = 1, \dots, N$. Then, we have $Pr[\tilde{R} = \tilde{r}_j] = Pr[\tilde{R} = \sum_{i=1}^n x_i r_i^j] = p_j$, $i = 1, \dots, N$. Therefore, we can reformulate Problem P1 in to a mixed integer

programming problem as follows (Problem P2):

$$\min_{x, tr, y, \eta^\epsilon} \quad \eta^\epsilon \quad (5.8)$$

$$\text{s.t.} \quad \tilde{r}_j = \sum_{i=1}^n x_i r_i^j, \quad j = 1, \dots, N \quad (5.9)$$

$$\sum_{i=1}^n x_i \omega_i \geq \omega^* \quad (5.10)$$

$$\tilde{r}_j + M(1 - y_j) \geq -\eta^\epsilon, \quad j = 1, \dots, N \quad (5.11)$$

$$\sum_{j=1}^N p_j(1 - y_j) \leq 1 - \epsilon \quad (5.12)$$

$$\sum_{i=1}^n x_i = 1 \quad (5.13)$$

$$y_j \in \{0, 1\}, \quad j = 1, \dots, N \quad (5.14)$$

$$x_i \geq 0, \quad i = 1, \dots, n \quad (5.15)$$

where M is a very large number and ω_i is the expected rate of return for each asset. Constraint (5.10) prevents the investment whose expected rate of return is less than ω^* . Variables \tilde{r}_j is the portfolio observed return of the j th realization. Constraint (5.11) and (5.12) prevent the choice of portfolios whose VaR is above the fixed threshold η^ϵ . Every time \tilde{r}_j is below $-\eta^\epsilon$, y_j must be equal to 0 and $1 - y_j = 1$ in constraint (5.11). Therefore, all probabilities of realization j whose returns are below $-\eta^\epsilon$ are summed up. If the result is greater than $1 - \epsilon$, then the portfolio is not feasible. We can see that Problem P2 is a standard mixed linear integer programming problem. If we have the joint distribution of the assets R_i , $i = 1, \dots, n$, we can use Monte Carlo method to simulate the realizations r_i^j and p_j , and Problem P2 can be solved efficiently.

We try to solve Problem 2 using parametric method. With the return data of each asset, we can fit them into a family of distributions such as GST and stable distribution. After that, we can get their joint distribution by copula method. We can also use this model together with time series models.

5.2 Numerical Results

5.2.1 Results without Time Dependence

For the same reason as in the last chapter, we use both GST distribution and stable distribution to fit the financial rate of returns. In addition, we will use student's t copula to describe the correlations between the assets. If we ignore the time dependence of the series for a moment, we use the historical data and the parametric to find the optimal VaR of different number of

assets for the portfolio. Table 5.1 and 5.2 summarize the optimal 95% VaR and the allocations using GST and stable distribution, respectively.

Table 5.1: Optimal VaR portfolio allocation with $\epsilon = 0.95$ and minimum return requirement $\omega^* = 0.001$ for GST distribution

	VaR	Portfolio allocation				
		PGN	IBM	BAC	F	HD
2 assets	0.8800	0.6778	0.3222			
3 assets	0.8385	0.6278	0.3251	0.0471		
4 assets	0.7836	0.6810	0.2507	0.0600	0.0083	
5 assets	0.7619	0.6503	0.2440	0.0451	0.0121	0.0485

Table 5.2: Optimal VaR portfolio allocation with $\epsilon = 0.95$ and minimum return requirement $\omega^* = 0.001$ for stable distribution

	VaR	Portfolio allocation				
		PGN	IBM	BAC	F	HD
2 assets	0.8535	0.6801	0.3199			
3 assets	0.8328	0.6442	0.3179	0.0379		
4 assets	0.8320	0.6596	0.2746	0.0373	0.0285	
5 assets	0.7978	0.7282	0.1676	0.0268	0.0164	0.0061

We ignore the estimation of parameters for the GST and stable distribution, as well as parameters of the copula, the reason we can do this is that these parameters are the same as those in Chapter 4. Comparing with the VaRs we get from the optimal CVaR in Table 4.2 and 4.5 in Chapter 4, we can see that the optimal VaRs in Table 5.1 and 5.2 are smaller. Therefore, we can say that, if we regard VaR as the risk measurement, the VaRs we get from the optimal CVaR model is not minimized. On the other hand, from Table 5.1 and 5.2, we can see that diversification of the investment from 2 asset to 5 assets will also reduce the VaR of the portfolios from 0.88 to 0.7619 for GST distribution and from 0.8535 to 0.7978 for stable distribution, respectively. In addition, the optimal VaRs we get from stable distribution is a little bit bigger than optimal VaRs we get from GST distribution. The reason is that

the tail of stable distribution is fatter than the tail of GST distribution. Furthermore, if we keep the number of assets in portfolio at 5 and take different values of the minimum return requirement ω^* from 0.001 to 0.0045, we will obtain the efficient frontiers of optimal VaR for both distributions, which are shown in Figure 5.1. From Figure 5.1, firstly, we can see that the frontier curve is not so smooth, it maybe because of the non-convex property of VaR. Secondly, even though it is not smooth, we can tell that the curve is still increasing. That means the bigger the return you require, the more risk you have to face. Thirdly, for the same return requirement, the optimal risk for GST distribution is smaller than that for stable distribution. The reason for this is that stable distribution has fatter tail than GST distribution.

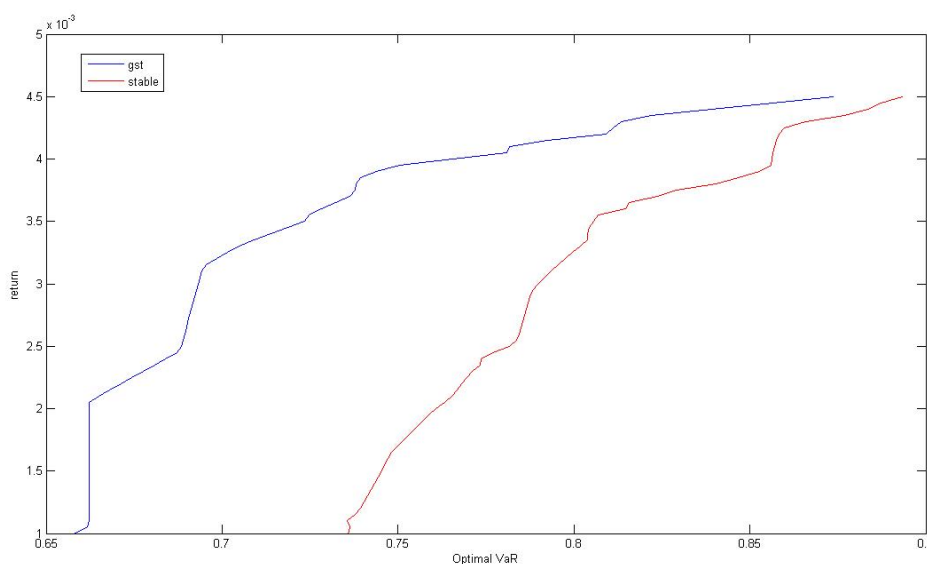


Figure 5.1: Efficient 95% VaR Frontier for both GST and stable distribution

5.2.2 Results with Time Dependence

If we consider the time dependence within each of the asset, we will use the time series ARMA-GARCH model which is introduced in Chapter 2. We use the simplest ARMA(0,0)-GARCH(1,1) model to fit the data of each assets. The correlations among the assets are also described by student's t copula as before. We use the most recent 2 years' data for any particular day to predict the return of that day. After we obtain the simulated data for all the assets using ARMA-GARCH and copula, we can solve the model (Problem 2) to predict the optimal

allocation and the optimal 95% VaR for that day. The following Figures 5.2 to 5.7 is obtained from the above procedure based on the data and we obtain the predicted daily optimal VaR for the time between Jan 1,2000 and Aug 10, 2009.

