University of Alberta

RISK MEASURE ESTIMATION IN FINANCE

by

Xupeng Wang

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of
Master of Science

in

Statistics

Department of Mathematical and Statistical Sciences

Fall, 2010
Edmonton, Alberta

Examining Committee

Dr. Ivan Mizera, Department of Mathematical and Statistical Sciences (Supervisor)

Dr. Peter Hooper, Department of Mathematical and Statistical Sciences (Committee chair)

Dr. Valentina Galvani , Department of Economics

Abstract

In financial market, risk management is very critical to a company. However, some risks in the market (market risk) can not be controlled or eliminated through management improvement or appropriate asset allocation. Thus, it is important to accurately measure these kinds of risks.

In this thesis, we introduce two most widely used risk measures: value-at-risk and expected shortfall. Their estimation from data is the issue we are concerned with in this thesis. We divide this thesis into two parts:

First, we survey the currently used estimation methods. We introduce these methods from the theoretical backgrounds. Then, we propose some criteria used to judge the performance of these methods.

Second, we apply all these methods to data. We use the criteria introduced to compare these methods. This empirical study can shed some light on the application of these methods, bringing us some guidelines about their use in the future.

Acknowledgements

I have been extremely fortunate to have had Professor Ivan Mizera as my supervisor throughout my master studies. I sincerely thank him for his great guidance and support on both my professional and personal developments. Without his help, I could not finish the thesis.

I am grateful to my committee members, Professor Peter Hooper and Professor Valentina Galvani, for reading and evaluating my thesis. Thank all my friends for their support and good cheer.

Finally, I am deeply thankful to my dear family for their unconditional love and support.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

In the financial industry, all institutions are exposed to a great variety of risks. Whether they can successfully deal with risks directly determines whether they could survive in this field and earn profits. From theory, risk in finance is defined as the degree of uncertainty about future net earning. Based on the sources of the risk, all kinds of risks can be classified into four categories[18]: default risk, liquidity risk, operational risk and market risk. Default risk is the potential loss for your counterpart's inability to meet the debt obligation, such as unable to repay the interest or principal. Liquidity risk is caused by the lack of cash or cash equivalent so that the corporation can not continue the business. If the corporation is facing liquidity troubles, it may be forced to sell some of its assets at discount to obtain cash. Operational risk accounts for possible losses deriving from execution of inadequate or failed internal processes, people and systems, or from external events. For example, accounting errors, theft of information, inappropriate business decision all belong to this kind of risk. Market risk estimates the uncertainty coming from changes of market condition, such as interest rate, economic environment or inflation rate. One

great distinction between market risk and the first three is that it cannot be greatly reduced or eliminated through diversification, appropriate asset allocation and management control. Therefore, the best we can do is to correctly estimate the market risk and take the proper safeguards.

Beyond dispute, accurate measures of market risk are of great importance. Many methods of measuring risks have been developed and the most commonly used one is Value at Risk (VaR). VaR is defined as the minimum amount of loss in value of a portfolio with a specified probability over a certain period. To put it in a simple way, it gives a lower bound on the loss made in the worst p percent of scenarios with a certain probability and period. From mathematical standpoint, VaR is actually a quantile of the distribution of future returns or losses. More specially, conditional on the information up to time t, the VaR for period $t+h$ of the financial return could be expressed as

$$\begin{aligned} VaR_p(t+h) &= Q_p(r_{t+h} \mid F_t) && (1.1) \\ &= \inf_r(r \in Re : Pr(r_{t+h} \leq r \mid F_t) \geq p), \quad 0 < p < 1, && (1.2) \end{aligned}$$

where $Q_p(*)$ denotes the quantile function; r_t is the financial return on an asset or portfolio in period t; and F_t represents the information available at time t.

This provides us the theoretical foundations upon which we can develop methods to estimate the VaR. The great popularity of VaR among financial institution lies in its theoretical simplicity. They reduce the market risk of any portfolio to just a number, which is easier for us to evaluate and compare.

VaR is widely used in different aspects, from risk management, evaluation of the performance of risk manager to regulatory requirements. What makes

the use of VaR so prevalent in financial industry is the requirement so-called "Basel rules"[19] imposed by securities commissions. All banks and financial institutions are required to report their estimate of VaR and hold capital reserve based on it. VaR has to be reported daily. Thus, precise estimate of this risk measure is quite important. If the underlying market risk is not correctly estimated, this may cause improper capital allocation with the result of losing profitability, incurring unnecessary loss or even bankruptcy.

However, VaR has been criticized for its failure to meet the coherent conditions set by Artzner, Delbaen and Eber in 1998[2], which is discussed in detail later. In addition, the practical usefulness of VaR is limited by the fact that it tells nothing about the potential size of loss beyond VaR. Therefore, the expected value of the loss or returns beyond VaR has been proposed as a supplement. This value is usually called expected shortfall (ES). From the view of mathematics, ES for a financial return can be defined as

$$ES_p(t+h) = E(r_{t+h}|r_{t+h} < VaR_{t+h}), \quad 0 < p < 1. \tag{1.3}$$

In contrast with its conceptual simplicity, the estimation of VaR and ES is a very challenging topic in the academic field. The fact that return distribution does not stay constant over time or belong to any typical distributions we have known, puts tremendous complications in the VaR and ES estimation. The main objective of this paper is to survey the most currently used methods of estimating VaR and ES, summarize their advantages and flaws from theoretical analysis and empirical applications, in the hope of providing some guidelines for our use of these methods.

Because the estimation of ES is based on that of VaR, we mainly discuss

the methods of estimating VaR. The performances of all these methods are evaluated through interval forecast test and empirical application, from aspects of accuracy, procedure simplicity and applicable conditions. Moreover, we use Monte Carlo simulation to study the performance of these methods on ES estimation.

The thesis is organized as follows: Chapter 2 is an introduction to characteristics of financial data and how they influence the estimation methods; Chapter 3 reviews most currently used methods for estimating VaR; Chapter 4 describes the criteria used to evaluate the performance of these methods; Chapter 5 discusses the properties of coherent risk measure and necessity of ES as a supplement to describe the risk; Chapter 6 is the core of the thesis, for it provides the details of how these methods are applied to data and reports the empirical results, which is the base for our conclusion. Finally, Chapter 7 briefly concludes the thesis.

Chapter 2

Properties of financial data

In order to build a model to predict risk in the financial industry, one needs first to understand the properties of financial data. These properties can lead us to create models which are able to reflect the realities in financial industry, and give us standards to evaluate the fitness of the methods. The most significant data in finance are prices, so we first research the dynamic distribution of prices. Even though financial practitioners usually prefer to work with the concepts of profit and loss, it is not well suited for risk management, with returns being a preferred measure. There are two equivalent ways to calculate returns[8],

$$r_t = \frac{P_t - P_{t-1}}{P_{t-1}} \quad and \tag{2.1}$$

$$r_t = \log(\frac{P_t}{P_{t-1}}), \tag{2.2}$$

where P_t is the price of stock or portfolio for time t. The compound return in (2.2) is generally preferred for risk analysis, due to its connection with common views of the distribution of returns, as well as the link with derivatives pricing.

Some empirical facts about financial returns are already known, due to the contribution of Mandelbrot (1963)[17] and Fama (1965)[12]. Financial returns usually exhibit the following characteristics:

(1) non-normality of returns,

(2) volatility clustering, and

(3) asymmetry in return distribution.

Now, we discuss these characteristics in more detail in following sections.

2.1 Non-normality

The fact that returns are not normally distributed is recognized both by risk managers and supervisory authorities. The non-normality property implies the following the relationship between the return distribution and a normal distribution with the same mean and variance[8]:

(1) the center of the return distribution is higher,

(2) the sides of the return distribution are lower, and

(3) the tails of the return distribution are higher.

Figure (2.1)and Figure(2.2) can evidently exhibit these features.

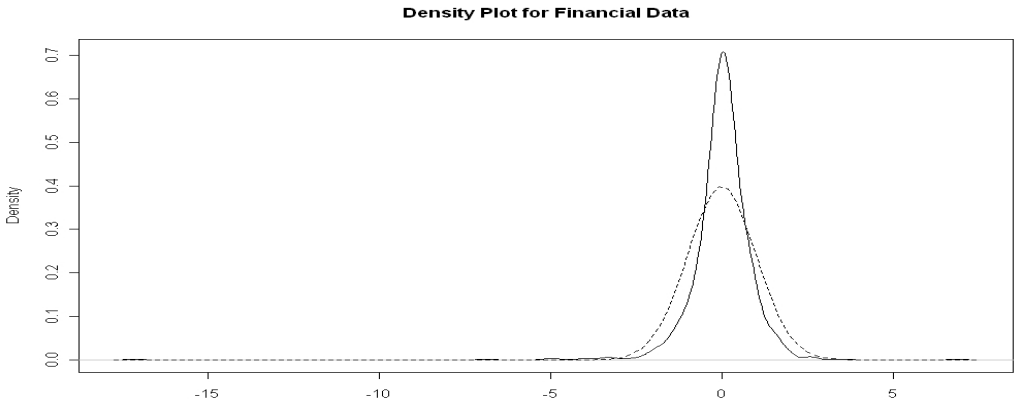

Figure 2.1: Solid line is the density plot for standardized daily returns for NYSE (New York Stock Exchange) Composite Index, from Jan 3rd,1986 to Dec 31st, 2002. It has mean=0 and standard deviation=1. Dashed line is the density plot for standard normal distribution. Non-normality, especially higher center and lower sides, can be clearly seen in this plot.

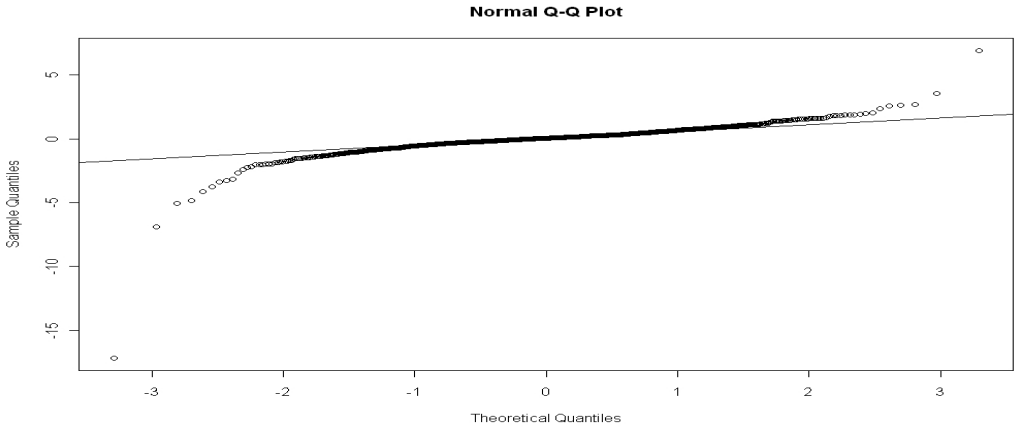

Figure 2.2: QQ-plot for standardized daily returns for NYSE (New York Stock Exchange) Composite Index, from Jan 3rd, 1986 to Dec 31st, 2002. It has mean=0 and standard deviation=1. Solid line is for standard normal distribution. Non-normality,especially higher tails, can be clearly seen in this plot.

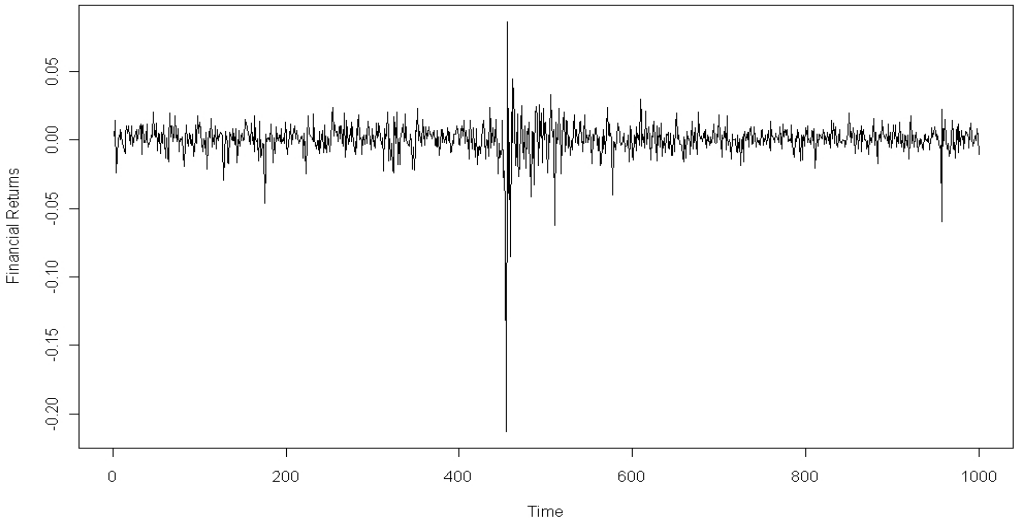

Figure 2.3: Plot for daily returns of NYSE (New York Stock Exchange) Composite Index, from Jan 3rd, 1986 to Dec 14th, 1989.

This implies that the market is either too quiet or too turbulent relative to the normal distribution. Of these three points in non-normality, the last one is most relevant to risk. The heavy-tailed property results in large losses and gains being more frequent than predicted by a normal model. An assumption of normality for the lower tail is increasingly inaccurate with the further going into the tail. For example, if one uses the normal distribution to forecast the probability of the crash in 1987 (Figure(2.3)), one would estimates that a crash of the 87's magnitude occurs only once in human's history. Most financial analysis, until recently, has been based on the assumption of normality of returns. The reason for that is the mathematical tractability of normal, since non-normal distributions are very difficult to work with. While normality may be a relatively innocuous assumption in many other situations, it is disastrous in risk management and needs to be addressed.

