Modeling the volatility of stock indexes and predicting VaR during COVID-19 pandemic

by

Xiang Zhang

A thesis submitted to the Faculty of Graduate and Postdoctoral Affairs

in partial fulfillment of the requirements for the degree of

Master of Science

in

Mathematics and Statistics

Carleton University

Ottawa, Ontario

© 2022

Xiang Zhang
Abstract

As COVID-19 spreads globally and becomes a pandemic disease, it has severely impacted the capital markets of major economies around the world, posing severe challenges to financial risk management. In this thesis, we study the ARIMA-GARCH model and the Markov switching GARCH model and its application in financial risk management.

This thesis verifies and compares the predictive ability of ARIMA-GARCH models and Markov switching GARCH models on value at risk through empirical research. Especially incorporating the drastic fluctuations caused by the COVID-19 into the forecasting scope, we found that the ARIMA-GARCH model has basically no predictive ability, and the Markov switching GARCH model still has a good forecasting ability. Then we try to further improve the predictive power of the model by adding one additional regime. However, the performance of the model did not improve significantly.

Key words: COVID-19, ARIMA-GARCH model, Markov switching GARCH model, value at risk, stock indexes
Acknowledgements

I would like to give my sincere gratitude to Prof Yiqiang Zhao, my supervisor, with extraordinary patience and consistent encouragement, gave me great help by providing me with necessary materials, advice of great value and inspiration of new ideas. In particular, Professor Zhao’s rigorous academic research and responsible spirit have benefited me a lot and will affect my future study and life.

I would also like to express my deep gratitude to my family, especially for my special study experience during the epidemic. My family’s encouragement helped me overcome the difficulties in study and life.
Glossary

**Value at Risk (VaR):** A statistic that quantifies the extent of possible financial losses within a firm, portfolio, or position over a specific time frame.

**Stock Index:** An index that measures a stock market, or a subset of the stock market, that helps investors compare current stock price levels with past prices to calculate market performance.

**Circuit Breaker or Trading Curb:** Circuit Breaker is a financial regulatory instrument that is in place to prevent stock market crashes from occurring, and is implemented by the relevant stock exchange organization. When triggered, circuit breakers either stop trading for a small amount of time or close trading early in order to allow accurate information to flow among market makers and for institutional traders to assess their positions and make rational decisions.

**GARCH Model:** GARCH models describe financial markets in which volatility can change, becoming more volatile during periods of financial crises or world events and less volatile during periods of relative calm and steady economic growth.

**Regime switching:** Regime switching models are most commonly used to model time series data that fluctuates between recurring states. If we are dealing with data that seem to cycle between behavioral cycles, we might want to consider a regime switching model.
# Contents

Abstract .......................................................... ii

Acknowledgements ........................................ iii

Glossary .......................................................... iv

1 Introduction ...................................................... 1

1.1 Motivation .................................................. 1

1.2 Background ............................................... 3

1.3 Contribution ............................................... 7

1.4 Organization ............................................... 7

2 Literature Review & Theory .................................. 9

2.1 Literature review ......................................... 9

2.2 Financial risk management theory ....................... 15

2.3 Volatility models ......................................... 28

3 Forecasting Models & Results ............................ 53

3.1 Data collection and preprocessing ....................... 53
# List of Figures

3.1 Equity market indexes for four different economies .......................... 54
3.2 Distribution of stock market indexes ........................................ 55
3.3 Daily returns of these stock indexes and their distribution .......... 57
3.4 Squared residuals of ARIMA models ............................................. 59
3.5 95%VaR predicted by ARIMA-GARCH model and delta-normal  
    method .................................................................................. 60
List of Tables

3.1 Correlation coefficients between different indexes .......................... 56
3.2 Preliminary analysis of index daily returns ................................. 58
3.3 Optimal ARIMA model for stock indexes ................................. 59
3.4 Backtest results of VaR estimated by ARIMA-GARCH model for recent 300 trading days ...................................................... 62
3.5 Backtest results of VaR estimated by ARIMA-GARCH model for recent 800 trading days ...................................................... 64
3.6 Backtest results of the S&P/TSX index in the recent 300 trading days ......................................................................................... 65
3.7 Backtest results of the S&P/TSX index in the recent 800 trading days ......................................................................................... 66
3.8 Backtest results of the S&P500 index in the recent 300 trading days 66
3.9 Backtest results of the S&P500 index in the recent 800 trading days 66
3.10 Backtest results of Nikkei225 index in the recent 300 trading days 68
3.11 Backtest results of Nikkei225 index in the recent 800 trading days 68
3.12 Backtest results of FTSE100 index in the recent 300 trading days 69
3.13 Backtest results of FTSE100 index in the recent 800 trading days 69
**LIST OF TABLES**

A.1 Backtest results of the S&P/TSX index in the recent 300 trading days with three regimes ........................................ 80
A.2 Backtest results of the S&P/TSX index in the recent 800 trading days with three regimes ........................................ 81
A.3 Backtest results of the S&P500 index in the recent 300 trading days with three regimes ........................................ 81
A.4 Backtest results of the S&P500 index in the recent 800 trading days with three regimes ........................................ 81
A.5 Backtest results of the Nikkei225 index in the recent 300 trading days with three regimes ........................................ 82
A.6 Backtest results of the Nikkei225 index in the recent 800 trading days with three regimes ........................................ 82
A.7 Backtest results of the FTSE100 index in the past 300 trading days with three regimes ........................................ 83
A.8 Backtest results of the FTSE100 index in the past 800 trading days with three regimes ........................................ 83
B.1 MSGARCH for S&P/TSX index ........................................ 85
B.2 The transition matrix of the S&P/TSX model ........................................ 85
B.3 MSGARCH for S&P500 index ........................................ 86
B.4 The transition matrix of the S&P500 model ........................................ 86
B.5 MSGARCH for Nikkei225 index ........................................ 87
B.6 The transition matrix of the Nikkei225 model ........................................ 87
B.7 MSGARCH for FTSE100 index ........................................ 88
B.8 The transition matrix of the FTSE100 model ........................................ 89
Chapter 1

Introduction

This chapter presents the specific background, importance and rationale of the research topic. The main contribution of this thesis to the literature is summarized. Finally, the structure of this thesis is presented before moving on to the next chapter.

1.1 Motivation

1.1.1 Global Pandemic of COVID-19

Coronavirus disease 2019 (COVID-19) is an infectious disease caused by the severe acute respiratory syndrome coronavirus (SARS-CoV-2) virus. The first known case was detected in Wuhan, China in December 2019. Due to its highly contagious and mutated ability, the disease has spread rapidly around the world, leading to the COVID-19 pandemic. Several major economies, such as the United States, Canada, Japan and the United Kingdom, all reported the first confirmed case of COVID-19 in their country in January 2020.
CHAPTER 1. INTRODUCTION

Some studies have shown that there is a positive statistical relationship between the number of confirmed cases in various countries and their level of economic development[20]. Countries with high per capita GDP also have a higher number of confirmed cases per million people. It is believed that there are three possible reasons: First, the mobility of people in relatively developed countries is usually larger, and the degree of internationalization is also higher, which may easily cause the infection and spread of the epidemic. Second, the population aging rate of some economically developed countries is relatively high, while current clinical data show that the elderly are more likely to be infected with the new coronavirus, which has also led to an increase in the number of confirmed cases in these countries. Third, the health care systems in developing countries are usually relatively backward, and the ability to conduct large-scale virus testing is insufficient, leading to underestimation of cases.

As the number of confirmed cases of the COVID-19 epidemic in the world continues to rise, it has caused a significant impact on the world economy. On the one hand, the spread of the epidemic has accelerated around the world, uncertainty has risen sharply, and investor confidence has been frustrated, causing financial and capital market turmoil. On the other hand, in order to control the spread of the epidemic, countries strictly restrict the movement of people and transportation, which has a negative impact on economic development.

1.1.2 2020 stock market crash

With the spread of the COVID-19 around the world, global financial markets have
experienced sharp fluctuations[21], and U.S. stocks have experienced four circuit breakers in less than two weeks (March 9-18). In addition to the panic caused by the epidemic, the recent general slowdown in world economic growth is also one of the important factors for stock market decline. Stock market indexes in several major economies fell by more than 30% in the two months of February and March 2020. For example, in February and March 2020, the S&P/TSX composite index fell from 17944 points to 11228 points with a maximum drawdown of more than 37%, the S&P 500 index fell from 3386.15 points to 2237.4 points with a maximum drawdown of nearly 34%, the Nikkei 225 index fell from 24083.51 points to 16552.83, with a maximum drawdown of more than 31%, and the UK’s FTSE 100 fell from 7,674.56 points to 4,993.89 points, with a maximum drawdown of nearly 35%.

This poses a huge challenge to financial risk management. So, how to use certain models to provide reliable forecasts for the financial risk management indicators of the volatile capital market since the outbreak of the COVID-19 has attracted more and more attention.

1.2 Background

1.2.1 Stock Market Indexes

The stock market index is compiled to measure and reflect the overall price level of the stock market and its changing trend. When the stock index rises, it indicates that the average price level of the stock has risen. When the stock index falls, it
CHAPTER 1. INTRODUCTION

indicates that the average price level of the stock has dropped. It is a barometer that reflects the social, political, and economic changes in the country (or region). Stock market indexes are leading indicators and often respond quickly and sensitively to major social and economic events. This thesis covers the following four major and influential stock market indexes.

The S&P/TSX Composite Index is the benchmark Canadian index, representing roughly 70% of the total market capitalization on the Toronto Stock Exchange (TSX) with about 250 companies included in it. The S&P/TSX indexes are calculated and managed by S&P Dow Jones Indexes. The TMX Group Inc. (TMX) is the owner and distributor of all S&P/TSX equity index data.

The Standard and Poor’s 500, or simply the S&P 500 is a stock market index tracking the stock performance of 500 large companies listed on exchanges in the United States. This stock index is created and maintained by S&P Dow Jones Indexes LLC. Compared to the Dow, the S&P 500 contains more companies, so risks are more spread out, it reflects broader market changes, and even the index is good enough to show the rise and fall of the U.S. economy.

The Nikkei 225, or the Nikkei Stock Average is a stock market index for the Tokyo Stock Exchange (TSE). It has been calculated daily by the Nihon Keizai Shimbun (The Nikkei) newspaper since 1950. It is a price-weighted index, operating in the Japanese Yen (JP¥), and its components are reviewed once a year. The Nikkei measures the performance of 225 large, publicly owned companies in Japan from
a wide array of industry sectors.

The Financial Times Stock Exchange 100 Index, also called the FTSE 100 Index, is a share index of the 100 companies listed on the London Stock Exchange with (in principle) the highest market capitalization. The index is maintained by the FTSE Group, a subsidiary of the London Stock Exchange Group.

The above stock indexes represent the major developed economies in the Americas, Europe and Asia, respectively. There is evidence that capital markets are more correlated in developed economies. We do find strong correlations with each other, especially the correlation coefficient between S&P/TSX Composite Index and S&P500 is even as high as 0.937.

1.2.2 Risk Management and VaR Approach

With the development of the economy, the international economic situation is unpredictable, and risk events often occur in a concentrated manner, such as the spread of COVID-19, the Russian-Ukrainian war, etc. Any investor and financial institution will inevitably bear certain risks in such a complex economic environment. Every violent fluctuation in financial market prices has attracted the attention of financial institutions and countless investors, and every risk case in a large financial institution has also attracted the attention of people all over the world. Therefore, the importance of financial risk management is getting more and more attention.
CHAPTER 1. INTRODUCTION

The initial research on risk issues can be traced back to the early 20th century, but the real development should be in the 1960s. By the 1980s, risk management has become a hot issue in economic research. The core content of financial risk management is the measurement of risk. Through the analysis of market factors, we can measure how they cause changes in financial assets. With the development of economic theory, the method of risk measurement has also made great progress. From the initial nominal value method, the quantitative method represented by value at risk (VaR) has been gradually developed[22].

Changes in market factors are random and difficult to measure and predict accurately. Therefore, the measurement of risk could be described by random variables. This leads to a new method of risk measurement, value at risk (VaR). The calculation method of VaR is to set an acceptable confidence level and calculate the maximum loss that the asset may suffer over a specific time frame. It comprehensively considers and quantifies all market factors and the risks that these market factors may cause, and finally expresses the risks in real numbers, which meets the needs of risk managers on this issue. However, because the global financial market is complex and continuously changing, although VaR can measure the possible losses caused by different risks, its calculation is relatively complicated. The most notable features of financial time series are the volatility clustering and the non-normality of return distribution, which prompts people to use GARCH models to calculate VaR.
1.3 Contribution

This thesis summarizes the core content of the financial risk field, and reviews the development and application of the GARCH model and the Markov switching GARCH model. Through empirical analysis, we draw the following conclusions:

1. For relatively stable market conditions, the ARIMA-GARCH model and the Markov Switching GARCH model both have good predictive ability for 95%VaR and 99%VaR respectively. But the predictive power of the Markov Switching GARCH model is still better.

2. This thesis is most concerned with the dramatic financial market volatility caused by the global spread of COVID-19. Incorporating this time period into the forecast horizon, we find that the ARIMA-GARCH model is no longer able to give reliable VaR predictions. VaR estimates based on this model will greatly underestimate the actual risk. But the Markov switching GARCH model still has good estimation ability for VaR.

3. Markov switching GARCH model with TGARCH and skewed and kurtosis distributions have better predictions.

4. Adding one more regime do not significantly improve predicted VaR accuracy, which may also require validation with more data.

1.4 Organization

The remaining portion of the thesis is organized as follows:

- Chapter 2 This thesis reviews the theoretical background, concepts and methods
of models and knowledge related to financial risk management. In addition, we systematize the literature of the VaR method and GARCH models with particular application to financial risk management, and summarize the development of studies in the field.

- Chapter 3 Compares the prediction effect of GARCH model and Markov switching GARCH models on the risk management index VaR of stock market indexes. We obtain relatively good models.
- Chapter 4 Summarizes and discusses potential research.
Chapter 2

Literature Review & Theory

In this chapter, the literature on financial risk management and volatility models are summarized and the development of the studies in the field is discussed. After that, the background knowledge of financial risk management and different estimation methods of risk indicators are introduced. And summarizes the characteristics and difficulties of financial time series modeling. Finally, we focus on related models and theoretical approaches.

2.1 Literature review

2.1.1 Financial risk management
Risk management is an unavoidable issue in the financial field. Since the middle of last century, financial risk management has become a research hotspot, and its theories and methods have been further improved. VaR method is the main method for studying risk today. Since it was proposed by the G30 group in July 1993, it has been widely used in various financial fields[1], and its calculation
methods have continued to appear. There are parametric (variance-covariance method) and nonparametric methods to estimate VaR.

As for the variance-covariance method, the development is relatively mature. Engle[2] first proposed the idea of using the ARCH model to model variance in 1982, and then Bollerslev, Tim[3] proposed the GARCH model based on the ARCH model in 1986. In order to better simulate financial time, many scholars have successively proposed methods to calculate VaR values based on other distributions. For example, Venkataraman[4] proposed a calculation method based on mixed normal distribution etc.

One of the advantages of the normal distribution is that it is relatively simple, and the VaR of the normal distribution can be directly expressed as a linear function of the mean and standard deviation. Although the normal distribution has facilitated the widespread use of VaR, assumptions based on the normal distribution ignore fat-tailed features, making it possible to underestimate financial risk, especially during the crisis time. In addition, most financial assets in real life show the characteristics of fat tails, and also show strong dependence on market volatility. Therefore, the variance-covariance method is may not suitable for some risk calculations[5].

The historical simulation method was first proposed by Butler and Schachter[6], and then improved by Boudoukh, J., Richardson, M. and Whitelaw, R.F[7]. This method assumes that the historical distribution of portfolio returns during the
observation period can represent the distribution of returns during future holding periods. Sort the possible virtual portfolio returns in ascending order to get the profit and loss distribution. Finally, the VaR of the current portfolio is determined by the quantile corresponding to the given confidence level.

The most significant advantage of the historical simulation method is that there are no restrictions on the distribution of returns on investments. For this reason, the historical simulation method is not affected by the fat-tailed characteristics of financial asset returns when calculating VaR, and overcomes the non-normality and nonlinearity that plague the variance-covariance method[8]. In addition, historical simulation methods with the properties of nonparametric methods do not need to estimate any parameters such as mean, variance, and correlation coefficient. In other words, compared to parametric methods, this method avoids the risk of model errors.

2.1.2 Modeling volatility

The modeling of financial time series has been extensively studied and has gradually evolved with numerous developments and applications. A topic that has received a lot of attention is the study of volatility in financial time series. Because investments are inherently risky, accurately modeling and predicting volatility is an important and challenging task.

The research focus on financial asset volatility dates back to the early 1960s. However, volatility models have attracted considerable attention in recent years[10].
GARCH models are widely used to model the volatility of time series data. While stock markets often experience volatility clustering, these GARCH-type models are believed to predict market volatility in a superior way than basic volatility models such as random walks and historical averages[11]. GARCH models are very popular among practitioners and generally perform well compared to other models used to estimate volatility and risk measures.

