ABSTRACT

ZHAO, YANG. A Simple and Robust Expected Shortfall Estimation Approach and A Comprehensive Comparison of Volatility Models. (Under the direction of Dr. Tao Pang.)

Measures of risk play essential roles in the financial risk management. Quantifying risks using suitable measures is undoubtedly inevitable for a business success. This dissertation, consisting of three essays, concentrates on two popular measures of risk: Expected Shortfall (ES) and asset return volatility. Particularly, we propose a simple and robust ES estimation approach for small samples and make a comprehensive comparison of volatility models under the physical and risk-neutral measures. Risk-neutral volatility models incorporating the leverage effect are investigated and a new Cumulative Return option pricing model is also put forward.

Estimating ES, though important and indispensable, is difficult when a sample size is small. The first essay makes efforts to create a recipe for such challenge. A *tail-based* normal approximation with explicit formulas is derived by matching a specific quantile and the mean excess square of the sample points. To enhance the estimation accuracy, we then introduce an adjusted *tail-based* normal approximation based on the sample's tail weight. The adjusted *tailed-based* normal approximation is robust and efficient in the sense that it can be applied to various heavy-tailed distributions, such as student's t, Lognormal, Gamma, Weibull, etc., and the errors are considerably small. In addition, compared to two prevalent ES estimators—the mean value of excessive losses and the extreme value theory estimator, our proposed approach achieves more accurate estimates with substantially smaller errors, especially at a high confidence level. Another appealing feature of the approach is that it works well with small samples. Effects of linear transformations on the proposed ES estimator are also investigated to guarantee its practicality and further validate our new approach.

The second essay conducts a comprehensive comparison of the GARCH(1, 1) and ARSV(1) models using S&P 500 Index. Under the physical measure, after fitting the historical return sequence, we calculate the likelihoods and test the normality for the error terms of these two

models. In addition, two robust loss functions, MSE and QLIKE, are adopted for the one-step-ahead volatility forecast comparison. On the other hand, under the risk-neutral measure, the in-sample and out-of-sample average option pricing errors of the two models are explored. We find the ARSV(1) model outperforms the GARCH(1, 1) model in terms of in-sample and out-of-sample performances under the physical measure. Under the risk-neutral model, these two models are considerably close when pricing call options while the ARSV(1) model is significantly superior to the GARCH(1, 1) model for put options.

In the third essay, enlightened by the Constant Elasticity of Variance model, a continuous-time option pricing model is proposed. In this model, we straightforwardly relate the volatility to the exponentially decayed weighted average cumulative asset return such that the leverage effect is captured. Then we investigate the performances of option pricing models with the leverage effect. Compared to the original risk-neutral GARCH(1, 1) and ARSV(1) models, both of the two models exhibit remarkable improvements in terms of the in-sample and out-of-sample pricing errors after incorporating the leverage effect. As for the put option pricing errors, the differences among most examined models are considerably insignificant. Surprisingly, our Cumulative Return model that has only two parameters dominates other sophisticated models when predicting call option prices and has a robust performance on an outlier day.

A Simple and Robust Expected Shortfall Estimation Approach and A Comprehensive Comparison of Volatility Models

by
Yang Zhao

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

Operations Research

Raleigh, North Carolina

2018

APPROVED BY:

Dr. Zhilin Li	Dr. Negash Medhin
Dr. Sujit Ghosh	Dr. Tao Pang Chair of Advisory Committee

DEDICATION

To my father Jiafu Zhao, my mother Minzhi Wang

and my girlfriend Weiran Cai

BIOGRAPHY

Yang Zhao was born and raised in Wuxi, an old eastern city in China. He graduated from Wuxi Big Bridge Experimental High School in 2009 and then attended Huazhong University of Science and Technology in Wuhan, China. In 2013, he received his Bachelor of Science degree in financial management. After that, he was admitted to the Operations Research Program in Department of Mathematics at North Carolina State University to pursue his Ph.D. degree under the supervision of Dr. Tao Pang. His research focuses on financial mathematics and risk management.

ACKNOWLEDGEMENTS

The first night in the United States is still vivid in my memory. I missed the last flight to Raleigh and got stuck in Chicago. The long-time travel had exhausted me while the jet lag kept me from falling asleep. At that time, I was nervous about how many things would be out of control in the years to come and how long it would take to find the right path for me. Maybe I would get lost forever in this totally brand-new environment like a drunk bird. However, thanks to the assistance, suggestion, and encouragement I received during my Ph.D. career, in retrospect, those anxieties looked unnecessary.

Firstly, I am truly grateful for my advisor, Dr. Tao Pang, for his inspiring instruction and constant support throughout my Ph.D. career. As an academic supervisor, he points out my problems and gives me directions when I run into difficulties. He inculcates me with rigorous and earnest attitudes towards the research. Since early 2014, we meet once a week. His precious feedbacks always broaden my horizon and consolidate my knowledge. As a friend, he enlightens me about the skills in communication and leadership through his kindness, courtesy, and integrity. It is indeed my privilege to have such an advisor.

Moreover, I would like to extend my sincere appreciation to Dr. Zhilin Li, Dr. Sujit Ghosh, and Dr. Negash Medhin, for being my advisory committee members. Their detailed questions and suggestions help improve my research. I am grateful for Dr. Li for becoming my graduate school representative. He is also an exceptional instructor. His humorous and lively lecture facilitates my understanding of the difficult course on PDE. I am also grateful for Dr. Ghosh for his invaluable comments on my preliminary exam. In addition, he teaches me how to solve problems from a Bayesian point of view. Especially, I want to thank Dr. Medhin for his great contributions to the Operations Research Program. He makes efforts to help students apply for funding and job market opportunities. Despite his busy schedule, he was glad to spend time explaining regulations and requirements of our Ph.D. program when I was a first-year student at NCSU.

My sincere thanks also go to Dr. Shu-Cherng Fang, Dr. Yunan Liu and Dr. Michael Kay for their instruction and help. I want to thank Department of Mathematics for providing me with a full scholarship. I also want to express my gratitude to the other faculty and staff in Operations Research Program and Department of Mathematics, especially Mrs. Linda Smith and Mrs. Denise Seabrooks, for their kind assistance these years.

I am sincerely grateful for Dr. Xilong Chen and Dr. Mark Little in SAS Institute for offering me the internship and full-time job opportunities. Their guidance benefits me a lot in terms of coding and applying what I have learned to the industry.

Many thanks go to my friends: Rui Hu, Hao Chen, Di Miao, Binghui Li, Jingwen Zhao, Hengrui Hu, Rui Yang, Ruihao Ni, Cagatay Karan, Dave Kwon, Ling Zhang, Peiqian Zhong, Chi Zhang, Benjun Miao and many others. They let me laugh; let me cry; give me motivation and keep me calm down. We have a lot of happy memories.

Last but not least, I want to express my deepest gratitude to my parents for their unconditional love. When I am depressed, anxious or hesitant, they always stand behind me. Without their support, I would not have been who I am today. I am also deeply grateful for my girlfriend, Weiran Cai, for her company and continuous caring about me. It is not easy for a couple to live far away from each other particularly in two different countries with a 12-hour time difference. She spent all the holidays with me. I owe her too much.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

Introduction

1.1 Background

Generally speaking, all financial activities can be and have to be connected with risk management. Investors diversify their portfolios; traders monitor the 'Greek Letters' and adjust hedges periodically; senior managers analyze possible total losses; regulators set requirements. Any gain is sure to be followed by the corresponding exposure to risks. Only through identifying threats and evaluating all scenarios can we reduce losses in unexpected situations and maximize the realized profits. That is why the valid financial risk management is inevitable for a business success.

It seems people are more inclined to remember unfortunate events but not willing to think about them. In 'good times', we always take it for granted that our risk management is already successful and overestimate our capability of recognizing extreme events with small probabilities. The longer the 'good time' lasts, the more confident we are about what have been proposed: models, algorithms, and assumptions. Ironically, after any financial crisis, the first question people

asked is why the risk management failed. In fact, the prosperousness may be just spontaneous or at most times it is because the rare events are too infrequent to occur. However, when Black Swan events happen, they really happen no matter how highly improbable they should be. As Albert Einstein said, "As far as the laws of mathematics refer to reality, they are not certain, and as far as they are certain, they do not refer to reality." Reality is so complicated that people cannot fully grasp or even predict it. Everyone knows the next financial crisis, sooner or later, is sure to come. Therefore, rather than judge when and where it will occur, we are supposed to ask ourselves: whether our risk management is still solid under an assumed crisis. To answer that question, it is necessary to enhance the awareness of pros and cons of each model used in risk management and what is the suitable situation to adopt a specific method. Moreover, exploring and validating feasible alternatives is also indispensable for the progress in risk management.

Fortunately, that is what we are doing. Increasingly more attention has been paid to the field of risk management theses years. People learn lessons from crises so the Basel Accord is kept updating. Researchers are developing estimation methods such that the risks can be recognized more preciously. Once a new method is put forward, it needs to be revised based on market feedbacks. Furthermore, comparing it to the existing alternatives also helps us determine which is preferred under a given circumstance. This dissertation handles such tasks and focuses on one of the essential areas in the financial risk management: measures of risk.

Measures of risk quantify the risk so the abstract conception is transformed to an amount that is comprehensible and applicable in the management. One of the most commonly-used measures of risk is Value-at-Risk (VaR), which aggregates the total capital loss at a given confidence level with a time period. VaR is not only easy to calculate but also widely fit for all kinds of risks. Moreover, the confidence level of VaR can be selected based on the specific risk preference. However, it is criticized for the three major problems: inability to capture tail risk, model risk, and non-sub-additivity. Therefore, Basel II introduced stressed VaR and researchers also proposed many other alternatives among which the most promising one is Expected Shortfall (ES). ES, also referred to as conditional VaR and expected tail loss, is defined

to be the conditional expected loss during a period given the loss is greater than a certain quantile. Compared to VaR, ES is a coherent measure and remedies the tail risk problem. The Basel Committee also suggests a shift from VaR to ES under stress when calculating minimum capital requirements. The calculation of ES, however, is not as straightforward as VaR. Although so far there are a large number of ES estimation methods, selecting a suitable ES estimator is a tough decision in practice especially for small samples. Moreover, current ES estimators are unstable for a heavy-tailed loss distribution where they are easily affected by whether the infrequent losses would occur in a realized sample.

In the first essay, we propose a simple and robust ES estimation approach on a basis of the *tail-based* normal approximation that is determined by a relatively large number of tail dataset. This *tail-based* feature proves to be effective to alleviate the original 'underestimation' problem when estimating ES. Moreover, relevant regression models are subsequently adopted to enhance our estimator's accuracy. After the regression adjustment, our ES estimation approach not only works well for heavy-tailed loss distributions but also outperforms the widely-used arithmetic average and Extreme Value Theory ES estimators for small loss samples simulated from heavy-tailed distributions. The properties of our ES estimator on linear transformation further facilitate its application in the portfolio management.

Another measure of risk that has drawn much attention is the return volatility that describes the fluctuations of asset prices. In financial risk management, there is never too much emphasis on monitoring the market volatility. Though volatility cannot be observed or measured directly, people dare not ignore it no matter whether they are sellers or buyers, hedgers or speculators, participants or regulators. Market volatility rises, as does the risk of collapse. The global financial crisis in 2008, the tech bubble burst in the late 1990s, the stock market crash on October 19, 1987, and the recent stock plunge on February 5, 2018, are all examples of such knock-on effect. Furthermore, volatility is a fundamental parameter when calculating many prevalent measures of risk such as the variance-covariance approach for VaR. It also plays an indispensable role in pricing derivatives. As a result, how to precisely estimate and then forecast volatility has been

an enduringly popular field of research.

So far, there are three main types of volatility models: ARCH/GARCH-family models, stochastic volatility (SV) models and realized volatility models. The calculation of the realized volatility needs to be supported by high-frequency data which is not always available. Therefore, in this dissertation, we limit our attention to the first two types. The crucial difference between these two volatility models is that in ARCH/GARCH-family models the volatility is a deterministic response variable of the previous information set while in SV models, the volatility is assumed to be a latent variable with unexpected innovation terms. The ARCH/GARCH-family models appeal to a wider variety of studies because they are straightforward to estimate by maximum likelihood estimation. Estimating the SV models, by contrast, is indeed time-consuming. Despite extra flexibility in capturing stylized facts of financial series, SV models have a troublesome problem resulted from the latent volatility process: no analytical likelihood function. Estimating the SV models relies on simulation methods but the computation burden often impedes its application. Thanks to the amazing breakthroughs in computing capacity and simulation methods in recent years, the SV models are considerably faster to estimate. Hence, given the two choices, people are interested in which model can bring about more accurate volatility estimate and forecast.