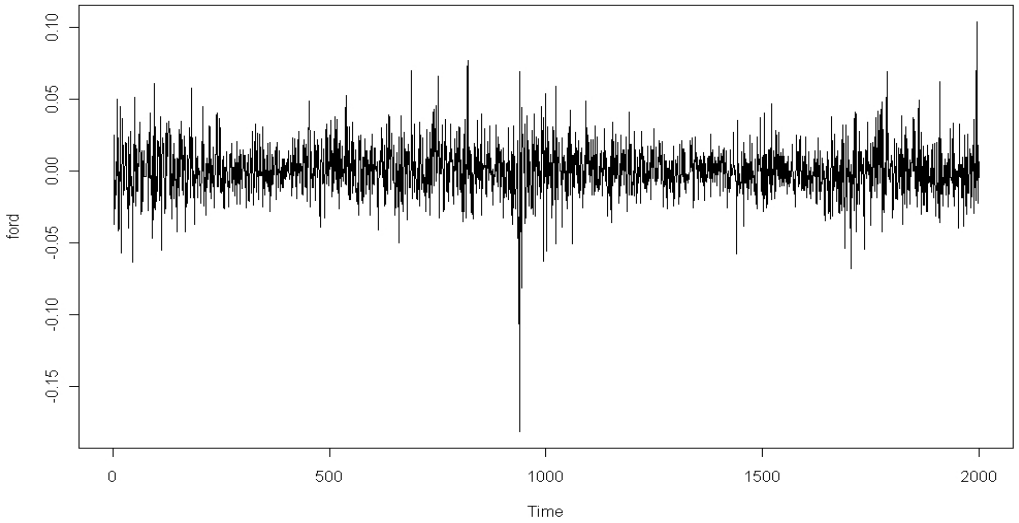

Figure 2.4: Plot for daily returns of auto manufacturer Ford's stock, from Feb 2nd, 1984 to Dec 31st, 1991.

2.2 Volatility clustering

The second obvious fact is the returns go through periods when volatility is high and period when volatility is low. This means that large volatilities tend to cluster together and small volatilities tend to cluster together. This phenomenon could be apparently seen from Figure (2.3) and (2.4). The natural inspiration from this fact is that we would make a forecast with large volatility if we know the market is in a high volatility state, and vice versa. It requires our risk measure model should adjust the volatility instantly according to the market. Any risk measure model which cannot reflect this volatility change will fail in its prediction.

A direct consequence of volatility clustering is that financial return exhibits dependence upon its anterior ones. Extreme returns, either positive or negative, tend to occur in a period of high volatility. The chance of extreme return

happening after a previous extreme return is much greater than the chance of its happening after a regular(non-extreme) return.

2.3 Asymmetry

One feature of conditional volatility models is the implicit assumption of symmetry of the return distribution. As discussed below, this is not correct. Usually, one of the tails is heavier than the other. For example, for equities, the lower tails is commonly thicker than the upper tail. In general, if the market trend is upward, the upper tail is thinner than the lower tail, it means the market moves in small steps in the direction of the market trend, and in large jumps away from this trend. Thus, if we are using symmetrical distribution, such as normal or student-t distribution, we would underpredict losses relative to gains.

We should say the best way to predict risk would be to use skewed conditional distributions. However, this is not commonly done because of the difficulty of finding an appropriate distribution and estimating its parameters.

2.4 Violation Clustering

When the concepts of volatility clustering applies to the return data, it is also of interest to consider clustering in violations of the VaR. Violation clustering is the phenomenon that violations of VaR (the real return is smaller than VaR) are likely to occur in series, not independently.

While financial extreme returns have a tendency to cluster, the job of risk measure is to forecast this trend. It is a very important feature for the risk

measure to adjust quickly to the new returns. Therefore, violation clustering is a phenomenon we should avoid when building a risk measure. Instead, we should set a model which can exhibit independent violation.

2.5 Summary

These properties we have discussed, should give us some insight and guidance to build an appropriate VaR estimation model. In brief, these points can be summarized:

(1) the model should exhibit non-normality of the return distribution ,

(2) the model should capture the asymmetry in return distribution, and

(3) the model should adjust to volatility quickly, exhibiting independent violation.

Chapter 3

VaR Estimation Methodologies

While VaR is a very easy and intuitive concept, its estimation is a very challenging statistical problem due to the facts discussed in last chapter that financial returns exhibit "non-standard" statistical characteristics. Although the existing methods for estimating VaR employ different methodologies, they all follow a common structure, which can be summarized in three steps:

(1) Record the returns of the portfolio,

(2) Estimate the distribution of portfolio returns, and

(3) Compute the risk measure of the portfolio.

The major difference among these methods is reflected in the second step, which deals with the way of estimating the change in the portfolio returns. Based on this point, we can classify the current models into three broad categories: Non-Parametric, Parametric and Semi-parametric.

These methods yield very different results. Some methods outperform in certain circumstances while others show superiority in other situations. Our

objective here is to analyze these differences and provide some guidelines for future selection of these methods. The models are designed to account for the three characteristics of financial returns. And each method is introduced in the following procedure:

(1) Provide a formula for calculating VaR(t) as a function of variables known at time t-1 and a set of parameters that need to be estimated,

(2) Provide a procedure (a loss function and a suitable optimization procedure) to estimate the set of unknown parameters,

(3) Analyze the properties of this method from the theoretical aspect.

Finally, in chapter 6, we compare their performance from the empirical application.

3.1 Nonparametric Method

3.1.1 Historical Simulation

One of the most common methods for VaR estimation is Historical Simulation. This method drastically simplifies the procedure for computing the VaR, since it does not make any distribution assumption about portfolio returns. Historical Simulation is based on the concept of a rolling window. First, one needs to choose a window of observations, which generally ranges from 6 months to two years. Then portfolio returns with this window are sorted in ascending order and the p -quantile of interest is given by the return that leaves p of the observations on its left side and $1-p$ on its right side. If such a number falls between two consecutive returns, then some interpolation rule is applied. To

compute the VaR the following day, the whole window is moved forward by one observation and the entire procedure is repeated. This estimation can be expressed as:

$$\widehat{VaR}(p) = |M - pK| r_{[M+1]} + |pK + 1 - M| r_{[M]}, \tag{3.1}$$

where K is the number of observations in the window and M is the lower integer of pK.

Even if this approach makes no explicit assumptions on the distribution of portfolio returns, an underlying assumption is hidden behind this procedure: the distribution of portfolio returns does not change within the window. From this implicit assumption several problems arise.

First, this method is simply controversial. If all the returns within the window are assumed to have the same distribution, then the logical consequence must be that all the returns of the time series must have the same distribution. Second, the empirical quantile estimator can be accurate only if K, the window size, goes to infinity. Third, the window size is a disputable issue. If the size is too large, the measure will respond very slowly to the latest change in the market. If it is too small, it will adversely affect the estimation accuracy because there are too few observations contained in the sample. These are two contradictory points very hard to balance.

Moreover, assume that the market is moving from a period of relatively low volatility to a period of relatively high volatility. In this case, VaR estimates based on the historical simulation will be biased upward, since it will take some time before the observation with the low volatility to leave.

3.2 Parametric Method-GARCH model

3.2.1 Introduction

Fully parametric methods are based on the assumption that financial returns belong to a location-scale family of probability distribution in the form

$$r_t = u_t + \epsilon_t = u_t + \sigma_t z_t, \tag{3.2}$$

where location u_t and scale σ_t are parameters and $z_t \sim^{iid} f_z$, where f_z is a zero-location, unit-scale probability density distribution. The h-period-ahead VaR forecast based on information up to time t is

$$\widehat{VaR}_{t+h} = \widehat{u}_{t+h} + \widehat{\sigma}_{t+h} Q_p(z), \tag{3.3}$$

where $Q_p(z)$ is the p-quantile implied by f_z.

Different approaches use different specifications on the conditional location u_t, conditional scale σ_t and especially z_t.

Unconditional parametric models set $u_t \equiv u$ and $\sigma_t \equiv \sigma$, which will come to Ito's lemma process after taking derivative to t.

Conditionally homoskedastic parametric models account for a changing conditional mean, which is captured by an ARMA(p,q) models:

$$u_t = a_0 + \sum_{i=1}^{p} a_i r_{t-i} + \sum_{j=1}^{q} b_j \epsilon_{t-j} \tag{3.4}$$

with $\sigma_t \equiv \sigma, t = 1, 2, \cdots, T$.

In view of volatility clustering in financial data, this model needs some

improvements for its inability to capture the volatility change. Naturally, conditionally heteroskedastic would be our choice. This model allows us to describe the scale parameter with a function of past information. If we update the volatility change with the form:

$$\sigma_t^2 = c_0 + \sum_{i=1}^{r} c_i \epsilon_{t-i}^2 + \sum_{j=1}^{s} d_j \sigma_{t-1}^2, \tag{3.5}$$

which would become the most widely used parametric method, generalized autoregressive conditional heteroskedasticity, GARCH(r, s) model.

3.2.2 GARCH Model

In econometrics and finance, a model featuring autoregressive conditional heteroskedasticity considers the variance of the current residual term to be a function of the actual sizes and variance of the previous time periods' residual terms. Such models are often called GARCH models (Bollerslev,1986)[4], although a variety of other forms are applied to particular structures of model which have a similar basis. GARCH models perform very well in modeling financial time series that exhibit time-varying volatility clustering[10]. That is one reason why it is selected to be the base for estimating the risk measure.

To give us a clear view of this method, we list the structure of this method again. We model the financial returns in the form:

$$r_t = u_t + \epsilon_t = u_t + \sigma_t z_t. \tag{3.6}$$

And we obtain the above u_t and σ_t through following processes:

$$u_t = a_0 + \sum_{i=1}^{p} a_i r_{t-i} + \sum_{j=1}^{q} b_j \epsilon_{t-j}, \tag{3.7}$$

$$\sigma_t^2 = c_0 + \sum_{i=1}^{r} c_i \epsilon_{t-i}^2 + \sum_{j=1}^{s} d_j \sigma_{t-j}^2, \tag{3.8}$$

where a_i, b_i, c_i *and* d_i are parameters we need to estimate.

Then the one-day-ahead p-VaR for r_t can be estimated by:

$$VaR_{t+1} = u_{t+1} + \sigma_{t+1} Q_p(z). \tag{3.9}$$

This model lies on two important points: the accurate expression of the variance equation and the assumption that the standardized residuals are i.i.d. The first one was triggered by the characteristics of financial data discussed above. The assumption of i.i.d standardized residuals, however, is just a necessary condition to estimate the unknown parameters in the model. A further step to implement any GARCH method is to determine the specification of the distribution of z_t. It is commonly assumed that z_t follows the standard normal distribution $N(0, 1)$, because it is easier for us to derive some conclusion in theory and estimate the unknown parameters in the model.

As we know, GARCH model is actually composed of two parts: we use ARMA to describe the conditional mean and GARCH model to capture the changing volatility. For simplicity, we just call it "GARCH" model. Therefore, the estimation of the parameters is also to be conducted in two steps. We first estimate the parameters in ARMA by minimizing the conditional sum-of-

square error, the details of which can be referred to [13]. Then, we utilize the maximum likelihood method to estimate the parameters in GARCH model[21]. Since we assume the error terms follow standard normal distribution, the log-likelihood of the models for the observed series r_t with length n is

$$\log(L) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{n}\frac{(r_t - u_t)^2}{\sigma_t^2} - \frac{1}{2}\sum_{t=1}^{n}\log\sigma_t^2, \tag{3.10}$$

where σ_t^2 is in the form of (3.8).

Thus the parameters could be obtained by maximizing $\log(L)$. Normally, GARCH(1,1) is good enough to capture the change of volatility in financial data and convenient for us to estimate the parameters. In the empirical study later, we use GARCH(1,1) as well.

The primary advantage of this model is that it allows a complete characterization of the distribution of returns and there may be room for improving their performance by avoiding the normality assumption. The main problem of this approach is that it tends to underestimate the VaR, because the normality assumption of the standardized residuals seem to be at odds with the real behavior of financial returns. Furthermore, this method is subject to three different sources of error: the specification of the variance equation and the distribution chosen to build the likelihood may be wrong, and the standardized residuals may not be i.i.d. Whether or not these misspecification issues are relevant for VaR estimation purpose is mainly an empirical issue.

Fortunately, Bollerslev and Woolridge (1992)[5] have showed that the maximization of the normal GARCH likelihood is able to deliver consistent estimates, provided that the variance equation is correctly specified, even if the standardized residuals are not normally distributed. Based on this important

result, we can use other distributions, like student-t or exponential based distributions instead of standard normal distribution, which can better reflect the skewness and fat-tail of financial returns, to improve the performances of GARCH method.

3.3 Semi-parametric Method

Included in the semi-parametric category are methods based on Extreme Value Theory (EVT) or Quantile Regression.

3.3.1 Extreme value theory

Origin of Extreme Value Theory

Extreme value theory is concerned with the distribution of the smallest and largest order statistics. While the models discussed in previous subsections specify the entire return distribution, the EVT approach focuses only the tails of the return distribution. Thus, it is very natural to consider employing extreme value theory when extreme VaR (normally $p \leq 0.05$) estimation is our interest.

In this subsection, we first review the extreme value theory in the statistical literature[23]. Denote the return of an asset or portfolio, measured in a fixed time interval such as daily, by r_i. Consider the collection of n returns $r_1, r_2 \cdots r_n$. The minimum return of the collection is $r_{(1)}$, that is, the smallest order statistic. We focus on properties of the minimum return $r_{(1)}$ because of the convenience for discussion. And this minimum corresponds to VaR calculation for a long position. However, the theory discussed here also applies to

the maximum return of an asset or portfolio over a given time period because properties of the maximum return can be obtained from those of the minimum by a simple sign change. The maximum return is relevant to holding a short financial position.