Bollerslev[3] developed GARCH as a generalization of Engle’s[2] original ARCH volatility modeling technique. Because ARCH models often contain a lot of parameters, and GARCH models reduce the number of parameters in the model, which also reduces the amount of calculation for parameter estimation. Nelson[31] proposed an exponential GARCH (EGARCH) model based on the logarithmic expression of the conditional variability of the analyzed variables. Later, this method was further developed. For example, Glosten, Jagannathan, and Runkle[34] proposed the GJR-GARCH model, which can accommodate the asymmetry in the response of the variance to a shock.

In addition, the volatility of many capital markets also showed structural changes. In these cases, volatility modeling and forecasting using GARCH-type models cannot fully capture volatility changes[12][13][14]. One way to solve this problem is to allow the parameters of the GARCH model to vary according to the latent variables of the Markov process, and thus Markov switching GARCH model(MS-GARCH model) was born. A generalization to Markov switching GARCH models was developed by Gray[15] and subsequently modified by Klaassen[16]. The
model allows for different GARCH behavior in different regime, that is, differences in variance dynamics can be captured during periods of low and high volatility. For example, the market may switch between a bull market and a bear market. If the market switches to a bear market, the risk of the market and the possible loss of investment will increase.

Some important literatures on the application of Markov switching models to the financial field include: Maheu & McCurdy\cite{17} divide the stock market into two states: bear market and bull market, and bear market has the characteristics of greater volatility.

Hamilton & Susmel\cite{18} introduced three states to describe the market, namely low, medium and high volatility regimes, where high volatility tends to be associated with recessions. Additionally, others have established that regime-switching models are able to identify structurally distinct periods of volatility.

Ardia, Bluteau, Boudt, and Catania\cite{45} implement their different models using the MS-GARCH R Package. They applied these models to predict various risk management indicators; for example, value-at-risk (VaR) and expected shortfall and found that the MS-GARCH model provided better results compared to the different single-regime GARCH/GJR models.

David Suda and Anna Bonello\cite{46} modeled the volatility of the highly volatile Bitcoin and then predicted Bitcoin’s VaR. They applied MSGARCH models with
two regimes, and the single-regime counterpart for comparison, on Bitcoin/US Dollar using both normal and $t$ distribution. In the frequentist approach, it was found that the $tMSG2$ was the best of the models considered where goodness-of-fit is concerned. It was also found to have the best risk forecasting performance for one-step VaR, and also performed well for multistep VaR and in explaining tail behaviors.

Since the beginning of 2020, COVID-19 has become a global pandemic, which has caused a huge impact on the global society, economy and financial market. Nobuhle Mthethwa et al.[47] applied a Markov Switching-Volatility Model Combined with Heavy-Tailed Distributions to Estimate the VaR of COVID-19 Death for South Africa.

Ibrahim, Omar[48] aims at evaluating among market risk measures to equity exposures on the Egyptian stock market, while utilizing a variety of parametric and non-parametric methods to estimating volatility dynamics. Comparing the prediction results of EWMA (RiskMetrics), GARCH, GJR-GARCH, and Markov-Regime switching GARCH models for Value at Risk and Conditional Value at Risk shows that the combination of the symmetric GARCH model and the Markov switching model is the best predictor of risk indicators.

N M S Dwipa and B Wicaksono[49] modeling the price of world oil using the Markov switching GARCH model. The paper concludes that the best model for describing international oil prices is Markov Regime Switching-GARCH. Modeling
results can be used as alternative data for investor’s consideration for determining their investment decisions.

Dinghai Xu[50] models and analyzes the Canadian S&P/TSX Composite Index, an index of Canadian capital markets, using the Time-Varying REGARCH (TV-REGARCH) model. The paper find that the global pandemic of COVID-19 drove a sharp increase in the level of volatility in Canada’s S&P/TSX composite index when Canada recorded its first confirmed case of coronavirus. After March 20, 2020, the model observed a slight decrease in market volatility.

2.2 Financial risk management theory

The core of the financial industry is risk. Western economics regards the time value of assets, asset pricing and risk management as the three pillars of modern financial theory. Risk management covers many fields, including market risk, credit risk, operational risk and many other contents. In today’s complex financial markets, risks are often difficult to grasp. In times of financial market distress or crisis, managing risks effectively is often the key to business success.

2.2.1 The basic characteristics of financial risk

The following are basic characteristics of financial risk:

Uncertainty: It is difficult to fully grasp the factors affecting financial risk in advance.

Relevance: The particularity of the commodity-currency operated by financial in-
Institutions determines that financial institutions are closely related to the economy and society.

High leverage: Financial enterprises have high debt ratios and large financial leverage, resulting in large negative externalities. In addition, financial instrument innovation and derivative financial instruments are also accompanied by high financial risks.

Contagiousness: Financial institutions assume the function of intermediary institutions, cutting off the corresponding relationship of original lending. Risks on any party in this intermediary network may have an impact on other parties, and even industry and regional financial risks may occur, leading to financial crises.

2.2.2 Important distributions

People often assume that the distribution of investment returns follows a normal distribution. However, the data in the financial field often show special characteristics, so more and more other distributions are introduced into the modeling of financial data to show better model effects. This thesis will use the following important distributions.

Normal distribution

In statistics, a normal distribution (also known as Gaussian, Gauss, or Laplace–Gauss distribution) is a type of continuous probability distribution for a real-valued random variable. The general form of its probability density function is

$$f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{x-\mu}{\sigma} \right)^2}$$  \hspace{1cm} (2.1)
The parameter $\mu$ is the mean or expectation of the distribution (and also its median and mode), while the parameter $\sigma$ is its standard deviation. The variance of the distribution is $\sigma^2$.

**Skewed normal distribution**

In probability theory and statistics, the skew normal distribution\[35\] is a continuous probability distribution that generalises the normal distribution to allow for non-zero skewness. The probability density function (pdf) of the skew-normal distribution with parameter $\alpha$ is given by

$$f(x) = 2\phi(x)\Phi(\alpha x) \quad (2.2)$$

where $\phi(x)$ denotes the standard normal probability density function and $\Phi(x)$ denote the cumulative distribution function given by

$$\Phi(x) = \int_{-\infty}^{x} \phi(t)dt \quad (2.3)$$

**Student-t distribution**

Bollerslev\[29\] suggested replacing the assumption of normality in the error term with the assumption of the t distribution (StD). The density of the distribution can be written as

$$f \left( u_t / Y_{t-1}^p \right) = \frac{\Gamma \left[ \frac{1}{2} (v + 1) \right]}{\pi^{\frac{1}{2}} \Gamma \left[ \frac{1}{2} v \right]} \left[ (v - 2)\sigma_t^2 \right]^{-\frac{1}{2}} \left[ 1 + \frac{u_t^2}{(v - 2)\sigma_t^2} \right]^{-\frac{1}{2}(v+1)} \quad (2.4)$$
where \( v \) (number of degrees of freedom), \( u_t \) and \( \sigma_t^2 \) are the parameters, and \( \Gamma(\cdot) \) is the gamma function.

**Skewed student distribution**

Fernandez and Steel[36] extended StD by introducing skewness parameter. The density function can be expressed as:

\[
\begin{align*}
    f(\epsilon | \xi) &= \frac{2}{\xi + \frac{1}{\xi}} \left[ g \left( \frac{\epsilon}{\xi} \right) I_{[0,\infty)}(\epsilon) + g(\xi \epsilon) I_{(\infty,0)}(\epsilon) \right] \\
    \text{(2.5)}
\end{align*}
\]

where \( 0 < \xi < \infty \) is a shape parameter which describes the degree of asymmetry. \( \xi = 1 \) means that the skewed Student distribution is symmetric.

Considering the random variable, \( z_t = \frac{\epsilon_t - \mu_s}{\sigma_t} \) with mean equal to zero and variance equal to one, the standardized skewed Student-t log-likelihood is:

\[
\begin{align*}
    L_{SkSt}(\theta) &= \ln \left[ \Gamma \left( \frac{v + 1}{2} \right) \right] - \ln \left[ \Gamma \left( \frac{v}{2} \right) \right] - \frac{1}{2} \ln[\pi(v - 2)] + \ln \left( \frac{2}{\xi + \frac{1}{\xi}} \right) \\
    \quad - \frac{1}{2} \sum_{t=1}^{T} \ln \sigma_t^2 + (1 + v) \ln \left( 1 + \frac{s \epsilon_t + m}{v - 2 - \xi - \mu_s} \right) \\
    \text{(2.6)}
\end{align*}
\]

where \( v \) (number of degrees of freedom), \( m \) and \( s \) are the parameters.

**Generalized error distribution**

In Nelson[31], the Exponential GARCH model was developed, referred to the
Generalized Error Distribution (GED) and provides this density function:

\[ f(z; v) = \frac{v \cdot \exp \left[ -\frac{1}{2} \left| \frac{z}{\lambda} \right|^v \right]}{\lambda^{2(1+1/v)} \Gamma(1/v)} \] (2.7)

where \( \lambda \equiv \sqrt{\frac{2(-2/v) \Gamma(1/v)}{\Gamma(3/v)}} \), \( v \) is the degree of freedom and \( \Gamma(\cdot) \) is the gamma function.

**Skewed generalized error distribution**

The probability density function of the Skewed Generalized Error Distribution (SGED)[37] is

\[ f(y | \mu, \sigma, k, \lambda) = \frac{C}{\sigma} \exp \left( -\frac{1}{\left[ 1 - \text{sign}(y - \mu + \delta \sigma) \lambda \right]^{k \theta} \sigma^k |y - \mu + \delta \sigma|^k} \right) \] (2.8)

where

\[
C = \frac{k}{2\theta} \Gamma\left(\frac{1}{k}\right)^{-1} \\
\theta = \Gamma\left(\frac{1}{k}\right)^{\frac{1}{2}} \Gamma\left(\frac{3}{k}\right)^{-\frac{1}{2}} S(\lambda)^{-1} \\
\delta = 2\lambda AS(\lambda)^{-1} \\
S(\lambda) = \sqrt{1 + 3\lambda^2 - 4A^2\lambda^2} \\
A = \Gamma\left(\frac{2}{k}\right) \Gamma\left(\frac{1}{k}\right)^{-\frac{1}{2}} \Gamma\left(\frac{3}{k}\right)^{-\frac{1}{2}}
\]

\( \mu = \mathbb{E}(y) \) and \( \sigma \) are the expected value and the standard deviation of the random variable \( y \), \( \lambda \) is a skewness parameter, \( \text{sign} \) is the sign function, and \( \Gamma(\cdot) \) is the gamma function.

**2.2.3 Value at risk**

Value at Risk (VaR) is an important financial risk management indicator used to
quantify the degree of financial loss that may occur to a company, portfolio or position over a specified time frame. This indicator is most commonly used by investment and commercial banks to determine the magnitude and probability of potential losses in their institutional portfolios.

Risk managers use VaR to measure and control risk exposure levels. One can apply VaR calculations to specific positions or entire portfolios, or use them to measure company-wide exposure. Global regulatory authorities have recognized Value at Risk (VaR) and Expected Shortfall (ES) as standard or best risk indicators. Even without mandatory regulation, risk management practitioners are increasingly using such risk indicators to reduce exposure to internal risk management tools[24].

**Parametric VaR**

Parametric VaR [25](also known as Variance/Covariance VaR) calculation is the most common form used by financial risk managers in practice. Using the parametric method, we assume that the returns of the portfolio are satisfied by a known distribution, and then estimate the parameter values of the distribution. When the financial risk management is not mature enough, people tend to assume that portfolio returns follow a normal distribution. In practice, this assumption of return normality has proven to be very dangerous. In fact, it is the biggest mistake made by long-term capital management firms, which grossly underestimate their portfolio risk. So, people start introducing some leptokurtic fat tails or asymmetry distribution to avoid underestimating risk.
Parametric VaR with normal distribution

Assumptions based on normal distribution have been shown to often underestimate financial risk. However, VaR based on the assumption of normal distribution sometimes still has reference value. In this case assume that arithmetic returns are normally distributed with mean $\mu$ and standard deviation $\sigma$.

$$\text{VaR} = - (\mu - z_\alpha \sigma) \quad (2.9)$$

A Z-score is a numerical measurement that describes a value’s relationship to the mean of a group of values. Z-score is measured in terms of standard deviations from the mean. For example, a Z-score of 1.65 would indicate a value that is 1.65 times the standard deviation from the mean.

Parametric VaR with $t$ distribution

The Leptokurtic distribution can be described as having a wider or flatter shape and a thicker tail, resulting in a greater chance of extreme positive or negative events. In a symmetric unimodal distribution, i.e., one whose density function has only one peak, leptokurtosis is indicated by a positive excess kurtosis. The impact of leptokurtosis on VaR is non-negligible. Firstly, assuming high significance levels, e.g., $\alpha \leq 0.05$ or $\alpha \leq 0.01$, the estimation of VaR for Normal and leptokurtic distribution would return:

$$\text{VaR}_{\text{lepto}, \alpha \leq 0.01} > \text{VaR}_{\text{Norm}, \alpha \leq 0.01} \quad (2.10)$$
This is because at a high level of significance, the probability of rare events is higher in a Leptokurtic distribution. This leads to a larger value of VaR, which also shows that VaR based on the assumption of Leptokurtic distribution is less likely to underestimate risk.

However, for low significance levels the situation may (not necessarily) change:

\[ \text{VaR}_{\text{lept}} \leq 0.05 < \text{VaR}_{\text{Norm}}, \alpha \geq 0.05 \]  

(2.11)

A good example of the leptokurtic distribution is student distribution\[26\]. If it describes the data better and displays significant excess kurtosis, there is a good chance that VaR for high significance levels will be a much better measure of the tail risk.

If it is assumed that the return \( r \) is the standard Student distribution with degrees of freedom \( v \), then the quantile of this distribution is \( q = \mu + t_v^*(\alpha) \), where \( t_v^*(\alpha) \) is the quantile to \( \alpha \) of standard Student distribution with degrees of freedom \( v \). The relationship between the quantile of standard Student distribution with degrees of freedom is \( v \), denoted by \( t_v \), and the standard distribution denoted by \( t_v^* \), is:

\[
p = P(t_v \leq q) = P\left( \frac{t_v}{\sqrt{v/(v-2)}} \leq \frac{q}{\sqrt{v/(v-2)}} \right) = P\left( t_v^* \leq \frac{q}{\sqrt{v/(v-2)}} \right)
\]

(2.12)

where \( v > 2 \)

So, if the quantile to \( \alpha \) of standard Student distribution with degrees of freedom \( v \), then \( q/\sqrt{v/(v-2)} \) is quantile to \( \alpha \) of standard Student distribution\[27\]. There-
fore, if give probability $\alpha$ and an initial investment of $W_0$, the value at risk (VaR) can be calculated using the equation:

$$VaR = -W_0 \times q = -W_0 \left( \mu + \frac{t_v(\alpha)\sigma}{\sqrt{v/(v-2)}} \right)$$

(2.13)

Where $t_v(\alpha)$ is the quantile to $\alpha$ of standard Student distribution with degrees of freedom $v$.

Non-parametric methods

If you are uncertain about the distribution of your data, historical VaR is a better approach. This calculation is even simpler than the parametric VaR calculation because all you need to do is to take all of your past historical returns in the order from the lowest to highest and calculate the lowest return in your history with a predetermined level of confidence. This means that if you had 100 past returns and you wanted to know with 95% confidence what the worst thing you could do, you would go to the 5th data point in your ranking series and know that 95% of the time you could do no worse than that number.

Historical VaR seems too simple, in fact that is the biggest criticism of the method. Without a distribution to help determine future returns, you are assuming that the past will exactly replicate the future, which in itself is highly unlikely. The advantage of the method is that all past data have been fully incorporated into the risk calculation without forcing the assumption of a normal distribution and without the need for a variance/covariance matrix to calculate the standard de-
viation of the portfolio. This avoids the risk of matrix variation over time as described in the weaknesses of the parametric VaR paragraph. Unfortunately, this historical VaR calculation is only as powerful as the number of data points you have available to measure it, and the time it takes to collect the data can prove cumbersome, if not impossible. In theory, this approach is better than parametric VaR if there are enough data to fully represent all the crisis events that occur and the changing business cycle.

**Expected shortfall**

Expected shortfall (ES) is a risk measure—a concept used in the field of financial risk measurement to evaluate the market risk or credit risk of a portfolio. The ”expected shortfall at $q\%$ level” is the expected return on the portfolio in the worst $q\%$ of cases. ES indicates the potential loss if the portfolio is “hit” beyond VaR.

Because ES is an average of the tail loss, one can show that it qualifies as a sub-additive risk measure. ES is an alternative to value at risk that is more sensitive to the shape of the tail of the loss distribution.