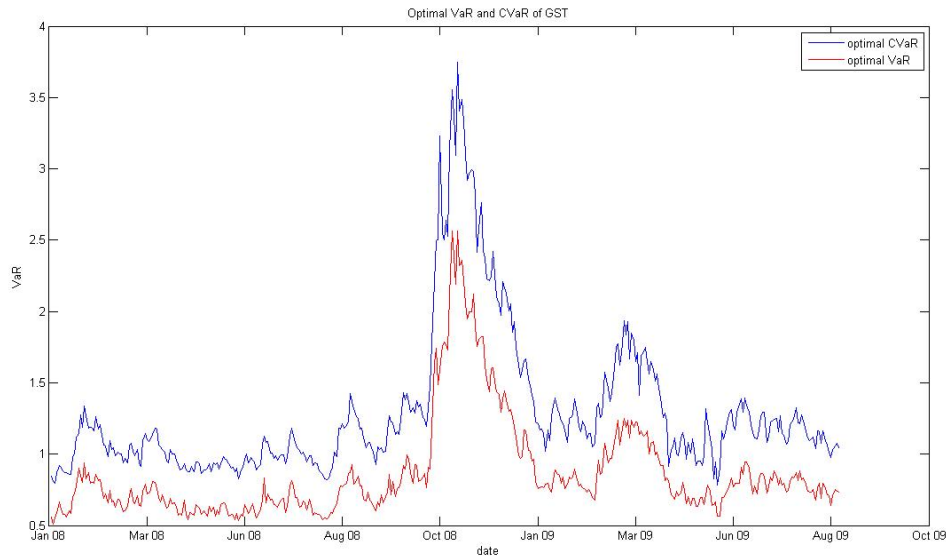


Figure 5.2: Predicted optimal daily 95% VaR compared with optimal 95% CVaR for GST distribution

In Figure 5.2 and 5.3, we plot the optimal daily VaR we obtain and compare it with the optimal CVaR we obtain from the model in Chapter 4 for both GST and stable distribution, respectively. From these two figures, we can see that the pattern for optimal VaR is almost the same as the pattern for optimal CVaR. That is, when the optimal CVaR is increasing, the corresponding optimal VaR is also increasing and vice-versa. In addition, when the optimal CVaR is big, the corresponding optimal VaR is also big and vice-versa. As the conclusion we state in Chapter 4 that the pattern of daily optimal CVaR is quite consistent with the real world market, we can conclude that the pattern of daily optimal VaR we obtain from this mix integer linear programming problem is also consistent with the real world market. Therefore, the model we set up is really good.

In Figure 5.4 and 5.5, we compares the optimal daily VaR we obtain and the VaR we obtain

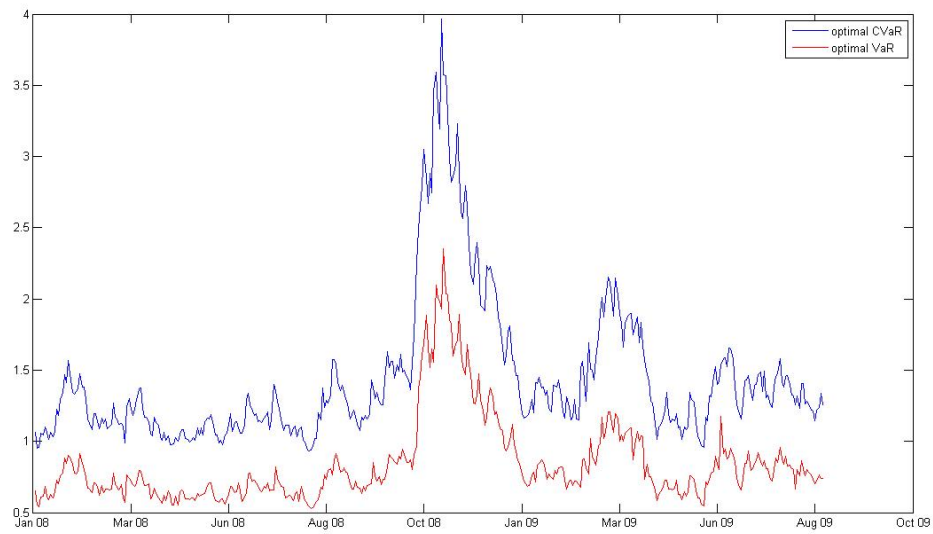


Figure 5.3: Predicted optimal daily 95% VaR compared with optimal 95% CVaR for stable distribution

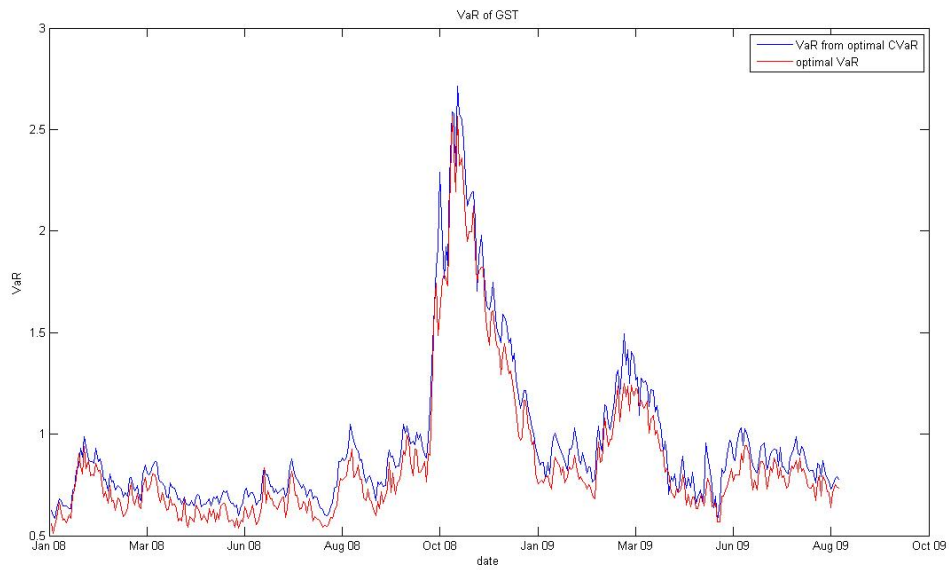


Figure 5.4: Predicted optimal daily 95% VaR compared with 95% VaR from optimal CVaR for GST distribution

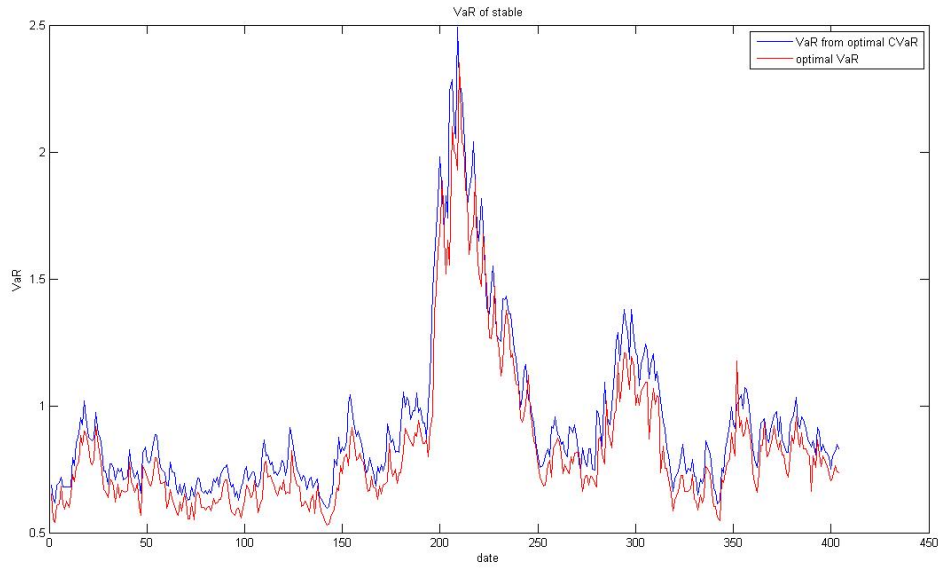


Figure 5.5: Predicted optimal daily 95% VaR compared with 95% VaR from optimal CVaR for stable distribution

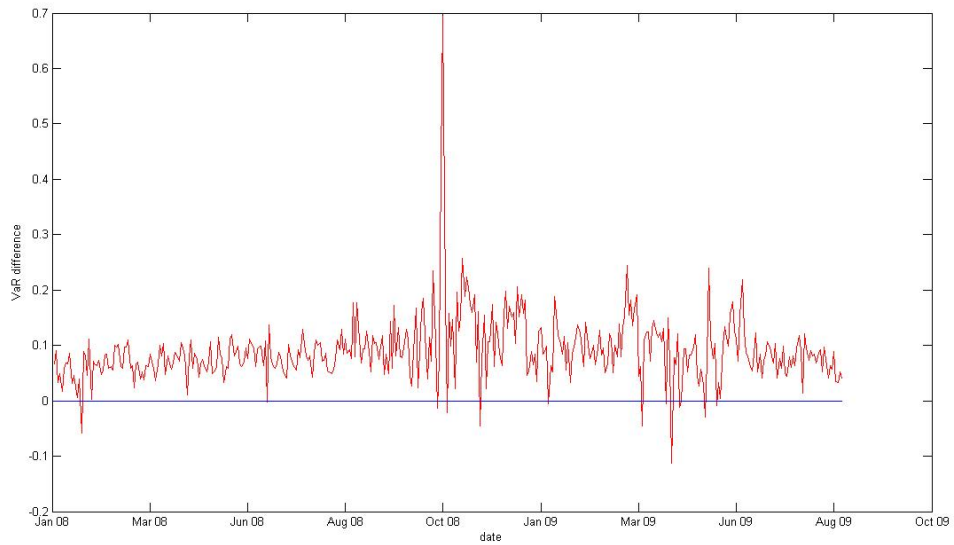


Figure 5.6: The difference between VaR from optimal CVaR and optimal VaR for GST distribution