When applying extreme value theory, we must assume that the returns r_t are serially independent with a common cumulative distribution function $F(x)$ and that the range of the return r_t is $[L, U]$. For log returns, we have $L = -\infty$ and $U = \infty$. Then the CDF(Cumulative Distribution Function) of $r_{(1)}$, denoted by $F_{n,1}(x)$, is given by

$$\begin{aligned} F_{n,1}(x) &= Pr[r_{(1)} \leq x] = 1 - Pr[r_{(1)} > x] \\ &= 1 - Pr[r_1 > x, r_2 > x, \cdots, r_n > x] \\ &= 1 - \prod_{j=1}^{n} Pr(r_j > x) \quad (\textit{by independence}) \\ &= 1 - \prod_{j=1}^{n} [1 - Pr(r_j \leq x)] \\ &= 1 - \prod_{j=1}^{n} [1 - F(x)] \\ &= 1 - [1 - F(x)]^n. \end{aligned}$$

In practice, the CDF $F(x)$ of r_t is unknown and, hence, $F_{n,1}(x)$ of $r_{(1)}$ is also unknown. However, as n increases to infinity, $F_{n,1}(x)$ becomes degenerate, $F_{n,1}(x) \longrightarrow 0$ if $x \leq L$ and $F_{n,1}(x) \longrightarrow 1$ if $x > L$ as n goes to infinity.

This degenerate CDF has no practical value. Therefore, the extreme value theory is concerned with finding two sequences c_n and d_n, where $c_n > 0$, such that the distribution of $(r_{(1)} - d_n)c_n^{-1}$ converges to a non-degenerate distribution H as n goes to infinity. The sequence d_n is a location series and c_n is a series

of scaling factors. Under the independence assumption for r_i, Fisher-Tippett theorem says that if there exist c_n and d_n, such that the distribution of the normalized minimum $x = (r_{(1)} - d_n)c_n^{-1}$ converges to limiting distribution H as n goes to infinity, then H is in the form

$$H \equiv H_k(x) = 1 - exp(-(1+kx)^{1/k}) \quad with\ 1 + kx > 0. \tag{3.11}$$

The special case $H_0(x)$ is interpreted as $\lim_{k\to 0} H_k(x)$. H is called the Generalized Extreme Value (GEV) distribution and describes the limit distribution of normalized minimum($x = (r_{(1)} - d_n)c_n^{-1}$). The parameter k is referred to as the shape parameter that governs the tail behavior of the limiting distribution. An important concept for the application of EVT to VaR estimation is Maximum Domain of Attraction: the random variable X belongs to the Maximum Domain of Attraction of the extreme value distribution H (and we write $X \in MDA(H)$) if and only if Fisher-Tippett theorem holds for X, with limit distribution H.

It has been proved by Gnedenko (1943) that $F(x)$ of r_t is strongly associated with the types of limiting distribution. Simply speaking, the tail behavior of $F(x)$, not the specific distribution, determines the limiting distribution H of the normalized minimum. However, the corresponding c_n and d_n may depend on the CDF $F(x)$. Thus, this theory is applicable to a wide range of distributions for the return r_t. To be specific, heavy tailed distributions(such as Cauchy, Student-t, Pareto, log-gamma) belong to $MDA(H)$, for $k < 0$. Gamma, normal, log-normal and exponential distributions belong to $MDA(H_0)$, while distributions with finite endpoints (such as the uniform and beta) belong to $MDA(H_0)$, for $k > 0$.

Although this extreme value theory seems applicable to practice, it encounters some problems due to the estimation of parameters (k, c_n and d_n). Thus, this method is not the one directly used for estimating VaR. Instead, it is the foundation of the advanced method.

A New Approach on EVT

In this part, a new Extreme Value Theory[23] based on last section is introduced. Here, we use maximum instead of minimum of r_t for ease of our presentation. Thus, it is necessary to derive the GEV (Generalized Extreme Value) distribution for normalized maximum ($x = (r_{(n)} - d_n)c_n^{-1}$), which becomes

$$H_* \equiv H_{*k}(x) = exp(-(1-kx)^{1/k}) \quad with\ 1-kx > 0, \tag{3.12}$$

where the special case $H_{*0}(x)$ is also interpreted as $\lim_{k\to 0} H_{*k}(x)$.

The basic theory of this new approach is to consider the conditional distribution of $r = x + u$ given $r > u$ for the limiting distribution of the maximum given in (3.12). Then the conditional distribution of $r \leq x + u$ given $r > u$ is

$$\begin{aligned} \Pr(r \leq x+u \mid r > u) &= \frac{\Pr(u \leq r \leq x+u)}{\Pr(r > u)} \\ &= \frac{\Pr(r \leq x+u) - \Pr(r \leq u)}{1 - \Pr(r \leq u)}. \end{aligned} \tag{3.13}$$

Using the CDF H_* given in (3.12) and the approximation $e^{-z} \approx 1 - z$, we

can get

$$\begin{aligned}
\Pr(r \le x+u \mid r > u) &= \frac{H_*(x+u) - H_*(u)}{1 - H_*(u)} \\
&= \frac{\exp(-[1 - \frac{k(x+u-d)}{c}]^{1/k}) - \exp(-[1 - \frac{k(u-d)}{c}]^{1/k})}{1 - \exp(-[1 - \frac{k(u-d)}{c}]^{1/k})} \\
&= 1 - \frac{1 - \exp(-[1 - \frac{k(x+u-d)}{c}]^{1/k})}{1 - \exp(-[1 - \frac{k(u-d)}{c}]^{1/k})} \\
&\approx 1 - \frac{[1 - \frac{k(x+u-d)}{c}]^{1/k}}{[1 - \frac{k(u-d)}{c}]^{1/k}} \\
&= 1 - [1 - \frac{kx}{c - k(u-d)}]^{1/k}, \qquad (3.14)
\end{aligned}$$

where $x > 0$ and $c - k(u-d) > 0$.

In (3.14), if we make $\beta = c - k(u-d)$ and $\xi = -k$, we will have the expression as:

$$G_{\xi,\beta}(x) = 1 - [1 + \frac{\xi x}{\beta}]^{-1/\xi}, \qquad (3.15)$$

which is the Generalized Pareto Distribution(GPD).

Based on the CDF in (3.15), we easily derive the expression of PDF(Probability Density Function) for Generalized Pareto Distribution as below:

$$g_{\xi,\beta}(x) = \frac{1}{\beta}[1 + \frac{\xi x}{\beta}]^{-1/\xi - 1}, \qquad (3.16)$$

Combining (3.13),(3.14) and (3.15), we conclude that the conditional distribution of $r \le x + u$ given $r > u$ can be approximated by Generalized Pareto Distribution(GPD). And this conclusion was verified by Pickands(1975)

through the property:

$$F \in MDA(H_k) \quad \Leftrightarrow \quad \limsup_{u \to x_F \;\; 0<x<x_F-u} \mid F_u(x) - G_{k,\beta}(x) \mid = 0, \qquad (3.17)$$

where x_F is the right endpoint (which could be infinite) of the distribution F. This result states that if F is in the Maximum Domain of Attraction of GPD, as the threshold u approaches the endpoint of F, GPD becomes an accurate approximation of the excess distribution function F_u.

The parameters (ξ and β) in GPD can be estimated by maximum likelihood, once the threshold u is given. These parameters could be obtained by maximizing the following log-likelihood function based on (3.16):

$$\max_{\xi,\beta}(-N \log \beta - (1 + 1/\xi) \sum_{i=1}^{N} \log(1 + \xi/\beta(X_{k_i} - u))), \qquad (3.18)$$

where $N = N(u)$ denotes the number of observations exceeding the threshold u and $X_{k_1}, X_{k_2} \cdots, X_{k_N}$ are the upper order statistics exceeding this threshold.

Meanwhile, from the conditional probability principle

$$\Pr(r \geq x + u) = \Pr(r > u) * \Pr(r \geq x + u \mid r > u)$$

and using the (3.14) and (3.15), we can have:

$$\begin{aligned} 1 - F(u + x) &= [1 - F(u)][1 - F_u(x)] \\ &\approx [1 - F(u)][1 - G_{\xi,\beta}(x)]. \end{aligned} \qquad (3.19)$$

Here, an estimator for the p- quantile of $F(u + x)$ for $x > 0$ can be easily derived. The term $1 - F(u)$ can be estimated using the empirical distribution

function: $[1 - F(u)] = N/n$. Substituting the maximum likelihood estimator of ξ and β in the GPD, we get:

$$[1 - F(u + x)] = (1 + \xi x/\beta)^{-1/\xi} N/n. \tag{3.20}$$

Inverting the above formula and making $F(u + x) = p$, we get:

$$x = [((1 - p)\frac{n}{N})^{-\xi} - 1]\frac{\beta}{\xi}. \tag{3.21}$$

And the real p-quantile estimator is

$$x_p = u + [((1 - p)\frac{n}{N})^{-\xi} - 1]\frac{\beta}{\xi}. \tag{3.22}$$

Note that this estimator is valid only for very high p, as the approximation is valid only asymptotically.

To summarize, EVT seems to be a very general approach to tail estimation. The main strength is that the use of a GEV distribution does not seem to be a very restrictive assumption, as it covers most of the commonly used distributions. On the other hand, there are several problems that need to be considered.

First, the assumption of i.i.d observations seems to be at odds with the characteristic of financial data. Although generalizations to dependent observations have been proposed, they either estimate the marginal unconditional distribution or impose conditions that rule out the volatility clustering behaviors typical of financial data.

Second, EVT works only for very extreme probability levels. How extreme these probability must be is hard to tell on a priori ground. A Monte Carlo

study might help to shed some light on how fast the performance of the EVT estimators deteriorates as we more away from tail.

Closely related to this issue is the selection of the cut-off point that determines the number of order statistics to be used in the estimation procedure. The choice of the threshold u presents the same problems encountered in the choice of the number of k extreme order statistics. If the threshold is too high, there are too few exceptions and the result is a high variance estimator. On the other hand, a threshold too low produces a biased estimator, because the asymptotic approximation might become very poor. Unfortunately, there is no statistical method to choose u and we have to rely on simulations and graphical devices.

Threshold choice for EVT by simulation

As we have already pointed out, a critical aspect of the implementation of EVT is the determination of the threshold beyond which the observations are assumed to follow a Generalized Pareto Distribution. To address this issue we need to refer to a Monte Carlo simulation[18].

We generated 1000 samples of 1000 observations each, using three different distribution, the standard normal and Student-t(5) and Exponential(1). For each sample, we estimated the quantile p= 0.005, 0.01 and 0.05 using the formula of the extreme value theory provided above and different threshold value. We chose the threshold values indirectly, by choosing the number of observations to be included in the maximum likelihood estimation. We started with k=50 and we increase it by 10 until it reached 500. To compare the different estimates, we computed the bias and the mean squared error of the estimators as follows:

$$bias_p = \sum_{j=1}^{1000} \hat{q}_p/1000 - q_p, \tag{3.23}$$

$$MSE_p = \sum_{j=1}^{1000} (\hat{q}_p - q_p)^2/1000. \tag{3.24}$$

Our goal is to determine how sensitive these estimatsor are to the choice of the parameter k, to the underlying distribution and to p-level of the estimated quantile. The results are reported in Figure (3.1,3.2 and 3.3). We plot the bias the MSE of the EVT estimates against k, together with historical simulation quantile estimates, which do not depend on k.

From these plots, we can notice several points:

First, the $p = 0.005$ EVT estimators appear to be very accurate compared with historical simulation, for all three different distributions, both from bias and MSE. This verifies that EVT works very well for extreme tails. But its performance deteriorates with k increase, especially for the normal distribution. We prefer to keep k below 150.

Second, for $p = 0.01, 0.05$, the superiority of using EVT method is not so evident as for $p = 0.005$, which is not surprising given that EVT works well only for extreme tails. For Exp(1) distribution, the bias and MSE of EVT are not stable but always smaller than the ones of historical simulation. For normal distribution, the bias and MSE show clear trend that they grow rapidly with k increase. They become greater than the ones of historical simulation when k is larger than 130. And the t(5) distribution exhibits very similar behaviors as normal distribution. Thus, the k ranging from 70 to 130 seems to be appropriate for EVT estimation. In our following empirical study, we

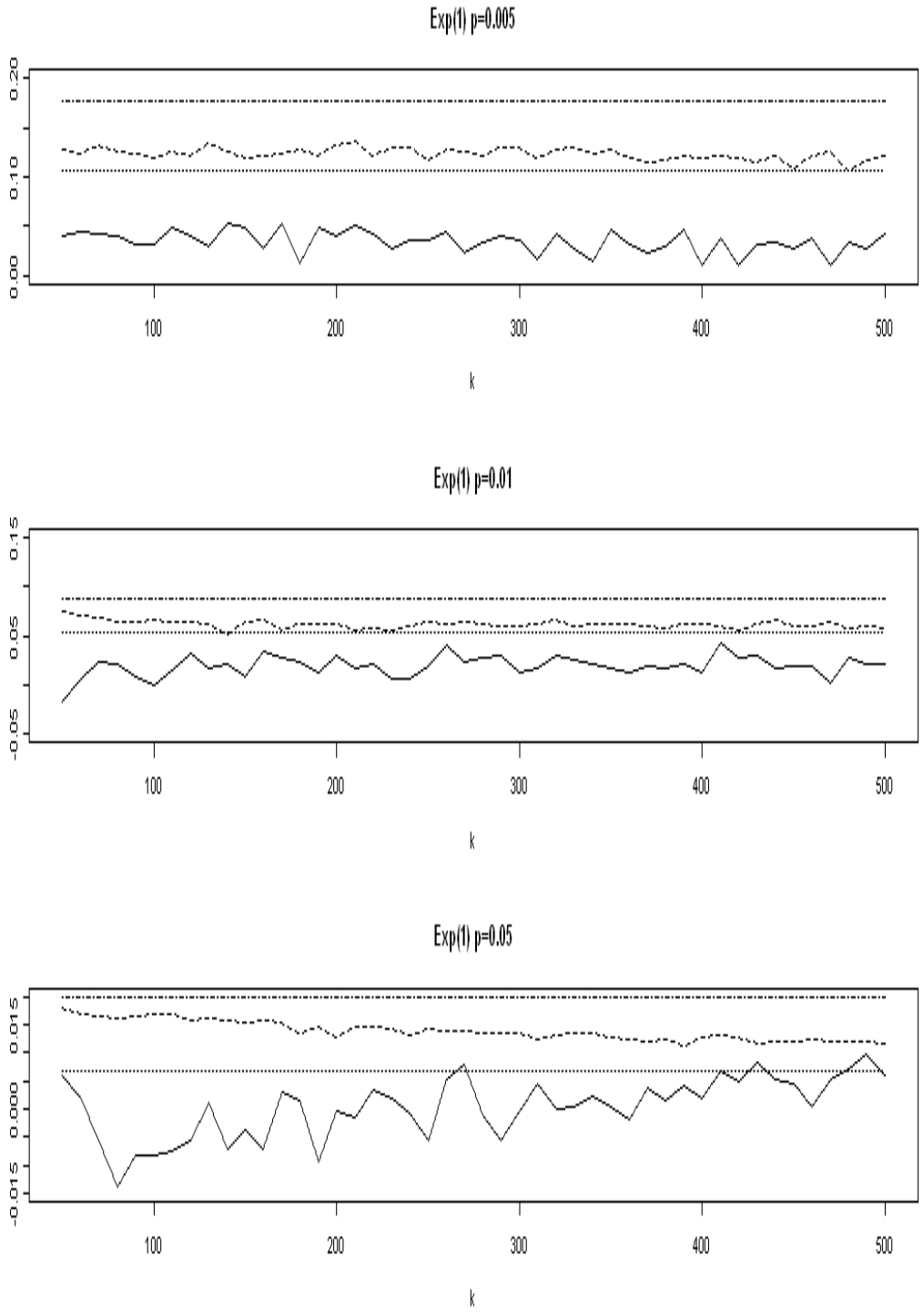

Figure 3.1: Simulation result for k selection. We estimate the exponential(1) 0.5%,1% and 5% quantile. Solid line is the bias of EVT method; dashed line is MSE of EVT method; dotted line is bias of Historical simulation and dashdotted line is MSE of Historical simulation.