**2.2.4 Kupiec’s POF test**

Kupiec’s POF test[28] was the earliest proposed VaR back test. The test is concerned with whether or not the reported VaR is violated more or less than $100\alpha$ percent of the time.
The null hypothesis for the POF-test is

$$H_0 : p = \hat{p} = \frac{x}{T} \quad (2.14)$$

Where $x$ is the number of VaR violations and $T$ is the total number of data observations.

The idea is to find out whether the observed failure rate $\hat{p}$ is significantly different from $p$, the failure rate suggested by the confidence level. According to Kupiec[28], the POF-test is best conducted as a likelihood-ratio (LR) test. The test statistic takes the form

$$LR_{POF} = -2 \ln \left( \frac{(1 - p)^{T-x}p^x}{\left[1 - (\frac{x}{T})\right]^{T-x} \left(\frac{x}{T}\right)^x} \right) \quad (2.15)$$

Under the null hypothesis that the model is correct, $LR_{POF}$ is asymptotically $\chi^2$ (chi-squared) distributed with one degree of freedom. If the value of the $LR_{POF}$ statistic exceeds the critical value of the $\chi^2$ distribution, the null hypothesis will be rejected and the model is deemed as inaccurate.

One advantage of Kupiec’s POF test is that it is simple to implement and use. It is statistically weak if a sample size is not large enough. Also, the test only considers the frequency of losses and not the time when they occur. Therefore, it may fail to reject a model that produces clustered violations.

2.2.5 Stationary time series
There have been much debate and active research on time series analysis for decades, with applications in fields such as economics, finance, medicine, and insurance. Time series involve primarily drawing inferences, understanding how they evolve over time, and replicating motion in some stochastic model. Essentially, time series analysis involves building a hypothetical model to represent the given data, estimating its parameters, and hopefully using the model to understand the data better. Another important goal in this field is to make predictions ahead of the data. Researchers have been using time series analysis as an important tool in trying to predict tangible things like market indexes, financial asset prices and insurance losses etc.

Denote \( \{ x(t) \mid t = \ldots, -1, 0, 1, 2, \ldots \} \) as a time series, in which time \( t \) can take positive and negative integers and zero values. This not only conforms to most real situations, but also facilitates theoretical analysis. Let the mean of \( x(t) \) be \( m(t) = E[x(t)] \), obviously it is a function of \( t \), similarly, the mean of \( x(s) \) is \( m(s) = E[x(s)] \), the covariance is \( \gamma(t, s) = E[x(t) - m(t)][x(s) - m(s)] \), and the correlation coefficient \( \rho(t, s) = \frac{\gamma(t, s)}{\sqrt{\gamma(t, t)\gamma(s, s)}} \) is a bivariate function of \( t \) and \( s \). If its mean \( m(t) \) is a constant value \( m \), \( \gamma(t, s) \) and \( \rho(t, s) \) depends only on the value of \( t - s \), the series is called a stationary time series. Then the correlation coefficient between \( x(t) \) and \( x(t + k) \) has nothing to do with \( t \), so it can be written as \( \gamma(t, s) = \gamma(t - s), \rho(t, s) = \rho(t - s) \).

Stationary time series is the most important special type of time series in analysis. So far, time series analysis is basically based on stationary time series, and
its methods and theories are limited for statistical analysis of non-stationary time series.

2.2.6 Characteristics of financial data

Financial data have unique characteristics relative to other time series data that make modeling more difficult. In this part those characteristics such as high kurtosis, volatility clustering, leverage effects and spillover effects will be discussed.

**High kurtosis**

Data with high kurtosis characteristics will have more data points far from the mean than a distribution with lower kurtosis. In the past, financial theory assumed that returns on financial assets were normally distributed, which proved to be underperforming in finance. Therefore, some distributions with asymmetric or fat-tailed characteristics are used in this field, such as $t$ distribution\[^{29}\], Skewed generalized $t$ distribution\[^{37}\] generalized error distribution\[^{31}\], Skewed Generalized Error Distribution (SGED)\[^{37}\].

**Volatility clustering**

In finance, volatility clustering refers to the observation, first noted by Mandelbrot \[^{33}\], that large changes tend to be followed by large changes, and small changes tend to be followed by small changes. Modeling processes with volatility clustering involves estimating the volatility at any point in time. For this reason, ARCH and GARCH models can be used to model conditional variance.
Leverage effects
The leverage effect describes the difference in the impact of good news and bad news on stock volatility. In general, bad news usually leads to greater volatility in the market than good news. Several models model this asymmetry, including the EGARCH model[31], the GJR-GARCH model[34].

Spillover effects
The spillover effect is when an event in a country has an effect on the economy of another, usually more dependent country. Financial markets often influence each other, and one financial market can influence the performance of another. The spillover effects among advanced economies are more pronounced. There is evidence that the degree of connectivity between developed countries is higher than that between developing countries.

2.3 Volatility models
ARIMA-GARCH is a commonly used model for modeling time series volatility, and it is also a popular model in the field of financial risk management. However, unexpected changes can occur in the market, such as a change from a bull market to a bear market and vice versa. In order to capture the style switching of the market and improve the analytical ability of the model. Markov switching GARCH model may have better forecasting effect when the market condition switches.
2.3.1 ARIMA model

Models for time series data can take many forms and represent different stochastic processes. When modeling variations in the level of a process, three broad classes of practical importance are the autoregressive (AR) models, the integrated (I) models, and the moving average (MA) models. The three categories are linearly dependent on previous data points. The combination of these ideas produces the Autoregressive Moving Average (ARMA) and Autoregressive Integrated Moving Average (ARIMA)[38] models.

ARIMA model has been used widely in many areas of time series analysis.

ARIMA\((p, d, q)\) model where \(p\) is the number of autoregressive (AR) terms, \(d\) is the number of differences taken and \(q\) is the number of moving average (MA) terms. ARIMA models always assume the variance of data to be constant.

ARIMA\((p, d, q)\) model can be represented by the following equation.

If

\[
Y_t = (1 - B)^d X_t
\]

(2.16)
is a sequence of ARMA\((p, q)\), it indicates that \(X_t\) is a sequence of ARMA\((p, q)\) and the model is

\[
\phi(B)(1 - B)^d X_t = \theta(B)\varepsilon_t, t \in \mathbb{Z}
\]

(2.17)

where \(B\) represents the operator, \((1 - B)\) represents finite difference operator, \(\varepsilon_t\) represents the random error with zero-mean, and real polynomial \(\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p\) and \(\theta(z) = \theta_0 + \theta_1 z + \cdots + \theta_q z^q\) meet the requirements of stationarity and reversibility, respectively.
CHAPTER 2. LITERATURE REVIEW & THEORY

The modeling steps of ARIMA($p, d, q$) model are as follows:

1. The stationarity test is carried out on the original time series. If the series is not stationary, the difference transformation is needed to make the series meet the stationarity condition, thereby obtaining the value of $d$ in the model.

2. The values of $p$ and $q$ in the model are determined by using ACF and PACF.

3. The unknown parameters of the model are estimated, and the significance of the parameters and the applicability of the diagnostic model are tested.

2.3.2 GARCH model

Bollerslev[19] proposed an important generalized model of the ARCH model, called the GARCH model. For a logarithmic return series $r_t$, let $a_t = r_t - \mu_t = r_t - E(r_t | F_{t-1})$, \{a_t\} is said to GARCH($m, s$) model, if $a_t$ satisfies

$$a_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = \alpha_0 + \sum_{i=1}^{m} \alpha_i a_{t-i}^2 + \sum_{j=1}^{s} \beta_j \sigma_{t-j}^2$$

where \{\varepsilon_t\} is a i.i.d white noise sequence, $\alpha_0 > 0, \alpha_i \geq 0, \beta_j \geq 0, 0 < \sum_{i=1}^{m} \alpha_i + \sum_{j=1}^{s} \beta_j < 1$

This last condition is used to guarantee that the unconditional variance of $a_t$ of the model is finite and constant, while the conditional variance $\sigma_t^2$ can vary with time $t$.

Next, take GARCH(1,1) as an example to study the properties of the GARCH model. Let $F_{t-1}$ denote the information contained in $a_{t-i}$ and $\sigma_{t-j}$ up to time $t - 1$. 
The model is
\[ a_t = \sigma_t \varepsilon_t, \quad \varepsilon_t \text{ i.i.d. } \mathcal{N}(0,1) \] (2.19)
\[ \sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \]

To calculate the unconditional mean \( Ea_t \), first calculate the conditional expectation
\[ E (a_t | F_{t-1}) = E (\sigma_t \varepsilon_t | F_{t-1}) = \sigma_t E (\varepsilon_t | F_{t-1}) = 0 \] (2.20)
where \( \sigma_t \in F_{t-1} \), \( \varepsilon_t \) and \( F_{t-1} \) are independent. So,
\[ Ea_t = E [E (a_t | F_{t-1})] = 0 \] (2.21)

To calculate the unconditional variance of \( a_t \).
\[
\begin{align*}
\text{Var} (a_t) &= E (a_t^2) = E \left[ E (a_t^2 | F_{t-1}) \right] = E \left[ E (\sigma_t^2 \varepsilon_t^2 | F_{t-1}) \right] \\
&= E \left[ \sigma_t^2 E (\varepsilon_t^2 | F_{t-1}) \right] = E \left[ \sigma_t^2 E (\varepsilon_t^2) \right] \\
&= E \left[ \sigma_t^2 \right] = E \left[ \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \right] \\
&= \alpha_0 + \alpha_1 E (a_{t-1}^2) + \beta_1 E \left( E (a_{t-1}^2 | F_{t-2}) \right) \\
&= \alpha_0 + (\alpha_1 + \beta_1) E (a_{t-1}^2)
\end{align*}
\] (2.22)

Let \( Ea_t^2 = Ea_{t-1}^2 \),
\[
\begin{align*}
\text{Var} (a_t) &= Ea_t^2 = \frac{\alpha_0}{1 - \alpha_1 - \beta_1} \tag{2.23}
\end{align*}
\]

Properties of the GARCH(1,1) model:
First, like the ARCH model, \( a_t \) has volatility clustering, a larger \( a_{t-1} \) or \( \sigma_{t-1}^2 \) makes the conditional variance larger, and thus tends to have larger log returns.
Second, when \( \varepsilon_t \) is a standard normal distribution, \( a_t \) has an unconditional fourth-
order moment under the following conditions:

\[ 1 - 2\alpha_1^2 - (\alpha_1 + \beta_1)^2 > 0 \] (2.24)

Then the excess kurtosis is

\[ \frac{Ea_t^4}{(Ea_t^2)^2} - 3 = \frac{2 [1 - (\alpha_1 + \beta_1)^2 + \alpha_1^2]}{1 - (\alpha_1 + \beta_1)^2 - 2\alpha_1^2} > 0 \] (2.25)

This means that the distribution of \( a_t \) has fat tails. However, even using the conditional \( t \) distribution when modeling real data, the fat-tailed fit of the data may still be insufficient.

Third, GARCH model gives a relatively simple volatility model.

Fourth, GARCH model cannot reflect the leverage effect.

This thesis uses the most common GARCH(1,1) to establish a volatility forecasting model. Taking the GARCH(1, 1) model as an example, we discuss making one-step ahead prediction based on observations up to time \( h \):

\[ \sigma_{h+1}^2 = \alpha_0 + \alpha_1 a_h^2 + \beta_1 \sigma_h^2 \in F_h \] (2.26)

So,

\[ \sigma_h^2(1) = E (\sigma_{h+1}^2 \mid F_h) = \sigma_{h+1}^2 = \alpha_0 + \alpha_1 a_h^2 + \beta_1 \sigma_h^2 \] (2.27)
By \( \sigma_t^2 = \sigma_{t+1}^2 \), we have

\[
\begin{align*}
\sigma_{h+2}^2 &= \alpha_0 + \alpha_1 \sigma_{h+1}^2 + \beta_1 \sigma_h^2 \\
&= \alpha_0 + \alpha_1 \sigma_{h+1}^2 + \beta_1 \sigma_h^2 \\
&= \alpha_0 + (\alpha_1 \sigma_{h+1}^2 + \beta_1) \sigma_h^2 
\end{align*}
\]

(2.28)

So,

\[
\sigma_h^2(2) = E (\sigma_{h+2}^2 \mid F_h) = \alpha_0 + E (\alpha_1 \sigma_{h+1}^2 + \beta_1 \mid F_h) \sigma_h^2 \\
= \alpha_0 + (\alpha_1 + \beta_1) \sigma_h^2(1). 
\]

(2.29)

Similarly, for \( l \geq 2 \) we have

\[
\sigma_{h+l}^2 = \alpha_0 + \alpha_1 \sigma_{h+l-1}^2 + \beta_1 \sigma_h^2 l-1 = \alpha_0 + (\alpha_1 \sigma_{h+l-1}^2 + \beta_1) \sigma_h^2 l-1 
\]

(2.30)

Then

\[
\begin{align*}
\sigma_h^2(l) &= E \{ \sigma_{h+l}^2 \mid F_h \} = \alpha_0 + E \{ (\alpha_1 \sigma_{h+l-1}^2 + \beta_1) \sigma_h^2 l-1 \mid F_h \} \\
&= \alpha_0 + E \{ E [(\alpha_1 \sigma_{h+l-1}^2 + \beta_1) \sigma_h^2 l-1 \mid F_{h+l-2}] \mid F_h \} \\
&= \alpha_0 + E \{ \sigma_h^2 l-1 E [\alpha_1 \sigma_{h+l-1}^2 + \beta_1 \mid F_{h+l-2}] \mid F_h \} \\
&= \alpha_0 + E \{ \sigma_h^2 l-1 (\alpha_1 + \beta_1) \mid F_h \} \\
&= \alpha_0 + (\alpha_1 + \beta_1) \sigma_h^2 (l-1) 
\end{align*}
\]

(2.31)

The prediction formula is the same as the advance prediction formula of ARMA(1,1) with the autoregressive coefficient \((\alpha_1 + \beta_1)\). Iteratively calculated from \( l = 2 \)

\[
\sigma_h^2(l) = \frac{\alpha_0 \left[ 1 - (\alpha_1 + \beta_1)^{l-1} \right]}{1 - (\alpha_1 + \beta_1)} + (\alpha_1 + \beta_1)^{(l-1)} \sigma_h^2(1) 
\]

(2.32)
If $\alpha_1 + \beta_1 < 1$, we have

$$\sigma_n^2(l) \rightarrow \frac{\alpha_0}{1 - \alpha_1 - \beta_1} = \text{Var}(a_t) \quad (2.33)$$

That is, the multi-step conditional variance prediction in advance tends to the unconditional variance of $a_t$.

In order to solve the problems of fat tail and asymmetry, some extended GARCH models are introduced.

**EGARCH**

The exponential general autoregressive conditional heteroskedastic (EGARCH) is another form of the GARCH model. E-GARCH model was proposed by Nelson[31] to overcome the weakness in GARCH handling of financial time series. In particular, to allow for asymmetric effects between positive and negative asset returns.

Formally, an E-GARCH(p,q):

$$x_t = \mu + a_t$$

$$\ln \sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i (|\epsilon_{t-i}| + \gamma_i \epsilon_{t-i}) + \sum_{j=1}^{q} \beta_j \ln \sigma_{t-j}^2$$

$$a_t = \sigma_t \times \epsilon_t$$

$$\epsilon_t \sim P_v(0, 1)$$

(2.34)

where:

$x_t$ is the time series value at time $t$. 
μ is the mean of the GARCH model.

\( a_t \) is the model’s residual at time \( t \).

\( \sigma_t \) is the conditional standard deviation (i.e. volatility) at time \( t \).

\( p \) is the order of the ARCH component model.

\( \alpha_0, \alpha_1, \alpha_2, \ldots, \alpha_p \) are the parameters of the ARCH component model.

\( q \) is the order of the GARCH component model.

\( \beta_1, \beta_2, \beta_3, \ldots, \beta_q \) are the parameters of the GARCH component model.

\( \epsilon_t \) are the standardized residuals:

\[
\epsilon_t \sim i.i.d
\]

\[
E(\epsilon_t) = 0
\]

\[
VAR(\epsilon_t) = 1
\]

\( P_v \) is the probability distribution function for \( \epsilon_t \). Currently, the following distributions are supported:

1. Normal Distribution

\[
P_v = N(0, 1)
\]

(2.35)

2. Student’s t-Distribution

\[
P_v = t_v(0, 1)
\]

(2.36)

3. Generalized Error Distribution (GED)

\[
P_v = GED_v(0, 1)
\]

(2.37)
**GJRGARCH**

Consider a return time series \( r_t = \mu + \varepsilon_t \), where \( \mu \) is the expected return and \( \varepsilon_t \) is a zero-mean white noise. Despite of being serially uncorrelated, the series \( \varepsilon_t \) does not need to be serially independent. For instance, it can present conditional heteroskedasticity. The Glosten-Jagannathan-Runkle GARCH (GJR-GARCH) model\[39\] assumes a specific parametric form for this conditional heteroskedasticity. More specifically, we say that \( \varepsilon_t \sim GJR - GARCH \) if we can write \( \varepsilon_t = \sigma_t z_t \), where \( z_t \) is standard Gaussian and:

\[
\sigma_t^2 = \omega + \sum_{i=1}^{p} (\alpha_i + \gamma_i I_{t-i}) \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 \tag{2.38}
\]

where

\[
I_{t-1} := \begin{cases} 
0 & \text{if } r_{t-1} \geq \mu \\
1 & \text{if } r_{t-1} < \mu 
\end{cases} \tag{2.39}
\]

The best model \((p \text{ and } q)\) can be chosen, for instance, by Bayesian Information Criterion (BIC), also known as Schwarz Information Criterion (SIC), or by Akaike Information Criterion (AIC). The former tends to be more parsimonious than the latter.