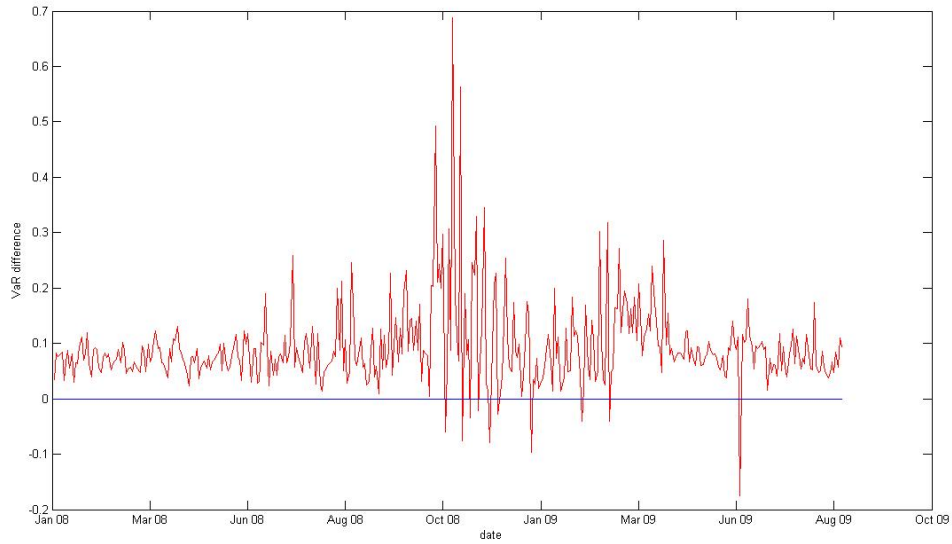


Figure 5.7: The difference between VaR from optimal CVaR and optimal VaR for stable distribution

from the optimal CVaR model in Chapter 4 for both GST and stable distribution. Theoretically, the optimal VaR we get should be the smallest among the VaRs of any allocation of the assets in the portfolio. From Figure 5.4 and 5.5, we can see that the optimal VaR is really smaller than the VaR we get from the optimal CVaR almost every time. The difference of the VaR from optimal CVaR and the optimal VaR is plotted in Figure 5.6 and 5.7, the black line is the 0 horizontal line. The number of points that the difference is below 0 is 12 and 11 for GST and Stable distribution, respectively and the total number of points we have is 404. Therefore, most of the difference is greater than 0, which is consistent with the theoretical fact. The reason that there are negative difference is because that we use different simulation realizations when we calculate the optimal CVaR and VaR.

Chapter 6

Independent Component Analysis and Its Application in Portfolio Risk Management

Throughout this paper, we can see that the most important thing in my research is the joint distribution of the returns/losses. In order to get the joint distribution, we firstly fit the marginal distributions of each asset. Then, use copula to connect the marginal distributions with the joint distribution. This approach does work because that the Sklar's theory ensures the existence of the copula for the joint distribution. However, there are also disadvantages. One of them is that it may take a long time to generate the simulated data. Therefore, in this chapter, we explore a method that can save time of simulation to solve the portfolio optimization problems using the model we set up in the previous chapters. The easiest situation to simulate multivariate data is when the data we are trying to simulate are independent with each other. If so, we can simulate the data independently. Therefore, we are considering if we can transform our original data into independent variables, and transform our models into independent variables we obtain. Inspired by this idea, we consider to use a statistical method called Independent Component Analysis (ICA), which can transform the mixed data into statistically independent sources, by linear transformation. Using this ICA method, we can transform the original models with dependent variables into the models with independent variable. As a result, this procedure may save us a lot of time of simulation.

The rest of this chapter is organized as follows. In Section 6.1, we will introduce the Independent Component Analysis (ICA) method in details. In Section 6.2, we will see how ICA is applied in risk management to solve portfolio optimization problems. Finally, in Section 6.3, we will summarize the numerical results if we use this ICA method in portfolio optimization

problems.

6.1 Independent Component Analysis Method

6.1.1 Introduction

Independent Component Analysis (ICA) is a statistical method, widely used in signal analysis. It is used to demix the mixed signals to find the independent sources, by linear combination of the mixed signals. Let $S \in \mathbb{R}^n$ be the n independent sources sending signals and $X \in \mathbb{R}^n$ be the mixture of these signals we obtain from the receiver. The question is can we isolate the independent sources Y from the received signals X . Suppose X is the linear combination of S , i.e. there exists a matrix $A \in \mathbb{R}^{n \times n}$ such that

$$X = AS.$$

We try to find a demixing matrix $W \in \mathbb{R}^{n \times n}$, such that $Y = WX = WAS$ and Y is statistical independent. We call Y Independent Components (ICs). If W is nonsingular and $W^{-1} = A$, we will get the original independent source S explicitly. The ICA procedure is shown in Figure 6.1. The original sources S are mixed through matrix A to form the observed signal X . The demixing matrix W transforms the observed signal X into the independent components Y .

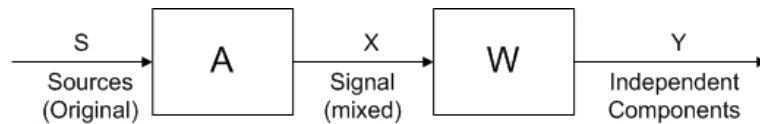


Figure 6.1: ICA Procedure

6.1.2 ICA Estimation

The ICA model is a generative model, which means that it describes how the observed data are generated by a process of mixing the components S . The independent components Y are latent variables, meaning that they cannot be directly observed. In addition, the mixing matrix A and the demixing matrix W are assumed to be unknown. All we can observe is the random vector X and we must estimate W using X . Once we know the demixing matrix W , we can obtain the independent component Y by $Y = WX$. In order to make the ICA method work, we must

assume that the independent component must have non-gaussian distribution. We will briefly introduce two methods to estimate W , Maximum Likelihood Estimation method and FastICA method.

Maximum Likelihood Estimation If we know the probability density function of the independent source S_i , $i = 1, 2, \dots, n$ is $f_i(s)$, then, since S_i , $i = 1, 2, \dots, n$ are independent, we can obtain the joint distribution of S_i , which is

$$f_S(s_1, \dots, s_n) = \prod_{i=1}^n f_i(s_i).$$

Therefore, the joint distribution of $X = AS = W^{-1}S$ should be

$$f_X(x_1, \dots, x_n) = \left(\prod_{i=1}^n f_i(WX) \right) |\det(W)|.$$

The log-likelihood function for the observation $x^j = (x_1^j, x_2^j, \dots, x_n^j)$, $j = 1, 2, \dots, T$ is

$$L = \sum_{t=1}^T \sum_{i=1}^n \ln(f_i(w_i^T x^j)) + T \ln(|\det(W)|) \quad (6.1)$$

where $W = (w_1, \dots, w_n)^T$ is unknown. Using the Maximum Likelihood Estimation method, we can estimate W to be the maximizer of the log-likelihood function (6.1).

FastICA Algorithm In 2000, Hyvärinen and Oja ([23]) proposed a very efficient algorithm to estimate the demixing matrix W , which is called FastICA algorithm. The algorithm is based on a fixed-point iteration scheme maximizing non-Gaussianity as a measure of statistical independence. It can be also derived as an approximative Newton iteration. The iterative algorithm finds the direction for the weight vector W by maximizing the non-Gaussianity of the projection $w^T x$ for the data x . The function $g(\cdot)$ is the derivative of a non-quadratic non-linearity function. For example, $g(x)$ could be the derivative of the function $G(x) = \frac{1}{a_1} \ln(\cosh(a_1 x))$, which is $g(x) = \tanh(a_1 x)$, where $1 \leq a_1 \leq 2$ is some suitable constant, often taken as $a_1 = 1$,

$$\cosh(x) = \frac{e^x + e^{-x}}{2}$$

and

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}.$$

Or $g(x)$ could be the derivative of the function $G(x) = -\exp(-\frac{x^2}{2})$, which is $g(x) = x \exp(-\frac{x^2}{2})$. The basic form of the FastICA algorithm is as follows:

1. Choose an initial (e.g. random) weight vector $w \in \mathbb{R}^n$.
2. Let $w^+ = \mathbf{E}[Xg(w^T X)] - \mathbf{E}[g'(w^T X)]w$.
3. Let $w = w^+ / \|w^+\|$.
4. If not converge, go back to 2.

In addition to the easy formulation, FastICA has another advantage that we do not need to know the distribution of the original source S or the data X . In this thesis, I use this FastICA algorithm to obtain the demixing matrix W .