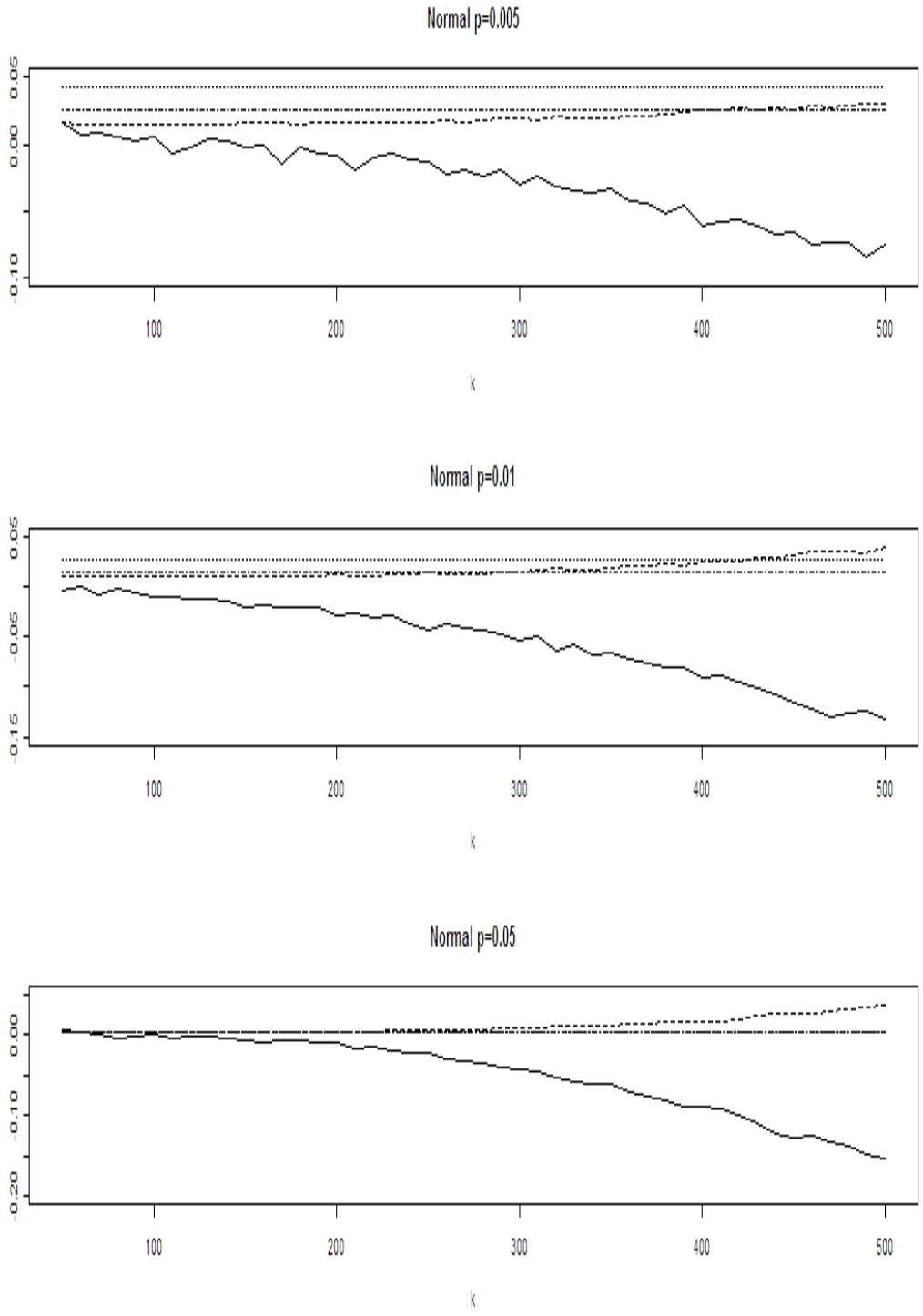

Figure 3.2: Simulation result for k selection. We estimate the normal 0.5%,1% and 5% quantile. Solid line is the bias of EVT method; dashed line is MSE of EVT method; dotted line is bias of Historical simulation and dashdotted line is MSE of Historical simulation.

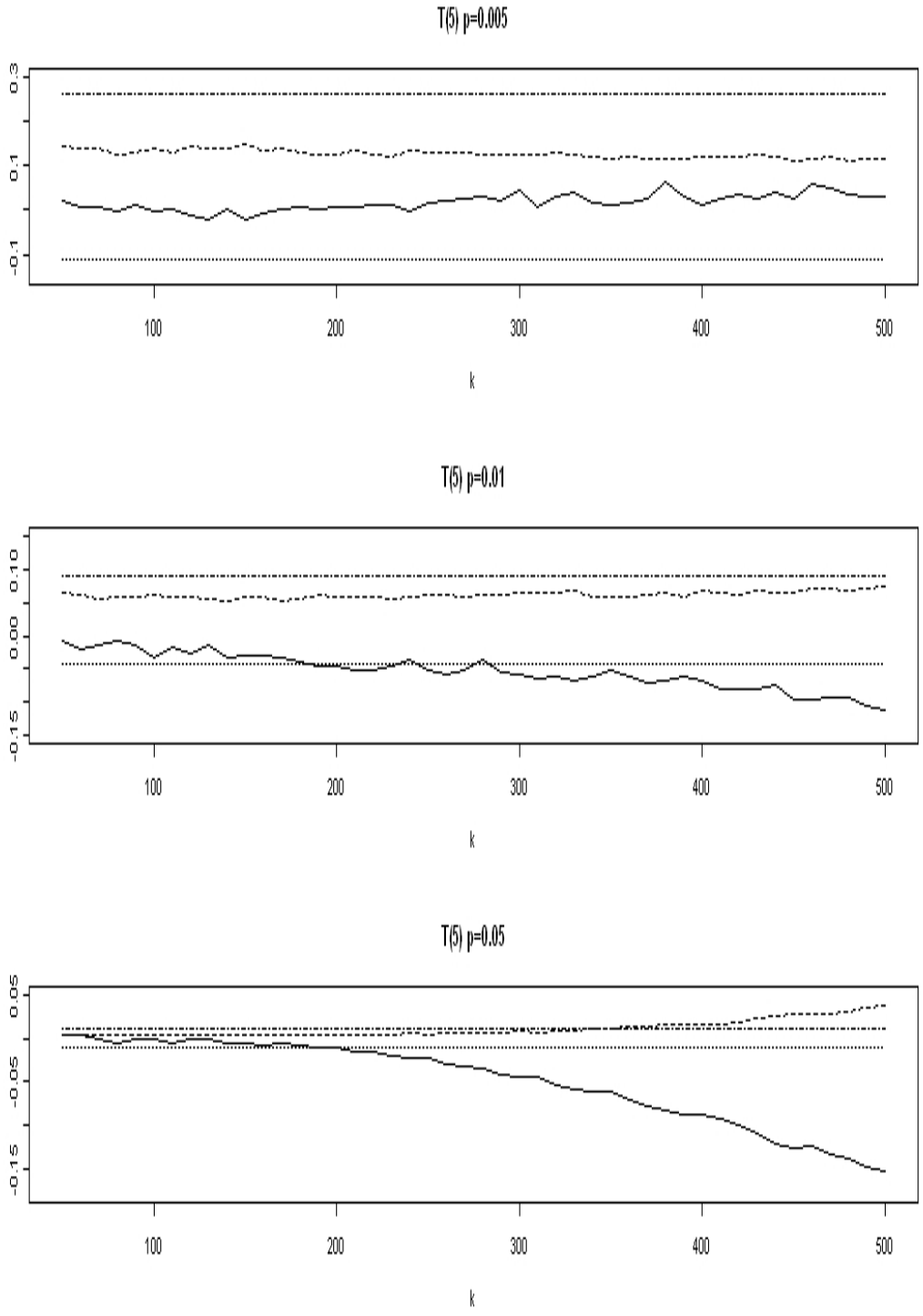

Figure 3.3: Simulation result for k selection. We estimate the t(5) 0.5%,1% and 5% quantile. Solid line is the bias of EVT method; dashed line is MSE of EVT method; dotted line is bias of Historical simulation and dashdotted line is MSE of Historical simulation.

choose k equal to 100.

3.3.2 Conditional Autoregressive Value at Risk

The definition of VaR naturally leads us to the concept of quantile regression. To estimate conditional quantiles, the time series of the specified quantile is explicitly modeled using any information deemed to be relevant. No distribution assumption for the time series behavior of returns is needed. The basic idea is to model the conditional p-quantile as some function of the information available. A good choice of relevant information and of the functional form should yield a close approximation to the population quantile. Koenker and Basset(1978)[14] generalize the common linear regression framework by shifting the focus from the conditional mean to conditional quantiles. As shown below, for example, in Koenker and Portnoy(1997), the unconditional sample p-quantile can be found as the solution to

$$\min_{\beta \in R}(\sum_{r_t \geq \beta} p \mid r_t - \beta \mid + \sum_{r_i < \beta}(1-p) \mid r_t - \beta \mid). \tag{3.25}$$

Extending this to the classical linear regression framework, Koenker and Bassett(1978)[14] define the p-regression quantile estimator by

$$\arg\min_{\beta \in R^k}(\sum_{r_t \geq x_t^{'}\beta} p \mid r_t - x_t^{'}\beta \mid + \sum_{r_t < x_t^{'}\beta}(1-p) \mid r_t - x_t^{'}\beta \mid), \tag{3.26}$$

where the x_t are non-random vectors containing information available to time t. The key assumption in the linear quantile regression model is that

$$r_t = x_t^{'}\beta_p + z_{t,p}. \tag{3.27}$$

Note that the distribution of the error term is left unspecified. The only assumption is that the conditional quantile function is given by $Q_p(r_t \mid x_t) = x_t'\beta$.

Because our focus is exclusively on one-step forecasting performance, we more closely examine the conditional VaR approaches formulated in Engle and Manganelli(2002)[11]. They link VaR to the conditional standard deviation of the returns such that an increase in the latter leads a more dispersed return distribution and thus, to a higher VaR. Their conditional autoregressive VaR specifications includes VaR_{t-1} as an explanatory variable in r_t, to adapt to serial dependence in volatility and mean. A function of r_{t-1} is also included to link the conditional quantile to return innovations.

As mentioned above, no explicit distributional assumptions need to be made, guarding against this source of model error. Although many specifications for regression model (3.27) are conceivable, we first adopt those put forth in Engle and Manganelli(2002). The baseline CAViaR models is give by

$$VaR_t = VaR_{t-1} + \beta[I(r_{t-1} \leq VaR_{t-1}) - p]. \tag{3.28}$$

As typically $p \leq 0.05$ for risk management purpose, we have an asymmetric response: VaR_t will jump upward when a violation occurs and will slowly decrease. In the baseline model, the adaptive process learns nothing from the actual size of returns, except that whether or not the returns exceed VaR. It is a very preliminary form of CAViaR model, and it cannot satisfy our requirements for the model. Thus, we make some improvement on it, as is the

case with the symmetric absolute value CAViaR specification

$$VaR_t = \beta_0 + \beta_1 VaR_{t-1} + \beta_2 \mid r_{t-1} \mid . \tag{3.29}$$

It allows the autoregressive parameter to be different from one, and introduces a direct response of the quantile to the return process, treating the effect of extreme returns on VaR -and implicitly, on volatility - symmetrically. In order to reflect the asymmetry in financial returns, we can use asymmetric slope CAViaR specification,

$$VaR_t = \beta_0 + \beta_1 VaR_{t-1} + \beta_2 max[r_{t-1}, 0] + \beta_3 min[r_{r-1}, 0], \tag{3.30}$$

which allows the VaR prediction to respond asymmetrically to positive and negative returns and so can accommodate the leverage effect.

The specifications for symmetric absolute value and asymmetric slope CAViaR are two very typical ones, and we would test their performance in real data in empirical study later. The specifications for CAViaR can be very flexible, as long as they can reflect the real situations.

3.4 Other Estimation Methods

In our discussion of GARCH models, we emphasized how the assumption of normality distributed standardized residuals seemed to be at odds with the fact that financial data tend to exhibit non-normality characteristic. It turns out, however, that the normality assumption might not be as restrictive as we think. This is due to a very important result Bollerslev and Woolridge(1992)[5] who showed that the maximization of the normal GARCH likelihood is able to deliver consistent estimates, provided that the variance equation is correctly specified, even if the standardized residuals are not normally distributed. We refer to this result as the quasi-maximum likelihood GARCH. Many papers have exploited this property. Engle and Magnanelli(2002)[11], suggests computing the VaR of a portfolio by first fitting GARCH and then multiplying the empirical quantile of the standardized residuals by the square root of the estimated variance. This estimation method is a mix of GARCH fitted to portfolio returns and historical simulation applied to the standardized residuals. As a consequence it retains some of the drawbacks of these approaches. First, the assumption of that the standardized residual are i.i.d is still required. Given this assumption, however, flaw of historical simulation still exists, as they use a window of the whole series of standardized residuals. On the other hand, the problem of discreteness of extreme returns remains. Historical simulation will provide very poor and volatile quantile estimation in the tails. A natural alternative seems to use EVT instead of historical simulation. GARCH-EVT is estimated by McNeil and Frey(2000)[19].