There is a fact that the GJR-GARCH model captures that is not contemplated by the GARCH model, which is the empirically observed fact that negative shocks at time \( t - 1 \) have a stronger impact in the variance at time \( t \) than positive shocks. This asymmetry used to be called leverage effect because the increase in risk was
believed to come from the increased leverage induced by a negative shock. Notice that the effective coefficient associated with a negative shock is \( \alpha + \gamma \) and the effective coefficient associated with a positive shock is \( \alpha \). In financial time series, we generally find that \( \gamma \) is statistically significant greater than zero.

**TGARCH**

Another volatility model commonly used to handle leverage effects is the threshold GARCH (or TGARCH) model[40]. A TGARCH(p, q) model assumes the form

\[
\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} (\alpha_i + \lambda_i 1_{t-i}) x_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2
\]  

(2.40)

where

\[
1_{t-1} = \begin{cases} 
1, & \text{if } \epsilon_{t-i} < 0, \text{ bad news} \\
0, & \text{if } \epsilon_{t-i} \geq 0, \text{ good news} 
\end{cases}
\]

(2.41)

\( \alpha_i > 0, \beta_j > 0 \) and \( \lambda_i > 0 \)

where \( \alpha \) and \( \alpha + \lambda \) denote the effect of good news and bad news respectively. \( \lambda > 0 \) is evidence that bad news upsurge volatility in the stock indexes. For an asymmetric news effect, \( \lambda \neq 0 \).

**2.3.3 Markov switching GARCH model**

**Markov chain**

Let \( X_n, n = 0, 1, 2, \ldots \) be a stochastic process.

a) It takes values from a finite or countable set \( S \) (called the state space). Example of countable sets: \( \mathbb{Z} \), etc.
b) Suppose

$$P(X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) = P_{ij}$$

for all $i_0, \ldots, i_{n-1}, i, j$ and n. Such a process is called a Markov chain.

Fact:

$$\sum_{j \in S} P_{ij} = 1, \forall i$$

(2.43)

For the Markov chain, $X_n$, suppose $S = \{0, 1, \ldots\}$.

Define the matrix

$$P = \begin{bmatrix}
P_{00} & P_{01} & \cdots \\
P_{10} & P_{11} & \cdots \\
\vdots & \vdots & \ddots
\end{bmatrix} =: (P_{ij})_{i,j \in S}$$

(2.44)

P is called the one-step transition matrix.

**Hidden Markov model**

Hidden Markov Model[41] is a probability model about time series, which describes the process of randomly generating an unobservable random sequence of states from a hidden Markov chain, and then generating an observation from each state to generate a random sequence of observations. The sequence of states randomly generated by the hidden Markov chain is called the state sequence; each state generates an observation, and the resulting random sequence of observations is called the observation sequence.
The hidden Markov model is determined by the initial probability distribution, state transition probability distribution, and observation probability distribution. The form of the hidden Markov Model is defined as follows: Let $S$ be the set of all possible states and $V$ is the set of all possible observations:

$$S = \{ s_1, s_2, \cdots, s_N \}, \quad V = \{ v_1, v_2, \cdots, v_M \}$$

where $N$ is the number of possible states and $M$ is the number of possible observations. $I$ is the state sequence with length $T$ and $O$ is the corresponding observation sequence.

$$I = (i_1, i_2, \cdots, i_T), \quad O = (o_1, o_2, \cdots, o_T)$$

$A$ is the state transition probability matrix:

$$A = [a_{ij}]_{N \times N}$$

Where $a_{ij} = P( i_{t+1} = s_j \mid i_t = s_i), \quad i = 1, 2, \cdots, N, \quad j = 1, 2, \cdots, N$ is the probability of transitioning to state $s_i$ at time $t+1$ while being in state $s_j$ at time $t$.

$B$ is the observation probability matrix

$$B = [b_j(k)]_{N \times M}$$

where

$$b_j(k) = P( o_t = v_k \mid i_t = s_j), \quad k = 1, 2, \cdots, M; \quad j = 1, 2, \cdots, N \quad (2.45)$$

is the probability of generating an observation $v_k$ given that it is in state $s_j$ at time $t$. $\pi$ is the initial state probability vector: $\pi = (\pi_i)$.

where $\pi_i = P( i_1 = s_i), \quad i = 1, 2, \cdots, N$ is the probability in state $s_i$ at time $t = 1$. 
The hidden Markov model is determined by the initial state probability vector $\pi$, the state transition probability matrix $A$ and the observation probability matrix $B$. $\pi$ and $A$ determine the sequence of states. $B$ determines the observation sequence. Therefore, the hidden Markov model lambda can be represented by a ternary notation $\lambda = (A, B, \pi)$, $A, B, \pi$ are called the elements of the hidden Markov model.

The state transition probability matrix $A$ and the initial state probability vector $\pi$ determine the hidden Markov chain and generate an unobservable state sequence. The observation probability matrix $B$ determines how the observations are generated from the states, and the combination with the state sequence determines how the observation sequence is generated.

From the definition, the hidden Markov model makes two basic assumptions (1) Homogeneous Markov hypothesis, it is assumed that the state of the hidden Markov chain at any time $t$ only depends on the state at time $t - 1$, and has nothing to do with the states and observations at other times.

$$ P(i_t \mid i_{t-1}, o_{t-1}, \ldots, i_1, o_1) = P(i_t \mid i_{t-1}), \quad t = 1, 2, \ldots, T \quad (2.46) $$

(2) The assumption of observation independence, it is assumed that the observation at any time only depends on the state of the Markov chain at that time, and has nothing to do with other observations and states.

$$ P(o_t \mid i_T, o_T, i_{T-1}, o_{T-1}, \ldots, i_{t+1}, o_{t+1}, i_t, i_{t-1}, o_{t-1}, \ldots, i_1, o_1) = P(o_t \mid i_t) \quad (2.47) $$
Markov switching autoregressive model

Markov-switching models (MSM) are a generalization of HMMs and it was first introduced by Hamilton[42]. A major difficulty in modeling financial time series data is that stationarity is often difficult to satisfy. The instability of economic (financial) models can be seen as structural changes in the regression equation in different subsample intervals. If the timing of the structural change is known, it can be verified by the Chow test, or by introducing dummy variables to simulate the change. However, whether it is theoretical research or practical operation, it is often difficult for people to determine the exact moment when the parameters change, and it is necessary to infer the moment when the turning point occurs. Markov switching models can be used to solve this dilemma.

For time series modeling in the financial field, the Markov switching autoregressive model is a commonly used model[43]. This model can be regarded as a combination of several autoregressive models. These autoregressive models are governed by a latent (or hidden) state that transforms into each other. A typical two-state MS-AR(1) model can be described as follows:

\[
y_t = \alpha_{S_t} + \beta y_{t-1} + \varepsilon_t \quad \varepsilon_t \sim N\left(0, \sigma_{S_t}^2\right) \tag{2.48}
\]

where

\[
\begin{aligned}
\alpha_{S_t} &= \alpha_1 \cdot S_{1t} + \alpha_2 \cdot S_{2t} \\
\sigma_{S_t}^2 &= \sigma_1^2 \cdot S_{1t} + \sigma_2^2 \cdot S_{2t}
\end{aligned} \tag{2.49}
\]
If $\alpha = \alpha_1 = \alpha_2$ and $\sigma = \sigma_1 = \sigma_2$, the above two-state MS-AR(1) degenerates into the traditional AR(1) model

$$y_t = \alpha + \beta y_{t-1} + \varepsilon_t \quad \varepsilon_t \sim N(0, \sigma^2)$$

(2.51)

The general form of a first-order Markov switching autoregressive model

$$y_t - \mu_{S_t} = \beta (y_{t-1} - \mu_{S_{t-1}}) + \varepsilon_t \quad \varepsilon_t \sim N(0, \sigma^2_{S_t})$$

(2.52)

This model is also known as the Markov mean-variance switching model.

We use $\mathcal{F}_t = \{y_0, y_1, \cdots, y_t\}$ to refer to all the information we have about the observations up to time $t$, and call it the information set or filtration.

$$y_t \sim N(\mu_{S_t} + \beta (y_{t-1} - \mu_{S_{t-1}}), \sigma^2_{S_t})$$

(2.53)

So,

$$y_t \mid S_{t-1} = i, S_t = j, \mathcal{F}_{t-1} \sim N(\mu_j + \beta (y_{t-1} - \mu_i), \sigma^2_j)$$

(2.54)

is equivalent to

$$f (y_t \mid S_{t-1} = i, S_t = j, \mathcal{F}_{t-1}) = \frac{1}{\sqrt{2\pi \sigma_j}} e^{-(y_t - \mu_j - \beta (y_{t-1} - \mu_i))^2 / 2\sigma^2_j}$$

(2.55)
In the hidden Markov chain model, the observations $X_t$ are independent of each other and are only determined by the state variable $C_t$, that is $X_t \neq f(X_1, X_2, \cdots, X_{t-1})$.

In the Markov switching transition model, the observed value $y_t$ is autocorrelated. In the two-regime MS-AR model, the transition probability matrix is a $2 \times 2$ square matrix

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} p & 1-p \\ 1-q & q \end{bmatrix}$$ (2.56)

**MS-GARCH model**

GARCH models are often used to describe asset return volatility. In practice, due to the adjustment of macro policy or the change of economic structure, the coefficients in the model will change, and the coefficients corresponding to different regimes may be different. The Markov switching model can describe this change. Markov switching GARCH models are mainly used in finance. Other fields of applications include the analysis of business cycles in economics[44].

**Definition 2.1** \(\{y_t : t = 1, 2, \cdots , T\}\) represent time series, \(y_t \in \mathbb{R}\), \(s_t\) represents the regime where the time series \(y_t\) is located. Assuming there are \(k (k \geq 1)\) regimes, \(s_t \in (1, 2, \cdots, k)\). If \(y_t\) satisfies the following formula

$$y_t = \mu_{s_t} + a_{s_t}$$ (2.57)

\(\mu_{s_t}\) is the mean equation truncated to order \(q\) for \(y_t\) in state \(s_t\)

$$\mu_{s_t} = \phi_{s_t0} + \phi_{s_t1}y_{t-1} + \cdots + \phi_{s_tq}y_{t-q}$$ (2.58)
\( a_{s,t} \) represents the variance equation of \( y_t \) in state \( s_t \), which satisfies the following GARCH model

\[
a_{s,t} = h_{s,t} \varepsilon_t, \quad h_{s,t}^2 = \alpha_{s,t}0 + \sum_{i=1}^{m} \alpha_{s,t}a_{s,t(i-1)}^2 + \sum_{j=1}^{n} \beta_{s,t}h_{s,t(j-1)}^2 \quad (2.59)
\]

where \( \varepsilon_t \) is i.i.d and \( \varepsilon_t \sim N(0,1) \), \( \alpha_{s,t} \geq 0 \), \( \beta_{s,t} \geq 0 \), \( \sum_{l=1}^{\max(m,n)} (\alpha_{s,l} + \beta_{s,l}) < 1 \).

Then this time series model is called a \( k \)-regimes Markov switching GARCH model, denoted as MS\((k)\)-GARCH model.

The parameter estimates for this model are described below. The parameters, which need to be estimated, are based on the observation sequence to estimate the corresponding most probable regime sequence (unobserved sequence), and then the probability transition matrix is estimated, and the parameters of the GARCH model corresponding to each regime.

Note that the parameter to be estimated in the mean equation is \( \theta_1 \), the parameter of the variance equation is \( \theta_2 \), and note \( \theta = (\theta'_1, \theta'_2)' \). Markov transition probability \( p = (p_{11}, p_{12}, \cdots, p_{kk})' \) is a \( k^2 \times 1 \) vector. The conditional density function of \( y_t \) can be expressed as

\[
f(y_t \mid s_t, s_{t-1}, \cdots, y_{t-1}, y_{t-2}, \cdots ; \theta) = f(y_t \mid s_t, s_{t-1}, \cdots, s_{t-q}, y_{t-1}, y_{t-2}, \cdots, y_{t-q}; \theta) \equiv f(y_t \mid z_t ; \theta) \quad (2.60)
\]

where

\[
z_t = (s_t, s_{t-1}, \cdots, s_{t-q}, y_{t-1}, y_{t-2}, \cdots, y_{t-q})
\]
The sequence \( \{y_t\} \) is the observed data, and \( s_t \) is unobserved, which we can infer \( s_t \) from \( y_t \). The maximum likelihood method can be used to estimate \( s_t \). Since the model involves the problem of nonlinear estimation and the calculation is difficult, the EM algorithm can be used for estimation in practice. Here, this thesis mainly introduces the estimation method of the state transition matrix and model parameters based on the premise that \( s_t \) is known.

The maximum likelihood estimation method is to make the likelihood function 
\[
f(y_T, y_{T-1}, \cdots, y_1 \mid p, \theta)
\]
reach the maximum to find the parameter vector \( (p', \theta')' \).

For the convenience of calculation, when \( q \)-order autoregression is involved, the conditional likelihood function 
\[
f(y_T, y_{T-1}, \cdots, y_{q+1} \mid y_q, y_{q-1}, \cdots, y_1)
\]
replaces the original likelihood function 
\[
f(y_T, y_{T-1}, \cdots, y_1)
\]. Note that the probability corresponding to the \( q \) states \( \{s_1, s_2, \cdots, s_q\} \) at the initial \( \{y_1, y_2, \cdots, y_q\} \) is

\[
p(s_q, s_{q-1}, \cdots, s_1 \mid y_q, y_{q-1}, \cdots, y_1) = \rho_{s_q, s_{q-1}, \cdots, s_1}
\]  

(2.61)

\( \rho_{s_q, s_{q-1}, \cdots, s_1} \) also need to be estimated. The different values of \( (s_1, s_2, \cdots, s_q) \) form a vector

\[
\rho = (\rho_{1,1}, \cdots, 1, \rho_{1,1}, \cdots, 2, \cdots, \rho_{k,k}, \cdots, k)'
\]  

(2.62)

when \( q=1 \)

\[
\rho = (\rho_1, \rho_2, \cdots, \rho_k) = \{p(s_1 = 1 \mid y_1), p(s_1 = 2 \mid y_1), \cdots, p(s_1 = k \mid y_1)\}
\]  

(2.63)
The sum of the elements in $\rho$ is 1. $\rho$ represents the probability at the initial moment, and can be estimated by maximum likelihood estimation together with $p, \theta$. The parameter to be estimated is represented by a vector $\lambda$, denoted by $\lambda \equiv (p', \theta', \rho')'$.

If the regimes $\{s_t\}$ at each moment of $\{y_t\}$ is known, then the conditional likelihood function corresponding to the model can also be determined. Once the conditional likelihood function is determined, the maximum likelihood estimate of the parameters can be obtained by making the first-order partial derivative equal to zero. In the parameter estimation in this section, it is also assumed that the state $S$ at all times $\{y_t\}$ is known (or an arbitrary set of numbers), and MLE $\hat{\lambda}$ is obtained by making the first-order partial derivative of the likelihood function equal to zero.

First, assuming that $S$ is known, let the first-order partial derivative of the log-likelihood function be equal to zero.

$$\frac{\partial \log p(Y, S \mid \lambda)}{\partial \lambda} \bigg|_{\lambda=\hat{\lambda}(S)} = 0 \quad (2.64)$$

An estimate of $\lambda$, $\hat{\lambda}(S)$ can be obtained. After $l$ iterations, after estimating $\hat{\lambda}_l$, $S$ can be estimated again by the following formula:

$$p(S \mid Y, \hat{\lambda}_l) = \frac{p(Y, S \mid \hat{\lambda}_l)}{p(Y \mid \hat{\lambda}_l)} \quad (2.65)$$
Iterate continuously until the change between $\hat{\lambda}_{t+1}$ and $\hat{\lambda}_t$ is small, or the difference between the two is less than a certain critical value. In this way, the maximum likelihood estimate of the parameter $\hat{\lambda}$ is obtained.