6.2 Applying ICA in Portfolio Optimization Problem

An early implementation of ICA in financial time series is given in Back (1998, [3]), where the author made comparisons of ICs and principal components (PCs) applied to 28 Japanese stock from 1986 to 1989. Hyvärinen and Oja (2000,[23]) stated that we can find hidden factors in financial data, but they did not describe in details. Few contributions, however, exist for the application of ICA in risk management. In this thesis, in order to apply ICA into portfolio risk optimization problem (4.1), we set $Y \in \mathbb{R}^n$ to be the returns/losses of the n assets. By ICA we can find a demixing matrix $W \in \mathbb{R}^{n \times n}$ to demix Y into n statistical independent components $S \in \mathbb{R}^n$, i.e.

$$S = WY$$

where W is a non-singular demixing matrix and S are statistically independent. Therefore, we have

$$Y = W^{-1}S.$$

We can reformulate the portfolio risk optimization problem (4.1) as:

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & \mathfrak{R}(x^T Y) = \mathfrak{R}(x^T W^{-1} S) \\ \text{st} \quad & x^T \omega \geq \omega^* \\ & \sum_{i=1}^n x_i = 1 \\ & x_i \geq 0, \quad i = 1, \dots, n \end{aligned} \tag{6.2}$$

where \mathfrak{R} represents the risk measure, $x \in \mathbb{R}^n$ is the investment allocation for the assets, $\omega = (\omega_1, \omega_2, \dots, \omega_n)^T$ is the vector of the expected return for the assets and ω^* is the minimum return requirement. Instead of the joint distribution of Y , what we need now switches to

the joint distribution of S , which is easy to calculate because of the independent property for S . Suppose that we have obtain the demixing matrix W by FastICA.

Portfolio Optimization with CVaR Let $Y = (L_1, L_2, \dots, L_n)$ be the rate of loss for the assets, \mathfrak{R} be the CVaR of the portfolio, W be the demixing matrix and S be the independent components of Y obtained from FastICA. Then, from problem (6.2), the portfolio optimization with CVaR is

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & CVaR(x^T Y) = CVaR(x^T W^{-1} S) \\ \text{s.t.} \quad & x^T \omega \geq \omega^* \\ & \sum_{i=1}^n x_i = 1 \\ & x_i \geq 0, \quad i = 1, \dots, n \end{aligned} \tag{6.3}$$

Following the procedures in Chapter 4, we can transform this problem into a linear programming problem:

$$\begin{aligned} \min_{(x, \xi) \in \mathbb{R}^n \times \mathbb{R}} \quad & \xi + (1 - \epsilon)^{-1} \frac{1}{N} \sum_{j=1}^N z_j \\ \text{s.t.} \quad & x^T \omega \geq \omega^* \\ & x^T W^{-1} s^j - \xi \leq z_j, \quad j = 1, \dots, N \\ & \sum_{i=1}^n x_i = 1 \\ & x_i, z_j \geq 0, \quad i = 1, \dots, n, \quad j = 1, \dots, N, \end{aligned} \tag{6.4}$$

where $s^j = (s_1^j, s_2^j, \dots, s_n^j)$, $j = 1, \dots, N$ is the N realizations of the independent components S . We use the parametric method to fit the independent components individually. As for W and s^j , $j = 1, \dots, n$, are known, we can see that (6.4) is also a linear programming problem, which is easy to solve.

Portfolio Optimization with VaR Let $Y = (R_1, R_2, \dots, R_n)$ be the rate of return for the assets, \mathfrak{R} be the VaR of the portfolio, W be the demixing matrix and S be the independent components of Y obtained from FastICA. Then, from problem (6.2), the portfolio optimization with VaR is

$$\begin{aligned}
& \min_{x \in \mathbb{R}^n} \text{VaR}(x^T Y) = \text{VaR}(x^T W^{-1} S) & (6.5) \\
& \text{s.t.} & \\
& & x^T \omega \geq \omega^* \\
& & \sum_{i=1}^n x_i = 1 \\
& & x_i \geq 0, \quad i = 1, \dots, n
\end{aligned}$$

Following the procedures in Chapter 5, we can transform this problem into a mixed integer linear programming problem:

$$\begin{aligned}
& \min_{x, tr, y, \eta^\epsilon} & \eta^\epsilon \\
& \text{s.t.} & \tilde{r}_j = \sum_{i=1}^n x_i \sum_{k=1}^n w_{ik} s_k^j, \quad j = 1, \dots, N \\
& & \sum_{i=1}^n x_i \omega_i \geq \omega^*, \\
& & \tilde{r}_j + M(1 - y_j) \geq -\eta^\epsilon, \quad j = 1, \dots, N \\
& & \sum_{j=1}^N p_j (1 - y_j) \leq 1 - \epsilon, \\
& & \sum_{i=1}^n x_i = 1, \\
& & y_j \in \{0, 1\}, \quad j = 1, \dots, N \\
& & x_i \geq 0, \quad i = 1, \dots, n & (6.6)
\end{aligned}$$

where $s^j = (s_1^j, s_2^j, \dots, s_n^j)$, $j = 1, \dots, N$ is the N realizations of the independent components S and $W^{-1} = (w_{ik})$, $i, k = 1, 2, \dots, n$. W and s^j , $j = 1, \dots, n$, can be obtained from FastICA and the parametric method. Therefore, problem (6.6) is a mixed integer linear programming problem. We can use some optimization software (like Cplex, SAS-OR package) to solve it. What we use here is SAS-OR package.

We can see that the only difference between (6.4), (6.6) and (4.6) in Chapter 4, (Problem 2) in Chapter 5 is that we replace Y with $W^{-1}S$. Because that W is invertible and the models are linear, if the we have the relation of realizations such that $Y^j = W^{-1}S^j$, $j = 1, 2, \dots, N$, the optimal solution with ICA must be the same as the optimal solution with copula.

6.3 Numerical Results

We use the same observed rate of return data as in the Chapter 2. As discussed in section 6.1, the original sources must be non-gaussian if we want to use ICA method. Let us first assume they are non-gaussian and we will see what happened if we use ICA. If we use FastICA to estimate the demixing matrix W , we will get

$$W = \begin{bmatrix} 0.1722 & 0.1584 & -0.6838 & 0.1271 & 0.2207 \\ -0.3295 & 0.0171 & -0.0131 & -0.1699 & 1.0255 \\ 0.3305 & 0.2070 & 0.0247 & -0.7233 & 0.0114 \\ 1.5698 & -0.0801 & -0.0125 & 0.0448 & -0.0216 \\ 0.3185 & -1.3265 & 0.0230 & -0.0079 & 0.3053 \end{bmatrix}$$

and

$$W^{-1} = \begin{bmatrix} -0.0123 & 0.0263 & 0.0319 & 0.6443 & -0.0353 \\ -0.0363 & 0.2434 & -0.0415 & 0.2211 & -0.7749 \\ -1.4974 & 0.4076 & -0.3332 & 0.3703 & -0.2481 \\ -0.0678 & 0.1115 & -1.3948 & 0.3747 & -0.2471 \\ -0.03376 & 1.0033 & -0.2244 & 0.2702 & -0.0425 \end{bmatrix}$$

Note that the sign of W does not matter. Furthermore, the statistics for the independent component is given in Table 6.1. From Table 6.1, we can see that the kurtosis of the independent

Table 6.1: Statistics for 5 Independent Components

	mean	variance	skewness	kurtosis
IC1	-0.0062	1.0004	-0.3834	91.8114
IC2	0.0117	1.0004	1.2468	25.1899
IC3	-0.0260	1.0004	-1.2613	33.6423
IC4	-0.0066	1.0004	0.0442	10.0029
IC5	0.0040	1.0004	-0.1611	11.4081

components are all really big. Therefore, the independent components are all non-gaussian, which is consistent with the assumption of ICA. We can use both GST and stable distribution to fit them. In addition, the correlation matrix of the five independent components is 5×5 identical matrix. Although identical correlation matrix is not sufficient condition for independent, but actually, the independent components are statistically independent. The series plots of independent components are given in Figure 6.2. From Figure 6.2, we can see that all

the independent components have volatility clustering property. Therefore, we may also use ARMA-GARCH model to fit the independent components and use this time series model to do simulation.

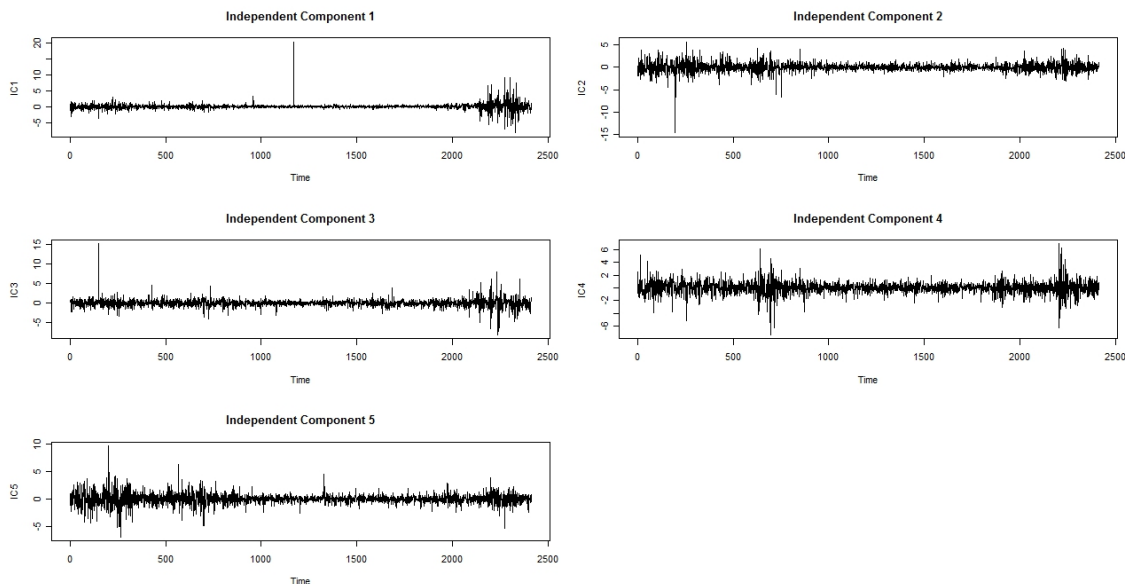


Figure 6.2: Independent Components series

6.3.1 Optimization with CVaR by ICA

To solve the Optimization with CVaR, we use the model (6.4) to find the optimal CVaR and allocations. We use the parametric method fitting independent components with both GST and stable distributions and simulate the realizations using the parameters we obtain.