Our second essay carries out a comprehensive comparison of the GARCH(1, 1) and autoregressive stochastic volatility (ARSV) models under both the physical and risk-neutral measures. As for the ARSV(1) model, we adopt a novel particle-based 'forward-only' version of the Forward Filtering Backward Smoothing algorithm with the off-line expectation-maximization method to estimate and subsequently maximize its log likelihood. Under the physical measure, we measure their log likelihoods[1] after fitting historical returns and test normality for their error sequences. In addition, we also refer to two robust loss functions, MSE and QLIKE, for a comparison of the one-step-ahead volatility forecasts of these two models. On the other hand, under the risk-neutral measure, the in-sample and out-of-sample option pricing errors are explored. Though

[1]These two models have the same number of parameters.

the ARSV(1) model has two innovation processes, we show only the volatility dynamics need to be simulated with the assistance of the famous Black-Scholes formula when pricing options. This discovery ensures that estimating the risk-neutral versions of such two models have almost same computation burdens. The results indicate the ARSV(1) model outperforms the GARCH(1, 1) model in terms of the in-sample and out-of-sample performances under the physical measure. Under the risk-neutral measure, the ARSV(1) model is significantly superior to the GARCH(1, 1) model for put options while they are considerably close to each other when pricing call options.

Financial times series possess some patterns that are crucial for volatility modeling. Such empirical patterns are called stylized facts in economics. One prevalent stylized fact is the volatility clustering, which is proved to get captured by both the GARCH(1, 1) and ARSV(1) models. It refers to the findings that large variations in price tend to be followed by large price variations. Another fundamental stylized fact is the leverage effect which indicates that the volatility is negatively correlated with the asset return. The ARSV(1) model is natural to incorporate the leverage effect through a pair of negatively correlated error terms in the return and volatility processes. By contrast, the GARCH(1, 1) model has to modify the specification of its conditional variance to capture the leverage effect. Two commonly-used candidates are the Exponential GARCH(1, 1) and Nonlinear Asymmetric GARCH(1, 1) models.

Enlightened by the Constant Elasticity of Variance (CEV) model (Cox, 1975), in the third essay, we put forward a continuous-time option pricing model that straightforwardly relates the volatility to the exponentially decayed weighted average (EDWA) cumulative asset return such that the leverage effect is captured. The new information has little influence on the EDWA cumulative asset return so our option pricing model is not sensitive to the initial conditions and expected to be robust for the out-of-sample prediction. In addition to Monte Carlo simulation, our Cumulative Return model is also capable of pricing options by solving the corresponding partial differential equations numerically. These favorable features raise the need to compare it to other sophisticated models regarding pricing options. Then we investigate the option pricing performances of the risk-neutral models with the leverage effect. The implied versions of the

GARCH(1, 1) and ARSV(1) models with the leverage effect, which parameterize the initial volatility, are also studied. Compared to the original risk-neutral GARCH(1, 1) and ARSV(1) models, both of the two models exhibit remarkable improvements regarding the in-sample and out-of-sample pricing errors after incorporating the leverage effect. Therefore, the leverage effect plays an indispensable role in option pricing. Surprisingly, our Cumulative Return model that has only two parameters are superior to other sophisticated models in terms of the out-of-sample price prediction for call options. When pricing put options, the differences among most examined models are considerably insignificant. Moreover, the performance of our option pricing model on an outlier day further highlights its robustness.

1.2 Contribution of the Research

This dissertation focuses on the ES estimator and volatility models. The contributions of our research are summarized as follows:

For the ES estimator:

- Propose an idea of *tail-based* normal approximation for ES estimation.
- Develop a regression-adjusted *tail-based* ES estimation approach based on the tail weight.
- The approach is easy to implement and robust for various heavy-tailed distributions.
- Our ES estimator is substantially accurate and works well for small loss samples. In addition, its properties on linear transformation are valuable in the portfolio management

For the volatility models:

- Demonstrate that only the volatility dynamics need to be simulated for option pricing using the ARSV(1) model which has two innovation processes no matter whether the leverage effect is incorporated or not.

- Complement the literature by conducting a comprehensive comparison of the GARCH(1, 1) and ARSV(1) models regarding in-sample fitting and out-of-sample prediction capabilities under both the physical and risk-neutral measures.

- Propose a continuous-time option pricing model that straightforwardly relates the volatility to the EDWA cumulative asset return.

- Investigate the option pricing performances of the risk-neutral models incorporating the leverage effect. Our proposed model is robust for the out-of-sample option pricing.

1.3 Outline of the Dissertation

The rest of this dissertation is organized as follows. Chapter 2 reviews the existing literature that is related to our research. In Chapter 3, we propose a simple and robust approach for expected shortfall estimation. Chapter 4 carries out a comprehensive comparison of the GARCH(1, 1) and ARSV(1) models under both the physical and risk-neutral measures. In Chapter 5, we investigate the option pricing performances of volatility models incorporating the leverage effect and put forward a new option pricing model that relates the volatility to the EDWA cumulative asset return.

CHAPTER 2

Literature Review

In this chapter, we review the literature that is relevant to our dissertation. Section 2.1 summarizes the existing Expected Shortfall (ES) estimation and back-test methods. In Section 2.2, we briefly review the ARCH/GARCH-family and stochastic volatility (SV) models with their extensions. The leverage effect, one of the most prevalent stylized facts, is also reviewed. In Section 2.3, we review the algorithms on particle filter, which is also referred to as sequential Monte Carlo simulation. Such methods can be applied to the estimation of the autoregressive stochastic volatility (ARSV) model under the physical measure.

2.1 Estimating and Back-Testing Expected Shortfall

Although Value-at-Risk (VaR) is a popular measure of risk, it fails to take into consideration the extreme losses and suffers the lack of subadditivity (Artzner, Delbaen, Eber, & Heath, 1997). Then ES, which remedies tail risk, is developed by Artzner, Delbaen, Eber, & Heath (1999) as a natural coherent alternative to VaR.

ES can be estimated by definition from a parametric loss distribution or by extreme value theory (EVT) approach (McNeil, 1997). In addition to Generalized Pareto distribution (GPD) supported by EVT, other asymmetric loss distributions such as skewed normal (Bernardi, 2013), asymmetric t and exponential power distributions (Zhu & Galbraith, 2011), Tukey g-and-h (Jiménez & Arunachalam, 2011), and Laplace (Q. Chen, Gerlach, & Lu, 2012) have also been studied. In light of the location-scale loss distributions, Bae & Iscoe (2012) propose various large-sample parametric confidence intervals for ES estimates. Moreover, Embrechts, Kaufmann, & Patie (2005) introduce a multiple-period ES estimation method within frameworks of random walks, autoregressive process or the GARCH(1, 1) model with t distributed innovations. Several quantile regression-based ES estimators are given by J. W. Taylor (2007) with the benefit of avoiding any distributional assumptions. As for the mixture distributions, normal and t mixture distributions and their closed-form ES estimates are investigated by Broda & Paolella (2011).

Compared to parametric methods, non-parametric ES estimators can avoid the errors of parameter estimation. A commonly-used non-parametric estimator is defined by S. X. Chen (2007) as the arithmetic average of losses that are beyond a specific VaR estimator, which is similar to the natural estimator for expected losses in the worst case proposed by Acerbi & Tasche (2002a). Hill (2013) suggests a few ES estimators for heavy-tailed data by combining the classic tail-trimming with an improved bias-correction method. Inui & Kijima (2005) demonstrate that ES is a priority in the class of coherent risk measures and propose an extrapolation method to estimate ES. Applying empirical likelihood, an asymptotically valid confidence interval for ES is derived by Baysal & Staum (2008). In addition, Scaillet (2004) comes up with a kernel-smoothed ES estimate that works well even in dependent situations. The kernel smoothing produces a VaR estimator with less mean square errors especially for small samples (S. X. Chen & Tang, 2005), while S. X. Chen (2007) shows such smoothing can not give a more accurate ES estimator. Therefore, taking advantage of its simplicity, the arithmetic average of exceedances is still one of the most popular non-parametric ES estimators.

Despite the favorable properties of ES, it is much more difficult to back-test an ES estimation

method than to back-test a VaR estimation method. Gneiting (2011) shows ES lacks elicitability that is a necessity for back-testing. Going on with this finding, Carver (2013) and many others also claim ES is not back-testable. Emmer, Kratz, & Tasche (2015), however, demonstrate ES is conditionally elicitable and point out back-testing ES, though not straightforward, is still possible. They also propose a back-test method based on an ES estimation approximated by VaRs at several confidence levels. Moreover, Wong (2008) adopts the saddlepoint technique to back-test ES. Acerbi & Szekely (2014) also introduce three non-parametric and model-independent back-test methods for ES. All the three methods can be easily implemented without assuming any asymptotic convergence. Costanzino & Curran (2015) discuss the coverage test for ES with a defined failure rate. With a sufficiently large sample size, they show the distribution of the defined ES failure rate at a given level can be well approximated by a normal distribution. Hence, they propose an ES back-test method that is analogous to the Traffic Light approach for back-testing VaR. Fissler, Ziegel, et al. (2016) argue that ES is jointly elicitable with VaR. Such conclusion makes it possible in theory to back-test ES together with VaR.

2.2 Volatility Models and Leverage Effect

Though volatility can be simply defined to equal the standard deviation of the returns, dynamic volatility models are preferred due to more accurate estimations. A typical group of such models belongs to the ARCH/GARCH family. Engle (1982) introduces the ARCH model that fits the conditional variance through a linear regression equation of previous squared errors. To avoid negative coefficients in the long lagged ARCH orders, Bollerslev (1986) offers the GARCH model so that a more flexible lag structure is considered. Though the GARCH process is able to capture excess kurtosis and volatility clustering, two important characteristics of market data, there are still some criticisms of its limitations. The strongest one is about its failure to exhibit the leverage effect (Black, 1976; Christie, 1982), which indicates an asset return is negatively correlated with the volatility. To overcome this inherent drawback, the EGARCH (Nelson, 1991) and GJR-GARCH (Glosten, Jagannathan, & Runkle, 1993) models are proposed. In addition,

Engle & Ng (1993) put forward the NGARCH and VGARCH models and examine empirical outputs of several popular asymmetric GARCH models. Leverage effect aside, the heavy-tailed feature of empirical returns also spark skepticism of the normally distributed error term in the GARCH model. Instead of the original normal distribution, some Leptokurtic distributions such as standardized t distribution (Bollerslev, 1987), exponential generalized beta family (Wang, Fawson, Barrett, & McDonald, 2001) and Generalized Error Distribution (Nelson, 1991) have been explored with the hope of fully capturing both the negative skewness and excess kurtosis.

Strictly speaking, given the previous information, volatility estimated from the ARCH/GARCH family is just a deterministic response variable. By contrast, stochastic volatility (SV) models, as its name implies, allow for a stochastic volatility process in a discrete-time or continuous-time form. Tauchen & Pitts (1983) indicate daily price volatility is independently linked to a Lognormal mixing variable after studying relationships between price variability and volume from the T-bills future market. Meanwhile, treating volatility as a latent variable, S. Taylor (1982) proposes a discrete-time ARSV model in which the log-variance follows a linear Gaussian process. Taylor's work is seen as the first published model that explicitly deals with volatility modeling in finance. Moreover, the ARSV model could capture the leverage effect naturally though the author himself does not explore such relevant traits. Continuous-time volatility models are unveiled afterwards especially in the field of option pricing. To handle volatility clustering, Hull & White (1987) generalize the famous Black-Scholes option pricing model and assume both price and variance follow Gaussian Ornstein-Uhlenbeck processes with correlated increments. Heston (1993) also builds an SV model where the increments for the variance process depend on the square root of variance with an extraordinary advantage—an analytical option pricing formula. As high-frequency data research is brought to center stage, more complicated SV models have been constructed. For example, multiple SV factors with jump components (Chernov, Gallant, Ghysels, & Tauchen, 2003) are considered and a fractional integration of the square root volatility process (Comte, Coutin, & Renault, 2012) is adopted.

In order to exploit the high-frequency intra-day data, which are not valuable for the traditional

volatility models at the daily level, realized volatility (RV) models have been developed in recent years (Andersen, Bollerslev, Diebold, & Labys, 2001). RV is proved to be an unbiased and efficient estimator of the return volatility. It is also straightforward to calculate RV by summing up squares of intra-day returns with short intervals. Such benefits spark researchers' interests in the application and extension of RV. For example, Andersen, Bollerslev, Diebold, & Ebens (2001) examine the realized return volatilities and the correlations of individual stocks in Dow Jones Industrial Average. They find the unconditional distributions of realized volatilities seem to be Lognormal while the daily returns standardized by realized volatilities nearly follow normal distributions. This finding is in line with the results on FTSE-100 index futures contracts obtained by Areal & Taylor (2002). Due to the lack of high-frequency data, we focus on the ARCH/GARCH-family and SV models in this dissertation.