Given the very general results of EVT, the GARCH augmented by the EVT requires very weak assumptions. The only required assumption is that

the variance is correctly specified and that the standardized residuals are i.i.d and in the maximum domain of attraction of some extreme value distribution.

Chapter 4

Testing the Fit of Value-at-risk Models

To assess the predictive performance of the methods under consideration, we follow the framework in paper [6] and [16], which is evaluating the accuracy of out-of-sample interval forecasts. By defining $H_t = I(r_t < VaR_t)$, Christoffersen(1998) terms the sequence of VaR forecasts efficient with respect to F_{t-1} if

$$E[H_t|F_{t-1}] = p,$$

which, by applying iterated expectations, implies that H_t is uncorrelated with any function of a variable in the information set available at t-1. If the above condition holds, then VaR violations will occur with the correct conditional and unconditional probability, and neither the forecast for VaR_t nor that for H_t can be improved.

Although a general test of this condition is desirable, we follow Christof-

fersen(1996) in using intermediate statistics for testing specific implications of the general hypothesis, so that particular inadequacies of a model can be revealed. The following specifies an LR test of correct unconditional coverage, an LR test of independence, and an LR test, which combines the two to form a complete test of the conditional coverage. By specifying F_{t-1} to include at least $H_1, H_2, \cdots, H_{t-1}$, it is straightforward to show (Christoffersen,1998) that efficiency implies

$$H_t|F_{t-1} \sim Ber(p), t = 1, 2, \cdots, T, \tag{4.1}$$

where $Ber(\cdot)$ denotes the Bernoulli distribution.

4.1 Test of Unconditional Coverage

From (4.1), we should first test the unconditional coverage through the following hypothesis

$$H_0 : E[H_t] = p \quad vs \quad H_A : E[H_t] \neq p. \tag{4.2}$$

This hypothesis could be done through likelihood ratio test. The likelihood under the null hypothesis is simply

$$L(p; H_1, H_2, \cdots, H_T) = (1-p)^{n_0} p^{n_1}, \tag{4.3}$$

and under the alternative

$$L(\pi; H_1, H_2, \cdots, H_T) = (1-\pi)^{n_0} \pi^{n_1}, \tag{4.4}$$

where n_0 is the number of violations that do not occur and n_1 is the number of violations that occur. Also, we have $n_0 + n_1 = T$ and $\pi = n_1/(n_0 + n_1)$ is the violation ratio.

Testing for the unconditional coverage can be formulated as a standard likelihood ratio test,

$$LR_{uc} = 2[L(\pi; H_1, H_2, \cdots, H_T] - L(p; H_1, H_2, \cdots, H_T] \sim^{asy} \chi^2_1. \qquad (4.5)$$

4.2 Test of Independence

The above test is just the first step because it does not take into account the possibility that a model that passes the unconditional coverage test(violation ratio is close to p), may have serious violation clustering problem. In the test above, the order of 0 and 1 in H_t does not matter, only the total number of 1 plays a role. Therefore, a test for violation independence should be employed.

Several tests for independence have been proposed in the literature, including the runs tests. Most recently, a test based on the time between violation was proposed in Danielsson and Morimoto(2000)[8]. Under the null, a violation today should have no influence on the probability of a violation tomorrow. Christoffersen(1998) models H_t as a binary first-order Markov chain with transition probability matrix

$$\Pi = \begin{pmatrix} 1 - \pi_{01} & \pi_{01} \\ 1 - \pi_{11} & \pi_{11} \end{pmatrix}, \pi_{ij} = P(H_t = j | H_{t-1} = i), \qquad (4.6)$$

as the alternative hypothesis of independence. The approximate joint likeli-

hood, conditional on the first observation, is

$$L(\Pi; H_2, H_3, \cdots, H_T|H_1) = (1-\pi_{01})^{n_{00}}\pi_{01}^{n_{01}}(1-\pi_{11})^{n_{10}}\pi_{11}^{n_{11}}, \tag{4.7}$$

where n_{ij} represents the number of transitions from stat i to stat j,

$$n_{ij} = \sum_{t=2}^{T} I(H_t = j|H_{t-1} = i),$$

and the maximum likelihood estimators under the alternative hypothesis are

$$\pi_{01} = n_{01}/(n_{00}+n_{01}) \;\; and \;\; \pi_{11} = n_{11}/(n_{10}+n_{11}).$$

Under the null hypothesis of independence, we have $\pi_{01} = \pi_{11} \equiv \pi_1$; and $\pi_1 = (n_{01}+n_{11})/(T-1)$, which is similar to π in the unconditional coverage test. The likelihood under the null hypothesis is

$$L(\pi_1; H_1, H_2, \cdots, H_T) = (1-\pi_1)^{(n_{00}+n_{10})}\pi_1^{(n_{01}+n_{11})}, \tag{4.8}$$

and the likelihood ratio test is given by

$$LR_{ind} = 2[L(\Pi; H_2, \cdots, H_T|H_1] - L(\pi_1; H_2, \cdots, H_T|H_1] \sim^{asy} \chi_1^2. \tag{4.9}$$

4.3 Test of Conditional Coverage

Because the tests for unconditional coverage and independence each examines one aspect of the problem, these two tests must be combined to form a complete test for conditional coverage. In fact, the null hypothesis of the

unconditional coverage test will be tested against the alternative hypothesis of the independence test. Based on Christoffersen(1998), we can utilize the likelihood ratio

$$LR_{cc} = 2[L(\Pi; H_2, \cdots, H_T|H_1] - L(p; H_2, \cdots, H_T|H_1] \sim^{asy} \chi^2_2 \qquad (4.10)$$

to test conditional coverage. Notice that if we condition on the first observation in the test for unconditional coverage, we can have $\pi = \pi_1$. It leads us, if we do not use the first observation in unconditional coverage test, to the following relationship among three likelihood ratio

$$LR_{cc} = LR_{uc} + LR_{ind}, \qquad (4.11)$$

which provides a means to check whether the violation series H_t fails the correct conditional property.

Chapter 5

Expected shortfall

Value-at-Risk has been criticized for the adequacy of being a risk measure for it lacks the coherent property defined by Artzner (1999)[2]. Because benefit and loss is an equivalent way to say financial return(positive return implies benefit and negative return implies loss), we use benefit and loss to introduce the definition of coherence for ease of explanation. For a risk measure ρ, we call it coherent, if for any random loss x and y, it can satisfy the following four axioms:

(1) Subadditivity - For all random loss x and y,

$\rho(x + y) \leq \rho(x) + \rho(y)$.

(2) Monotonicity - If x $\leq$ y for each scenario, then

$\rho(x) \leq \rho(y)$.

(3) Positive homogeneity - For all $\mu \geq 0$ and random loss x,

$\rho(\mu x) = \mu\rho(x)$.

(4) Translation Invariance - For any random loss x and constant α,

$\rho(x + \alpha) = \rho(x) + \alpha$.

The concept of coherence is so important to us because these four axioms represent the most basic requirements for the measure of risk.

Subadditivity reflects the diversification in portfolio management, which expresses the fact that the portfolio made up of sub-portfolios will suffer a risk no more than the sum of the risks of sub-portfolios. This is the most significant feature for a risk measure. Because diversification is the most common concept in financial and investment field, any measure of risk that does not meet this axiom will encounter some problems in the practice.

For example, in insurance field, we all know one saying" Do not put all the eggs in one basket". This is just talking about the diversification in the portfolio management. Therefore, any risk measure without subadditivity property can not capture the essence of how the portfolio will behave with the addition of other portfolios. Unfortunately, the most commonly used risk measure, Value at risk is not in line with this term.

Monotonicity is very easy to understand, which means that the greater measure of risk will correspond to the greater risks. This is the essential requirement of the risk measure.

Positive Homogeneity is a limiting case of subadditivity, showing what will happen when there is no diversification effect between the portfolios.

Translation Invariance is not so straightforward, but it can be explained in this way: when the capital is withdrawn from one portfolio, the risk this new portfolio will undertake is equal to the previous risk of this portfolio plus the capital already withdrawn. This is saying another common sense in banking and insurance industry, that the less reserve or margin there is at hand, the more risk the bank and insurance company will confront in the future.

Expected shortfall is the expected value of portfolio losses exceeding VaR. It has been proven by Carlo Acerbi(2002)[1] as a coherent measure and is usually proposed as a supplement to VaR. Since most methods to estimate ES are based on the estimation of VaR, we do not specially introduce them here. The procedures to ES estimation will be provided later in empirical study.

Chapter 6

Empirical study

6.1 Introduction to empirical study

In this chapter, we apply the methods discussed in preceding chapters to empirical study and evaluate their performance. We divide this chapter into two parts.

The first section is about estimations of VaR. We employed the methods discussed in Chapter 3 to estimate the VaR of two data sets: BMW stock price returns and NYSE(New York Stock Exchange) index returns. Then, we use the three criteria introduced in Chapter 4 to evaluate their performance. All these computations and programming are completed by myself in R(Details are provided in next section). I also did the computations of some methods(GARCH-normal, GARCH-HS) in Matlab, which give very similar results to ones from R. This verifies the validity of estimation from R.

The second section is related to estimation of ES. Because the ES is usually estimated after VaR, the methods we use to estimate VaR are still compared in this section. Since no very efficient ways to evaluate estimation of ES for

real data exist, we generate data in R from three different GARCH processes for comparison. Unlike VaR estimation, we use bias and MSE to determine the performance of different methods.

6.2 Empirical study for VaR

6.2.1 Procedures for different methods

In our empirical study, we use one-day-ahead VaR prediction to study the performance of these methods, because of its conceptual simplicity and wide application. We divide the whole data set(N observations) into two parts. The first part is called "window"(K observations), which is used to estimate the parameters in each model. The second part is used to evaluate the estimating performance of different methods. We move this window one day ahead each time, keeping the window length constant at K. At very advancement of the window, the first observation in the window will go out and the first new observation out of the window will come in. For each moving window, we recalculate model parameters and make the VaR prediction. This leaves us with $N-K$ one-step-ahead VaR forecasts to study the predictive performance of the models.

For example, if there are 6146 observations(N=6146), we first take observations from 1 to 1000 as the window, using them to estimate the parameters for each method. We use the estimated model to calculate the VaR and compare it with the real return(observation 1001). If this real returns is smaller than our predicted VaR, we would call this a "violation". And we would count the number of violation during the whole process. Then we move the window one

step ahead, which would be composed of observations from 2 to 1001. We use this new window to make the new prediction and compare it with observation 1002. This process would be repeated for 5146 times until the VaR prediction is made for last observation(6146).

In addition to regular GARCH approaches, which rely on the normal assumption for residual terms, we consider alternative conditional distributions which use the student's t(5) and modified Exponential(1) distribution instead of the normal distribution in order to better account for conditional asymmetry and heavy-tail. We also utilize the GARCH model combined with EVT and Historical Simulation.

We estimate the VaR at $p = 0.01, 0.025, 0.05, 0.1$ and 0.25 using the following methods:

(1) Historical simulation,

(2) Extreme value theory,

(3) GARCH-Normal,

(4) GARCH-Student's t(5),

(5) GARCH-Exponential(1),

(6) GARCH-Extreme value theory,

(7) GARCH-Historical simulation,

(8) CAViaR-Symmetric absolute value, and

(9) CAViaR-Asymmetric slope.

All the estimating and evaluating procedures are performed in R. Now, we provide the operation details of these methods in the order shown above.

(1)Historical simulation

From the discussion in chapter 3, we know that estimation of p-VaR by historical simulation is actually the unconditional p-quantile of the return distribution. The "window" is used to obtain this unconditional quantile. First, we sort all the observations($K = 1000$) in the window from smallest to greatest and denote them by $r_{[1]}, r_{[2]} \cdots r_{[1000]}$. Then, we would use following formula to compute p-quantile:

$$VaR(p) = |M - pK| r_{[M+1]} + |M - pK + 1| r_{[M]}, \tag{6.1}$$

where M is the floor integer of pK.

(2) Extreme value theory

We apply EVT to VaR estimation with help of package "evir" in R. Since the program in "evir" is using maximum order to estimate (as we discussed in chapter 3) while we need the VaR estimation on the minimum side, we should first take negative sign on all the observations $r_i,\ i = 1, 2 \cdots 1000$ in the window, and then apply the command "gpd" in R. After we get the result from R, we can get the real VaR by making another sign change.

(3) GARCH-Normal

We perform this estimation also in R , using the package of "tseries". In this model, we use ARMA(1,1) to compute the conditional mean for the financial return and GARCH(1,1) to capture the changing volatility. We use command "arma" and "garch" in package "tseries" to conduct the computation. The estimation of parameters in 'arma' is finished by conditional least square method

while the parameters in 'garch' are estimated by maximum likelihood method, as we discussed in chapter 3. Every time, we use observations in the "window" to calculate the parameters and then make the prediction. I also did the estimations for parameters in ARMA and GARCH in Matlab, because my supervisor, Pro. Mizera ever doubted about the computation validity from R. The results for these two software turned out to be very similar, probably because R has updated its package and made some progress. After we have fitted the ARMA(1,1)

$$u_t = a_0 + a_1 r_{t-1} + b_1 \epsilon_{t-j}, \tag{6.2}$$

and GARCH(1,1) model

$$\sigma_t^2 = c_0 + c_1 \epsilon_{t-1}^2 + d_1 \sigma_{t-1}^2, \tag{6.3}$$

the one-day-ahead VaR can be expressed as

$$VaR_t = u_t + \sigma_t Q_p(z), \tag{6.4}$$

where Q_p indicates the p-quantile and z_t follows the standard normal distribution.