Static Model If we ignore the time dependence for now, we fit the whole data using the distributions and simulate, solve (6.4) for two assets' portfolio up to five assets' portfolio as in Chapter 4, the results are shown in Table 6.2 and in Table 6.3 for GST and stable distribution, respectively, with the minimum return requirement to be $\omega^* = 0.001$.

Comparing the results in Chapter 4 Table 4.2 and Table 4.5, we can see that the optimal CVaRs and allocations are not exactly the same, but the difference is small. This is because of

Table 6.2: Optimal portfolio CVaR allocation with $\epsilon = 0.95$ and minimum return requirement $\omega^* = 0.001$ for GST distribution using ICA

	opt-CVaR	VaR	Portfolio allocation				
			PGN	IBM	BAC	F	HD
2 assets	1.4238	0.9382	0.6627	0.3373			
3 assets	1.3909	0.9286	0.6850	0.2283	0.0267		
4 assets	1.3877	0.9106	0.6772	0.2636	0.0178	0.0414	
5 assets	1.3557	0.9012	0.6257	0.2581	0.0019	0.0062	0.1081

Table 6.3: Optimal portfolio CVaR allocation with $\epsilon = 0.95$ and minimum return requirement $\omega^* = 0.001$ for stable distribution using ICA

	opt-CVaR	VaR	Portfolio allocation				
			PGN	IBM	BAC	F	HD
2 assets	1.8836	0.9221	0.8489	0.1511			
3 assets	1.8509	0.9213	0.7742	0.2247	0.0011		
4 assets	1.8112	0.9263	0.7894	0.1728	0.0019	0.0259	
5 assets	1.7595	0.9503	0.8884	0.0699	0.0000	0.0055	0.0362

the random simulation difference. On the other hand, from Table 6.2 and 6.3, we can still tell that diversification does reduce the risk of the portfolio. If we take ω^* from 0.001 to 0.0045 and make the portfolio to be the five assets we choose, we will have efficient frontier shown in Figure 6.3 and Figure 6.4 for GST and stable distribution, respectively. From these two frontier, we can see that as the minimum return requirement increase, the corresponding risk as CVaR also increase.

Dynamic Model As we saw in Figure 6.2, the independent components show volatility clustering properties, which means we can use time series ARMA-GARCH model. Here, we use ARMA(0,0)-GARCH(1,1) model for the independent component and use the most recent 2 years' data to estimate the parameters in the GARCH(1,1) model. With the parameters, we can simulate the realizations of the independent components at each time and solve the portfolio optimization problem (6.4) using those realizations. Since the independent components are statistically independent, we can set up the GARCH(1,1) model separately and do not need copula. The following Figure 6.5 and 6.6 shows the predicted daily optimal CVaRs with ICA (red curve) for GST and stable distribution, compared with those with copula (blue curve) at level .95. We can see that firstly, the daily optimal CVaRs with ICA is oscillating around

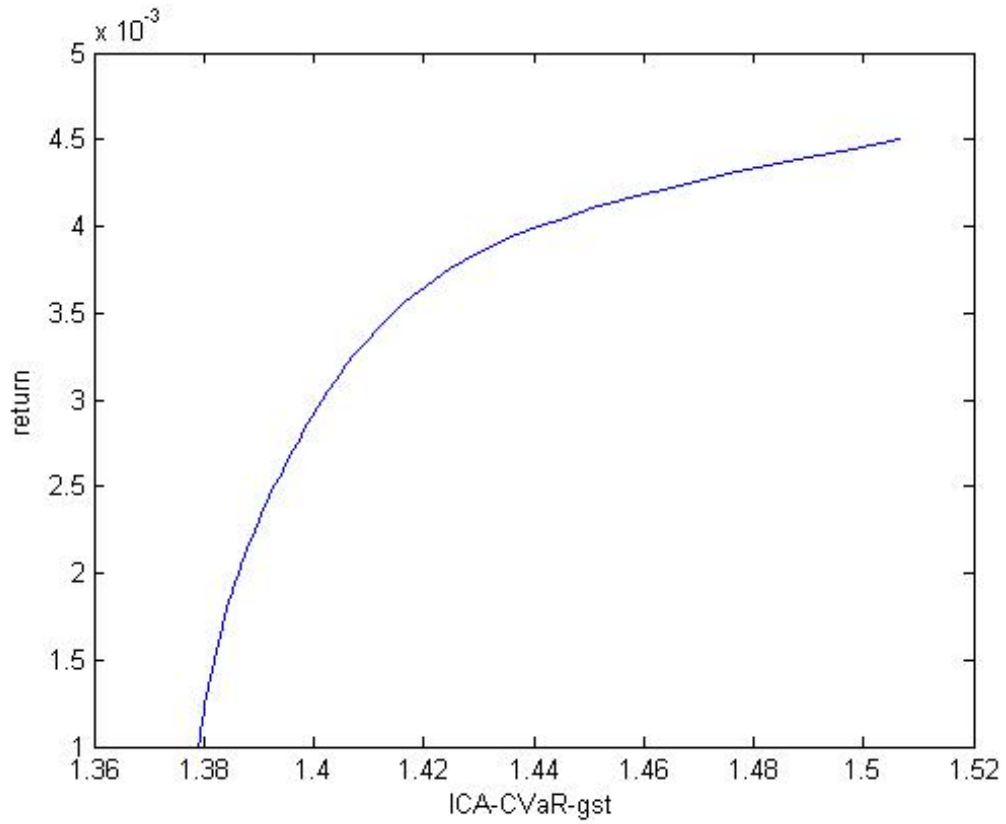


Figure 6.3: Efficient frontier with $\text{CVaR}_{0.95}$ for GST distribution using ICA

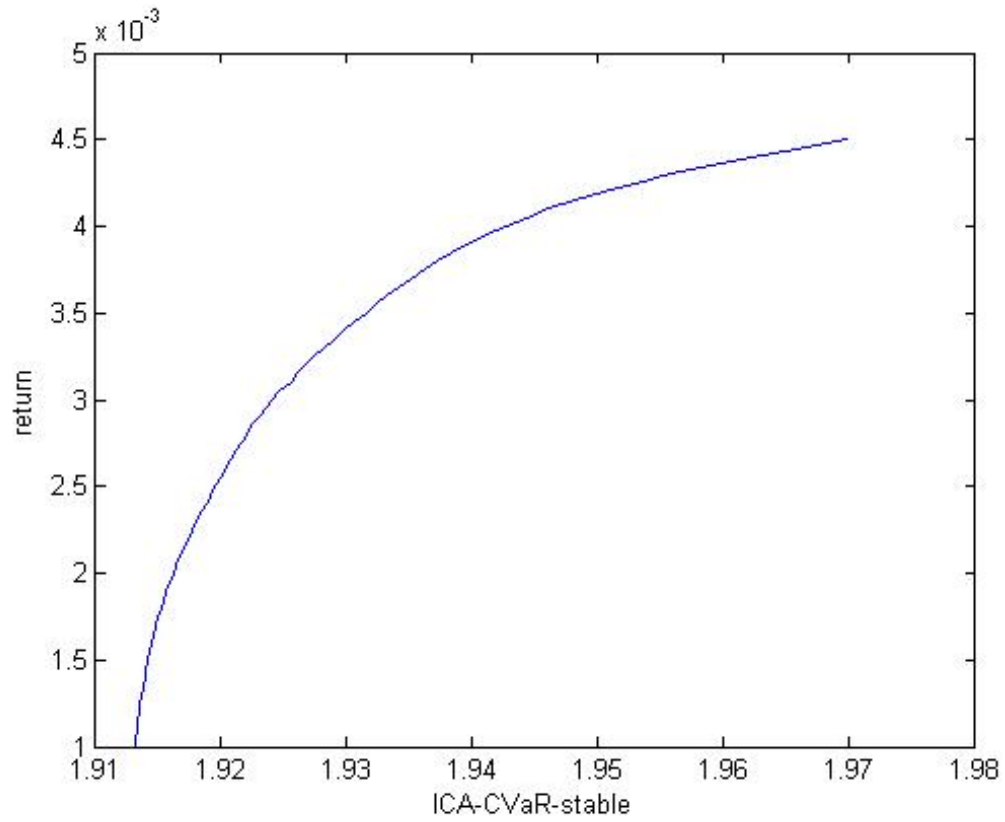


Figure 6.4: Efficient frontier with $\text{CVaR}_{0.95}$ for stable distribution using ICA