2.3 Particle Filter

Except for special cases such as linear Gaussian and hidden finite state space Markov chains, it is impossible to evaluate the posterior distributions analytically. Particle Filter is a class of methods for posterior distribution approximation and parameter estimation in the state space models such as the ARSV(1) model.

Particle filtering estimates the current state variable. Johannes & Polson (2009) propose the exact particle filter for linear Gaussian models where the direct sampling is feasible. For non-linear or non-Gaussian models, Kong, Liu, & Wong (1994) introduce the sequential importance sampling (SIS) method to estimate marginal distribution of the latent state variables using weighted particles. One problem with the SIS method is that it is easy to encounter a situation where almost all particles have nearly zero weights. This problem is referred to as the degeneracy problem and can be solved by a re-sampling procedure. Adding a re-sampling step to SIS leads to sequential importance re-sample (SIR) methods (Gordon, Salmond, & Smith, 1993; Kitagawa, 1996; Doucet, Godsill, & Andrieu, 2000). In many SIR methods, the re-sampling step is performed only when it is necessary. J. S. Liu & Chen (1995) use an 'effective number' of

particles to help judge whether a re-sampling is needed at a step. The bootstrap filter (BF) (Gordon, Salmond, & Smith, 1993) is a special case of SIR methods where the importance distribution is same as the transition distribution. This setting makes its implementation indeed easy. However, it may require a large number of simulation samples for accurate estimation due to the inefficiency of that special importance distribution. In addition, BF asks for re-sampling at each time step. Introducing an auxiliary index into the particle filter, Pitt & Shephard (1999) suggest the auxiliary particle filter that is more flexible and reliable than BF.

Particle smoothing approximates the distribution of the state variable in the past. It is straightforward to solve this problem by marginalizing posterior distributions obtained from particle filtering while some other alternatives such as fixed-lag smoother (Kitagawa, 1996), backward smoothing (Doucet, Godsill, & Andrieu, 2000), backward simulation (Doucet, Godsill, & West, 2000) and generalized two-filter smoother (Briers, Doucet, & Maskell, 2010) are capable of better solving the degeneracy problem.

Particle learning methods provide a particle-based parameter estimation. J. Liu & West (2001) see parameters as state variables and apply the shrinkage of kernel locations. Del Moral, Doucet, & Singh (2010) recursively estimate parameters by maximizing the likelihood using a particle-based 'forward-only' version of the Forward Filtering Backward Smoothing algorithm with the on-line or off-line expectation-maximization method. Other particle-based maximum likelihood approaches such as the gradient ascent method (Poyiadjis, Doucet, & Singh, 2011) and iterated filtering (Ionides, Bhadra, Atchadé, King, et al., 2011) are also studied by many researchers.

CHAPTER 3

A Simple and Robust Approach for Expected Shortfall Estimation

3.1 Introduction

Expected Shortfall (ES), which takes into account losses exceeding the corresponding Value-at-Risk (VaR), is a coherent measure of risk. As ES remedies tail risk and non-sub-additivity, problems VaR inherently suffers (Artzner, Delbaen, Eber, & Heath, 1999), it has been attracting more and more attention in the risk management recently.

Selecting a suitable ES estimator is a challenge in practice. As for the most commonly-used arithmetic average ES estimator, the given sample of losses, however, is not always large enough to acquire a convincing estimator. Assuming the sample size is 250, the amount of one year's daily observations, only the largest two losses are covered when calculating daily $ES_{99\%}$[1] and only the maximum one is valuable for daily $ES_{99.5\%}$. Moreover, according to Yamai & Yoshiba (2005), such ES estimator is quite unstable especially for a heavy-tailed loss distribution where it is easily affected by whether infrequent losses would occur in the realized sample.

[1] ES_β denotes the ES estimate at β confidence level during a given period.

Normal distributions have many nice properties with the presence justified by the Central Limit Theorem (CLT). That's why it remains popular when modeling data. Such distributions, however, cannot capture the heavy-tailed[2] behaviors properly. If a model building approach with normal distributions is applied to ES estimates, we usually suffer the so-called 'underestimation' problem. Heavy-tailed distributions such as student's t and other stable distributions have also been investigated but some inherent drawbacks impede their applications. For example, the sum of two t distributed random variables generally no longer belongs to t distribution. On the other hand, even though the sum of two stable random variables still follows a stable distribution, there is usually no general explicit formula for the probability density function. Therefore, some numerical methods must be used (Pang & Yang, 2015).

In this chapter, we first propose a new ES estimator on a basis of the *tail-based* normal approximation that is determined by a relatively large number of tail data. Compared to the *global-based* normal approximation, this *tail-based* feature proves to be quite effective to alleviate the 'underestimation' problem. To further improve the estimation accuracy, we then introduce an adjusted *tail-based* normal approximation, in which the sample tail weight is taken into consideration. Moreover, we carry out robustness tests and the results show our adjusted *tail-based* ES estimator works well for various heavy-tailed loss distributions. Moreover, it also outperforms the popular sample arithmetic average estimator and EVT estimator for small loss samples. Another appealing feature is that our ES estimator still works under linear transformations.

Hereafter, for the ease of presentation, the abbreviation 'r.v.' stands for random variable. The rest of this chapter is organized as follows: In Section 3.2, we propose a *tail-based* normal approximation and derive the explicit formula of its ES estimator. The accuracy analysis and the comparison with the traditional *global-based* normal approximation are also provided. In Section 3.3, we adjust the *tail-based* normal approximation based on the sample tail weight, and explore whether such adjustment leads to a more accurate ES estimation for loss r.v.'s

[2]In this chapter, a heavy-tailed distribution refers to any loss distribution that has a heavier right tail than the normal distribution.

with heavy-tailed distributions. In Section 3.4, we run self consistence and robustness tests to investigate the efficiency of the adjusted ES estimator. Section 3.5 compares the adjusted ES estimator to the widely-used arithmetic average estimator and EVT estimator for small samples. Moreover, we study the effects of linear transformation on the proposed ES estimator in Section 3.6. Section 3.7 summarizes this chapter and suggests the directions of the future work. All detailed derivations are included in the Appendix A.

3.2 Tail-Based Normal Approximation for ES Estimation

For a loss (or negative return) r.v. L, Acerbi & Tasche (2002b) defines its ES at the confidence level $\beta \in (0,1)$ for a corresponding period as follows:

$$\mathrm{ES}_\beta(L) = \frac{1}{1-\beta}\int_\beta^1 \mathrm{VaR}_\phi(L)d\phi, \tag{3.2.1}$$

where $\mathrm{VaR}_\phi(L)$, the VaR at level $\phi \in (0,1)$, is defined by

$$\mathrm{VaR}_\phi(L) = \inf\{z \in R | Pr(L \le z) \ge \phi\}. \tag{3.2.2}$$

If the loss r.v. L is continuously distributed with a PDF $f(\cdot)$, it can be shown Eq. (3.2.1) is also equivalent to Eq. (3.2.3) through a variable transformation.

$$\mathrm{ES}_\beta(L) = \mathbb{E}[L|L \ge \mathrm{VaR}_\beta(L)] = \frac{1}{1-\beta}\int_{\mathrm{VaR}_\beta(L)}^\infty xf(x)dx. \tag{3.2.3}$$

In traditional model building approaches, all available sample points are used when estimating the distribution parameters. For example, a normal approximation is usually obtained by matching its mean and variance to the sample mean and sample variance. This type of approaches are *global-based* in the sense that all sample points are taken into consideration.

ES, however, is a statistic that mainly depends on the tail data. Therefore, *global-based* approaches may not give efficient and accurate ES estimations. Actually, this is one of the

reasons why the *global-based* normal approximation typically underestimates the ES for real market data, especially at high confidence levels such as 99% and 99.5%.

In this section, we propose a *tail-based* approach that only considers the tail sample points. In particular, focusing on the excessive observations, we build a *tail-based* normal approximation by equating its specific quantile (e.g. 95%-quantile) and mean excess square to those of the objective sample. This approximation will be further improved in the next section through some adjustment factors related to the sample's tail weight.

We want to point out that the idea of *tail-based* approximation can be applied to distributions other than normal distributions, such as student's t, Gamma, etc. In this chapter, we only consider normal distributions for the following reasons. Firstly, normal distributions are simple and have many nice properties. For example, the sum of two normal r.v.'s is still a normal r.v., which is useful in risk management when calculating n-day ES (or VaR) based on daily estimates. Secondly, the *tail-based* normal approximation gives considerably accurate results so it might not be necessary to explore other distributions.

3.2.1 Explicit Formulas

Let $\mathbf{Y} = \{y_n\}_{n=1}^N$ denote a sample of losses with a given time unit and our goal is to develop a model-based approach to estimate its ES_β. β is usually close to 1, and two popular choices are 99% and 99.5%. Apparently, this approach should depends on the right-tail sample because we look at losses instead of returns.

Firstly, for any $\alpha \in (0,1)$, we define A_α, the α-quantile of the sample $\mathbf{Y}$, as follows:

$$A_\alpha \equiv (\lfloor N\alpha \rfloor + 1 - N\alpha) y_{(\lfloor N\alpha \rfloor)} + (N\alpha - \lfloor N\alpha \rfloor) y_{(\lfloor N\alpha \rfloor + 1)}, \tag{3.2.4}$$

where $\lfloor N\alpha \rfloor$ represents the greatest integer that is less than or equal to $N\alpha$ (i.e. $\lfloor \cdot \rfloor$ is the floor function), and $y_{(1)}, y_{(2)}, \cdots, y_{(N)}$ are the ascending order statistics of the sample $\mathbf{Y}$. In case of possible misunderstandings, 'α-quantile' ($0 < \alpha < 1$) in this chapter is equivalent to (100α)th

percentile. For example, 0.75-quantile or 75%-quantile is equivalent to 75th percentile.

Next, we choose a threshold level α that is less than β (e.g. $\alpha = 0.95$ when $\beta = 0.99$) and define a normal r.v. $X \sim N(\mu, \sigma^2)$ to approximate the right tail of the sample $\mathbf{Y}$ beyond the α-quantile[3] A_α; that is, we are going to find a *tail-based* normal approximation for the given sample. In particular, we solve μ and σ^2 such that the following two tail statistics, the α-quantile and the 'conditional tail variance' (mean excess square) match the corresponding sample statistics:

$$Pr(X \leq A_\alpha) = \alpha, \qquad \mathbb{E}[(X - A_\alpha)^2 | X > A_\alpha] = \frac{\sum_{n=1}^{N} (y_n - A_\alpha)^2 \mathbb{1}_{\{y_n > A_\alpha\}}}{\sum_{n=1}^{N} \mathbb{1}_{\{y_n > A_\alpha\}}}, \tag{3.2.5}$$

where $\mathbb{1}_{\{\cdot\}}$ is the indicator function.

We can derive a unique solution (μ, σ^2) for the system of equations listed above. Define $Z \equiv \frac{X-\mu}{\sigma}$, and then Z follows the standard normal distribution. Let $\Phi(\cdot)$ be the CDF and z_α be the z-score of the standard normal distribution; that is, $z_\alpha = \Phi^{-1}(\alpha)$. Thus the first equation in Eq. (3.2.5) can be transformed as follows:

$$Pr(X \leq A_\alpha) = Pr\left(\frac{X-\mu}{\sigma} \leq \frac{A_\alpha - \mu}{\sigma}\right) = Pr\left(Z \leq \frac{A_\alpha - \mu}{\sigma}\right) = \Phi\left(\frac{A_\alpha - \mu}{\sigma}\right) = \alpha. \tag{3.2.6}$$

So we can get

$$\frac{A_\alpha - \mu}{\sigma} = \Phi^{-1}(\alpha) = z_\alpha. \tag{3.2.7}$$

The conditional first and second moments of X (See Appendix A.1 for details) are calculated by

$$\begin{aligned} \mathbb{E}[X | X > A_\alpha] &= \mathbb{E}\left[\sigma Z + \mu \middle| Z > \frac{A_\alpha - \mu}{\sigma}\right] = \mu + \frac{\sigma}{(1-\alpha)\sqrt{2\pi}} e^{-\frac{(A_\alpha - \mu)^2}{2\sigma^2}}, \\ \mathbb{E}[X^2 | X > A_\alpha] &= \mathbb{E}\left[\mu^2 + \sigma^2 Z^2 + 2\mu\sigma Z \middle| Z > \frac{A_\alpha - \mu}{\sigma}\right] \end{aligned} \tag{3.2.8}$$

[3] In this chapter, A_α stands for the α-quantile for either a loss sample or a loss r.v. based on the specific situation.

$$= \mu^2 + \sigma^2 + \frac{\sigma(A_\alpha + \mu)}{(1-\alpha)\sqrt{2\pi}} e^{-\frac{(A_\alpha - \mu)^2}{2\sigma^2}}. \tag{3.2.9}$$