**Estimation of parameter $\lambda$**

The probability density function of $Y, S$ can be decomposed into the following form:

$$p(Y, S \mid \lambda) = p(y_T \mid z_T; \theta) \cdot p(s_T \mid s_{T-1}; p) \cdot p(y_{T-1} \mid z_{T-1}; \theta) \cdot p(s_{T-1} \mid s_{T-2}; p) \cdots (2.66)$$

First, estimate the transition probability $p$. Take the partial derivative with respect to $p(Y, S \mid \lambda)$

$$\frac{\partial p(Y, S \mid \lambda)}{\partial p_{ij}} = \sum_{t=p+1}^{T} \frac{\partial \log p(s_t \mid s_{t-1}; p)}{\partial p_{ij}} \cdot p(Y, S \mid \lambda) \quad (2.67)$$

$$\frac{\partial \log p(s_t \mid s_{t-1}; p)}{\partial p_{ij}} = \begin{cases} \frac{1}{p_{ij}}, & s_t = j, s_{t-1} = i \\ 0, & \text{otherwise.} \end{cases} \quad (2.68)$$

Denote the indicative function by $\delta$ ($\delta_{[A]} = 1$ when event $A$ occurs, otherwise $\delta_{[A]} = 0$), then we have

$$\frac{\partial p(Y, S \mid \lambda)}{\partial p_{ij}} = \frac{1}{p_{ij}} \left\{ \sum_{t=p+1}^{T} \delta_{[s_t = j; s_{t-1} = i]} \right\} \cdot p(Y, S \mid \lambda) \quad (2.69)$$
Equivalently,

\[
\frac{\partial \log p(Y, S | \lambda)}{\partial p_{ij}} = \frac{1}{p_{ij}} \left\{ \sum_{t=p+1}^{T} \delta_{[s_t=j, s_{t-1}=i]} \right\}
\]  \hspace{1cm} (2.70)

let \( Q(\lambda_{t+1} | \lambda_t, Y) \) is the expectation of the log-likelihood function

\[
Q(\lambda_{t+1} | \lambda_t, Y) = \int_S \log p(Y, S | \lambda_{t+1}) \cdot p(Y, S | \lambda_t)
\]  \hspace{1cm} (2.71)

Take the derivative with respect to \( Q(\cdot) \), that is

\[
\frac{\partial Q(\lambda_{t+1} | \lambda_t, Y)}{\partial p_{ij}^{(t+1)}} = \int_S \frac{\partial \log p(Y, S | \lambda_{t+1})}{\partial p_{ij}^{(t+1)}} \cdot p(Y, S | \lambda_t)
\]  \hspace{1cm} (2.72)

and

\[
\int_S \delta_{[s_t=j, s_{t-1}=i]} p(Y, S | \lambda_t) = p(s_t = j, s_{t-1} = i | S; \lambda_t) \cdot p(Y | \lambda_t)
\]  \hspace{1cm} (2.73)

Therefore

\[
\frac{\partial Q(\lambda_{t+1} | \lambda_t, Y)}{\partial p_{ij}^{(t+1)}} = \frac{1}{p_{ij}^{(t+1)}} \sum_{t=p+1}^{T} p(s_t = j, s_{t-1} = i | S; \lambda_t) \cdot p(Y | \lambda_t).
\]  \hspace{1cm} (2.74)

The purpose of maximum likelihood estimation is to find \( p_{t+1} \) by maximizing \( Q(\lambda_{t+1} | \lambda_t, Y) \).

Consider constraints
CHAPTER 2. LITERATURE REVIEW & THEORY

\[ \sum_{j=1}^{K} p_{ij}^{(l+1)} = 1 \]

From Lagrange multipliers, construct the following formula

\[ Q(\bm{\lambda}_{l+1} \mid \bm{\lambda}_l, Y) - \varphi_i \left( \sum_{j=1}^{k} p_{ij}^{(l+1)} - 1 \right) \quad (2.75) \]

Then take its partial derivative with respect to \( p_{ij} \) and make it equal to zero, i.e.

\[ \frac{\partial Q(\bm{\lambda}_{l+1} \mid \bm{\lambda}_l, Y)}{\partial p_{ij}^{(l+1)}} = \varphi_i, \quad j = 1, \ldots, k \quad (2.76) \]

Substituting equation (2.74) into equation (2.76), we have

\[ \sum_{t=q+1}^{T} p(s_t = j, s_{t-1} = i \mid Y; \lambda_l) = \frac{p_{ij}^{(l+1)} \varphi_i}{p(Y \mid \lambda_l)} \quad (2.77) \]

Summing (2.77) for \( j = 1, 2, \ldots, k \), we have

\[ \sum_{t=q+1}^{T} \sum_{j=1}^{k} p(s_t = j, s_{t-1} = i \mid Y; \lambda_l) = \sum_{j=1}^{k} p_{ij}^{(l+1)} \varphi_i / p(Y \mid \lambda_l) \quad (2.78) \]

Since \( \sum_{j=1}^{K} p_{ij}^{(l+1)} = 1 \),

\[ \sum_{t=q+1}^{T} p(s_{t-1} = i \mid S; \lambda_l) = \varphi_i / p(Y \mid \lambda_l) \quad (2.79) \]

Substituting equation (2.79) into equation (2.78), we get

\[ p_{ij}^{(l+1)} = \frac{\sum_{t=q+1}^{T} p(s_t = j, s_{t-1} = i \mid S; \lambda_l)}{\sum_{t=q+1}^{T} p(s_{t-1} = i \mid S; \lambda_l)}, \quad i, j = 1, \ldots, k \quad (2.80) \]
CHAPTER 2. LITERATURE REVIEW & THEORY

Then estimate the parameter $\theta$.

First ignore the GARCH effect (variance equation), then let $\lambda^1 = (p, \theta(1), \rho)$, estimate the parameter $\theta_1$, and write equation (2.66) as the following form:

$$p(Y, S \mid \lambda^1)$$

$$= p(y_t \mid z_t; \theta_1) \cdot p(s_t \mid s_{t-1}; p) \cdot p(y_{T-1} \mid z_{T-1}; \theta_1) \cdot p(s_{T-1} \mid s_{T-2}; p) \cdots$$

$$\cdot p(y_{q+1} \mid z_{q+1}; \theta_1) \cdot p(s_{q+1} \mid s_q; p) \cdot \rho_{s_q, s_{q-1}, \ldots, s_1}.$$  

(2.81)

By taking the partial derivative of the parameter $\theta_1$ from equation (2.81), we can get

$$\frac{\partial p(Y, S \mid \lambda^1)}{\partial \theta_1} = \sum_{t=q+1}^{T} \frac{\partial \log p(y_t \mid z_t; \theta_1)}{\partial \theta_1} \cdot p(Y, S \mid \lambda^1)$$  

(2.82)

In the model, the autoregressive terms are truncated to order $q$, i.e. $y_t$ is at most related to $y_{t-q}$ with lags of order $q$, so

$$\frac{\partial Q(\lambda^1_{t+1} \mid \lambda^1_1, Y)}{\partial \theta_1^{(l+1)}}$$

$$= \int_S \frac{\partial \log p(Y, S \mid \lambda^1_{t+1})}{\partial \theta_1^{(l+1)}} \cdot p(Y, S \mid \lambda^1_1)$$

$$= \sum_{t=q+1}^{T} \int_S \frac{\partial \log p(y_t, z_t \mid \theta_1^{(l+1)})}{\partial \theta_1^{(l+1)}} \cdot p(Y, S \mid \lambda^1_1)$$

(2.83)

$$= \sum_{t=q+1}^{T} \sum_{s_{t-1}=1}^{k} \sum_{s_{t-m}=1}^{k} \cdots \sum_{s_{t-m}=1}^{k} \left\{ \frac{\partial \log p(y_t, z_t \mid \theta_1^{(l+1)})}{\partial \theta_1^{(l+1)}} \right\} \cdot p(s_t, s_{t-1}, \ldots, s_{t-m} \mid Y, \lambda^1_1) \cdot p(Y \mid \lambda^1_1),$$
CHAPTER 2. LITERATURE REVIEW & THEORY

Note that \( p(Y | \lambda_1) \) has nothing to do with the state, so it can be taken outside the summation symbol. Let the above formula be equal to 0, we can get

\[
\sum_{t=q+1}^{T} \sum_{s_{t-q}=1}^{k} \sum_{s_{t-1}=1}^{k} \frac{\partial \log p(y_t | z_t; \theta_1)}{\partial \theta_1} \bigg|_{\theta_1 = \theta_1^{(t+1)}} \cdot p(s_t, s_{t-1}, \ldots, s_{t-q} | Y; \lambda_1) = 0
\]

After obtaining \( \theta_1 \) in the recursion of each step, the residual of each iteration can be calculated. \( \theta_2 \) can be estimated, and then \( \theta \) can be obtained.

Finally, taking the partial derivative of the parameter \( \rho \) in equation (2.66), we get

\[
\frac{\partial \log p(Y, S | \lambda)}{\partial \rho_{i_q,i_{q-1},\ldots,i_1}} = [\rho_{i_q,i_{q-1},\ldots,i_1}]^{-1} \cdot \delta_{[s_q = i_q, s_{q-1} = i_{q-1}, \ldots, s_1 = i_1]} \cdot p(Y, S | \lambda) .
\]

Integrate \( S \) in equation (2.85)

\[
\frac{\partial Q(\lambda_{t+1} | \lambda_t, Y)}{\partial \rho_{i_q,i_{q-1},\ldots,i_1}} = \int_S \left[ \rho_{i_q,i_{q-1},\ldots,i_1}^{t+1} \right]^{-1} \cdot \delta_{[s_q = i_q, s_{q-1} = i_{q-1}, \ldots, s_1 = i_1]} \cdot p(Y, S | \lambda) .
\]

let

\[
\frac{\partial Q(\lambda_{t+1} | \lambda_t, Y)}{\partial \rho_{i_q,i_{q-1},\ldots,i_1}} = \varphi
\]

From equation (2.86)

\[
p(s_q = i_q, s_{q-1} = i_{q-1} \ldots s_1 = i_1 | Y, \lambda_t) \cdot p(Y, \lambda_t) = \varphi \cdot \rho_{i_q,i_{q-1},\ldots,i_1}^{t+1} .
\]

Summing all the values of \( i_q, i_{q-1}, \ldots, i_1 \) in equation (2.88), we have

\[
\rho_{i_q,i_{q-1},\ldots,i_1}^{t+1} = 1
\]
So,
\[ p(Y, \lambda_l) = \varphi \]  

(2.90)

Substitute equation (2.90) into equation (2.88) to get the recursive expression for the parameter \( \rho \)

\[ \rho^{(l+1)}_{i_q, i_{q-1}, \ldots, i_1} = p(s_q = i_q, s_{q-1} = i_{q-1}, \ldots, s_1 = i_1 \mid S; \lambda_l) \]

\[ i_1, \ldots, i_q = 1, \ldots, k \]  

(2.91)

### 2.3.4 AIC and BIC criteria

Multiple statistical models have been proposed for data analysis. In order to evaluate the quality of statistical models, different criteria could be used. Akaike information criterion (AIC) and Bayesian information criterion (BIC) are two of the widely used criteria to measure the relative quality of statistical models.

The AIC and BIC are formally defined as

\[ AIC = -2 \log(\hat{L}) + 2K \]

\[ BIC = -2 \log(\hat{L}) + \log(N)K \]

\( \hat{L} \) is the maximized likelihood.

\( K \) is the number of parameters to be estimated in the model.

\( N \) is the overall sample size.

Some guidance to select a good model by AIC BIC method.

1. The better model has a smaller value of AIC, BIC.
2. BIC criterion may be more reasonable that AIC criterion.
Chapter 3

Forecasting Models & Results

3.1 Data collection and preprocessing

We selected the securities indexes of four major developed economies in the past decade (from May 1, 2012 to April 30, 2022). These indexes are the S&P/TSX Composite, S&P500, Nikkei225 and FTSE100. In February and March 2020, as COVID-19 became a pandemic around the world, it caused rare capital market shocks, especially in March 2020, the U.S. stock market triggered a market wide circuit breaker four times in eight consecutive trading days, which is an unprecedented situation. This poses a significant challenge for financial risk management, so we attempt to estimate the crucial risk management indicator VaR (value at risk) using GARCH model and Markov switching GARCH model, and test the estimation effect using out-of-sample data and compare the estimated effect. It is worth noting whether the VaR estimated by the model still has a good predictive ability for abnormal fluctuations during the epidemic.
Time Series of securities indexes
The time series chart of the four securities indexes in the past ten years is as follows:

![Figure 3.1: Equity market indexes for four different economies](image)

The vertical axis of the line chart is the value of the stock index, and the horizontal axis is the number of days, with May 1, 2012 as the first day. It can be seen that the stock indexes of the four economies generally showed an upward trend, but all of them experienced a severe decline in early 2020, and the decline was the largest in a decade.

Converting the data into a histogram and comparing it with the normal distribution curve, we can see that the time series data is not normally distributed. In the
field of financial risk management, assuming that the data is normally distributed often leads to an underestimation of financial risk.

As shown in the Figure 3.2, the horizontal axis is the value of the stock index, the vertical axis is the frequency, and the curve is a normal distribution curve. It can be seen that the values of these four securities indices may not satisfy the normal distribution.

Different capital markets often influence each other and it is obvious that the Canadian stock market must be greatly affected by the US stock market. In particular, the correlation between capital markets in advanced economies tends to be stronger. It is worth mentioning that we can see that S&P/TSX, S&P500 and
Nikkei225 are strongly correlated with each other. The correlation between the FTSE100 and these three securities indexes is relatively low.

**Augmented Dickey-Fuller Test**

Before building an ARIMA model for a time series, the time series is required to be stationary. The Augmented Dickey-Fuller Test is used first to determine whether the time series is stationary. As can be seen from figure 3.1, the stock indexes are all showing an upward trend, and the time series should not be stationary.

The stationarity test results of the original data are as follows:

Augmented Dickey-Fuller Test

S&P/TSX: Dickey-Fuller = -2.6902, p-value = 0.2861
S&P500: Dickey-Fuller = -2.3278, p-value = 0.4395
Nikkei225: Dickey-Fuller = -3.0833, p-value = 0.1197
FTSE100: Dickey-Fuller = -3.224, p-value = 0.08376

alternative hypothesis: stationary

The Augmented Dickey-Fuller test was performed on the original data of these four stock indexes, and all of the P value were greater than 0.05, so the null hy-
Hypothesis that the time series was not stationary could not be rejected.

In general, the original data can be made stationary by taking logarithms or first-order differences. In this thesis, the daily return of stock indexes is used to test the stationarity and establish the model.

The daily return of a stock index is defined as follows:

\[ r_t = \frac{\text{price}_t - \text{price}_{t-1}}{\text{price}_{t-1}} \]

Figure 3.3: Daily returns of these stock indexes and their distribution

From Figure 3.3 the time series plot of the daily return shows that the series may be stationary.

The stationarity test results of the daily return of stock indexes are as follows:
Augmented Dickey-Fuller Test

S&P/TSX \text{\_return}: Dickey-Fuller = -13.485, p-value < 0.01

S&P500 \text{\_return}: Dickey-Fuller = -13.79, p-value < 0.01

Nikkei225 \text{\_return}: Dickey-Fuller = -13.752, p-value < 0.01

FTSE100 \text{\_return}: Dickey-Fuller = -14.433, p-value < 0.01

Alternative hypothesis: stationary

The Augmented Dickey-Fuller Test is performed on the daily returns of these four stock indexes, rejecting the null hypothesis that the time series is not stationary. alternative hypothesis: stationary

The reason why this thesis chooses to model the daily rate of return, rather than taking the logarithm or first-order difference of the original data is because the daily rate of return is more realistic and does not affect the conclusions of the model.

Preliminary analysis of the daily return data. It is worth mentioning that the data are skewed and leptokurtic (the kurtosis is all well above 3).

<table>
<thead>
<tr>
<th>Securities Index</th>
<th>Min</th>
<th>Median</th>
<th>Mean</th>
<th>Max</th>
<th>std</th>
<th>skewness</th>
<th>kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>S&amp;P/TSX</td>
<td>-0.1234469</td>
<td>0.0007379</td>
<td>0.0002504</td>
<td>0.1195707</td>
<td>0.009206714</td>
<td>-1.15427</td>
<td>46.54899</td>
</tr>
<tr>
<td>S&amp;P500</td>
<td>-0.1198405</td>
<td>0.0005529</td>
<td>0.0004789</td>
<td>0.0938277</td>
<td>0.01047821</td>
<td>-0.62771</td>
<td>21.72051</td>
</tr>
<tr>
<td>Nikkei225</td>
<td>-0.0792156</td>
<td>0.0007573</td>
<td>0.0005193</td>
<td>0.0803810</td>
<td>0.01323683</td>
<td>-0.1325984</td>
<td>7.127068</td>
</tr>
<tr>
<td>FTSE100</td>
<td>-0.1087446</td>
<td>0.0005420</td>
<td>0.0001531</td>
<td>0.0905346</td>
<td>0.009948335</td>
<td>-0.6360464</td>
<td>14.75221</td>
</tr>
</tbody>
</table>

Table 3.2: Preliminary analysis of index daily returns
3.2 ARIMA-GARCH model

The optimal ARIMA model is chosen using the AIC BIC criterion (selecting the model with minimum AIC and BIC value), and the outcomes are as follows:

<table>
<thead>
<tr>
<th>index</th>
<th>optimal ARIMA model</th>
</tr>
</thead>
<tbody>
<tr>
<td>S&amp;P/TSX</td>
<td>ARIMA(5,0,2)</td>
</tr>
<tr>
<td>S&amp;P500</td>
<td>ARIMA(4,0,4)</td>
</tr>
<tr>
<td>Nikkei225</td>
<td>ARIMA(2,0,0)</td>
</tr>
<tr>
<td>FTSE100</td>
<td>ARIMA(4,0,3)</td>
</tr>
</tbody>
</table>

Table 3.3: Optimal ARIMA model for stock indexes

Then we analyze the residuals of the ARIMA model.