those with copula, but the difference is small most of the time. The difference is the result of the simulation error. As we said in section 6.2, if the simulated data with and without ICA have the relationship $Y^j = W^{-1}S^j$, $j = 1, 2, \dots, N$, the optimal solution with ICA must be the same as the optimal solution without ICA. However, in simulation, the realization with two scenarios can not guarantee this relationship. Therefore, there may be some simulation difference. Secondly, the pattern of the daily optimal CVaRs with ICA is almost the same as the pattern with copula. That is when the optimal CVaRs with copula is increasing, the corresponding optimal CVaRs with ICA is also increasing, vice verse. When the optimal CVaRs with copula is bigger, the optimal CVaRs with ICA is also bigger at almost the same time, vice verse. As we discussed in Chapter 4 that the pattern of optimal daily CVaRs with copula is quite consistent with the real world market and the model we set up works well. Therefore, we can conclude that the pattern of optimal daily CVaRs with ICA is also consistent with the real world market and the model we set up works well.

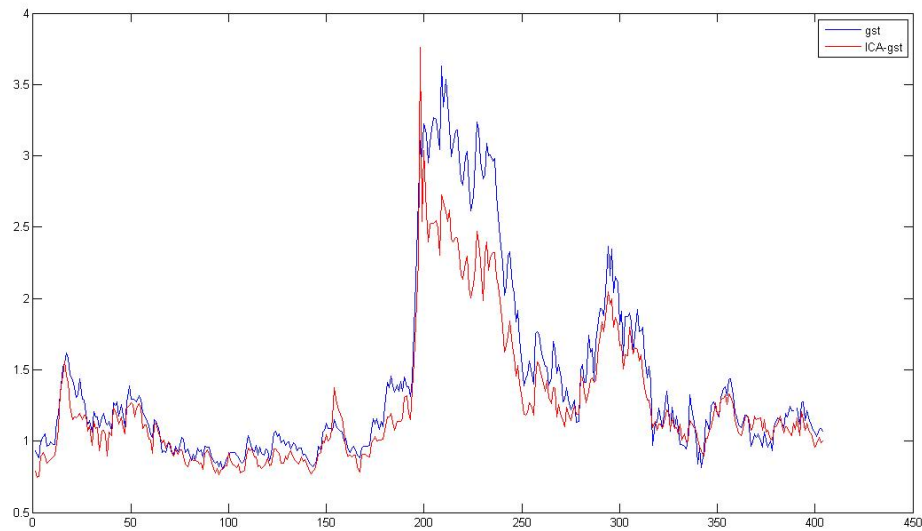


Figure 6.5: Predicted daily optimal portfolio CVaR from Jan 1, 2008 to Aug 10, 2009 for GST distribution with and without ICA at 0.95

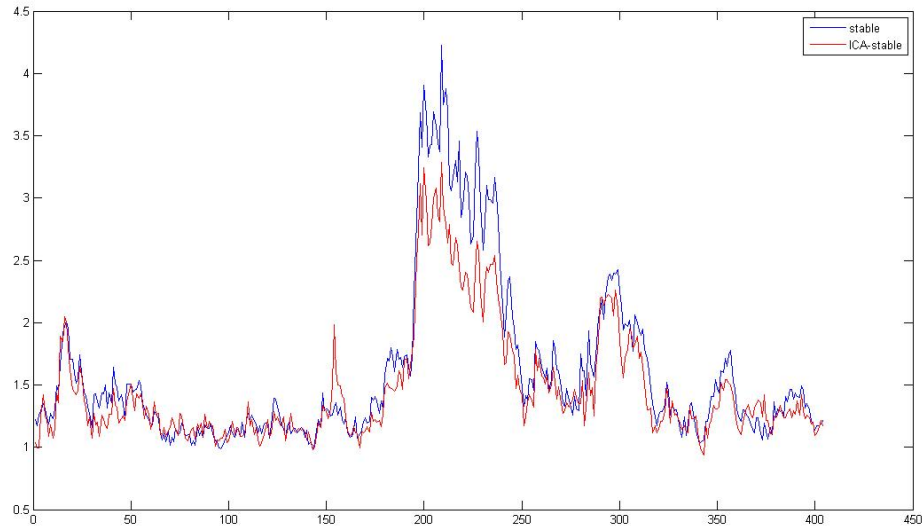


Figure 6.6: Predicted daily optimal portfolio CVaR from Jan 1, 2008 to Aug 10, 2009 for stable distribution with and without ICA at 0.95

6.3.2 Optimization with VaR by ICA

To solve the Optimization with VaR, we use the model (6.6) to find the optimal VaR and allocations. We use the parametric method fitting independent components with both GST and stable distributions and simulate the realizations using the parameters we obtain.

Static Model We ignore the time dependence first and fit the whole data using the distributions and simulate, solve (6.6) for two assets' portfolio up to five assets' portfolio as in Chapter 5. The results are shown in Table 6.4 and 6.5 for GST and stable distribution, respectively, with the minimum return requirement to be $\omega^* = 0.001$.

Comparing the results in Chapter 5 Table 5.1 and Table 5.2, we can see that the optimal VaRs and allocations are not exactly the same, but the difference is not so big. This is because of the random simulation difference. On the other hand, from Table 6.2 and 6.3, we can still tell that diversification does reduce the risk of the portfolio. If we take ω^* from 0.001 to 0.0045 and make the portfolio to be the five assets we choose, we will have efficient frontier shown in Figure 6.7 and Figure 6.8 for GST and stable distribution, respectively. From these two frontier, we can see that as the minimum return requirement increase, the corresponding risk as CVaR also

Table 6.4: Optimal portfolio VaR allocation with $\epsilon = 0.95$ and minimum return requirement $\omega^* = 0.001$ for GST distribution using ICA

	VaR	Portfolio allocation				
		PGN	IBM	BAC	F	HD
2 assets	0.8311	0.7838	0.2162			
3 assets	0.7874	0.7207	0.2326	0.0267		
4 assets	0.7728	0.7231	0.2212	0.0204	0.0253	
5 assets	0.7728	0.6923	0.2167	0.0166	0.0100	0.0644

Table 6.5: Optimal portfolio VaR allocation with $\epsilon = 0.95$ and minimum return requirement $\omega^* = 0.001$ for stable distribution using ICA

	VaR	Portfolio allocation				
		PGN	IBM	BAC	F	HD
2 assets	0.8289	0.7745	0.2255			
3 assets	0.8185	0.6899	0.2650	0.0451		
4 assets	0.7757	0.7218	0.2460	0.0232	0.0250	
5 assets	0.7668	0.7229	0.1844	0.0000	0.0167	0.076

increase, but the curve is not so smooth.

Dynamic Model We consider the time dependence of each independent component itself. Here, we use ARMA(0,0)-GARCH(1,1) model for the independent component and use the most recent 2 years' data to estimate the parameters in the GARCH(1,1) model. With the parameters, we can simulate the realizations of the independent components at each time and solve the portfolio optimization problem (6.6) using those realizations. The following Figure 6.9 and 6.10 shows the predicted daily optimal VaRs with ICA (red curve) for GST and stable distribution, compared with those with copula (blue curve) at level .95. We can see that firstly, the daily optimal VaRs with ICA is oscillating around those with copula, but the difference is small most of the time. That is because there may be some simulation difference. Secondly, the pattern of the daily optimal VaRs with ICA is almost the same as the pattern with copula. That is when the optimal VaRs with copula is increasing, the corresponding optimal VaRs with ICA is also increasing, vice verse and when the optimal VaRs with copula is bigger, the optimal VaRs with ICA is also bigger at almost the same time, vice verse. As we discussed in Chapter 5 that the pattern of optimal daily VaRs with copula is quite consistent with the real world market and the model we set up works well. Therefore, we can conclude that the pattern of