By virtue of Eq. (3.2.7)∼Eq. (3.2.9), we have

$$\begin{aligned}
\mathbb{E}[(X - A_\alpha)^2 | X > A_\alpha] &= \mathbb{E}[X^2 | X > A_\alpha] - 2A_\alpha \mathbb{E}[X | X > A_\alpha] + A_\alpha^2 \\
&= (\mu - A_\alpha)^2 + \sigma^2 + \frac{\sigma(\mu - A_\alpha)}{(1-\alpha)\sqrt{2\pi}} \exp\big(-\frac{(A_\alpha - \mu)^2}{2\sigma^2}\big) \\
&= \sigma^2 \left[z_\alpha^2 + 1 - \frac{z_\alpha}{(1-\alpha)\sqrt{2\pi}} e^{-\frac{1}{2}z_\alpha^2} \right].
\end{aligned} \tag{3.2.10}$$

From the above derivation (see Eq. (3.2.10), Eq. (3.2.7) and Eq. (3.2.5)), we can get the explicit formulas for the parameters, μ and σ, of the *tail-based* normal approximation using the sample dataset $\{y_n\}_{n=1}^N$ as follows:

$$\begin{aligned}
\hat{\sigma}^2 &= \frac{1}{\left[z_\alpha^2 + 1 - \frac{z_\alpha}{(1-\alpha)\sqrt{2\pi}} e^{-\frac{1}{2}z_\alpha^2} \right]} \left[\frac{\sum_{n=1}^N (y_n - A_\alpha)^2 \mathbb{1}_{\{y_n > A_\alpha\}}}{\sum_{n=1}^N \mathbb{1}_{\{y_n > A_\alpha\}}} \right], \\
\hat{\mu} &= A_\alpha - \hat{\sigma} z_\alpha.
\end{aligned} \tag{3.2.11}$$

In practice, when calculating the ES estimate of the loss sample $\mathbf{Y} = \{y_n\}_{n=1}^N$ at confidence level β (e.g. $\beta = 99\%$ or 99.5%), we first derive the *tail-based* normal approximation given by Eq. (3.2.11) with a threshold level α that is less than β (e.g. $\alpha = 95\%$).

Once the approximation normal distribution is solved, it can be used to calculate the ES_β by

$$\mathrm{VaR}_\beta = \hat{\mu} + \hat{\sigma} z_\beta, \tag{3.2.12}$$

$$\mathrm{ES}_\beta = \mathbb{E}[X | X > \mathrm{VaR}_\beta] = \hat{\mu} + \frac{\hat{\sigma}}{(1-\beta)\sqrt{2\pi}} e^{-\frac{1}{2}z_\beta^2}. \tag{3.2.13}$$

where $z_\beta = \Phi^{-1}(\beta)$ is the z-score of the standard normal distribution at confidence level β.

Furthermore, for the purpose of testing accuracy, we apply the *tail-based* normal approximation to a loss r.v. with an explicit distribution function. As for a loss r.v. W with CDF $F_W(\cdot)$

and PDF $f_W(\cdot)$, its α-quantile A_α and mean excess square are calculated by

$$A_\alpha = F_W^{-1}(\alpha), \qquad \mathbb{E}[(W - A_\alpha)^2 | W > A_\alpha] = \frac{1}{1-\alpha} \int_{A_\alpha}^{\infty} (x - A_\alpha)^2 f_W(x) dx. \tag{3.2.14}$$

Then we can obtain the *tail-based* normal approximation for the loss r.v., W, by virtue of Eq. (3.2.10), Eq. (3.2.7) and Eq. (3.2.5), and the explicit formulas for $\hat{\mu}$ and $\hat{\sigma}$ are

$$\begin{aligned} \hat{\sigma}^2 &= \frac{1}{\left[z_\alpha^2 + 1 - \frac{z_\alpha}{(1-\alpha)\sqrt{2\pi}} e^{-\frac{1}{2}z_\alpha^2}\right]} \mathbb{E}[(W - A_\alpha)^2 | W > A_\alpha], \\ \hat{\mu} &= A_\alpha - \hat{\sigma} z_\alpha, \end{aligned} \tag{3.2.15}$$

where A_α and $\mathbb{E}[(W - A_\alpha)^2 | W > A_\alpha]$ are given by Eq. (3.2.14).

3.2.2 Accuracy Analysis for Tail and Global Based Normal Approximations

Now let us test the accuracy of the *tail-based* normal approximation. Firstly, we compare the *global-based* and *tail-based* normal approximations for the daily loss sequence of S&P 500 Index between Jan 4, 2010, and Jun 30, 2017. The daily loss at Day m is calculated by $-\ln(P_m/P_{m-1})$, where P_m is the index value at Day m. As previously mentioned, the *global-based* normal approximation is determined by matching its mean and variance to the sample mean and sample variance. The *tail-based* normal approximation is obtained by Eq. (3.2.11). The results are displayed in the form of Q-Q probability plots in Figure 3.1.

As we can see from the figure, with respect to the right tail, the *tail-based* normal approximation is much closer to the given daily loss dataset than the *global-based* counterpart. Since only the right tail is used for ES estimation, the *tail-based* normal approximation is expected to give a more accurate ES estimator than the one from global-based normal approximation.

To further test the accuracy of our *tail-based* normal approximation, we examine its performances for some heavy-tailed loss distributions that are widely adopted in empirical studies. The idea is that, for a r.v. W, with a known distribution (such as t, Lognormal, etc.), we obtain its *tail-based* normal approximation based on Eq. (3.2.15), and then use formulas Eq. (3.2.12)

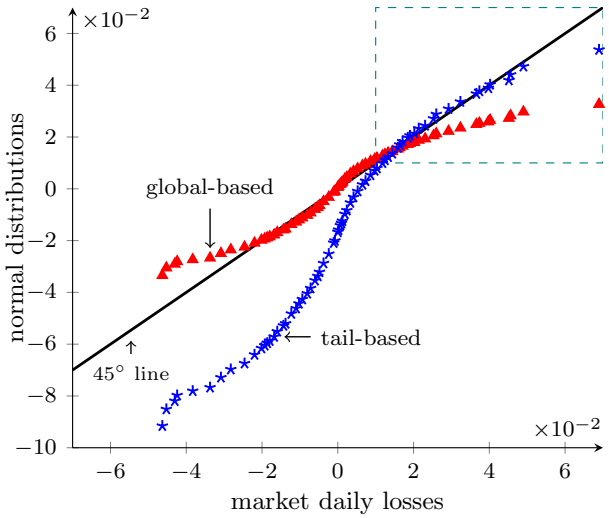

Figure 3.1: Q-Q Plots: *global* & *tail-based* normal approximations vs. realized losses with $\alpha = 95\%$.

and Eq. (3.2.13) to drive its VaR estimator and ES estimator which are denoted by VaR_β^t and ES_β^t, respectively. Here the superscript t stands for *tail-based*. The theoretical (true) values of the VaR and ES estimates can be calculated through the PDF of the actual distribution and they are denoted by VaR_β and ES_β, respectively.

To quantify the estimation errors, we define the relative errors for the *tail-based* normal approximation as follows:

$$\mathbf{e}_\beta^t(\text{ES}) \equiv \frac{\text{ES}_\beta - \text{ES}_\beta^t}{\text{ES}_\beta}, \qquad \mathbf{e}_\beta^t(\text{VaR}) \equiv \frac{\text{VaR}_\beta - \text{VaR}_\beta^t}{\text{VaR}_\beta}. \tag{3.2.16}$$

From the above definitions, we can see that if the relative error is positive, there is an underestimation while if it is negative, there is an overestimation. Moreover, for comparison reasons, we also obtain the traditional *global-based* normal approximation by matching its mean and variance to the corresponding statistics of W. Its VaR and ES estimators are denoted by VaR_β^g and ES_β^g where the superscript g stands for *global-based*. Similarly, the relative errors for the

global-based normal approximation are defined by

$$\mathbf{e}_\beta^g(\mathrm{ES}) \equiv \frac{\mathrm{ES}_\beta - \mathrm{ES}_\beta^g}{\mathrm{ES}_\beta}, \qquad \mathbf{e}_\beta^g(\mathrm{VaR}) \equiv \frac{\mathrm{VaR}_\beta - \mathrm{VaR}_\beta^g}{\mathrm{VaR}_\beta}. \tag{3.2.17}$$

To learn more about the performance of the *tail-based* normal approximation, we test it with heavy-tailed distributions including t, Gamma, Lognormal, GPD, and Weibull with different shape parameters. Relative errors for ES and VaR estimates at $\beta = 99\%$ or 99.5% of the *tail-based* and *global-based* normal distributions are given in Table 3.1, which shows the *tail-based* normal approximation has substantially smaller ES estimation errors than the *global-based* counterpart for all examined loss distributions.

Table 3.1: Estimation errors of *tail-based* and *global-based* normal approximations, $\alpha = 95\%$.

W	β (%)	ES_β	$\mathbf{e}_\beta^g(\mathrm{ES})$ (%)	$\mathbf{e}_\beta^t(\mathrm{ES})$ (%)	VaR_β	$\mathbf{e}_\beta^g(\mathrm{VaR})$ (%)	$\mathbf{e}_\beta^t(\mathrm{VaR})$ (%)
t, df=3.5	99	5.895	30.94	-4.848	4.061	12.49	-19.84
	99.5	7.290	39.41	3.152	5.086	22.63	-14.72
t, df=8	99	3.591	14.30	0.121	2.896	7.258	-4.023
	99.5	4.083	18.22	2.770	3.355	11.36	-2.380
Gamma(0.1, 1)	99	2.349	59.85	1.175	1.589	47.39	-9.720
	99.5	2.887	64.86	6.214	2.095	56.34	-3.526
Gamma(3, 1)	99	9.639	20.98	0.303	8.406	16.38	-1.225
	99.5	10.49	23.62	1.332	9.274	19.54	-0.489
Lognormal(0, 1.1^2)	99	20.15	53.75	-3.925	12.92	35.26	-23.92
	99.5	25.64	61.16	5.455	17.00	47.67	-15.51
Lognormal(0, 0.3^2)	99	2.235	14.91	0.225	2.010	10.79	-1.179
	99.5	2.391	17.42	1.237	2.166	13.52	-0.564
GPD(0.3,1)	99	15.62	52.33	-7.747	9.937	32.74	-29.38
	99.5	20.01	60.21	2.547	13.00	44.27	-21.38
GPD(0.1,1)	99	7.610	41.89	0.179	5.849	31.59	-6.576
	99.5	8.874	46.99	4.121	6.987	38.30	-3.584
Weibull(1,0.6)	99	17.99	52.45	0.339	12.75	39.92	-10.16
	99.5	21.77	57.96	5.711	16.10	48.34	-4.972
Weibull(1,1.4)	99	3.415	21.84	0.262	2.977	17.83	-0.859
	99.5	3.714	24.10	1.005	3.290	20.66	-0.290

The improvements in ES estimation using the *tail-based* normal approximation are associated with overestimated VaR values (see the last column in Table 3.1). The overestimates of VaR are

necessary to compensate for the originally underestimated ES estimates because the normal distribution has a lighter right tail than all the examined distributions.

Though the *tail-based* normal approximation gives more accurate ES estimates than the *global-based* counterpart, its errors are not considerably small and further improvements are thus needed. In addition, it seems that the errors have some dependence on the tail weight (shape) parameters of those distributions. Therefore, we consider an adjustment factor related to the tail weight statistics in the next section.

3.3 Adjusted Tail-Based Normal Approximation

For real-time financial data, loss distributions are usually heavy-tailed so extreme losses are more frequent than those modeled by normal distributions. Table 3.1 indicates the estimation errors of the *tail-based* normal approximation depend on the tail weight of its actual distribution. Therefore, to further decrease the estimation errors, now we propose an adjusted *tail-based* normal approximation.

Let T stand for a loss r.v. with a given distribution and β be the confidence level of the ES estimate. We pick a value of α such that $\alpha < \beta$ and the α-quantile of T is denoted by A_α again. Assuming X is the r.v. of the *tail-based* normal approximation of T obtained by Eq. (3.2.15), a ratio $R_{\alpha,\beta}$ that measures estimation errors is defined by

$$R_{\alpha,\beta} \equiv \frac{\mathrm{ES}_\beta(T) - A_\alpha}{\mathrm{ES}_\beta(X) - A_\alpha}, \tag{3.3.1}$$

where $\mathrm{ES}_\beta(X)$ is obtained by Eq. (3.2.13) and $ES_\beta(T)$ is the theoretical β-level ES estimate of T. The reason why we subtract A_α from both the numerator and denominator of $R_{\alpha,\beta}$ in Eq. (3.3.1) is that it ensures $R_{\alpha,\beta}$ stays unchanged when the loss r.v. is subject to a linear transformation based on properties of our *tail-based* normal approximation (see Section 3.6). This feature is desired as the relative distribution shapes of X and T are not changed by a linear transformation, neither should $R_{\alpha,\beta}$.