As shown in the Figure 3.4, the vertical axis is the square of the residual of the ARIMA model, and the horizontal axis is the number of days. From the squared
residual plot of the ARIMA model, it can be clearly seen that the residuals have obvious heteroscedasticity. In particular, the S&P/TSX, S&P500 and FTSE100 indexes show unusually high squared residuals at the same time. This is precisely when COVID-19 started spreading in these countries.

According to the AIC criterion, the optimal GARCH models for the four residuals are all GARCH (1,1) models. Then we estimate VaR using the Garch(1,1) model.

Figure 3.5: 95%VaR predicted by ARIMA-GARCH model and delta-normal method

Red line denotes 95%VaR produced by ARIMA-GARCH model and green line refers to delta-normal 95% VaR. A grey dot in the graph represents the one-day return. It can be seen that when the daily rate of return falls, the VaR predicted by
the GARCH model will also be adjusted simultaneously to avoid underestimating the risk.

After modelling the volatility, it can be seen that the ARIMA-GARCH model has a good ability to fit VaR. The VaR estimated by the model will react to the risk when the volatility is high.

Next, we check the model’s capacity for prediction. In order to confirm the model’s capacity for prediction, we first use the last 300 trading days as out-of-sample data. Then we included the largest risk fluctuation in the past decade due to the epidemic at the beginning of 2020 as out-of-sample data to verify the predictive ability of the model. We use the last 800 trading days as out-of-sample data. This will pose a greater challenge to the VaR estimated by the model. Finally, we introduce Kupiec’s POF test to perform a back test on the VaR prediction effect.

According to the definition of VaR, when the prediction effect of VaR for recently 300 trading days is tested, the optimal value is that the daily return of 15 trading days is lower than 95% VaR, and the daily return of 3 trading days is lower than 99% VaR. When testing the prediction effect of VaR for recently 800 trading days, the optimal value is that the daily return of 40 trading days is lower than 95% VaR, and the daily return of 8 trading days is lower than 99% VaR. If the value of the out-of-sample daily rate of return is less than the estimated value of VaR, the number is lower than the optimal value, which means that the estimated VaR is too low and the risk is overestimated, and vice versa.
Kupiec’s POF test helps us determine whether the observed failure rate \( \hat{p} \) (The number of days in which the daily return of the stock index is worse than the predicted VaR divided by the total number of days) is significantly different from \( p \) (\( p = 5\% \) or \( 1\% \)), the failure rate suggested by the confidence level.

The test results of the VaR predicted by the ARIMA-GARCH model against out-of-sample data are as follows:

<table>
<thead>
<tr>
<th>Conditional dist</th>
<th>S&amp;P/TSX 95% VaR</th>
<th>kupiec test for 95% VaR</th>
<th>99% VaR</th>
<th>kupiec test for 99% VaR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>15</td>
<td>accurate</td>
<td>7</td>
<td>accurate</td>
</tr>
<tr>
<td>Student-t</td>
<td>17</td>
<td>accurate</td>
<td>7</td>
<td>accurate</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Conditional dist</th>
<th>S&amp;P500 95% VaR</th>
<th>kupiec test for 95% VaR</th>
<th>99% VaR</th>
<th>kupiec test for 99% VaR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>24</td>
<td>not accurate</td>
<td>10</td>
<td>not accurate</td>
</tr>
<tr>
<td>Student-t</td>
<td>29</td>
<td>not accurate</td>
<td>4</td>
<td>accurate</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Conditional dist</th>
<th>Nikkei225 95% VaR</th>
<th>kupiec test for 95% VaR</th>
<th>99% VaR</th>
<th>kupiec test for 99% VaR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>21</td>
<td>accurate</td>
<td>8</td>
<td>accurate</td>
</tr>
<tr>
<td>Student-t</td>
<td>26</td>
<td>not accurate</td>
<td>4</td>
<td>accurate</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Conditional dist</th>
<th>FTSE100 95% VaR</th>
<th>kupiec test for 95% VaR</th>
<th>99% VaR</th>
<th>kupiec test for 99% VaR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>14</td>
<td>accurate</td>
<td>10</td>
<td>not accurate</td>
</tr>
<tr>
<td>Student-t</td>
<td>16</td>
<td>accurate</td>
<td>8</td>
<td>accurate</td>
</tr>
</tbody>
</table>

Table 3.4: Backtest results of VaR estimated by ARIMA-GARCH model for recent 300 trading days

We select the normal distribution and the \( t \) distribution to establish the ARIMA-GARCH model. By comparing the actual value of the out-of-sample data with the estimated value of VaR. We counted the number of days when the actual daily return was lower than the estimated VaR value, and then used the kupiec test to help judge whether the gap from the optimal number of days was acceptable.
CHAPTER 3. FORCASTING MODELS & RESULTS

After forecasting 95% VaR and 99% VaR for the recent 300 trading days, then back test the VaR. It can be found that the number of days with daily return lower than the predicted VaR is greater than the theoretical optimal value. This means that the VaR predicted by ARIMA-GARCH generally underestimates the risk. Among them, only the prediction results of the GARCH model of S&P500 did not pass the kupiec test. Specifically, using a GARCH model with $t$ distribution to estimate 95% VaR for 300 trading days, the number of days with daily returns below the predicted 95% VaR is 29. Through the Kupiec test, it is considered that $\hat{p} = \frac{29}{300}$ is significantly different from 5%, so the model underestimates the actual risk. For 99% VaR, the number of days with daily returns below the predicted 99% VaR is 4. $\hat{p} = \frac{4}{300}$ is considered not significantly different from 1%, so the predicted 99% VaR passes the Kupiec test.

Next, the data of last 800 trading days were included out of sample and remodeled to test the prediction results of the ARIMA-GARCH model. Back test results of estimated 95% VaR and 99% VaR in recent 800 trading days.

Incorporating the dramatic fluctuations caused by the spread of COVID-19 into the forecast range, it can be found that the ARIMA-GARCH model is basically unable to predict VaR. The model has severely underestimated risk, and only the S&P/TSX index passes the kupiec test. For example, using the GARCH model with normal distribution to estimate 95% VaR for 800 trading days, the number of days with daily returns below the predicted 95% VaR is 61. Through the Kupiec test, it is considered that $\hat{p} = \frac{61}{800}$ is significantly different from 5%, so it cannot
### Table 3.5: Backtest results of VaR estimated by ARIMA-GARCH model for recent 800 trading days

<table>
<thead>
<tr>
<th>Conditional dist</th>
<th>S&amp;P/TSX</th>
<th>S&amp;P500</th>
<th>Nikkei225</th>
<th>FTSE100</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>95% VaR</td>
<td>kupiec test for 95% VaR</td>
<td>99% VaR</td>
<td>kupiec test for 99% VaR</td>
</tr>
<tr>
<td>Normal</td>
<td>48</td>
<td>accurate</td>
<td>22</td>
<td>not accurate</td>
</tr>
<tr>
<td>Student-t</td>
<td>50</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>95% VaR</td>
<td>kupiec test for 95% VaR</td>
<td>99% VaR</td>
<td>kupiec test for 99% VaR</td>
</tr>
<tr>
<td>Normal</td>
<td>61</td>
<td>not accurate</td>
<td>31</td>
<td>not accurate</td>
</tr>
<tr>
<td>Student-t</td>
<td>67</td>
<td>not accurate</td>
<td>16</td>
<td>accurate</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>95% VaR</td>
<td>kupiec test for 95% VaR</td>
<td>99% VaR</td>
<td>kupiec test for 99% VaR</td>
</tr>
<tr>
<td>Normal</td>
<td>50</td>
<td>accurate</td>
<td>22</td>
<td>not accurate</td>
</tr>
<tr>
<td>Student-t</td>
<td>54</td>
<td>not accurate</td>
<td>8</td>
<td>accurate</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>95% VaR</td>
<td>kupiec test for 95% VaR</td>
<td>99% VaR</td>
<td>kupiec test for 99% VaR</td>
</tr>
<tr>
<td>Normal</td>
<td>49</td>
<td>accurate</td>
<td>30</td>
<td>not accurate</td>
</tr>
<tr>
<td>Student-t</td>
<td>49</td>
<td>accurate</td>
<td>21</td>
<td>not accurate</td>
</tr>
</tbody>
</table>

Passed the Kupiec test. For 99% VaR, the number of days with daily returns below the predicted 99% VaR is 31. \( \hat{p} = \frac{31}{800} \) is considered significantly different from 1%, so the model seriously underestimates the actual risk. This may be due to the high volatility in the first half of 2020 resulting in a large number of daily returns lower than the predicted VaR.

So, we can try to solve the above problems by introducing Markov switching GARCH model and introducing more conditional distributions. It is hoped that even in extreme cases, a model that still has a good prediction effect on the risk management indicator VaR can be found.
3.3 Markov switching GARCH model

We introduce the Markov switching GARCH model with several conditional volatility models and conditional distributions. The number of regimes equals two. This is consistent with the habit of people in finance to divide the capital market into bull markets and bear markets.

Regime=2

S&P/TSX

<table>
<thead>
<tr>
<th>Volatility models</th>
<th>Distributions</th>
<th>95% VaR</th>
<th>kupiec test for 95% VaR</th>
<th>99% VaR</th>
<th>kupiec test for 99% VaR</th>
</tr>
</thead>
<tbody>
<tr>
<td>GARCH</td>
<td>Normal</td>
<td>13</td>
<td>accurate</td>
<td>3</td>
<td>accurate</td>
</tr>
<tr>
<td>GARCH</td>
<td>Student-t</td>
<td>14</td>
<td>accurate</td>
<td>1</td>
<td>accurate</td>
</tr>
<tr>
<td>GARCH</td>
<td>Skewed normal</td>
<td>14</td>
<td>accurate</td>
<td>4</td>
<td>accurate</td>
</tr>
<tr>
<td>GARCH</td>
<td>Skewed Student-t</td>
<td>13</td>
<td>accurate</td>
<td>1</td>
<td>accurate</td>
</tr>
<tr>
<td>GARCH</td>
<td>Skewed GED</td>
<td>14</td>
<td>accurate</td>
<td>4</td>
<td>accurate</td>
</tr>
<tr>
<td>TGARCH</td>
<td>Normal</td>
<td>9</td>
<td>accurate</td>
<td>0</td>
<td>accurate</td>
</tr>
<tr>
<td>TGARCH</td>
<td>Student-t</td>
<td>9</td>
<td>accurate</td>
<td>0</td>
<td>accurate</td>
</tr>
<tr>
<td>TGARCH</td>
<td>GED</td>
<td>11</td>
<td>accurate</td>
<td>1</td>
<td>accurate</td>
</tr>
<tr>
<td>TGARCH</td>
<td>Skewed normal</td>
<td>11</td>
<td>accurate</td>
<td>1</td>
<td>accurate</td>
</tr>
<tr>
<td>TGARCH</td>
<td>Skewed Student-t</td>
<td>11</td>
<td>accurate</td>
<td>0</td>
<td>accurate</td>
</tr>
<tr>
<td>TGARCH</td>
<td>Skewed GED</td>
<td>10</td>
<td>accurate</td>
<td>0</td>
<td>accurate</td>
</tr>
<tr>
<td>GJR</td>
<td>Normal</td>
<td>15</td>
<td>accurate</td>
<td>4</td>
<td>accurate</td>
</tr>
<tr>
<td>GJR</td>
<td>Student-t</td>
<td>15</td>
<td>accurate</td>
<td>4</td>
<td>accurate</td>
</tr>
<tr>
<td>GJR</td>
<td>GED</td>
<td>15</td>
<td>accurate</td>
<td>4</td>
<td>accurate</td>
</tr>
<tr>
<td>GJR</td>
<td>Skewed normal</td>
<td>10</td>
<td>accurate</td>
<td>0</td>
<td>accurate</td>
</tr>
<tr>
<td>GJR</td>
<td>Skewed Student-t</td>
<td>11</td>
<td>accurate</td>
<td>0</td>
<td>accurate</td>
</tr>
<tr>
<td>GJR</td>
<td>Skewed GED</td>
<td>11</td>
<td>accurate</td>
<td>0</td>
<td>accurate</td>
</tr>
</tbody>
</table>

Table 3.6: Backtest results of the S&P/TSX index in the recent 300 trading days

It can be seen that when the MS-GARCH model is used, the backtest results of many models are good. However, the past 300 trading days have been relatively stable, and there has been no serious financial risk.

Back test results of estimated 95%VaR and 99%VaR in recent 800 trading days. The predictions of the above four models for nearly 300 days and 800 days all passed the kupiec test. It can be seen that MS(2)-TGARCH model has a better
CHAPTER 3. FORECASTING MODELS & RESULTS

<table>
<thead>
<tr>
<th>Volatility models</th>
<th>Distributions</th>
<th>95% VaR</th>
<th>kupiec test for 95% VaR</th>
<th>99% VaR</th>
<th>kupiec test for 99% VaR</th>
</tr>
</thead>
<tbody>
<tr>
<td>TGARCH</td>
<td>Student-t</td>
<td>41</td>
<td>accurate</td>
<td>15</td>
<td>accurate</td>
</tr>
<tr>
<td>TGARCH</td>
<td>GED</td>
<td>30</td>
<td>accurate</td>
<td>8</td>
<td>accurate</td>
</tr>
<tr>
<td>TGARCH</td>
<td>Skewed normal</td>
<td>32</td>
<td>accurate</td>
<td>10</td>
<td>accurate</td>
</tr>
<tr>
<td>GJR</td>
<td>Skewed GED</td>
<td>45</td>
<td>accurate</td>
<td>14</td>
<td>accurate</td>
</tr>
</tbody>
</table>

Table 3.7: Backtest results of the S&P/TSX index in the recent 800 trading days

The prediction effect. This is probably because the threshold GARCH (or TGARCH) model is used to handle leverage effects[40].

S&P500

Back test results of estimated 95%VaR and 99%VaR in recent 300 trading days.

<table>
<thead>
<tr>
<th>Volatility models</th>
<th>Distributions</th>
<th>95% VaR</th>
<th>kupiec test for 95% VaR</th>
<th>99% VaR</th>
<th>kupiec test for 99% VaR</th>
</tr>
</thead>
<tbody>
<tr>
<td>TGARCH</td>
<td>Student-t</td>
<td>17</td>
<td>accurate</td>
<td>2</td>
<td>accurate</td>
</tr>
<tr>
<td>TGARCH</td>
<td>GED</td>
<td>17</td>
<td>accurate</td>
<td>2</td>
<td>accurate</td>
</tr>
<tr>
<td>TGARCH</td>
<td>Skewed normal</td>
<td>14</td>
<td>accurate</td>
<td>1</td>
<td>accurate</td>
</tr>
<tr>
<td>TGARCH</td>
<td>Skewed Student-t</td>
<td>13</td>
<td>accurate</td>
<td>0</td>
<td>accurate</td>
</tr>
<tr>
<td>GJR</td>
<td>Skewed normal</td>
<td>17</td>
<td>accurate</td>
<td>1</td>
<td>accurate</td>
</tr>
<tr>
<td>GJR</td>
<td>Skewed Student-t</td>
<td>18</td>
<td>accurate</td>
<td>1</td>
<td>accurate</td>
</tr>
<tr>
<td>GJR</td>
<td>Skewed GED</td>
<td>16</td>
<td>accurate</td>
<td>0</td>
<td>accurate</td>
</tr>
</tbody>
</table>

Table 3.8: Backtest results of the S&P500 index in the recent 300 trading days

Back test results of estimated 95%VaR and 99%VaR in recent 800 trading days.