(4) GARCH-t(5)

The procedure to use GARCH-t(5) is basically the same to GARCH-Normal, except we assume that z_t follows t(5) distribution instead of standard normal

distribution. And the form of VaR estimation has some little modification as

$$VaR_t = u_t + \sqrt{\frac{3}{5}}\sigma_t Q_p(z), \tag{6.5}$$

where Q_p indicates the p-quantile and z_t follows t(5) distribution. Here, $\sqrt{\frac{3}{5}}$ is used to normalized z_t because t(v) distribution has variance of $(\frac{v}{v-2})$.

(5) GARCH-Exponential(1)

Using GARCH-Exponential(1) follows the same as GARCH-Normal except that we replace standard normal distribution with Exponential(1). The reason for doing this is that we want to use Exponential(1) to account for heavy-tail property of financial returns. The one-day-ahead VaR estimation is in the form of

$$VaR_t = u_t + \sigma_t(1 + Q_p(-z)), \tag{6.6}$$

where Q_p indicates the p-quantile and z_t follows Exponential(1) distribution. We take negative of z in Q_p to have infinite left end. We add 1 to $Q_p(-z)$ because we want the error term to have mean 0 and variance 1.

(6) GARCH-Extreme value theory

This method combines GARCH model and extreme value theory together and we still implement it in R with the package "tseris" and "evir". We do not assume any distribution for error terms z_t. Because extreme value theory can be applied to any $i.i.d$ observations, we use it for normalized residuals

$\frac{r_1-u_1}{\sigma_1}, \frac{r_2-u_2}{\sigma_2} \cdots, \frac{r_K-u_K}{\sigma_K}$. Then the one-day-ahead VaR estimation is

$$VaR_t = u_t + \sigma_t Q_p(z), \tag{6.7}$$

where Q_p indicates the p-quantile we get from normalized residuals by using extreme value theory.

(7) GARCH-Historical simulation

Instead of applying extreme value theory, we get the p-quantile of normalized residuals through historical simulation. We still sort the normalized residuals $\frac{r_1-u_1}{\sigma_1}, \frac{r_2-u_2}{\sigma_2} \cdots, \frac{r_K-u_K}{\sigma_K}$ from smallest to greatest and then use formula(6.1) to calculate the p-quantile. Then the one-day-ahead VaR estimation is

$$VaR_t = u_t + \sigma_t Q_p(z), \tag{6.8}$$

where Q_p indicates the p-quantile we get from normalized residuals by using historical simulation.

(8) CAViaR-Symmetric absolute value

In order to conduct the quantile regression estimation, we would refer to package "quantreg" in R. And the relationship we want to find is

$$VaR_t = \beta_0 + \beta_1 VaR_{t-1} + \beta_2 \mid r_{t-1} \mid, \tag{6.9}$$

where $\beta_0, \beta_1, \beta_2$ are the parameters we need to estimate.

These parameters could be found by minimizing the following equation

$$\arg \min_{\beta_0,\beta_1,\beta_2 \in R} (\sum_{r_t \geq VaR_t} p \mid r_t - VaR_t \mid + \sum_{r_t < VaR_t} (1-p) \mid r_t - VaR_t \mid), \quad (6.10)$$

which could be done by 'rq" command in package "quantreg" in R.

In order to fit the quantile regression model, we should first have the data for independence factors and response factor. In practice, we can fit the regression relationship $r_t \sim |r_{t-1}| + VaR_{t-1}$ by using the following "window" of data:

r_t	r_{t-1}	VaR_{t-1}
r_{L+1}	$\|r_L\|$	$VaR_L = p - quantile\ of\ data[r_1, r_L]$
r_{L+2}	$\|r_{L+1}\|$	$VaR_{L+1} = p - quantile\ of\ data[r_2, r_{L+1}]$
$\vdots$	$\vdots$	$\vdots$
r_{L+K}	$\|r_{L+K-1}\|$	$VaR_{L+K-1} = p - quantile\ of\ data[r_K, r_{L+K-1}]$

Table 6.1: Window of data for regression of CAViaR-Symmetric absolute value method

We use this K groups of data to estimate the parameters $\beta_0, \beta_1, \beta_2$. Then we calculate VaR_{L+K} and compare it with r_{L+K}. If $r_{L+K} < VaR_{L+K}$, we call it "violation".

Then we use the new estimated VaR_{L+K} to update the data as below, keeping the "window" length as K.

r_t	r_{t-1}	VaR_{t-1}
r_{L+2}	$\|r_{L+1}\|$	$VaR_{L+1} = p - quantile\ of\ data[r_2, r_{L+1}]$
r_{L+3}	$\|r_{L+2}\|$	$VaR_{L+2} = p - quantile\ of\ data[r_3, r_{L+2}]$
$\vdots$	$\vdots$	$\vdots$
r_{L+K+1}	$\|r_{L+K}\|$	$VaR_{L+K} just\ estimated\ from\ CAViaR\ model$

Table 6.2: Updated window of data for regression of CAViaR-Symmetric absolute value method

We keep updating the "window" until we estimate the VaR for last return

r_N, which gives us $N - K - L$ forecasts for VaR. These forecasts are used to evaluate the performance of CAViaR-Symmetric absolute value method.

(9) CAViaR-Asymmetric slope

The way to employ asymmetric slope method is quite similar to symmetric absolute value method, except that the specification of regression relationship is

$$VaR_t = \beta_0 + \beta_1 VaR_{t-1} + \beta_2 max[r_{t-1}, 0] + \beta_3 min[r_{r-1}, 0]. \tag{6.11}$$

We follow the same procedure to build the "window" of data, estimate the parameters and make the forecasts.

6.2.2 Empirical results for BMW stock returns

We first examine the performance of various methods for estimating VaR by using the stock price return of BMW. The data comprise daily closing price, p_t, of the BMW stock from January 1 ,1973 to July 22, 1996, yielding a total of N=6146 observations of percentage log-returns, $r_t = 100(\log p_t - \log p_{t-1})$. Table 6.3 presents the relevant summary statistics. The sample skewness indicates considerable asymmetry, which is taken together with the sample kurtosis, indicating a substantial violation of normality.

Sample Size	Mean	Std.Dev.	Skewness	Kurtosis	Min	Max
6146	0.034	1.476	-0.0458	10.161	-14.061	11.719

Table 6.3: Summary Statistics for BMW equity returns

These statistical characteristics for BMW stock returns can be more clearly seen from Figures (6.1, 6.2, 6.3).

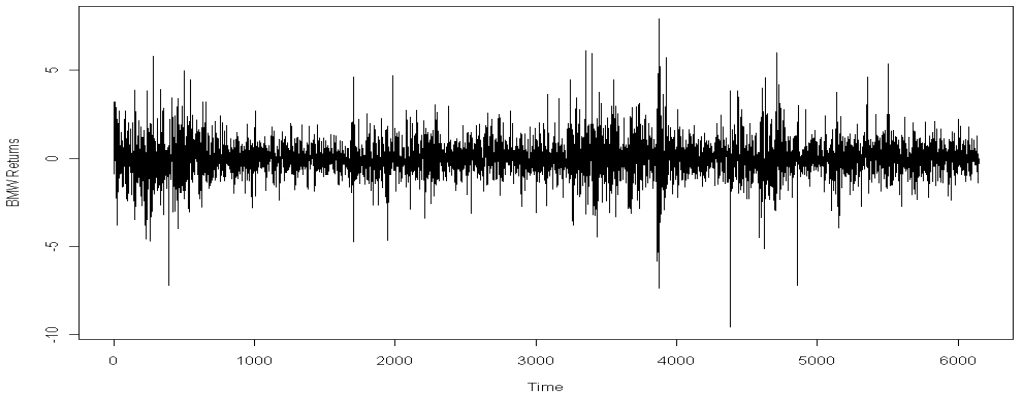

Figure 6.1: Plot for daily returns of BMW equity, from January 1, 1973 to July 22, 1996.

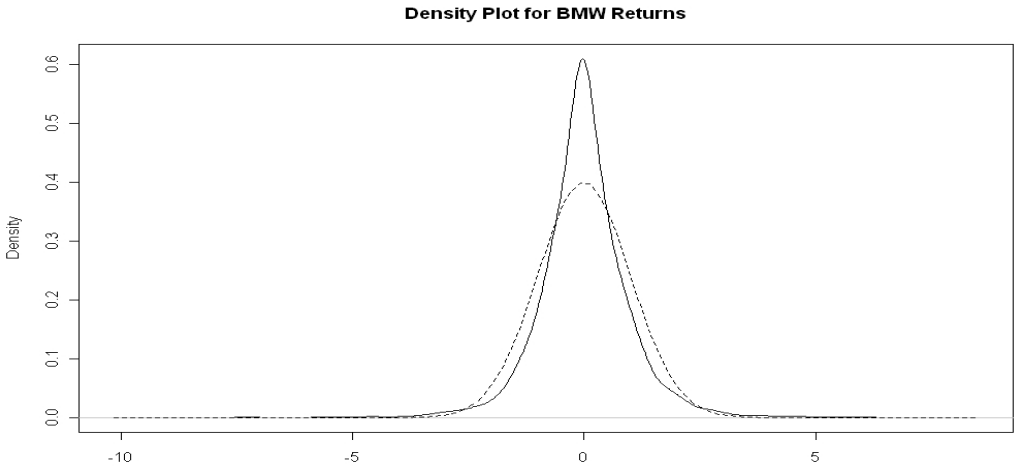

Figure 6.2: Solid line is the density plot for standardized daily returns for BMW equity, from January 1, 1973 to July 22, 1996. It has mean=0 and standard deviation=1. Dashed line is the density plot for standard normal distribution. Non-normality, especially higher center and lower sides, can be clearly seen in this plot.

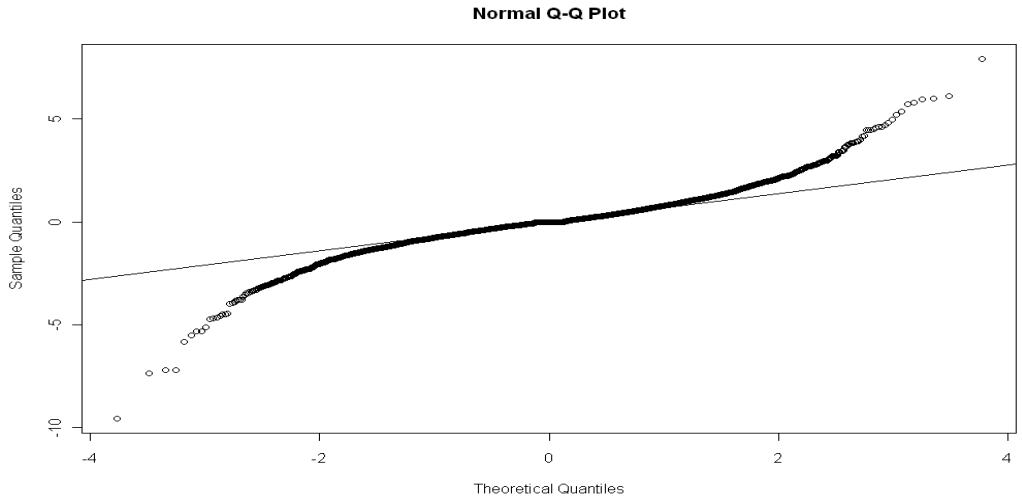

Figure 6.3: QQ-plot for standardized daily returns for BMW equity, from January 1 ,1973 to July 22, 1996. It has mean=0 and standard deviation=1. Solid line is for standard normal distribution. Non-normality, especially higher tails, can be clearly seen in this plot.

The performance of these methods are evaluated by the criteria we discussed in chapter 4. In the following tables, p denotes the theoretical probability with which the observations should fall below our VaR estimation; Violation rate is the actual probability with which the return is smaller than our VaR estimation; LR_{uc} is the p-value for unconditional coverage likelihood ratio test; LR_{ind} is the p-value for independence likelihood ratio test; LR_{cc} is the p-value for conditional coverage likelihood ratio test. The best scenario for one method is that it can pass all these three tests, with all the p-value greater than 0.05.

Because historical simulation and extreme value theory methods are both unconditional methods, we put their results together in Table(6.4) and make some comparisons first. From this table, we can easily notice these points:

First, as the unconditional models do not account for volatility change,

none of them is able to produce $i.i.d$ VaR violations(none of them pass the independence test, with p-value greater than 0.05), causing us to strongly reject independence of the H_t sequences for all unconditional models.

Second, at almost p-levels, the historical simulation performs well with respect to violation rate. For EVT, the estimation accuracy is much better when p is small than when p is large, which can be seen from the p-value of unconditional coverage test. This is not surprising for the fact that EVT only works well in extreme tails.

Model	P	Vio Rate	LR_{uc}	LR_{ind}	LR_{cc}
Historical Simulation	0.01	0.0113	0.369	0.004	0.0109
	0.025	0.024	0.6113	0.00	0.00
	0.05	0.0487	0.6858	0.00	0.00
	0.1	0.1007	0.9851	0.00	0.00
	0.25	0.2503	0.9615	0.00	0.00
EVT	0.01	0.0118	0.9485	0.0145	0.0505
	0.025	0.0289	0.8123	0.00	0.00
	0.05	0.0585	0.7337	0.00	0.00
	0.1	0.1207	0.7776	0.00	0.00
	0.25	0.2755	0.00	0.00	0.00

Table 6.4: VaR Prediction Performance: Unconditional Models

Methods (3),(4) and (5) are typical fully-parametric approaches to estimating VaR, which assume the distributions for financial returns. The success of this method largely depend on whether the distribution we have assumed can accurately reflect the properties of real financial data. Comparing these three methods, we can notice these points:

First, although they use GARCH(1,1) model to reflect the changing volatility of the financial return, none of them pass the independence test and exhibit independent "violation". The possible reason is that these assumptions for return distribution do not hold for the real data.

Second, we can note that the violation rate from GARCH-Normal and GARCH-T(5) methods are always higher than the target probability, which is the result of the "light tail" of normal and t(5) distributions compared with real financial return. However, the violation rate from GARCH-Exponential(1) is less than the target probability(except p=0.25), which can be explained its "heavier tail".