If the estimation error is 0 or very small, then the value of $R_{\alpha,\beta}$ will be 1 or very close to 1 and vice versa. Therefore, $R_{\alpha,\beta}$ can be seen as a measure of the estimation error. Moreover, the results in Table 3.1 imply that this ratio depends on the tail weight of T.

To further navigate the relation between the ratio $R_{\alpha,\beta}$ (estimation error) and the tail weight, we need a variable to generally quantify the tail weight. Here a *conditional skewness* is defined, as in Eq. (3.3.2):

$$\gamma_\alpha \equiv \frac{\mathbb{E}[(T - A_\alpha)^3 | T > A_\alpha]}{(\mathbb{E}[(T - A_\alpha)^2 | T > A_\alpha])^{\frac{3}{2}}}. \tag{3.3.2}$$

A similarly defined conditional kurtosis can be considered, too. Since results with the *conditional skewness* are already satisfying, we stick to it in this chapter. Our idea is to develop a regression model between the $R_{\alpha,\beta}$ and γ_α such that the expected value of $R_{\alpha,\beta}$, denoted by $\hat{R}_{\alpha,\beta}$, satisfies the following equation:

$$\hat{R}_{\alpha,\beta} = f_{\alpha,\beta}(\gamma_\alpha). \tag{3.3.3}$$

Subsequently, we can have a more accurate ES estimator by adjusting the *tail-based* normal approximation based on Eq. (3.3.1).

Now we take the student's t distribution for the r.v. T as the training distribution. By changing values of the degree of freedom, different tail weights γ_α from the corresponding distributions of T are obtained. Therefore, a relevant regression model can be built. We want to point out that, although our model is developed based on student's t distribution, it can be applied to other heavy-tailed distributions (see Remark 1 and the robustness test results in Subsection 3.4.2).

To illustrate the relation between $R_{\alpha,\beta}$ and γ_α, we show some values in Figure 3.2 (the small circles) for $\beta = 99\%$ (left) and $\beta = 99.5\%$ (right) with $\alpha = 95\%$. A regression analysis is subsequently conducted to develop an equation that predicts $R_{\alpha,\beta}$ from the *conditional skewness*

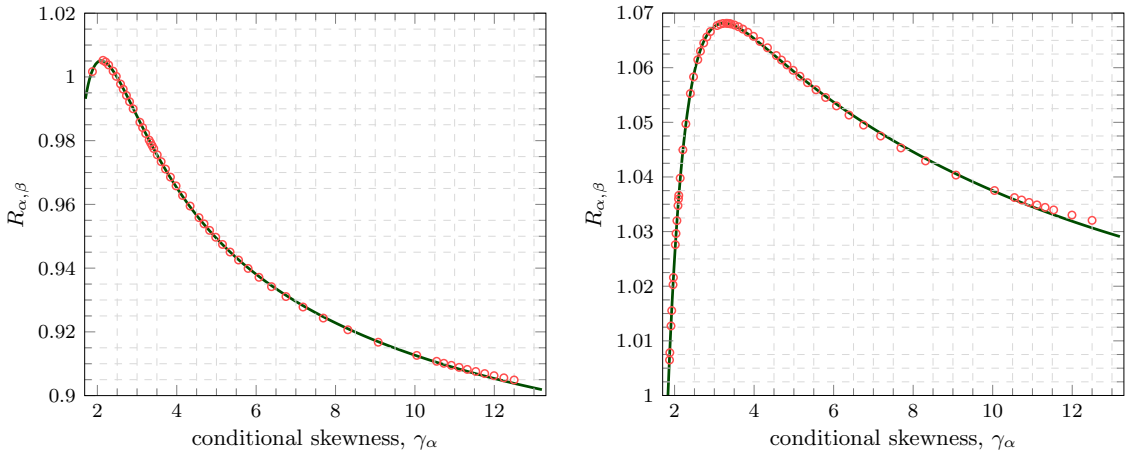

Figure 3.2: Scatter plots of $R_{\alpha,\beta}$ on γ_α with regression lines, $\alpha = 95\%$, $\beta = 99\%$ (left) & 99.5% (right).

of the t distributed r.v. T. According to Figure 3.2, the scatter plots[4] of $R_{\alpha,\beta}$ on γ_α, the following nonlinear regression lines are devised (See the solid lines in Figure 3.2):

$$\hat{R}_{\alpha,\beta} = f_{\alpha,\beta}(\gamma_\alpha) = b_0 + b_1 e^{-b_2\gamma_\alpha} + b_3\gamma_\alpha^{-1} + b_4\gamma_\alpha^{-2}, \quad \text{for } \alpha = 95\%,\ \beta = 99\% \text{ or } 99.5\%. \tag{3.3.4}$$

In fact, the regression model is not unique. Any prediction equation works as long as it precisely describes the relation between $R_{\alpha,\beta}$ and γ_α. Summary of the coefficients in Eq. (3.3.4) is reported in Table 3.2, which shows all variables are useful to predict $R_{\alpha,\beta}$.

By virtue of Eq. (3.3.1) and Eq. (3.3.3), we can obtain a more accurate ES estimator based on the adjusted *tail-based* normal approximation as follows:

$$\widehat{\mathrm{ES}}_\beta(X) \equiv [\mathrm{ES}_\beta(X) - A_\alpha] f_{\alpha,\beta}(\gamma_\alpha) + A_\alpha. \tag{3.3.5}$$

The adjustment factor $f_{\gamma,\beta}(\gamma_\alpha)$ may depend on the level of α and β, but it is usually robust on the distribution of T.

Remark 1. *Though the two regression models ($\beta = 99\%$ or 99.5%) in Eq.* (3.3.4) *are built on a*

[4]To ensure the regression line is shown clearly, only a few points are displayed in Figure 3.2 while regression analysis covers a sufficiently large number of points to guarantee the effectiveness.

basis of student's t distributions, we can adopt them to adjust the tail-based normal approximation obtained from loss samples or loss r.v.'s from other distributions. This adjustment is feasible since as a general statistic, the corresponding γ_α is defined by Eq. (3.3.6) *or Eq.* (3.3.7)*. The accuracy of our ES estimator after the adjustment will be explored in the next two sections.*

For a sample of losses $\mathbf{Y} = \{y_n\}_{n=1}^N$, its *conditional skewness* γ_α can be acquired through

$$\gamma_\alpha \equiv \frac{\frac{\sum_{n=1}^N (y_n - A_\alpha)^3 \mathbb{1}_{\{y_n > A_\alpha\}}}{\sum_{n=1}^N \mathbb{1}_{\{y_n > A_\alpha\}}}}{\left[\frac{\sum_{n=1}^N (y_n - A_\alpha)^2 \mathbb{1}_{\{y_n > A_\alpha\}})}{\sum_{n=1}^N \mathbb{1}_{\{y_n > A_\alpha\}}}\right]^{\frac{3}{2}}}. \tag{3.3.6}$$

where A_α, the sample's α-quantile, is given by Eq. (3.2.4). Moreover, for a loss r.v., W, with a given distribution whose α-quantile equals A_α, its γ_α is defined by

$$\gamma_\alpha \equiv \frac{\mathbb{E}[(W - A_\alpha)^3 | W > A_\alpha]}{(\mathbb{E}[(W - A_\alpha)^2 | W > A_\alpha])^{\frac{3}{2}}}. \tag{3.3.7}$$

Eq. (3.3.7) will be used later for error estimation and robustness tests.

Table 3.2: Summary for regression models with $\alpha = 95\%, \beta = 99\%$ and $\alpha = 95\%, \beta = 99.5\%$.

	$\alpha = 95\%, \beta = 99\%$			$\alpha = 95\%, \beta = 99.5\%$		
Coefficient	Value	SE	p-value	Value	SE	p-value
b_0	0.8611	1.7×10^{-4}	0.00	0.9919	2.2×10^{-4}	0.00
b_1	0.5191	7.0×10^{-3}	0.00	0.6681	8.5×10^{-3}	0.00
b_2	0.9747	3.8×10^{-3}	0.00	0.9607	3.7×10^{-3}	0.00
b_3	0.6099	2.0×10^{-3}	0.00	0.6022	2.5×10^{-3}	0.00
b_4	-0.9413	6.9×10^{-5}	0.00	-1.4623	6.2×10^{-3}	0.00
	Adjusted $R^2 = 0.9999$			Adjusted $R^2 = 0.9998$		

For a given loss sample $\mathbf{Y}$, the algorithm to obtain our regression-adjusted ES estimator from the *tail-based* normal approximation can be summarized as follows:

Algorithm for the adjusted tail-based normal ES estimation

Step 1. Choose a value for α (e.g. $\alpha = 95\%$), and calculate A_α using Eq. (3.2.4);

Step 2. Determine the *tail-based* normal r.v. X by Eq. (3.2.5);

Step 3. Compute sample's *conditional skewness* γ_α by virtue of Eq. (3.3.6);

Step 4. Compute the adjustment factor $f_{\alpha,\beta}(\gamma_\alpha)$ by virtue of Eq. (3.3.4);

Step 5. Obtain the adjusted *tail-based* normal ES estimator at β level by Eq. (3.3.5).

3.4 Consistence and Robustness Tests

Although the adjustment factor $R_{\alpha,\beta}$ is developed based on student's t distribution, in this section, we will demonstrate that it works well for many other heavy-tailed distributions, such as Gamma, Lognormal, GPD, Weibull, etc. Therefore, the adjusted *tail-based* normal approximation is a simple and robust approach for ES estimation.

3.4.1 Consistence Test

Before examining the accuracy of the adjusted *tail-based* normal ES estimator, we first conduct a self-consistence test. Suppose the loss r.v. W follows a normal distribution, $W \sim N(\mu, \sigma^2)$. Under this circumstance, in theory, the corresponding *tail-based* normal r.v. X is exactly same as W, i.e., $X \sim N(\mu, \sigma^2)$, and no further adjustment based on tail weights is needed. That is, the adjustment factor $R_{\alpha,\beta}$ should be equal to 1. Let us define

$$z_\alpha \equiv \Phi^{-1}(\alpha), \quad q_\alpha \equiv \frac{1}{(1-\alpha)\sqrt{2\pi}} e^{-\frac{z_\alpha^2}{2}}. \tag{3.4.1}$$

Let A_α be the α-quantile of W and we can get

$$A_\alpha = \mu + \sigma z_\alpha. \tag{3.4.2}$$

In addition, we have (see Appendix A.1 for details)

$$\mathbb{E}[W \ |W > A_\alpha] \ = \ \mu + \sigma q_\alpha, \tag{3.4.3}$$

$$\mathbb{E}[W^2|W > A_\alpha] \ = \ \mu^2 + \sigma^2 + \sigma(A_\alpha + \mu) q_\alpha, \tag{3.4.4}$$

$$\mathbb{E}[W^3|W > A_\alpha] \ = \ \mu^3 + 3\mu\sigma^2 + [\sigma(A_\alpha^2 + \mu^2 + A_\alpha\mu) + 2\sigma^3] q_\alpha. \tag{3.4.5}$$

Define

$$s_2 \equiv \mathbb{E}[(W - A_\alpha)^2|W > A_\alpha], \quad r_3 \equiv \mathbb{E}[(W - A_\alpha)^3|W > A_\alpha].$$

By virtue of Eq. (3.4.2)∼Eq. (3.4.5), we have

$$\begin{aligned}
s_2 \ &= \ (\mu - A_\alpha)[\sigma q_\alpha + (\mu - A_\alpha)] + \sigma^2, \\
r_3 \ &= \ (\mu - A_\alpha)^3 + 3(\mu - A_\alpha)\sigma^2 + [2\sigma^2 + (A_\alpha - \mu)^2]\sigma q_\alpha, \\
\frac{s_2}{\sigma^2} \ &= \ -z_\alpha(q_\alpha - z_\alpha) + 1, \quad \frac{r_3}{\sigma^3} \ = \ -z_\alpha^3 - 3z_\alpha + (2 + z_\alpha^2) q_\alpha.
\end{aligned}$$

As we can see, r_3/σ^3 and s_2/σ^2 depend only on α. Therefore, we can get the square of the *conditional skewness* of W by

$$\gamma_\alpha^2 = \frac{r_3^2}{s_2^3} = \frac{\left(\frac{r_3}{\sigma^3}\right)^2}{\left(\frac{s_2}{\sigma^2}\right)^3} = \frac{[-z_\alpha^3 - 3z_\alpha + (2 + z_\alpha^2) q_\alpha]^2}{[-z_\alpha(q_\alpha - z_\alpha) + 1]^3}. \tag{3.4.6}$$

So γ_α is a constant that only depends on α and it is independent of μ and σ^2.