<table>
<thead>
<tr>
<th>Volatility models</th>
<th>Distributions</th>
<th>95% VaR</th>
<th>kupiec test for 95% VaR</th>
<th>99% VaR</th>
<th>kupiec test for 99% VaR</th>
</tr>
</thead>
<tbody>
<tr>
<td>TGARCH</td>
<td>normal</td>
<td>36</td>
<td>accurate</td>
<td>5</td>
<td>accurate</td>
</tr>
<tr>
<td>TGARCH</td>
<td>GED</td>
<td>50</td>
<td>accurate</td>
<td>5</td>
<td>accurate</td>
</tr>
<tr>
<td>TGARCH</td>
<td>Skewed Student-t</td>
<td>44</td>
<td>accurate</td>
<td>9</td>
<td>accurate</td>
</tr>
</tbody>
</table>

Table 3.9: Backtest results of the S&P500 index in the recent 800 trading days

Combining the results of the above two tables, the prediction results of the VaR of S&P500 by the models MS(2)-TGARCH with GED distribution and MS(2)-TGARCH with skewed student distribution both passed the kupiec test. It can be seen that the results of the Markov switching GARCH model(MSGARCH) for the VaR estimation of S&P500 are significantly better than the ARIMA-GARCH.
It is worth mentioning that the MS(2)-TGARCH with GED distribution model slightly underestimates risk for 95% VaR and slightly overestimates risk for 99% VaR. When forecasting for nearly 300 trading days, 17 days were lower than 95% VaR, which was higher than the optimal value of 15 days, and there were 2 trading days where the rate of return was lower than 99% VaR, which was lower than the optimal value of 3. Similarly, when forecasting for nearly 800 trading days, there are 50 days where rate of return is lower than 95% VaR, which are higher than the optimal value of 40 days. And there are 5 trading days where the rate of return is lower than 99% VaR and lower than the optimal value of 8. In addition, MS(2)-TGARCH with skewed t distribution is closer to the optimal value, which is the best model for the value of risk of S&P500.

Similarly, the MS(2)-TGARCH model has better predictive ability and the model that has the conditional distribution with skewness and leptokurtic has better predictive ability, which is consistent with the prediction results of other stock indexes.
### Nikkei225

<table>
<thead>
<tr>
<th>Volatility models</th>
<th>Distributions</th>
<th>95% VaR</th>
<th>kupiec test for 95% VaR</th>
<th>99% VaR</th>
<th>kupiec test for 99% VaR</th>
</tr>
</thead>
<tbody>
<tr>
<td>GARCH</td>
<td>Normal</td>
<td>21</td>
<td>accurate</td>
<td>1</td>
<td>accurate</td>
</tr>
<tr>
<td>GARCH</td>
<td>Student-t</td>
<td>14</td>
<td>accurate</td>
<td>1</td>
<td>accurate</td>
</tr>
<tr>
<td>GARCH</td>
<td>GED</td>
<td>18</td>
<td>accurate</td>
<td>1</td>
<td>accurate</td>
</tr>
<tr>
<td>GARCH</td>
<td>Skewed normal</td>
<td>14</td>
<td>accurate</td>
<td>1</td>
<td>accurate</td>
</tr>
<tr>
<td>GARCH</td>
<td>Skewed Student-t</td>
<td>11</td>
<td>accurate</td>
<td>1</td>
<td>accurate</td>
</tr>
<tr>
<td>TGARCH</td>
<td>Normal</td>
<td>9</td>
<td>accurate</td>
<td>1</td>
<td>accurate</td>
</tr>
<tr>
<td>TGARCH</td>
<td>Student-t</td>
<td>9</td>
<td>accurate</td>
<td>1</td>
<td>accurate</td>
</tr>
<tr>
<td>TGARCH</td>
<td>Skewed normal</td>
<td>9</td>
<td>accurate</td>
<td>1</td>
<td>accurate</td>
</tr>
<tr>
<td>TGARCH</td>
<td>Skewed Student-t</td>
<td>9</td>
<td>accurate</td>
<td>1</td>
<td>accurate</td>
</tr>
<tr>
<td>GJR</td>
<td>Student-t</td>
<td>12</td>
<td>accurate</td>
<td>1</td>
<td>accurate</td>
</tr>
<tr>
<td>GJR</td>
<td>GED</td>
<td>12</td>
<td>accurate</td>
<td>1</td>
<td>accurate</td>
</tr>
<tr>
<td>GJR</td>
<td>Skewed normal</td>
<td>11</td>
<td>accurate</td>
<td>1</td>
<td>accurate</td>
</tr>
<tr>
<td>GJR</td>
<td>Skewed Student-t</td>
<td>9</td>
<td>accurate</td>
<td>1</td>
<td>accurate</td>
</tr>
</tbody>
</table>

Table 3.10: Backtest results of Nikkei225 index in the recent 300 trading days

<table>
<thead>
<tr>
<th>Volatility models</th>
<th>Distributions</th>
<th>95% VaR</th>
<th>kupiec test for 95% VaR</th>
<th>99% VaR</th>
<th>kupiec test for 99% VaR</th>
</tr>
</thead>
<tbody>
<tr>
<td>GARCH</td>
<td>Normal</td>
<td>38</td>
<td>accurate</td>
<td>8</td>
<td>accurate</td>
</tr>
<tr>
<td>GARCH</td>
<td>GED</td>
<td>29</td>
<td>accurate</td>
<td>6</td>
<td>accurate</td>
</tr>
<tr>
<td>GARCH</td>
<td>Skewed normal</td>
<td>34</td>
<td>accurate</td>
<td>6</td>
<td>accurate</td>
</tr>
<tr>
<td>GJR</td>
<td>Student-t</td>
<td>29</td>
<td>accurate</td>
<td>6</td>
<td>accurate</td>
</tr>
<tr>
<td>GJR</td>
<td>GED</td>
<td>30</td>
<td>accurate</td>
<td>6</td>
<td>accurate</td>
</tr>
</tbody>
</table>

Table 3.11: Backtest results of Nikkei225 index in the recent 800 trading days

As can be seen from the above table, although all of the above model passed the kupiec test, the overall value was lower than the optimal value, indicating that the model overestimated the risk. In the field of financial risk management, a reasonable degree of overestimation of risk is acceptable relative to underestimation of risk.

### FTSE100

Back test results of estimated 95%VaR and 99%VaR in recent 300 and 800 trading days.
The FTSE 100 fell sharply as COVID-19 entered the UK and began to spread. Fortunately, there is still a model (MS(2)-TGARCH with GED distribution) that passes the Kupiec test and predicts 95% VaR, 99% VaR with excellent prediction effects in both the 300-day and 800-day ranges.

In the financial industry, the market can also be divided into low, medium and high volatility, especially when the market is high volatility is often related to economic recession. Therefore, we try to see if adding a regime can improve the model’s ability to predict VaR. The model results are in the appendix. However, we did not find that adding a regime significantly improved the prediction results of the model, but increased the number of parameters and the complexity of the model.

<table>
<thead>
<tr>
<th>Volatility models</th>
<th>Distributions</th>
<th>95% VaR</th>
<th>Kupiec test for 95% VaR</th>
<th>99% VaR</th>
<th>Kupiec test for 99% VaR</th>
</tr>
</thead>
<tbody>
<tr>
<td>GARCH</td>
<td>Normal</td>
<td>16</td>
<td>accurate</td>
<td>8</td>
<td>accurate</td>
</tr>
<tr>
<td>GARCH</td>
<td>Student-t</td>
<td>13</td>
<td>accurate</td>
<td>3</td>
<td>accurate</td>
</tr>
<tr>
<td>GARCH</td>
<td>Skewed normal</td>
<td>15</td>
<td>accurate</td>
<td>6</td>
<td>accurate</td>
</tr>
<tr>
<td>GARCH</td>
<td>Skewed Student-t</td>
<td>15</td>
<td>accurate</td>
<td>3</td>
<td>accurate</td>
</tr>
<tr>
<td>TGARCH</td>
<td>Normal</td>
<td>15</td>
<td>accurate</td>
<td>4</td>
<td>accurate</td>
</tr>
<tr>
<td>TGARCH</td>
<td>Student-t</td>
<td>15</td>
<td>accurate</td>
<td>4</td>
<td>accurate</td>
</tr>
<tr>
<td>TGARCH</td>
<td>GED</td>
<td>15</td>
<td>accurate</td>
<td>4</td>
<td>accurate</td>
</tr>
<tr>
<td>TGARCH</td>
<td>Skewed normal</td>
<td>15</td>
<td>accurate</td>
<td>4</td>
<td>accurate</td>
</tr>
<tr>
<td>GJR</td>
<td>Student-t</td>
<td>16</td>
<td>accurate</td>
<td>6</td>
<td>accurate</td>
</tr>
<tr>
<td>GJR</td>
<td>GED</td>
<td>17</td>
<td>accurate</td>
<td>6</td>
<td>accurate</td>
</tr>
<tr>
<td>GJR</td>
<td>Skewed normal</td>
<td>16</td>
<td>accurate</td>
<td>6</td>
<td>accurate</td>
</tr>
<tr>
<td>GJR</td>
<td>Skewed Student-t</td>
<td>15</td>
<td>accurate</td>
<td>4</td>
<td>accurate</td>
</tr>
<tr>
<td>GJR</td>
<td>Skewed GED</td>
<td>15</td>
<td>accurate</td>
<td>4</td>
<td>accurate</td>
</tr>
</tbody>
</table>

Table 3.12: Backtest results of FTSE100 index in the recent 300 trading days

<table>
<thead>
<tr>
<th>Volatility models</th>
<th>Distributions</th>
<th>95% VaR</th>
<th>Kupiec test for 95% VaR</th>
<th>99% VaR</th>
<th>Kupiec test for 99% VaR</th>
</tr>
</thead>
<tbody>
<tr>
<td>TGARCH</td>
<td>GED</td>
<td>36</td>
<td>accurate</td>
<td>3</td>
<td>accurate</td>
</tr>
<tr>
<td>TGARCH</td>
<td>Skewed Student-t</td>
<td>50</td>
<td>accurate</td>
<td>9</td>
<td>accurate</td>
</tr>
</tbody>
</table>

Table 3.13: Backtest results of FTSE100 index in the recent 800 trading days
Chapter 4

Conclusion & Potential Research

4.1 Conclusion

In this thesis, the GARCH model and the Markov switching GARCH model were used to model the volatility of the stock indexes of the four major economies and predict VaR. The validity of VaR predictions was tested on out-of-sample data. In particular, the thesis first forecasts data for the recent 300 trading days, and then rebuilt the model to include the first half of 2020 in the forecast. This is because in the first half of 2020, the severe market decline due to COVID-19 created a severe challenge to the forecasts of risk indicators. Through empirical analysis, this thesis found that when the market is relatively stable, the forecasting effect of the GARCH model and the Markov switching GARCH model is not much different. However, when the market has experienced switching of market conditions such as changing from a bull market to a bear market, the GARCH model has insufficient predictive ability for the risk management indicator VaR, but the Markov
switching GARCH model can still give reliable VaR predictions.

Furthermore, since financial data often have leptokurtic and fat tails, we compared the VaR’s predicted results by selecting different conditional distributions and different conditional volatility models for the model. This thesis found the MS(2)-TGARCH model has relatively good prediction performance. This may be because compared to the GARCH model, MS(2)-TGARCH model has two switching regimes for data with changes in market conditions, so that when the model switch to the different regimes, the model parameters are different, which can be more efficient modeling market conditions switching. At the same time, the TGARCH model can more effectively model the asymmetry caused by leverage effects.

Finally, we tried to add one more regime to analyze whether the accuracy of the prediction could be improved. The results of the model can be seen in the appendix, we did not find a significant improvement in the prediction results of the model.

### 4.2 Potential Research

This thesis is particularly concerned with forecasting financial risk management indicators in the context of the extremely severe shock to financial markets as COVID-19 becomes a global pandemic. Based on Markov switching GARCH model, the prediction effect in this thesis is good. It is worth noting that machine
learning methods are increasingly used in practice, such as long short-term memory (LSTM) methods to analyze financial time series have proved to have good prediction effects. Whether some machine learning methods can further improve the prediction effect of risk management indicators is worth studying.

In addition, it is not enough to focus on VaR alone, because VaR hardly reflects the extreme tail risk. Based on the definition of expected shortfall (ES), further prediction of ES is meaningful only when the model’s prediction of VaR is reliable. Expected shortfall can be further predicted based on the model in this thesis, and then both VaR and ES can be used as risk control indicators to better measure extreme tail risk. However, it is relatively difficult to verify the estimated effect of ES. Due to the extreme market fluctuations caused by the epidemic, whether we can also give a high-accuracy prediction for ES deserves further study.

In the capital market, the extreme volatility of individual stocks is often more severe than that of indexes. The estimation of risk management indicators for individual stocks or investment portfolios is a more challenging and deserves further study.

It can be seen that different conditional volatility models and different distributions have an impact on the results of the model predictions. Introducing more conditional volatility models and distributions, especially asymmetric fat-tailed distributions, may further improve the predictive power of the model.
Bibliography


[5] AIP Conference Proceedings 1827, 02006 (2017); https://doi.org/10.1063/1.4979422 Published Online: 30 March 2017


Appendix A

Markov switching GARCH model
with three regimes

A.1 S&P/TSX

<table>
<thead>
<tr>
<th>Volatility models</th>
<th>Distributions</th>
<th>95% VaR</th>
<th>kupiec test for 95% VaR</th>
<th>99% VaR</th>
<th>kupiec test for 99% VaR</th>
</tr>
</thead>
<tbody>
<tr>
<td>GARCH</td>
<td>Normal</td>
<td>14</td>
<td>accurate</td>
<td>1</td>
<td>accurate</td>
</tr>
<tr>
<td>GARCH</td>
<td>Student-t</td>
<td>14</td>
<td>accurate</td>
<td>4</td>
<td>accurate</td>
</tr>
<tr>
<td>GARCH</td>
<td>GED</td>
<td>14</td>
<td>accurate</td>
<td>1</td>
<td>accurate</td>
</tr>
<tr>
<td>GARCH</td>
<td>Skewed normal</td>
<td>15</td>
<td>accurate</td>
<td>4</td>
<td>accurate</td>
</tr>
<tr>
<td>GARCH</td>
<td>Skewed Student-t</td>
<td>9</td>
<td>accurate</td>
<td>0</td>
<td>accurate</td>
</tr>
<tr>
<td>GARCH</td>
<td>Skewed GED</td>
<td>15</td>
<td>accurate</td>
<td>4</td>
<td>accurate</td>
</tr>
<tr>
<td>TGARCH</td>
<td>Student-t</td>
<td>10</td>
<td>accurate</td>
<td>0</td>
<td>accurate</td>
</tr>
<tr>
<td>TGARCH</td>
<td>Skewed normal</td>
<td>11</td>
<td>accurate</td>
<td>0</td>
<td>accurate</td>
</tr>
<tr>
<td>TGARCH</td>
<td>Skewed Student-t</td>
<td>11</td>
<td>accurate</td>
<td>0</td>
<td>accurate</td>
</tr>
<tr>
<td>GJR</td>
<td>Normal</td>
<td>16</td>
<td>accurate</td>
<td>5</td>
<td>accurate</td>
</tr>
<tr>
<td>GJR</td>
<td>Student-t</td>
<td>15</td>
<td>accurate</td>
<td>4</td>
<td>accurate</td>
</tr>
<tr>
<td>GJR</td>
<td>GED</td>
<td>15</td>
<td>accurate</td>
<td>4</td>
<td>accurate</td>
</tr>
<tr>
<td>GJR</td>
<td>Skewed normal</td>
<td>11</td>
<td>accurate</td>
<td>4</td>
<td>accurate</td>
</tr>
<tr>
<td>GJR</td>
<td>Skewed Student-t</td>
<td>10</td>
<td>accurate</td>
<td>2</td>
<td>accurate</td>
</tr>
<tr>
<td>GJR</td>
<td>Skewed GED</td>
<td>10</td>
<td>accurate</td>
<td>2</td>
<td>accurate</td>
</tr>
</tbody>
</table>

Table A.1: Backtest results of the S&P/TSX index in the recent 300 trading days with three regimes
APPENDIX A. MARKOV SWITCHING GARCH MODEL WITH THREE REGIMES

### Volatility models

<table>
<thead>
<tr>
<th>Volatility models</th>
<th>Distributions</th>
<th>95% VaR</th>
<th>kupiec test for 95% VaR</th>
<th>99% VaR</th>
<th>kupiec test for 99% VaR</th>
</tr>
</thead>
<tbody>
<tr>
<td>TGARCH</td>
<td>Student-t</td>
<td>49</td>
<td>accurate</td>
<td>14</td>
<td>accurate</td>
</tr>
<tr>
<td>TGARCH</td>
<td>Skewed normal</td>
<td>32</td>
<td>accurate</td>
<td>10</td>
<td>accurate</td>
</tr>
<tr>
<td>TGARCH</td>
<td>Skewed Student-t</td>
<td>40</td>
<td>accurate</td>
<td>15</td>
<td>accurate</td>
</tr>
<tr>
<td>GJR</td>
<td>Skewed Student-t</td>
<td>40</td>
<td>accurate</td>
<td>10</td>
<td>accurate</td>
</tr>
</tbody>
</table>

Table A.2: Backtest results of the S&P/TSX index in the recent 800 trading days with three regimes