Third, when p becomes larger than 0.1, the VaR estimation from GARCH-normal method comes to be very accurate, which can be verified by p-value of unconditional likelihood ratio test. This implies that GARCH-normal method is an option for moderate-tail VaR estimation.

Model	P	Vio Rate	LR_{uc}	LR_{ind}	LR_{cc}
GARCH-Normal	0.01	0.0205	0.00	0.00	0.00
	0.025	0.0400	0.00	0.00	0.00
	0.05	0.0602	0.001	0.00	0.00
	0.1	0.1063	0.538	0.00	0.00
	0.25	0.2507	0.9167	0.00	0.00
GARCH-t(5)	0.01	0.0132	0.04	0.00	0.00
	0.025	0.0362	0.00	0.00	0.00
	0.05	0.0721	0.007	0.00	0.00
	0.1	0.124	0.026	0.00	0.00
	0.25	0.2855	0.00	0.00	0.00
GARCH-Exponential(1)	0.01	0.0033	0.00	0.00	0.00
	0.025	0.0128	0.00	0.00	0.00
	0.05	0.0334	0.00	0.00	0.00
	0.1	0.0987	0.7586	0.00	0.00
	0.25	0.3461	0.00	0.00	0.00

Table 6.5: VaR Prediction Performance: CAViaR Models

Then, we compare the two methods combining GARCH(1,1) model respectively with EVT and historical simulation. With the test result shown in Table(6.6), we can note that the performances from GARCH-EVT and

GARCH-Historical simulation are very excellent: (1) Their estimations are very accurate. From the table, we can see that the "violation rate" is very close to "target probability" for GARCH-EVT with $p < 0.05$, and for GARCH-HS with almost all p levels. (2) These two methods pass all three tests with $p = 0.025, 0.05$ for GARCH-EVT and with $p \geq 0.025$ for GARCH-HS methods. Their performances are much superior to other methods, exhibiting independent violation, which means they successfully capture the changing volatility of financial returns. Comparatively, GARCH-EVT estimation is more credible for extreme p and GARCH-HS is more appropriate for moderate p, which is natural result given their theoretical differences.

To conclude the GARCH-based methods, introducing GARCH volatility dynamics gives us the possibility to improves VaR prediction performance, because GARCH model enables our VaR forecasts to adjust more quickly to the new situation. This can be clearly seen from Figure(6.4). However, we can also see, GARCH-Normal, GARCH-t(5) and GARCH-Exponential(1) methods do not bring us the convincing results. Their failure can be greatly attributed to the inappropriate assumptions for return distribution. Meanwhile, the success of GARCH-Hs and GARCH-EVT can be perceived as the consequence of utilizing GARCH model and the proper ways to deal with the dynamic factor z_t. If we want to improve the performance of regular GARCH models, we need to find a distribution which can better reflect the asymmetry and heavy-tail of the financial returns, such as Skew-t distribution.

Comparing the results in Table (6.7), we can draw these conclusions: (1) Their estimations are both accurate for $p \geq 0.05$; (2) CAViaR-asymmetric slope method almost pass all the violation independence test (except $p = 0.001$) while CAViaR just pass for $p = 0.05$, which is resulted from the fact

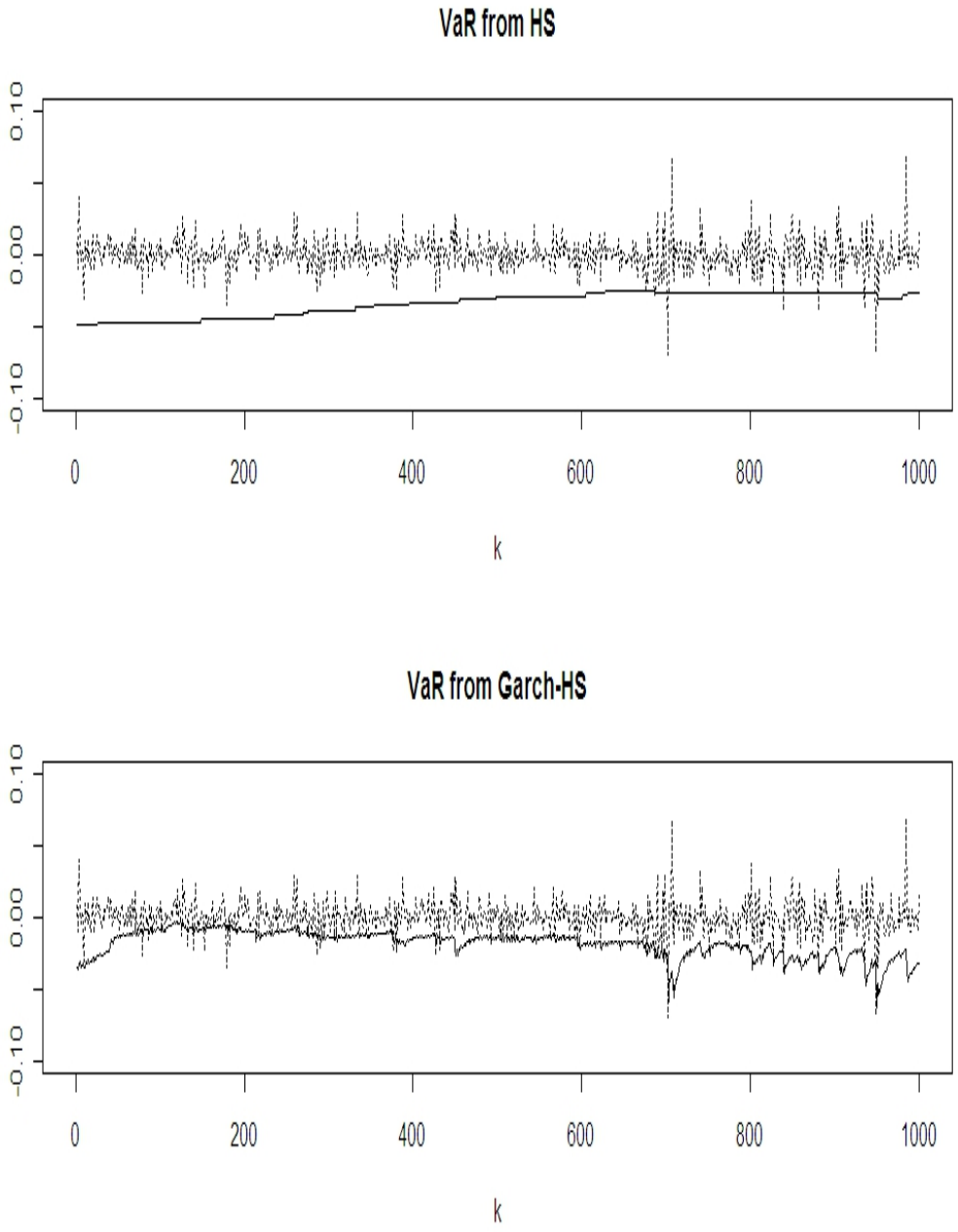

Figure 6.4: Plot 1 is BMW stock return and its estimation of VaR for $p = 0.05$ from Historical Simulation; Plot 2 is BMW stock return and its estimation of VaR for $p = 0.05$ from GARCH-Historical simulation method. Solid line is the VaR plot and dotted line is financial return plot. Apparently, we can tell that VaR from GARCH-HS adjusts more quickly to the changing volatility.

Model	P	Vio Rate	LR_{uc}	LR_{ind}	LR_{cc}
GARCH-EVT	0.01	0.0091	0.5259	0.00	0.00
	0.025	0.0233	0.4347	0.5766	0.6307
	0.05	0.0503	0.9135	0.3699	0.6651
	0.1	0.117	0.00	0.00	0.00
	0.25	0.2265	0.00	0.1922	0.00
GARCH-HS	0.01	0.0079	0.1709	0.00	0.00
	0.025	0.0234	0.5037	0.7741	0.7674
	0.05	0.0494	0.8695	0.1228	0.3000
	0.1	0.0950	0.2826	0.5986	0.4888
	0.25	0.2499	0.4611	0.0648	0.1386

Table 6.6: VaR Prediction Performance: CAViaR Models

Model	P	Vio Rate	LR_{uc}	LR_{ind}	LR_{cc}
Symmetric Abs.Value	0.01	0.0139	0.0099	0.8471	0.0353
	0.025	0.02905	0.0840	0.0072	0.0061
	0.05	0.05258	0.4494	0.2683	0.4070
	0.1	0.1041	0.3706	0.0037	0.01
	0.25	0.2551	0.4416	0.00	0.00
Asymmetric slope	0.01	0.0159	0.0004	0.00	0.00
	0.025	0.0323	0.0038	0.04631	0.0021
	0.05	0.05282	0.4085	0.3732	0.4781
	0.1	0.1013	0.7802	0.1021	0.2528
	0.25	0.2522	0.7335	0.4432	0.7033

Table 6.7: VaR Prediction Performance: CAViaR Models

that asymmetric method uses two parameters (positive return and negative return) to describe the volatility change while symmetric just employ one parameter (absolute value of return) to do the job.

To sum up, the performance of these two methods is better than unconditional and GARCH regular methods, but a little inferior compared with GARCH-HS and GARCH-EVT methods.

6.2.3 Empirical results for NYSE index returns

To strengthen the persuasion, we can apply these methods to another data set, NYSE (New York Stock Exchange) Composite Index from Jan 3rd, 1986 to Dec 31st, 2002. NYSE Composite Index is the price mathematic average of all the common stocks listed on NYSE, weighted by the number to stocks issued by each issuer. We still utilize the $r_t = 100(\log p_t - \log p_{t-1})$ as returns, which yields 4290 observations.

To avoid the repetition, we do not describe the data by plots here again. We implement all the procedures we did for last data, and have the results in Table(6.8). We can notice that GARCH-HS and CAViaR-Asymmetric slope still perform best among these methods, both giving us very accurate estimation for $p \geq 0.05$. But for extreme p, none of these methods give us very satisfied estimation, including GARCH-EVT. Although the results for these two data are not exactly the same, we can see that their conclusions are consistent.

Because of the limit of pages, we do not present many empirical examples here, but all these procedures can be repeated very easily to as many data as we like.

Model	P	Vio Rate	LR_{uc}	LR_{ind}	LR_{cc}
(1)Historical Simulation	0.01	0.0164	0.00	0.00	0.00
	0.025	0.0313	0.0257	0.00	0.00
	0.05	0.0586	0.0263	0.00	0.00
	0.1	0.1064	0.2266	0.00	0.00
	0.25	0.2465	0.6428	0.00	0.00
(2)EVT	0.01	0.0152	0.1716	0.0144	0.0197
	0.025	0.0370	0.0697	0.0116	0.0197
	0.05	0.0766	0.0074	0.00	0.00
	0.1	0.1371	0.2054	0.00	0.00
	0.25	0.2993	0.1628	0.00	0.00
(3)GARCH-Normal	0.01	0.0255	0.00	0.00	0.00
	0.025	0.0379	0.0319	0.00	0.00
	0.05	0.0583	0.001	0.00	0.00
	0.1	0.1006	0.9075	0.00	0.00
	0.25	0.2237	0.0004	0.00	0.00
(4)GARCH-t(5)	0.01	0.0179	0.00	0.00	0.00
	0.025	0.0371	0.00	0.00	0.00
	0.05	0.0665	0.00	0.00	0.00
	0.1	0.1249	0.00	0.00	0.00
	0.25	0.2568	0.3664	0.00	0.00
(5)GARCH-Exponential(1)	0.01	0.0048	0.001	0.0659	0.0008
	0.025	0.0158	0.0003	0.0009	0.00
	0.05	0.0371	0.0004	0.00	0.00
	0.1	0.0978	0.6832	0.00	0.00
	0.25	0.3158	0.00	0.00	0.00
(6)GARCH-EVT	0.01	0.0155	0.0033	0.0509	0.0020
	0.025	0.0319	0.0147	0.0073	0.0014
	0.05	0.0568	0.0778	0.094	0.0521
	0.1	0.1072	0.1675	0.0294	0.0360
	0.25	0.2051	0.00	0.0041	0.00
(7)GARCH-HS	0.01	0.0164	0.0007	0.0696	0.0006
	0.025	0.0297	0.0876	0.0027	0.0026
	0.05	0.0544	0.2524	0.0894	0.1228
	0.1	0.1051	0.3268	0.0336	0.0648
	0.25	0.2499	0.8878	0.0796	0.2130
(8)Symmetric Abs.Value	0.01	0.0114	0.4458	0.0541	0.1169
	0.025	0.0287	0.2243	0.0088	0.0155
	0.05	0.0527	0.5182	0.2251	0.3888
	0.1	0.1061	0.2876	0.0008	0.0021
	0.25	0.2530	0.7105	0.0001	0.0005
(9)Asymmetric slope	0.01	0.0262	0.00	0.5812	0.00
	0.025	0.0367	0.00	0.1476	0.0012
	0.05	0.0572	0.1216	0.4901	0.2377
	0.1	0.1113	0.0747	0.5682	0.1736
	0.25	0.2524	0.7908	0.1139	0.0470

Table 6.8: VaR Prediction Performance for NYSE

6.3 Empirical study for expected shortfall

6.3.1 Procedures for different methods

In this section, we still employ the methods for VaR estimation. Because ES estimation is not easy to evaluate, we compare its performance by the simulated data instead of real data. We respectively generate 5000 observations for 3 different processes. They all follow ARMA process with parameters [2.6,-0.06,0.9]. But the error terms are assumed to come from 3 different distributions: (1) standard normal, (2) Student-t with 5 degrees of freedom, (3) Exponential with parameter 1. Note that in order to have error terms with mean zero and variance 1(as required by GARCH model), the distributions above had to be standardized. For the Student-t distribution, we simply divide the random numbers from t(5) by $\sqrt{5/3}$, the standard deviation of t(5) distribution. For the Exponential(1) distribution, instead, we first subtracted the mean 1, then reversed the sign in order to have an infinite left tail.