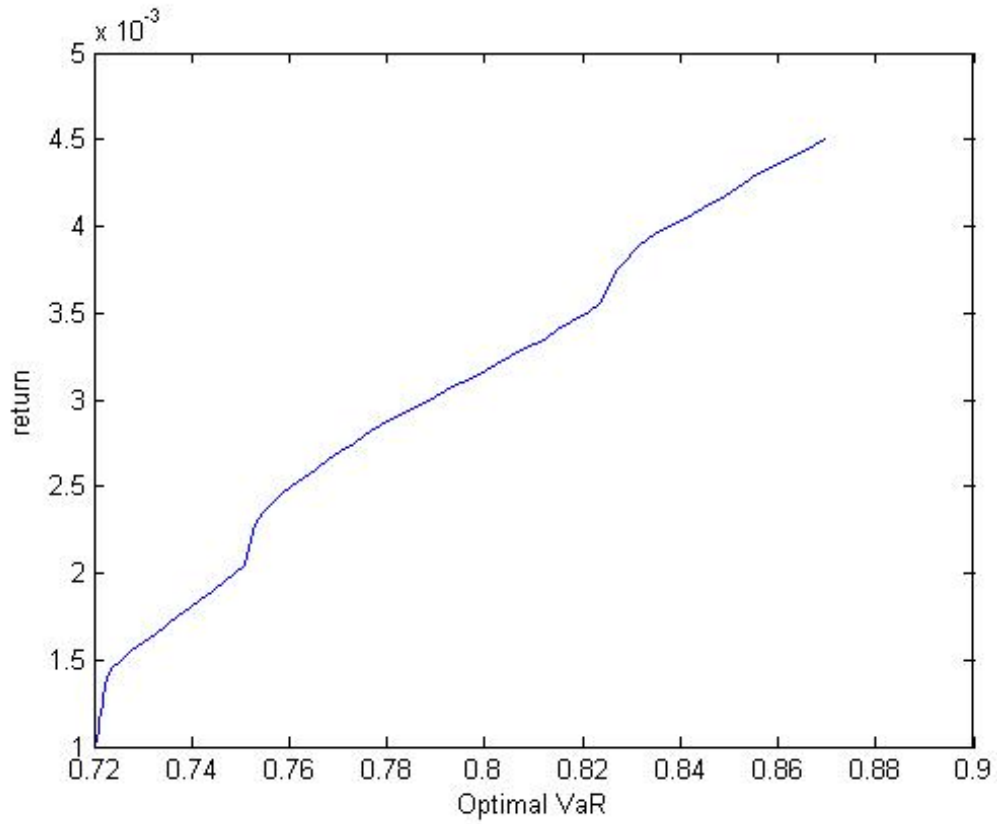


Figure 6.7: Efficient frontier with $\text{VaR}_{0.95}$ for GST distribution using ICA

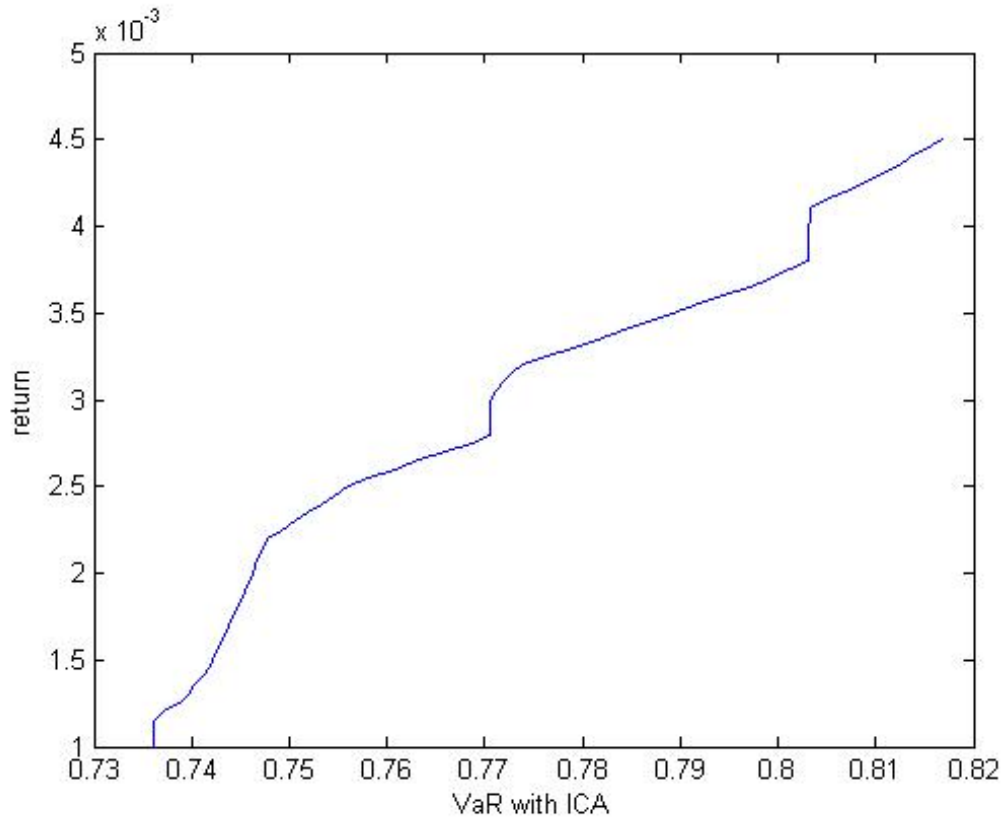


Figure 6.8: Efficient frontier with $\text{VaR}_{0.95}$ for stable distribution using ICA

optimal daily VaRs with ICA is also consistent with the real world market and the model we set up works well.

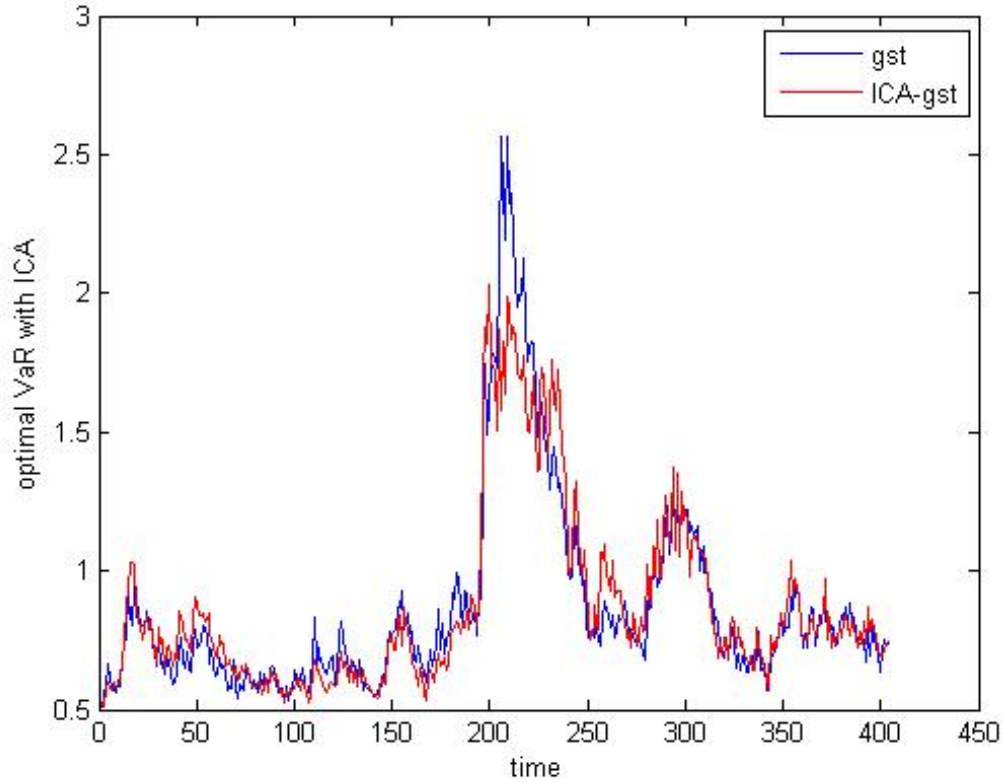


Figure 6.9: Predicted daily optimal portfolio VaR from Jan 1, 2008 to Aug 10, 2009 for GST distribution with and without ICA at 0.95

6.3.3 Running Time for Simulations

As we discussed at the beginning of the chapter, the reason we introduce ICA method is that we want to save the time of simulation by not using copula. The ICA can generate independent components which are independent with each other. Therefore, we can save time from simulating independent components, individually. After the realizations are generated, the procedure to solve the model is the same. Therefore, ICA can only save the time of simulation. In Table 6.6, we show the average time of simulation with two different method: copula method and ICA method. We can see that the time of simulation for ICA method is dramatically