#### A.2 S&P500

<table>
<thead>
<tr>
<th>Volatility models</th>
<th>Distributions</th>
<th>95% VaR</th>
<th>kupiec test for 95% VaR</th>
<th>99% VaR</th>
<th>kupiec test for 99% VaR</th>
</tr>
</thead>
<tbody>
<tr>
<td>GARCH</td>
<td>GED</td>
<td>14</td>
<td>accurate</td>
<td>4</td>
<td>accurate</td>
</tr>
<tr>
<td>TGARCH</td>
<td>GED</td>
<td>14</td>
<td>accurate</td>
<td>0</td>
<td>accurate</td>
</tr>
<tr>
<td>TGARCH</td>
<td>Skewed normal</td>
<td>15</td>
<td>accurate</td>
<td>2</td>
<td>accurate</td>
</tr>
<tr>
<td>TGARCH</td>
<td>Skewed Student-t</td>
<td>12</td>
<td>accurate</td>
<td>1</td>
<td>accurate</td>
</tr>
<tr>
<td>GJR</td>
<td>Skewed normal</td>
<td>18</td>
<td>accurate</td>
<td>2</td>
<td>accurate</td>
</tr>
<tr>
<td>GJR</td>
<td>Skewed Student-t</td>
<td>18</td>
<td>accurate</td>
<td>1</td>
<td>accurate</td>
</tr>
<tr>
<td>GJR</td>
<td>Skewed GED</td>
<td>18</td>
<td>accurate</td>
<td>2</td>
<td>accurate</td>
</tr>
</tbody>
</table>

Table A.3: Backtest results of the S&P500 index in the recent 300 trading days with three regimes

<table>
<thead>
<tr>
<th>Volatility models</th>
<th>Distributions</th>
<th>95% VaR</th>
<th>kupiec test for 95% VaR</th>
<th>99% VaR</th>
<th>kupiec test for 99% VaR</th>
</tr>
</thead>
<tbody>
<tr>
<td>TGARCH</td>
<td>GED</td>
<td>46</td>
<td>accurate</td>
<td>7</td>
<td>accurate</td>
</tr>
</tbody>
</table>

Table A.4: Backtest results of the S&P500 index in the recent 800 trading days with three regimes
### A.3 Nikkei225

<table>
<thead>
<tr>
<th>Volatility models</th>
<th>Distributions</th>
<th>95% VaR</th>
<th>kupiec test for 95% VaR</th>
<th>99% VaR</th>
<th>kupiec test for 99% VaR</th>
</tr>
</thead>
<tbody>
<tr>
<td>GARCH</td>
<td>Normal</td>
<td>20</td>
<td>accurate</td>
<td>1</td>
<td>accurate</td>
</tr>
<tr>
<td>GARCH</td>
<td>Student-t</td>
<td>14</td>
<td>accurate</td>
<td>1</td>
<td>accurate</td>
</tr>
<tr>
<td>GARCH</td>
<td>GED</td>
<td>16</td>
<td>accurate</td>
<td>0</td>
<td>accurate</td>
</tr>
<tr>
<td>GARCH</td>
<td>Skewed normal</td>
<td>12</td>
<td>accurate</td>
<td>1</td>
<td>accurate</td>
</tr>
<tr>
<td>GARCH</td>
<td>Skewed Student-t</td>
<td>11</td>
<td>accurate</td>
<td>1</td>
<td>accurate</td>
</tr>
<tr>
<td>TGARCH</td>
<td>Normal</td>
<td>10</td>
<td>accurate</td>
<td>1</td>
<td>accurate</td>
</tr>
<tr>
<td>GJR</td>
<td>Student-t</td>
<td>13</td>
<td>accurate</td>
<td>1</td>
<td>accurate</td>
</tr>
<tr>
<td>GJR</td>
<td>GED</td>
<td>14</td>
<td>accurate</td>
<td>1</td>
<td>accurate</td>
</tr>
<tr>
<td>GJR</td>
<td>Skewed normal</td>
<td>11</td>
<td>accurate</td>
<td>1</td>
<td>accurate</td>
</tr>
<tr>
<td>GJR</td>
<td>Skewed GED</td>
<td>9</td>
<td>accurate</td>
<td>1</td>
<td>accurate</td>
</tr>
</tbody>
</table>

Table A.5: Backtest results of the Nikkei225 index in the recent 300 trading days with three regimes

<table>
<thead>
<tr>
<th>Volatility models</th>
<th>Distributions</th>
<th>95% VaR</th>
<th>kupiec test for 95% VaR</th>
<th>99% VaR</th>
<th>kupiec test for 99% VaR</th>
</tr>
</thead>
<tbody>
<tr>
<td>GARCH</td>
<td>Normal</td>
<td>33</td>
<td>accurate</td>
<td>6</td>
<td>accurate</td>
</tr>
<tr>
<td>GARCH</td>
<td>Student-t</td>
<td>33</td>
<td>accurate</td>
<td>6</td>
<td>accurate</td>
</tr>
<tr>
<td>GARCH</td>
<td>Skewed normal</td>
<td>31</td>
<td>accurate</td>
<td>6</td>
<td>accurate</td>
</tr>
<tr>
<td>TGARCH</td>
<td>GED</td>
<td>32</td>
<td>accurate</td>
<td>6</td>
<td>accurate</td>
</tr>
<tr>
<td>GJR</td>
<td>Student-t</td>
<td>29</td>
<td>accurate</td>
<td>7</td>
<td>accurate</td>
</tr>
<tr>
<td>GJR</td>
<td>GED</td>
<td>30</td>
<td>accurate</td>
<td>6</td>
<td>accurate</td>
</tr>
</tbody>
</table>

Table A.6: Backtest results of the Nikkei225 index in the recent 800 trading days with three regimes
A.4 FTSE100

<table>
<thead>
<tr>
<th>Volatility models</th>
<th>Distributions</th>
<th>95% VaR</th>
<th>kupiec test for 95% VaR</th>
<th>99% VaR</th>
<th>kupiec test for 99% VaR</th>
</tr>
</thead>
<tbody>
<tr>
<td>GARCH</td>
<td>Normal</td>
<td>14</td>
<td>accurate</td>
<td>3</td>
<td>accurate</td>
</tr>
<tr>
<td>GARCH</td>
<td>Student-t</td>
<td>15</td>
<td>accurate</td>
<td>6</td>
<td>accurate</td>
</tr>
<tr>
<td>GARCH</td>
<td>Skewed normal</td>
<td>14</td>
<td>accurate</td>
<td>5</td>
<td>accurate</td>
</tr>
<tr>
<td>GARCH</td>
<td>Skewed Student-t</td>
<td>15</td>
<td>accurate</td>
<td>3</td>
<td>accurate</td>
</tr>
<tr>
<td>TGARCH</td>
<td>Normal</td>
<td>15</td>
<td>accurate</td>
<td>4</td>
<td>accurate</td>
</tr>
<tr>
<td>TGARCH</td>
<td>Student-t</td>
<td>13</td>
<td>accurate</td>
<td>2</td>
<td>accurate</td>
</tr>
<tr>
<td>TGARCH</td>
<td>GED</td>
<td>9</td>
<td>accurate</td>
<td>0</td>
<td>accurate</td>
</tr>
<tr>
<td>TGARCH</td>
<td>Skewed normal</td>
<td>15</td>
<td>accurate</td>
<td>4</td>
<td>accurate</td>
</tr>
<tr>
<td>GJR</td>
<td>Student-t</td>
<td>17</td>
<td>accurate</td>
<td>7</td>
<td>accurate</td>
</tr>
<tr>
<td>GJR</td>
<td>GED</td>
<td>17</td>
<td>accurate</td>
<td>6</td>
<td>accurate</td>
</tr>
<tr>
<td>GJR</td>
<td>Skewed normal</td>
<td>17</td>
<td>accurate</td>
<td>7</td>
<td>accurate</td>
</tr>
<tr>
<td>GJR</td>
<td>Skewed Student-t</td>
<td>11</td>
<td>accurate</td>
<td>3</td>
<td>accurate</td>
</tr>
<tr>
<td>GJR</td>
<td>Skewed GED</td>
<td>15</td>
<td>accurate</td>
<td>3</td>
<td>accurate</td>
</tr>
</tbody>
</table>

Table A.7: Backtest results of the FTSE100 index in the past 300 trading days with three regimes

<table>
<thead>
<tr>
<th>Volatility models</th>
<th>Distributions</th>
<th>95% VaR</th>
<th>kupiec test for 95% VaR</th>
<th>99% VaR</th>
<th>kupiec test for 99% VaR</th>
</tr>
</thead>
<tbody>
<tr>
<td>TGARCH</td>
<td>GED</td>
<td>40</td>
<td>accurate</td>
<td>5</td>
<td>accurate</td>
</tr>
</tbody>
</table>

Table A.8: Backtest results of the FTSE100 index in the past 800 trading days with three regimes

No significant improvement was observed in the VaR back test results predicted by adding one more regime model.
Appendix B

Relatively good models

We present a summary of the models that perform relatively well in predicting VaR for each stock index.

B.1 S&P/TSX

Specification type: Markov-switching
Specification name: gjrGARCH_sged
Number of parameters in each variance model: 4
Number of parameters in each distribution: 2
Fitted parameters:
APPENDIX B. RELATIVELY GOOD MODELS

|                | Estimate | Std. Error | Pr(>|t|) |
|----------------|----------|------------|----------|
| alpha0_1       | 0.0000   | 0          | <1e-16   |
| alpha1_1       | 0.0007   | 0          | <1e-16   |
| alpha2_1       | 0.1198   | 0          | <1e-16   |
| beta_1         | 0.9222   | 0          | <1e-16   |
| nu_1           | 1.5924   | 0          | <1e-16   |
| xi_1           | 0.7942   | 0          | <1e-16   |
| alpha0_2       | 0.0000   | 0          | <1e-16   |
| alpha1_2       | 0.0085   | 0          | <1e-16   |
| alpha2_2       | 0.3734   | 0          | <1e-16   |
| beta_2         | 0.7385   | 0          | <1e-16   |
| nu_2           | 3.1497   | 0          | <1e-16   |
| xi_2           | 0.2451   | 0          | <1e-16   |
| P_1_1          | 0.9941   | 0          | <1e-16   |
| P_2_1          | 0.0654   | 0          | <1e-16   |

Table B.1: MSGARCH for S&P/TSX index

<table>
<thead>
<tr>
<th></th>
<th>t+1</th>
<th>k=1</th>
<th>t+1</th>
<th>k=2</th>
</tr>
</thead>
<tbody>
<tr>
<td>t</td>
<td>k=1</td>
<td>0.9941</td>
<td>0.0059</td>
<td></td>
</tr>
<tr>
<td>t</td>
<td>k=2</td>
<td>0.0654</td>
<td>0.9346</td>
<td></td>
</tr>
</tbody>
</table>

Table B.2: The transition matrix of the S&P/TSX model

The parameters of the model in the table are all significant. It can be seen from the transition matrix that if the current state is in state 1, the probability of remaining in state 1 is 99.41%, and the probability of transitioning to state 2 is 0.59%.
B.2 S&P500

Specification type: Markov-switching

Specification name: TGARCH_sstd

Number of parameters in each variance model: 4 4

Number of parameters in each distribution: 2 2

Fitted parameters:

|            | Estimate | Std. Error | Pr(>|t|) |
|------------|----------|------------|----------|
| alpha0_1   | 0.0008   | 0.0000     | <1e-16   |
| alpha1_1   | 0.0001   | 0.0000     | <1e-16   |
| alpha2_1   | 0.2236   | 0.0044     | <1e-16   |
| beta_1     | 0.8289   | 0.0014     | <1e-16   |
| nu_1       | 98.2885  | 0.0163     | <1e-16   |
| xi_1       | 0.0298   | 0.0002     | <1e-16   |
| alpha0_2   | 0.0002   | 0.0000     | <1e-16   |
| alpha1_2   | 0.0030   | 0.0000     | <1e-16   |
| alpha2_2   | 0.3075   | 0.0023     | <1e-16   |
| beta_2     | 0.8677   | 0.0000     | <1e-16   |
| nu_2       | 4.3716   | 0.0215     | <1e-16   |
| xi_2       | 0.9385   | 0.0085     | <1e-16   |
| P_1_1      | 0.3925   | 0.0012     | <1e-16   |
| P_2_1      | 0.1470   | 0.0023     | <1e-16   |

Table B.3: MSGARCH for S&P500 index

<table>
<thead>
<tr>
<th></th>
<th>t+1</th>
<th>k=1</th>
<th>t+1</th>
<th>k=2</th>
</tr>
</thead>
<tbody>
<tr>
<td>t</td>
<td>k=1</td>
<td>0.3925</td>
<td>0.6075</td>
<td></td>
</tr>
<tr>
<td>t</td>
<td>k=2</td>
<td>0.1470</td>
<td>0.8530</td>
<td></td>
</tr>
</tbody>
</table>

Table B.4: The transition matrix of the S&P500 model
B.3 Nikkei225

Specification type: Markov-switching

Specification name: SGARCH_snorm

Number of parameters in each variance model: 3 3

Number of parameters in each distribution: 1 1

| Estimate | Std. Error | Pr(|t|) |
|----------|------------|--------|
| alpha0_1 | 0.0000     | 0.0000 | 1.810e-01 |
| alpha1_1 | 0.0321     | 0.0260 | 1.079e-01 |
| beta_1   | 0.8743     | 0.0325 | <1e-16    |
| xi_1     | 0.9249     | 0.1037 | <1e-16    |
| alpha0_2 | 0.0000     | 0.0000 | 2.700e-04 |
| alpha1_2 | 0.1672     | 0.0889 | 3.005e-02 |
| beta_2   | 0.8223     | 0.0063 | <1e-16    |
| xi_2     | 0.8038     | 0.0367 | <1e-16    |
| P_1_1    | 0.2996     | 0.1346 | 1.299e-02 |
| P_2_1    | 0.4158     | 0.0886 | 1.356e-06 |

Table B.5: MSGARCH for Nikkei225 index

<table>
<thead>
<tr>
<th>t+1</th>
<th>k=1</th>
<th>t+1</th>
<th>k=2</th>
</tr>
</thead>
<tbody>
<tr>
<td>t</td>
<td>k=1</td>
<td>0.2996</td>
<td>0.7004</td>
</tr>
<tr>
<td>t</td>
<td>k=2</td>
<td>0.4158</td>
<td>0.5842</td>
</tr>
</tbody>
</table>

Table B.6: The transition matrix of the Nikkei225 model
## B.4 FTSE 100

Specification type: Markov-switching

Specification name: TGARCH\_ged

Number of parameters in each variance model: 4

Number of parameters in each distribution: 1

| Parameter   | Estimate | Std. Error | Pr(>|t|)  |
|-------------|----------|------------|----------|
| alpha0\_1   | 0.0000   | 0.0002     | 4.929e-01|
| alpha1\_1   | 0.2954   | 11.8704    | 4.901e-01|
| alpha2\_1   | 0.0961   | 4.6778     | 4.918e-01|
| beta\_1     | 0.0828   | 4.1350     | 4.920e-01|
| nu\_1       | 0.7000   | 0.0001     | <1e-16   |
| alpha0\_2   | 0.0019   | 0.1080     | 4.929e-01|
| alpha1\_2   | 0.0004   | 0.0290     | 4.948e-01|
| alpha2\_2   | 0.9312   | 4.5590     | 4.191e-01|
| beta\_2     | 0.5128   | 0.0000     | <1e-16   |
| nu\_2       | 0.7000   | 0.00       | <1e-16   |
| alpha0\_3   | 0.0006   | 0.0002     | 1.039e-03|
| alpha1\_3   | 0.0000   | 0.0003     | 4.924e-01|
| alpha2\_3   | 0.2678   | 0.0700     | 6.575e-05|
| beta\_3     | 0.8663   | 0.0160     | <1e-16   |
| nu\_3       | 0.7000   | 0.0001     | <1e-16   |
| P\_1\_1     | 0.5754   | 0.0000     | <1e-16   |
| P\_1\_2     | 0.4246   | 0.0000     | <1e-16   |
| P\_2\_1     | 0.0000   | 19.4038    | 5.000e-01|
| P\_2\_2     | 0.0002   | 0.0000     | <1e-16   |
| P\_3\_1     | 0.0000   | 0.0002     | 5.000e-01|
| P\_3\_2     | 0.0000   | 0.0092     | 5.000e-01|

Table B.7: MSGARCH for FTSE 100 index
### Table B.8: The transition matrix of the FTSE 100 model

<table>
<thead>
<tr>
<th></th>
<th>t+1</th>
<th>k=1</th>
<th>t+1</th>
<th>k=2</th>
<th>t+1</th>
<th>k=3</th>
</tr>
</thead>
<tbody>
<tr>
<td>t</td>
<td>k=1</td>
<td>0.5754</td>
<td>0.4246</td>
<td>0.0000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>t</td>
<td>k=2</td>
<td>0.0000</td>
<td>0.0002</td>
<td>0.9998</td>
<td></td>
<td></td>
</tr>
<tr>
<td>t</td>
<td>k=3</td>
<td>0.0000</td>
<td>0.0000</td>
<td>1.0000</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table B.8: The transition matrix of the FTSE 100 model