Obviously, the original estimation error of the unadjusted *tail-based* normal approximation is zero because X is a replication of W. As for $\alpha = 95\%$, we have $\gamma_\alpha = 1.838$ by Eq. (3.4.6) and $f_{\alpha,\beta}(\gamma_\alpha) \approx 1.00$ by Eq. (3.3.4) for both 99% and 99.5% β levels. When $f_{\alpha,\beta}(\gamma_\alpha)$ is near 1, there is almost no adjustment so the estimation error of the adjusted *tail-based* normal approximation is still close to zero. Therefore, the adjusted *tail-based* normal ES estimation method is verified to be self-consistent.

3.4.2 Robustness Test and Estimation Errors for Heavy-Tailed Loss Distributions

To test the accuracy of our proposed adjusted *tail-based* normal ES estimator, we calculate the corresponding ES estimates of some heavy-tailed distributions. The results will be compared to the theoretical values derived from the distribution functions.

In particular, for a loss r.v. W with a known distribution, its corresponding *tail-based* normal r.v. X and *conditional skewness* γ_α can be determined respectively by Eq. (3.2.15) and Eq. (3.3.7). In addition to the original ES estimation error that has been defined in Eq. (3.2.16), a measure of the estimation error for our regression-adjusted ES estimator at β level is defined by

$$\widehat{\mathbf{e}}^t_\beta(\mathrm{ES}) \equiv \frac{\mathrm{ES}_\beta(W) - \widehat{\mathrm{ES}}_\beta(X)}{\mathrm{ES}_\beta(W)}, \tag{3.4.7}$$

where $\widehat{\mathrm{ES}}_\beta(X)$, the β-level adjusted *tail-based* normal approximation, is given in Eq. (3.3.5).

Table 3.3: Regression-adjusted errors for various heavy-tailed loss distributions, $\alpha = 95\%$.

W	γ_α	$\mathbf{e}^t_{99\%}(\mathrm{ES})$ (%)	$\widehat{\mathbf{e}}^t_{99\%}(\mathrm{ES})$ (%)	$\mathbf{e}^t_{99.5\%}(\mathrm{ES})$ (%)	$\widehat{\mathbf{e}}^t_{99.5\%}(\mathrm{ES})$ (%)
t, df=3.5	7.181	-4.848	-0.028	3.152	-0.036
t, df=5	3.165	-0.919	-0.003	3.924	-0.004
t, df=8	2.359	0.121	-0.001	2.770	-0.001
Gamma(0.1,1)	2.441	1.175	1.100	6.214	2.002
Gamma(0.3,1)	2.249	0.819	0.572	3.954	0.985
Gamma(1.5,1)	2.086	0.430	0.218	1.938	0.354
Lognormal(0,1.1^2)	4.560	-3.925	-0.660	5.455	1.073
Lognormal(0,0.9^2)	3.416	-1.316	0.104	5.417	1.116
Lognormal(0,0.3^2)	2.098	0.225	0.091	1.237	0.158
GPD(0.3, 1)	11.23	-7.747	-0.689	2.547	0.065
GPD(0.25, 1)	5.196	-3.926	-0.245	4.395	0.452
GPD(0.1, 1)	2.571	0.179	0.274	4.121	0.652
Weibull(0.5, 1)	3.109	-0.574	0.525	6.919	2.092
Weibull(0.9, 1)	2.192	0.584	0.352	2.936	0.612
Weibull(1.4, 1)	1.967	0.262	0.114	1.005	0.166

The results are summarized in Table 3.3, in which the original and adjusted errors are both reported for various heavy-tailed loss distributions. The location parameter is set to zero and the scale parameter is set to one if they exist. For each kind of distributions, we choose three different parameter sets, and the most heavy-tailed one is listed first.

From Table 3.3, we can see that the adjusted *tail-based* normal approximation gives us a substantially better accuracy than the original one. Not surprisingly, the relative errors after adjustment are less than 0.05% for t distributions since the adjustment factors are obtained on a basis of that distribution. Moreover, for other heavy-tailed distributions, the adjusted *tail-based* normal approximation also has good results with considerably small relative errors $(0 - 2\%)$, even though the adjustment factors are initially developed from t distributions.

Therefore, the regression models work well for distributions other than student's t and the adjusted *tail-based* normal approximation offers an effective and robust ES estimation method for loss r.v.'s with various heavy-tailed distributions.

3.5 Comparison with the AA and EVT Estimators

To further test the accuracy of our method, we compare the dispersion and accuracy of the adjustment tail-based normal ES estimator with the arithmetic average (AA) of excessive losses and extreme value theory (EVT) ES estimators for small loss samples simulated from various heavy-tailed distributions.

We adopt the AA estimator proposed by S. X. Chen (2007): Assuming $\tilde{Y}_{(1)}, \ \ldots, \ \tilde{Y}_{(N-1)}, \ \tilde{Y}_{(N)}$ is a sample of losses sorted in ascending order, its ES estimator at level β is defined as

$$\widetilde{\mathrm{ES}}_{\beta} = \frac{\sum_{n=\lceil N\beta \rceil}^{N} \tilde{Y}_{(n)}}{N + 1 - \lceil N\beta \rceil}, \tag{3.5.1}$$

where $\lceil \cdot \rceil$ denotes the ceiling function; that is, $\lceil x \rceil$ gives the smallest integer greater than or equal to x.

In light of the EVT estimator, suppose that $F(w)$ is the CDF for a loss r.v. W. The

distribution function of excesses beyond a threshold v is defined by McNeil (1997) as follows:

$$F_v(y) = Pr(W - v \leq y | W > v) = \frac{F(v+y) - F(v)}{1 - F(v)}. \tag{3.5.2}$$

According to Gnedenko (1943), as the threshold v increases, $F_v(y)$ converges to a GPD whose CDF and PDF are listed as below:

$$G_{\xi,\sigma}(y) = 1 - \left(1 + \xi\frac{y}{\sigma}\right)^{-1/\xi}, \qquad g_{\xi,\sigma}(y) = \frac{1}{\sigma}\left(1 + \xi\frac{y}{\sigma}\right)^{-1/\xi - 1}, \quad \xi \neq 0,\ \sigma > 0, \tag{3.5.3}$$

where ξ is a shape parameter and σ is a scale parameter. Assuming there are n_v losses, $\{w_k\}_{k=1}^{n_v}$, greater than the threshold v, then the sequence $\{w_k - v\}_{k=1}^{n_v}$ will show Generalized Pareto behaviors (Davison & Smith, 1990). If $\xi < 0$, an extra condition is $1 + \xi\frac{(w_i - v)}{\sigma} > 0$ for $i = 1, \ldots, n_v$. When $\xi = 0$, the GPD becomes the exponential distribution so we have

$$G_\sigma(y) = 1 - \exp\left(-\frac{y}{\sigma}\right), \qquad g_\sigma(y) = \frac{1}{\sigma}\exp\left(-\frac{y}{\sigma}\right), \quad \xi = 0,\ \sigma > 0. \tag{3.5.4}$$

Moreover, ξ and σ in GPD can be estimated by maximizing its log-likelihood function:

$$(\xi^*, \sigma^*) = \arg\max_{\xi,\sigma} \sum_{k=1}^{n_v} \ln\big(g_{\xi,\sigma}(w_k - v)\big), \tag{3.5.5}$$

where the function g is given by Eq. (3.5.3) or Eq. (3.5.4), depending on the value of ξ. We need to compare all the three situations where $\xi < 0$, $\xi = 0$ or $\xi > 0$ and choose the one with the largest maximum log-likelihood.

$F_v(y)$ in Eq. (3.5.2) is a conditional probability that $W \leq v + y$ given that $W > v$. The unconditional distribution of excesses beyond the threshold v is thus derived by

$$Pr(W \leq w) = (1 - F(v))G_{\xi,\sigma}(w - v) + F(v), \quad w > v. \tag{3.5.6}$$

Based on Eq. (3.5.6), EVT β-level ($\beta > F(v)$) VaR and ES estimators of W can be derived and

the results are showed in Table 3.4 (see Appendix A.2 for derivation details). If the loss sample size is N, $1 - F(v)$ can be approximated by n_v/N (Hull, 2012).

Table 3.4: EVT estimators for VaR and ES at β-level.

	$\text{VaR}_\beta(W)$	$\text{ES}_\beta(W)$
$\xi \neq 0$	$v + \frac{\sigma}{\xi}\left[\left(\frac{1-\beta}{1-F(v)}\right)^{-\xi} - 1\right]$	$\frac{\text{VaR}_\beta(W)+\sigma-\xi v}{1-\xi}$
$\xi = 0$	$v - \sigma \ln\left(\frac{1-\beta}{1-F(v)}\right)$	$\text{VaR}_\beta(W) + \sigma$

Remark 2. *In practice, the threshold v is usually set as the 95th percentile of the loss sample. We keep this selection so that the threshold v of the EVT estimator is same as A_α ($\alpha = 0.95$) in our ES estimation method.*

To get more information about the accuracy of our adjusted *tail-based* normal approximation, we apply the Monte Carlo simulation and compare our adjusted ES estimates to the counterparts of the AA method and EVT method. In particular, small random loss samples are generated from various heavy-tailed distributions. For every examined distribution, 10^3 loss samples are produced in which the same sample size of 250 (or 500) is adopted. With regard to each simulated sample, our regression-adjusted estimator, AA estimator and EVT estimator for ES at 99% and 99.5% levels will be respectively computed and recorded. Repeating this step for all the 10^3 samples, we get the corresponding mean, relative standard deviation (RSD) and mean squared error (MSE) of each estimator.

The MSE of the ES estimator for a given underlying distribution is computed by

$$\text{MSE of } \text{ES}_{\beta,j} = \frac{1}{1000}\sum_{i=1}^{1000}(\text{ES}_{\beta,j}(i) - \text{true } \text{ES}_\beta)^2, \quad j = 1, 2, 3. \tag{3.5.7}$$

where 'true ES_β' is the theoretical ES value at the confidence level β of the underlying distribution, $j = 1$ stands for the adjusted tail-based estimator, $j = 2$ stands for the AA estimator, $j = 3$

stands for the EVT estimator and i is the sample index for all the 10^3 samples.

To better compare the three ES estimation methods, based on the extent of heavy-tailed behaviors, the underlying distributions are split up into three groups, namely highly, moderately, and slightly heavy-tailed distributions. This categorization is conducted by shape parameters within each distribution family and the criteria are vague.

Table 3.5: Comparisons of our estimator, AA and EVT estimators for ES, (a).

		$\beta = 99\%$			$\beta = 99.5\%$		
Sample	Estimator	Mean	RSD	MSE	Mean	RSD	MSE
t, df=3.5	$\widehat{\mathrm{ES}}_\beta$	5.663	**0.331**	**3.558**	6.474	**0.383**	**6.792**
	AA	6.035	0.360	4.746	7.105	0.481	11.70
	EVT	5.830	0.340	3.935	7.087	0.463	10.77
Gamma(0.1, 1)	$\widehat{\mathrm{ES}}_\beta$	2.231	**0.288**	**0.427**	2.585	**0.318**	**0.766**
	AA	2.384	0.310	0.548	2.783	0.377	1.111
	EVT	2.330	0.311	0.526	2.796	0.417	1.368
Lognormal(0, 1.1^2)	$\widehat{\mathrm{ES}}_\beta$	19.39	0.390	**57.59**	22.61	**0.442**	**109.1**
	AA	20.89	0.413	74.85	25.23	0.551	193.1
	EVT	20.18	**0.382**	59.20	25.49	0.500	162.0
GPD(0.3, 1)	$\widehat{\mathrm{ES}}_\beta$	14.89	0.440	43.40	17.38	**0.501**	**82.67**
	AA	16.08	0.478	59.26	19.61	0.618	146.9
	EVT	15.26	**0.410**	**39.19**	19.26	0.538	107.7
Weibull(0.6, 1)	$\widehat{\mathrm{ES}}_\beta$	17.06	**0.276**	**22.96**	19.46	**0.314**	**42.74**
	AA	18.15	0.310	31.71	21.02	0.378	63.76
	EVT	17.69	0.317	31.51	21.22	0.429	82.93
sample size = 250, ***highly*** heavy-tailed distributions.							

The results are summarized in Table 3.5, Table 3.6 and Table 3.7, in which the adjusted tail-based estimator is denoted by $\widehat{\mathrm{ES}}_\beta$, and the smallest RSD and MSE of the three estimators are highlighted in bold. From those tables, we can see that when $\beta = 99\%$, our adjusted *tail-based* normal approximations dominates the other two estimators except RSD and MSE of GPD(0.3, 1) and RSD of Lognormal(0, 1.1^2). Even for the scenarios containing those exceptions, the results of our adjusted *tail-based* normal approximation are quite close to the best results. Moreover, for a

high confidence level with $\beta = 99.5\%$, our adjusted *tail-based* normal ES estimator substantially outperforms the other two in terms of RSD and MSE of all underlying distributions.