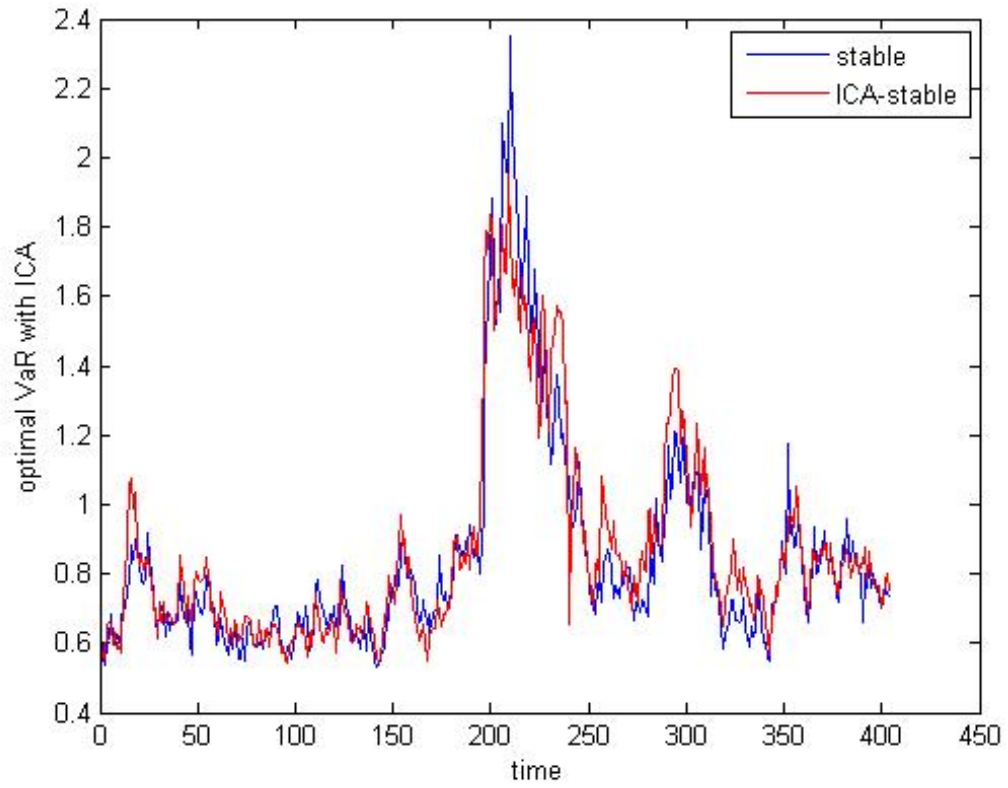


Figure 6.10: Predicted daily optimal portfolio VaR from Jan 1, 2008 to Aug 10, 2009 for stable distribution with and without ICA at 0.95

shorter than the time for copula method.

Table 6.6: Average time ellipse for ICA and copula in seconds

	time for ICA	time for copula
GST	16.6443	31.8276
stable	21.7145	41.5858

REFERENCES

- [1] Andersson, F., Mausser, H., Rosen, D. and Uryasev, S. (2001), Credit Risk Optimization with Conditional Value-at-Risk Criterion. *Mathematical Programming*, 89(2), pp. 273-291.
- [2] Artzner, P., Delbaen, F., Eber, J.M. and Heath, D. (1999), Coherent Measure of Risk. *Math of Finance*, 9(3), pp. 203-228.
- [3] Back, A.D. and Weigend A.S. (1998), A First Applicaiton of Independent Component Analysis to Extracing Structure From Stock Returns. *International Journal of Neural Systems*, 8, pp. 473-484.
- [4] Bali, T.G. and Theodossiou,P. (2007), A Conditional-SGT-VaR Approach with Alternative GARCH Models. *Annal Operations Research*, 151, pp. 241-267.
- [5] Benati, S. and Rizzi, R. (2007), A mixed integer linear programming formulation of the optimal mean/Value-at-Risk portfolio problem. *Europian Journal of Operations Research*, 176, 423-434.
- [6] Brocklebank, J.C. and Dickey, D.A. (2003), *SAS for Forecasting Time Series, 2nd Edition*. Wiley.
- [7] Bollerslev, T. (1986) Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, 31, pp. 307C327.
- [8] Box G.E.P., Tiao G.C. (1962) A Further Look at Robustness via Bayes Theorem. *Biometrika*, 49, pp. 419C432.
- [9] Charpentier, A. and Oulidi, A. (2009), Estimating Allocations for Value-at-Risk Portfolio Optimizaiton. *Mathematical Methods of Operations Research*, 69, pp. 395-410. Wiley.
- [10] Curto, J.D., Pinto, J.C., and Tavares, G.N. (2009), Modeling Stock Markets Volatility Using GARCH Models with Normal, Student's t and Stable Paretian Distributions. *Statistical Papers*, 50, pp. 311-321.
- [11] Dimakos, X.K. and Aas, K. (2004), Integrated Risk Modelling. *Statistical Modelling*, 4, pp. 265-277.
- [12] Dobrić, J. and Schmid, A (2005), Testing Goodness of Fit for Parametric Families of Copulas-Application to Financial Data. *Communications in Statistics-Simulation and Computation*, 34, pp. 1053-1068.
- [13] Engle, R.F. (1982) Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, 50(4), pp. 987C1006
- [14] Fama, E.F. (1965), The Behavior of Stock Prices. *The Journal of Business*, 38(1), pp. 34-105.
- [15] Franke, J., Härdle, W.K. and Hafner, C.M. (2008), *Statistics of Financial Markets*. Springer.
- [16] Gaivoronski, A.A. and Pflug, G. (2005), Value-at-Risk in Portfolio Optimization: Properties and Computational Approach. *Journal of Risk*, 7(2).
- [17] Gilli,M. and Kellezi, E. (2006), An Application of Extreme Value Theory for Measuring Financial Risk. *Computational Economics*, 27, pp. 207-228.

- [18] Glasserman, P., Heidelberger, P. and Shahabuddin, P. (2002), Portfolio Value-at-Risk with Heavy-Tailed Risk Factors. *Mathematical Finance*, 12, pp. 239-269.
- [19] Hansen, B.E. (1994), Autoregressive Conditional Density Estimation. *International Economic Review*, 35(3), pp. 705-730.
- [20] Harmantzis, F.C., Miao, L. and Chien, Y. (2006), Empirical Study of Value-at-Risk and Expected Shortfall Models with Heavy Tails. *The Journal of Risk Finance*, 7(2) pp. 117-135.
- [21] He, X. and Gong, P. (2009), Measuring the Coupled Risks: A Copula-based CVaR Model. *Journal of Computational and Applied Mathematics*, 233, pp.1066-1080.
- [22] Hull, J. and White, A. (1998), Value at Risk When Daily Changes in Market Variables are not Normal Distributed. *Journal of Derivatives*, 5, pp. 9-19.
- [23] Hyvärinen A. and Oja E. (2000), Independent Component Analysis: Algorithms and Application. *Neural Networks*, 13(4-5), pp. 411-430.
- [24] Khindanova, I., Rachev, S. and Schwartz, E. (2001), Stable Modeling of Value at Risk. *Mathematical and Computer Modelling*, 34, pp. 1223-1259.
- [25] Krokmal, P. and Uryasev, S. (2002), Portfolio Optimization with Conditional Value-at-Risk Objective and Constraints. *Journal of Risk*, 4(2).
- [26] Laurini, F. and Tawn, J.A. (2009), Regular Variation and Extremal Dependence of GARCH Residuals with Application to Market Risk Measures. *Econometric Reviews*, 28(1), pp. 146-169.
- [27] Liu, S.M., Brorsen, B.W. (1995), Maximum Likelihood Estimation of a GARCH-Stable Model. *Journal of Applied Economics*, 10, pp. 273-285.
- [28] Markowitz, H.M. (1952), Portfolio Selection. *Journal of Finance*, 7(1), pp. 77-91.
- [29] McNeil, A.J. and Frey, R. (2000), Estimation of Tail-related Risk Measures for Heteroscedastic Financial Time Series : An Extreme Value Approachs. *Journal of Empirical Finance*, 7, pp. 271-300.
- [30] Medova, E.A. and Smith, R.G. (2003), A Framework to Measure Integrated Risk. *Judge Institute of Management Working Paper*.
- [31] Mittnik, S., Rachev, S.T., Doganoglu, T. and Chen, Y.D (1999), Maximum Likelyhood Estimation of Stable Paretian Models. *Mathematical and Computer Modelling*, 29, pp. 275-293.
- [32] Nelson, D.B. (1991), Conditional Heteroskedasticity in Asset Returns: A New Approach. *Econometrica*, 59(2), pp. 347-370.
- [33] Nolan, J.P. (1997), Numerical Calculation of Stable Densities and Distribution Functions. *Commun. Statist. -Stochastic Models*, 13, pp. 759-774.
- [34] Nolan J.P. (2010), *Stable Distributions - Models for Heavy Tailed Data*. Birkhäuser.
- [35] Rachev, S. and Mittnik, S. (2000), *Stable Paretian Models in Finance*. Wiley.
- [36] Panorska, A., Mittnik, S., and Rachev, S.T. (1995), Stable ARCH Models for Financial Time Series. *Applied Mathematics Letters*, 8, pp. 33-37.

- [37] Rockafellar, R.T. and Uryasev, S. (2000), Optimization of Conditional Value-at-Risk. *Journal of Finance*, 7(1), pp. 77-91.
- [38] Rockafellar, R.T. and Uryasev, S. (2002), Conditional Value-at-Risk for General Loss Distributions. *Journal of Banking and Finance*, 26, 1443-1471.
- [39] Shumway, R.H. and Stoffer, D.S. (2006) *Time Series Analysis and Its Applications With R Examples, 2nd Edition*. Springer.
- [40] Theodossiou, P. (1998), Financial Data and the Skewed Generalized T Distribution. *Management Science*, 44(12), 1650-1661.
- [41] Trivedi, P.K. and Zimmer, D.M. (2007), *Copula Modeling: An Introduction for Practitioners*. NOW.