To be specific, the data is generated through the following process:

$$u_t = 2.6 - 0.06u_{t-1} + 0.9z_{t-1}, \tag{6.12}$$

$$r_t = u_t + z_t, \tag{6.13}$$

where we set $u_0 = 1$ and z_t is the random number from the specified distribution: N(0,1), t(5) or modified Exponential(1). Repeating this process for 5000 times will gives us 5000 observations. Although we call it a GARCH process, we actually do not utilize a change volatility. Instead, we assume that $\sigma = 1$ in the process. We are doing it because this process is much simpler but good enough to make the comparison.

Thus, we can easily obtain the real expected shortfall for each observation:

$$ES(p) = u_t + \sigma_t CE_p(z_t), \tag{6.14}$$

where $CE_p(z)$ is the conditional expectation of error term z_t and $\sigma_t = 1$. It can be calculated as

$$CE_p(z) = p^{-1} \int_{-\inf}^{VaR_p} z_t f_t(z_t) dz, \tag{6.15}$$

where f_t is the PDF of the corresponding error distribution.

Therefore, we evaluate the performance of various methods through bias and MSE:

$$bias_p = \sum_{j=1}^{4000} \widehat{ES}(p)/4000 - ES(p), \tag{6.16}$$

$$MSE_p = \sum_{j=1}^{4000} (\widehat{ES}(p) - ES(p))^2/4000. \tag{6.17}$$

The estimation of ES is directly related to the estimation of VaR. Therefore, we still compare the expected shortfall estimations based on VaR from the 9 models we used in last section. The operation details for each methods are provided below.

(1) Historical simulation

After we have calculated the $VaR(p)$ from formula(6.1), we can get the expected shortfall through the following formula

$$ES(p) = E[r_{[i]} | r_{[i]} < VaR(p)], \tag{6.18}$$

which is actually a conditional expectation.

(2) Extreme value theory

If we assume that the standardized residuals follow an EVT distribution, EVT provides a very simple formula to compute the expected shortfall, which has be proved by McNeil and Frey(2000)[19]. It can be expressed as

$$E[r_t|r_t > VaR(p)] = VaR(p)(\frac{1}{1-\xi} + \frac{\beta - \xi u}{(1-\xi)VaR(p)}). \tag{6.19}$$

Because we were interested in the minimum side of the returns and already reversed the sign before applying the extreme value theory, this conditional expectation formula can be utilized to compute the expected shortfall.

(3) GARCH-Normal

The expected shortfall for this models is relatively simple because they assume specific distributions for the error term. Referring to formula (6.4), the estimation for expected shortfall is

$$ES(p) = \widehat{u}_t + \widehat{\sigma}_t CE_p(z), \tag{6.20}$$

where $CE_p(z)$ is the conditional expectation of error term z. It can be calculated as

$$CE_p(z) = p^{-1} \int_{-\inf}^{VaR_p} z_t f_t(z_t) dz, \tag{6.21}$$

where f_t is the PDF of standard normal distribution.

(4) GARCH-t(5)

(5) GARCH-Exponential(1)

The procedures of these two methods are basically the same with GARCH-normal except that we replace standard normal distribution with t(5) and exponential(1) distribution for error terms. Here we do not repeat again.

(6) GARCH-EVT

Because in this Monte Carlo simulation, all the error terms are generated from standardized distribution, we can directly apply extreme value theory to these error terms. After we get the $CE_p(z)$ by formula (6.19), the expected shortfall is

$$ES(p) = \widehat{u}_t + \widehat{\sigma}_t CE_p(z), \tag{6.22}$$

(7) GARCH-Historical simulation

As the GARCH-EVT method, we apply the historical simulation not to the original data, but to the error terms. Using the formula (6.18) to calculate $CE_p(z)$, the expected shortfall is

$$ES(p) = \widehat{u}_t + \widehat{\sigma}_t CE_p(z), \tag{6.23}$$

(8) CAViaR-Symmetric absolute value

(9) CAViaR-Asymmetric slope

Since there are still not good ways to compute ES for VaR estimated from these two models, we can try a regression method. One can calculate the p-ES

by simply regressing the returns less than the p-VaR against the corresponding estimated p-VaR.

$$r_t = \beta VaR_t(p) + \epsilon_t, \quad for \quad r_t < VaR_t(p). \tag{6.24}$$

Now the expected shortfall is estimated by

$$ES(p) = \widehat{\beta VaR_t}(p), \tag{6.25}$$

where β is the regression coefficient in (6.24).

6.3.2 Empirical results for expected shortfall

We generated the data through the specified GARCH process with three different error distributions. We compare the estimated and real expected shortfall, and then calculate the bias and MSE. In order to present the results more clearly, we multiply all the bias and MSE we obtained with 100. The results are displayed in Table (6.9, 6.10, 6.11). Obviously, the method always with small bias and MSE is the one we need.

In the table, the number still indicates the selected method as before.

(1) Historical simulation,

(2) Extreme value theory,

(3) GARCH-Normal,

(4) GARCH-Student's t(5),

(5) GARCH-Exponential(1),

(6) GARCH-Extreme value theory,

(7) GARCH-Historical simulation,

(8) CAViaR-Symmetric absolute value, and

(9) CAViaR-Asymmetric slope.

From Table(6.9),(6.10) and (6.11), we can notice several points:

First, there is always a method which works best. For data generated with normal error terms, (3) performs best. For data with t(5) and Exponential(1) error terms, (4) and (5) methods respectively give most accurate estimation. It is an expected result. But these methods's performance would be very poor when they are applied to data from other distributions. For example, the ES estimations of GARCH-normal method on data from t(5) are very terrible.

Second, (6) and (7) perform very well, giving very small bias and MSE. And (6)'s estimations do not deteriorate with p increase, which is a behavior different from VaR estimation.

Third, (1) and (2) methods usually give us the largest bias and MSE, as they did for VaR estimation. It is not surprising because in nature, these two are just unconditional method.

Fourth, (8) and (9) methods behave superior to unconditional methods, but inferior to (6) and (7) methods. Compared with methods(6) and (7), they have one more source of error from regression, which is probably the reason for their inferiority. Meanwhile, between them two, the estimation from (9)asymmetric slope method is better than the one from (8)symmetric absolute value method, because method (9) can produce a more accurate estimation of VaR.

Last, the performance of all methods gets better when p becomes large, which is resulted from more observations contained in the estimation.

Bias	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
p=0.01	-80.40	-81.51	0.08	60.16	-193.91	2.52	-1.93	49.62	35.99
p=0.025	-69.23	-77.89	0.08	52.78	-135.02	-0.55	-0.48	35.17	33.73
p=0.05	-63.23	-71.78	0.08	46.57	-93.21	-1.64	-1.76	30.94	30.76
p=0.1	-56.57	-62.17	0.08	39.64	-54.67	-1.96	-2.47	27.99	27.31
p=0.25	-43.20	-43.61	0.08	28.73	-11.43	-2.73	-2.62	22.82	21.28
MSE	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
p=0.01	140.03	143.05	0.3	37.16	376.98	1.38	2.06	64.93	58.85
p=0.025	121.24	136.45	0.3	28.83	183.28	1.02	1.21	53.92	50.39
p=0.05	113.23	127.41	0.3	22.66	87.86	0.92	0.97	49.17	43.82
p=0.1	105.15	114.71	0.3	16.68	30.86	0.87	0.89	44.26	42.98
p=0.25	91.61	95.23	0.3	9.23	2.28	0.11	0.82	42.52	40.35

Table 6.9: ES Prediction Performance: Normal

Bias	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
p=0.01	-83.26	-74.50	76.91	0.39	-116.96	-2.92	-8.74	44.51	20.24
p=0.025	-74.05	-71.18	37.43	0.39	-97.55	-4.73	-7.09	50.18	34.05
p=0.05	-65.97	-65.01	16.01	0.39	-77.16	-4.19	-6.46	42.59	32.07
p=0.1	-57.60	-56.89	1.27	0.39	-53.37	-3.05	-3.75	38.58	30.36
p=0.25	-43.60	-45.83	8.37	0.39	-19.77	-3.15	-3.95	27.61	24.66
MSE	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
p=0.01	150.25	137.34	59.79	1.14	139.93	4.69	6.55	71.05	56.47
p=0.025	132.71	128.04	14.65	1.14	98.29	1.96	2.18	72.51	53.38
p=0.05	119.57	118.23	3.21	1.14	62.67	1.37	1.76	61.29	48.54
p=0.1	108.41	107.58	0.66	1.14	31.61	1.09	1.13	56.30	46.04
p=0.25	93.97	96.24	1.34	1.14	7.04	1.06	1.02	48.03	42.11

Table 6.10: ES Prediction Performance: Student-t(5)

Bias	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
p=0.01	-72.68	-61.69	193.85	115.08	-0.48	10.43	4.93	58.17	7.86
p=0.025	-63.80	-58.34	134.96	95.67	-0.48	6.17	5.16	45.96	18.47
p=0.05	-59.36	-55.59	93.16	75.28	-0.48	3.52	2.87	44.30	15.98
p=0.1	-52.46	-51.71	54.61	51.48	-0.48	1.71	0.73	33.70	17.54
p=0.25	-41.99	-43.27	11.37	17.89	-0.48	1.08	-0.51	24.11	17.17
MSE	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
p=0.01	175.24	163.24	376.31	134.33	1.89	12.70	15.08	94.52	84.41
p=0.025	125.84	119.48	182.67	93.42	1.89	4.49	5.19	72.49	46.51
p=0.05	110.92	106.01	87.31	58.57	1.89	2.41	2.11	64.82	45.86
p=0.1	99.71	98.92	30.35	28.41	1.89	1.51	1.50	56.81	39.68
p=0.25	89.13	91.23	1.81	5.09	1.89	0.88	1.01	55.05	43.42

Table 6.11: ES Prediction Performance: Exponential(1)

Chapter 7

Conclusion

For VaR estimation, from the empirical study in last chapter, we can clearly see GARCH-Historical Simulation and Asymmetric Slope have the best performance, especially when p is greater than 0.05. Both of them passed all the tests for moderate quantile. Their superiorities derive from their successful captures of the volatility clustering in their models. Comparatively, their performances when p is equal or smaller than 0.01, are much poorer, probably because there are too few observations contained in the estimations.

GARCH-EVT, however, works very well for low quantile estimation. It passed all the tests for p smaller and equal to 0.05, which is a sharp contrast with the GARCH-HS. This is not surprising for the property of EVT.

The other methods behave poorly, either due to their inability to describe the change in volatility, like historical simulation and EVT, or to their inaccurate assumption about the financial returns, such as GARCH-Normal.

For ES estimation, after comparing the bias and MSE, we can clearly conclude that GARCH-HS and GARCH-EVT always work best, no matter what distribution we are dealing with.

To conclude, GARCH-EVT and GARCH-HS perform best for both VaR and ES estimation. CAViaR-Asymmetric slope method can give us very precise estimation for VaR, but its estimation on ES still needs a great improvement.

Bibliography

[1] C. Acerbi and D. Dasche. On coherence of expected shortfall. *Journal of Banking & Finance*, 26:1487–1503, 2002.

[2] P. Artzner, F. Delbaen, J. Eber, and D. Heath. Coherent measures of risk. *Mathematical Finance*, 9:203–228, 1999.

[3] J. Berkowitz and J. O'Brien. How accurate are Value-at-risk models at commercial banks. *Journal of Finance*, 57:1093–1111, 2002.

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

[5] T. Bollerslev and J. M. Wooldridge. Quasi-Maximum likelihood estimation and inference in dynamic models with Time-varying covariance. *Econometric Reviews*, 11:143–172, 1992.

[6] P. F. Christofferson. Evaluating interval forecasts. *International Economic Review*, 39:841–862, 1998.

[7] P. Csoka, P. Jean, J. Herings, and L. A. Koczy. Coherent measure of risk from a general equilibrium perspective. *Journal of Banking & Finance*, 31:2517–2534, 2006.

[8] J. Danielsson and Y. Morimoto. Forecasting extreme financial risk: a critical analysis of practical methods for Japanese market. *Monetary and Economic Studies*, 18:25–48, 2000.

[9] F. Delbaen. Coherent risk measures on general probability spaces. *Advances in finance and stochastics: essays in honour of Dieter Sondermann*, 2000.

[10] R. Engle. GARCH 101: the use of ARCH/GARCH models in applied econometrics. *Journal of Economic Perspectives*, 15:157–168, 2001.

[11] R. F. Engle and S. Manganelli. CAViaR: Conditional Autoregressive value at risk by regression quantiles. *NBER Working Papers*, 7341, 1999.

[12] E. F. Fama. The behavior of stock market prices. *The Journal of Business*, 38:34–105, 1965.

[13] E. J. Hannan and J. Rissanen. Recursive Estimation of Mixed Autoregressive-Moving Average Order. *Biometrika*, 69:81–94, 1982.

[14] R. Koenker and G. Bassett. Regression Quantiles. *Econometrica*, 46:33–50, 1978.

[15] R. Koenker and K. F. Hallock. Quantile regression. *Journal of Economic Perspectives*, 15:143–156, 2001.

[16] K. Kuester, S. Mittnik, and M. S. Paolella. Value-at-Risk Prediction:A Comparison of Alternative Strategies. *Journal of Financial Econometrics*, 4:53–89, 2006.

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

[18] S. Manganelli and R. F. Engle. Value at risk models in finance. *European Central Bank Working Paper No. 75*, 2001.

[19] A. J. Mcneil and R. Frey. Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of Empirical Finance*, 7:271–300, 2000.

[20] F. Peracchi and A. V. Tanase. On estimating the conditional expected shortfall. *Applied Stochastic Models in Business and Industry*, 24:471–493, 2008.

[21] P. Posedel. Properties and estimation of GARCH(1,1) model. *Metodoloski zvezki*, 2:243–257, 2005.

[22] J. W. Taylor. Using exponentially weighted quantile regression to estimate value at risk and expected shortfall. *Journal of Financial Econometrics*, 6:382–406, 2008.

[23] R. S. Tsay. *Analysis of financial time series*. John Wiley&Sons, New Jersey, 2005.

[24] Z. Xiao and R. Koenker. Conditional quantile estimation for Garch models. *Boston College Working Papers in Economics*, 725, 2009.