Table 3.6: Comparisons of our estimator, AA and EVT estimators for ES, (b).

Sample	Estimator	$\beta = 99\%$ Mean	RSD	MSE	$\beta = 99.5\%$ Mean	RSD	MSE
t, df=5	$\widehat{\mathrm{ES}}_\beta$	4.278	**0.236**	**1.048**	4.769	**0.276**	**1.963**
	AA	4.492	0.264	1.408	5.080	0.343	3.060
	EVT	4.381	0.261	1.311	5.050	0.367	3.480
Gamma(0.3, 1)	$\widehat{\mathrm{ES}}_\beta$	3.343	**0.220**	**0.565**	3.740	**0.249**	**0.992**
	AA	3.507	0.244	0.729	3.924	0.294	1.362
	EVT	3.425	0.246	0.714	3.917	0.341	1.817
Lognormal(0, 0.9^2)	$\widehat{\mathrm{ES}}_\beta$	11.12	**0.294**	**10.84**	12.68	**0.339**	**20.38**
	AA	11.83	0.320	14.37	13.79	0.424	34.25
	EVT	11.52	0.313	13.00	13.92	0.425	35.04
GPD(0.2, 1)	$\widehat{\mathrm{ES}}_\beta$	10.21	**0.306**	**10.01**	11.63	**0.355**	**18.99**
	AA	10.87	0.344	13.95	12.69	0.439	31.13
	EVT	10.53	0.331	12.18	12.70	0.450	32.77
Weibull(0.9, 1)	$\widehat{\mathrm{ES}}_\beta$	6.557	**0.174**	**1.361**	7.190	**0.203**	**2.429**
	AA	6.810	0.198	1.821	7.510	0.242	3.363
	EVT	6.651	0.196	1.717	7.439	0.272	4.163
sample size = 250, ***moderately*** heavy-tailed distributions.							

The reason why our ES estimator defeats AA estimator is that the latter one only uses a very small percentage ($1 - \beta$, with $\beta = 99\%, 99.5\%$) of all observations while our ES estimator always considers much more data whatever β is: the largest 5% ($1 - \alpha$, with $\alpha = 95\%$) of the losses. Moreover, compared to the EVT estimator, our ES estimator is also preferable though both of them make use of the largest 5% of the sample observations in this chapter. The EVT ES estimator is thus more sensitive to the largest losses than our adjusted *tail-based* normal estimator.

To further compare the three estimators, let us double the size of the simulated samples to 500 and investigate those three ES estimators again. Results are reported in Table 3.8, Table

Table 3.7: Comparisons of our estimator, AA and EVT estimators for ES, (c).

Sample	Estimator	$\beta = 99\%$ Mean	RSD	MSE	$\beta = 99.5\%$ Mean	RSD	MSE
t, df=8	$\widehat{\mathrm{ES}}_\beta$	3.473	**0.171**	**0.365**	3.794	**0.202**	**0.670**
	AA	3.595	0.191	0.469	3.969	0.251	1.007
	EVT	3.533	0.184	0.425	3.904	0.264	1.091
Gamma(1.5, 1)	$\widehat{\mathrm{ES}}_\beta$	6.541	**0.132**	**0.788**	7.043	**0.155**	**1.379**
	AA	6.730	0.148	0.991	7.263	0.185	1.843
	EVT	6.587	0.145	0.939	7.147	0.200	2.155
Lognormal(0, 0.3^2)	$\widehat{\mathrm{ES}}_\beta$	2.206	**0.084**	**0.036**	2.313	**0.102**	**0.062**
	AA	2.245	0.095	0.046	2.361	0.124	0.087
	EVT	2.251	0.097	0.047	2.351	0.131	0.097
GPD(0.1, 1)	$\widehat{\mathrm{ES}}_\beta$	7.306	**0.217**	**2.596**	8.121	**0.254**	**4.802**
	AA	7.658	0.247	3.575	8.618	0.310	7.182
	EVT	7.478	0.249	3.478	8.605	0.348	9.003
Weibull(1.4, 1)	$\widehat{\mathrm{ES}}_\beta$	3.337	**0.109**	**0.138**	3.547	**0.129**	**0.237**
	AA	3.411	0.124	0.179	3.632	0.152	0.311
	EVT	3.397	0.114	0.150	3.592	0.161	0.350
sample size = 250, ***slightly*** heavy-tailed distributions.							

3.9 and Table 3.10. We can see our adjusted *tail-based* normal ES estimator is still superior to the other two at 99.5% level while the three estimators are considerably close to each other regarding RSD and MSE at 99% level.

3.6 Effects of Linear Transformations

In this section, we investigate the effects of linear transformation on the ES estimator of a loss r.v. or a loss sample. Linear transformation is common in portfolio management and risk management, such as the exchanges among currencies and the change of time units.

Assuming an underlying distribution is transformed linearly with a positive scale multiplier m ($m > 0$) and a constant summand c, then any random loss sample generated from it will be subject to the same transformation. In what follows, we use superscript τ for the new variables after transformation. Suppose the transformation is in forms of $\mathbf{Y}^\tau = m\mathbf{Y} + c$, where

Table 3.8: Comparisons of our estimator, AA and EVT estimators for ES, (d).

Sample	Estimator	$\beta = 99\%$ Mean	RSD	MSE	$\beta = 99.5\%$ Mean	RSD	MSE
t, df=3.5	$\widehat{\mathrm{ES}}_\beta$	5.781	0.248	2.074	6.475	**0.291**	**4.132**
	AA	5.808	**0.234**	**1.860**	7.434	0.344	6.551
	EVT	5.929	0.254	2.258	7.360	0.342	6.344
Gamma(0.1, 1)	$\widehat{\mathrm{ES}}_\beta$	2.265	**0.203**	**0.218**	2.681	**0.225**	**0.406**
	AA	2.301	0.207	0.228	2.920	0.261	0.579
	EVT	2.343	0.241	0.319	2.916	0.316	0.852
Lognormal(0, 1.1^2)	$\widehat{\mathrm{ES}}_\beta$	19.89	0.294	34.29	23.73	**0.336**	**67.20**
	AA	19.97	**0.267**	**28.46**	26.52	0.384	104.6
	EVT	20.65	0.287	35.29	26.70	0.375	101.4
GPD(0.3, 1)	$\widehat{\mathrm{ES}}_\beta$	15.30	0.344	27.85	18.29	**0.394**	**54.94**
	AA	15.28	**0.306**	**22.02**	20.64	0.455	88.38
	EVT	15.74	**0.306**	23.22	20.43	0.401	67.18
Weibull(0.6, 1)	$\widehat{\mathrm{ES}}_\beta$	17.35	**0.201**	**12.57**	20.19	**0.231**	**24.30**
	AA	17.54	0.203	12.83	22.00	0.276	36.75
	EVT	17.93	0.232	17.25	21.97	0.311	46.80

sample size = 500, ***highly*** heavy-tailed distributions.

Table 3.9: Comparisons of our estimator, AA and EVT estimators for ES, (e).

Sample	Estimator	$\beta = 99\%$ Mean	RSD	MSE	$\beta = 99.5\%$ Mean	RSD	MSE
t, df=5	$\widehat{\mathrm{ES}}_\beta$	4.376	0.181	0.634	4.971	**0.216**	**1.231**
	AA	4.411	**0.177**	**0.608**	5.334	0.258	1.900
	EVT	4.469	0.196	0.765	5.285	0.275	2.105
Gamma(0.3, 1)	$\widehat{\mathrm{ES}}_\beta$	3.403	**0.154**	**0.284**	3.878	**0.177**	**0.516**
	AA	3.438	0.161	0.308	4.118	0.206	0.721
	EVT	3.460	0.179	0.383	4.052	0.241	0.954
Lognormal(0, 0.9^2)	$\widehat{\mathrm{ES}}_\beta$	11.34	0.218	6.119	13.18	**0.254**	**11.94**
	AA	11.42	**0.208**	**5.639**	14.40	0.298	18.56
	EVT	11.69	0.232	7.388	14.40	0.315	20.63
GPD(0.2, 1)	$\widehat{\mathrm{ES}}_\beta$	10.40	0.230	5.826	12.09	**0.270**	**11.52**
	AA	10.47	**0.221**	**5.395**	13.27	0.323	18.37
	EVT	10.72	0.246	6.930	13.21	0.334	19.43
Weibull(0.9, 1)	$\widehat{\mathrm{ES}}_\beta$	6.627	**0.126**	**0.723**	7.379	**0.148**	**1.324**
	AA	6.668	0.130	0.770	7.764	0.177	1.890
	EVT	6.697	0.143	0.931	7.610	0.196	2.247

sample size = 500, ***moderately*** heavy-tailed distributions.

Table 3.10: Comparisons of our estimator, AA and EVT estimators for ES, (f).

Sample	Estimator	$\beta = 99\%$ Mean	RSD	MSE	$\beta = 99.5\%$ Mean	RSD	MSE
t, df=8	$\widehat{\mathrm{ES}}_\beta$	3.553	**0.129**	**0.210**	3.945	**0.155**	**0.392**
	AA	3.570	0.131	0.220	4.141	0.189	0.613
	EVT	3.587	0.143	0.263	4.063	0.201	0.663
Gamma(1.5, 1)	$\widehat{\mathrm{ES}}_\beta$	6.583	**0.091**	**0.382**	7.172	**0.108**	**0.696**
	AA	6.617	0.095	0.414	7.439	0.130	0.931
	EVT	6.614	0.101	0.465	7.289	0.140	1.082
Lognormal$(0,0.3^2)$	$\widehat{\mathrm{ES}}_\beta$	2.218	0.060	**0.018**	2.346	**0.074**	**0.032**
	AA	2.223	0.063	0.020	2.401	0.088	0.045
	EVT	2.243	**0.059**	**0.018**	2.405	0.081	0.038
GPD(0.1, 1)	$\widehat{\mathrm{ES}}_\beta$	7.401	**0.159**	**1.412**	8.370	**0.188**	**2.727**
	AA	7.455	0.160	1.451	8.948	0.227	4.136
	EVT	7.539	0.180	1.851	8.819	0.249	4.827
Weibull(1.4, 1)	$\widehat{\mathrm{ES}}_\beta$	3.360	**0.078**	**0.072**	3.610	**0.094**	**0.126**
	AA	3.368	0.082	0.078	3.715	0.111	0.171
	EVT	3.366	0.085	0.085	3.640	0.116	0.182
sample size = 500, ***slightly*** heavy-tailed distributions.							

$\mathbf{Y}$ represents the original loss sample dataset (or the original loss r.v.).

The three β-level ES estimators for the i-th original sample set are denoted by $\{\mathrm{ES}_{\beta,j}(i)\}_{j=1}^3$ ($j = 1$ for adjusted *tail-based* normal approximation, $j = 2$ for AA, and $j = 3$ for EVT). As for the corresponding i-th linearly-transformed sample, the new estimators are denoted by $\{\mathrm{ES}^\tau_{\beta,j}(i)\}_{j=1}^3$. Apparently, the β-level theoretical ES estimate for the new sample is

$$\text{true } \mathrm{ES}^\tau_\beta = m(\text{true } \mathrm{ES}_\beta) + c.$$

Firstly, we examine the effects of the linear transformation on our adjusted *tail-based* normal ES estimator. We have the following proposition:

Proposition 1. *If a loss sample or a loss r.v. is transformed linearly, then the corresponding tail-based normal r.v. X and the adjusted tail-based normal ES estimator are subject to the same linear transformation. Moreover, the conditional skewness stays unchanged.*

Proof. Suppose the original loss sample is denoted by $\{y_n\}_{n=1}^N$ whose α-quantile is A_α and *conditional skewness* is γ_α. After undergoing a linear transformation, the loss sample becomes a new one denoted by $\{y_n^\tau\}_{n=1}^N$ such that $y_n^\tau = my_n + c$ for $n = 1, \ldots, N$ where m, c are both constants and $m > 0$.

The α-quantile for the new sample after the linear transformation is $mA_\alpha + c$. The new *conditional skewness* is

$$\gamma_\alpha^\tau = \frac{\frac{\sum_{n=1}^N (y_n^\tau - mA_\alpha - c)^3 \mathbb{1}_{\{y_n^\tau > mA_\alpha + c\}}}{\sum_{n=1}^N \mathbb{1}_{\{y_n^\tau > mA_\alpha + c\}}}}{\left[\frac{\sum_{n=1}^N (y_n^\tau - mA_\alpha - c)^2 \mathbb{1}_{\{y_n^\tau > mA_\alpha + c\}}}{\sum_{n=1}^N \mathbb{1}_{\{y_n^\tau > mA_\alpha + c\}}}\right]^{\frac{3}{2}}} = \frac{m^3 \left[\frac{\sum_{n=1}^N (y_n - A_\alpha)^3 \mathbb{1}_{\{y_n > A_\alpha\}}}{\sum_{n=1}^N \mathbb{1}_{\{y_n > A_\alpha\}}}\right]}{m^3 \left[\frac{\sum_{n=1}^N (y_n - A_\alpha)^2 \mathbb{1}_{\{y_n > A_\alpha\}}}{\sum_{n=1}^N \mathbb{1}_{\{y_n > A_\alpha\}}}\right]^{\frac{3}{2}}} = \gamma_\alpha. \tag{3.6.1}$$

That is, after the linear transformation, γ_α^τ is equal to γ_α. Furthermore, by Eq. (3.2.11), the parameters of the original *tail-based* normal r.v. $X \sim N(\mu, \sigma^2)$ is solved by

$$\begin{aligned} \sigma &= \left(z_\alpha^2 + 1 - z_\alpha q_\alpha\right)^{-\frac{1}{2}} \left[\frac{\sum_{n=1}^N (y_n - A_\alpha)^2 \mathbb{1}_{\{y_n > A_\alpha\}}}{\sum_{n=1}^N \mathbb{1}_{\{y_n > A_\alpha\}}}\right]^{\frac{1}{2}}, \\ \mu &= A_\alpha - \sigma \Phi^{-1}(\alpha). \end{aligned} \tag{3.6.2}$$

where z_α and q_α are defined in Eq. (3.4.1). Assuming the *tail-based* normal r.v. for $\{y_n^\tau\}_{n=1}^N$ is $X^\tau \sim N(\mu^\tau, (\sigma^\tau)^2)$, we have

$$\begin{aligned} \sigma^\tau &= \left(z_\alpha^2 + 1 - z_\alpha q_\alpha\right)^{-\frac{1}{2}} \left[\frac{\sum_{n=1}^N (y_n^\tau - mA_\alpha - c)^2 \mathbb{1}_{\{y_n^\tau > mA_\alpha + c\}}}{\sum_{n=1}^N \mathbb{1}_{\{y_n^\tau > mA_\alpha + c\}}}\right]^{\frac{1}{2}} \\ &= \left(z_\alpha^2 + 1 - z_\alpha q_\alpha\right)^{-\frac{1}{2}} \left[\frac{m^2 \sum_{n=1}^N (y_n - A_\alpha)^2 \mathbb{1}_{\{y_n > A_\alpha\}}}{\sum_{n=1}^N \mathbb{1}_{\{y_n > A_\alpha\}}}\right]^{\frac{1}{2}}. \end{aligned}$$

Therefore, $\sigma^\tau = m\sigma$ and $\mu^\tau = mA_\alpha + c - m\sigma\Phi^{-1}(\alpha) = m\mu + c$. So we can get that $X^\tau \sim N(m\mu + c, m^2\sigma^2)$ and X^τ has the same distribution as $mX + c$. By virtue of Eq. (3.3.5), we can get the new β-level ES estimator as follows:

$$\widehat{\mathrm{ES}}_\beta(X^\tau) = [\mathrm{ES}_\beta(X^\tau) - mA_\alpha - c] f_{\alpha,\beta}(\gamma_\alpha^\tau) + mA_\alpha + c$$

$$= [m\mathrm{ES}_\beta(X) + c - mA_\alpha - c]f_{\alpha,\beta}(\gamma_\alpha) + mA_\alpha + c$$
$$= m[\mathrm{ES}_\beta(X) - A_\alpha]f_{\alpha,\beta}(\gamma_\alpha) + mA_\alpha + c$$
$$= m\widehat{\mathrm{ES}}_\beta(X) + c. \tag{3.6.3}$$

where $\mathrm{ES}_\beta(X^\tau) = m\mathrm{ES}_\beta(X) + c$ is derived from Eq. (3.2.13). The other scenario that a loss r.v. W is transformed linearly can be proved similarly. **Q.E.D**

From Proposition 1, we have $\mathrm{ES}^\tau_{\beta,1}(i) = m\,\mathrm{ES}_{\beta,1}(i) + c$. As for AA estimator, $\mathrm{ES}^\tau_{\beta,2}(i)$ equals $m\,\mathrm{ES}_{\beta,2}(i) + c$ obviously. Next we consider the effects on the EVT estimator; in particular, we examine whether $\mathrm{ES}^\tau_{\beta,3}(i)$ equals $m\,\mathrm{ES}_{\beta,3}(i) + c$.

Suppose $\{w_k\}_{k=1}^{n_v}$ denote the n_v losses that are greater than the threshold v in the original sample. After the linear transformation, the new threshold is $m\,v + c$ and losses beyond it become $\{m\,w_k + c\}_{k=1}^{n_v}$. Assuming $(\hat{\xi}, \hat{\sigma})$ are the maximum likelihood estimation (MLE) GPD parameters of the i-th original sample, we have

$$(\hat{\xi}, \hat{\sigma}) = \arg\max_{\xi,\sigma} \sum_{k=1}^{n_v} -\ln\sigma - \left(\frac{1}{\xi} + 1\right)\ln\left(1 + \frac{\xi}{\sigma}(w_k - v)\right), \quad \xi \neq 0. \tag{3.6.4}$$

Suppose $(\hat{\xi}^\tau, \hat{\sigma}^\tau)$ are the MLE GPD parameters after linear transformation. Then they satisfy:

$$\begin{aligned}(\hat{\xi}^\tau, \hat{\sigma}^\tau) &= \arg\max_{\xi,\sigma} \sum_{k=1}^{n_v} -\ln\sigma - \left(\frac{1}{\xi} + 1\right)\ln\left(1 + \frac{\xi}{\sigma}(m\,w_k + c - m\,v - c)\right), \quad \xi \neq 0 \\ &= \arg\max_{\xi,\sigma} \sum_{k=1}^{n_v} -\ln(\sigma/m) - \left(\frac{1}{\xi} + 1\right)\ln\left(1 + \frac{\xi}{\sigma/m}(w_k - v)\right), \quad \xi \neq 0. \end{aligned} \tag{3.6.5}$$

Comparing Eq. (3.6.5) to Eq. (3.6.4), we conclude that $\hat{\xi}^\tau = \hat{\xi}$ and $\hat{\sigma}^\tau = m\,\hat{\sigma}$. According to

Table 3.4, the EVT β-level ES estimator ($\xi \neq 0$) for the sample after the linear transformation is

$$\begin{aligned} \mathrm{ES}^{\tau}_{\beta,3}(i) &= \frac{m\,v + c + \frac{\hat{\sigma}^{\tau}}{\hat{\xi}^{\tau}}\left[(\frac{1-\beta}{n_v/N})^{-\hat{\xi}^{\tau}} - 1\right] + \hat{\sigma}^{\tau} - \hat{\xi}^{\tau}(m\,v + c)}{1 - \hat{\xi}^{\tau}} \\ &= \frac{m\,v + m\frac{\hat{\sigma}}{\hat{\xi}}\left[(\frac{1-\beta}{n_v/N})^{-\hat{\xi}} - 1\right] + m\hat{\sigma} - m\hat{\xi}v + (1-\hat{\xi})c}{1 - \hat{\xi}} \\ &= m\,\mathrm{ES}_{\beta,3}(i) + c. \end{aligned} \tag{3.6.6}$$

The other situation when $\xi = 0$ can be demonstrated similarly. Therefore, all three ES estimators are subject to the same linear transformation as the simulated sample. Furthermore, based on the definition of MSE, we have

$$\begin{aligned} & (\mathrm{ES}^{\tau}_{\beta,j}(i) - \text{true } \mathrm{ES}^{\tau}_{\beta})^2 \\ =\ & (m\mathrm{ES}_{\beta,j}(i) + c - m(\text{true } ES_{\beta}) - c)^2, \\ =\ & m^2(\mathrm{ES}_{\beta,j}(i) - \text{true } \mathrm{ES}_{\beta})^2, \quad j = 1,2,3,\ i = 1,2,\ldots,10^3. \end{aligned} \tag{3.6.7}$$

Therefore, all the squared errors of the new sample are proved to be proportional to the original ones with a constant multiple m^2 by Eq. (3.6.7). Every simulated sample has this property, so does the MSE for any examined underlying distribution. We can conclude that the superiority of our estimator in MSE will be maintained even after a linear transformation ($m \neq 0$) is applied to the underlying distribution.

3.7 Conclusion and Future Work

In this chapter, we propose a simple and robust ES estimation method on a basis of the *tail-based* normal approximation. The regression models related to sample's tail weight are also introduced to make the ES estimates more accurate. For various heavy-tailed loss distributions, the regression-adjusted estimation errors are all considerably small. Moreover, compared to the commonly-used arithmetic average and EVT estimators, our ES estimator is preferred, especially

at a high confidence level such as 99.5% with a small sample size of 250 or 500. In addition, we also show that our method works well under linear transformations, which further adds to its practicality in portfolio management.

Nonetheless, we only consider the scenario that $\beta = 99\%$ or 99.5% with $\alpha = 95\%$ and other combinations may have better performances. It is also likely our regression models would fail to work for an extremely large conditional skewness γ_α ($\gamma_\alpha > 12$, etc.) even though that situation is very rare in practice. Furthermore, instead of fitting the tailed data, the *tail-based* normal approximation matches specific statistics of the excessive losses. That normal approximation itself cannot describe the tail behaviors correctly. It needs to work jointly with the regression model to give the ES estimate. These deficiencies are worth looking into in the future.

CHAPTER 4

GARCH vs. ARSV under the Physical and Risk-Neutral Measures

4.1 Introduction

Although return volatility cannot be observed directly, there is never too much emphasis on monitoring it. As a fundamental parameter, return volatility plays crucial roles in pricing derivatives and estimating measures of risk. Hence, how to precisely estimate and forecast market return volatilities has been an enduringly popular field of research. There are three prevalent types of volatility models: the ARCH/GARCH-family models, the stochastic volatility (SV) models and the realized volatility (RV) model. The RV model is developing fast these years but it depends on high-frequency intra-daily data which is not always available. The scope of this chapter is thus limited to the first two types of models—specifically, the basic GARCH(1, 1) and ARSV(1) models. The fundamental difference between these two models is that the ARSV(1) model considers the volatility to be a latent variable with the unexpected noise while the volatility is deterministic in the GARCH(1, 1) model.

Despite the extra adaptability, the latent volatility process in the SV models adds to

the difficulty of parameter estimation. Due to the improvement in computing capacity along with more efficient estimation methods, it is increasingly likely to take advantage of the SV models when estimating or forecasting volatilities. As alternatives to the ARCH/GARCH-family models, the SV models generally better explain some stylized facts of financial time series. For example, M. Á. Carnero, Peña, & Ruiz (2001) show the ARSV(1) model is more flexible than the GARCH(1, 1) model in terms of excess kurtosis, low first-order autocorrelation and high persistence of volatility. Yu (2002) makes a comparison of nine models in terms of predicting volatilities using New Zealand stock data and demonstrate the ARSV(1) model outperforms the rival candidates. Lehar, Scheicher, & Schittenkopf (2002) also compare the performance of the GARCH and Hull-White models in terms of the out-of-sample option valuation errors and Value-at-Risk forecasts. Moreover, the abilities to reproduce the stylized facts in financial series of the GARCH(1, 1), exponential GARCH(1, 1) and ARSV(1) models are investigated by Malmsten, Teräsvirta, et al. (2004) and they conclude that none of the models dominates the others.

Generally speaking, measurements of return volatility of financial assets can be classified into two main categories based on the probability measures. The first one directly takes into account historical market data and tracks volatility by fitting asset return series. Volatility models that give such measurements are known to be built under the real-world or physical probability measure. The second measurement derives the volatility of the underlying asset from its derivative market prices. This measurement follows volatility under the so-called risk-neutral probability measure because of its relationship with the derivative pricing.

Until now, the risk-neutral GARCH(1, 1) and ARSV(1) models have not been directly compared in the previous empirical studies. This chapter complements the literature by a comprehensive comparison of the GARCH(1, 1) and ARSV(1) models regarding in-sample fitting and out-of-sample prediction capabilities under both the physical and risk-neutral measures. Under the physical measure, we compare their log-likelihoods after fitting historical returns and test normality for the error terms. Moreover, two robust loss functions, MSE and QLIKE,

are referred to for the one-step-ahead volatility forecast comparison. On the other hand, under the risk-neutral measure, the in-sample and out-of-sample option pricing errors of these two models are investigated. When pricing options, we consider two choices of the initial volatility: a casually determined value derived from historical returns and an optimal value that minimizes the in-sample pricing error. The latter choice leads to the implied versions of the two models. The results show the ARSV(1) model outperforms the GARCH(1, 1) model in terms of the in-sample and out-of-sample performances under the physical measure. Under the risk-neutral measure, these two models obtain considerably similar results when pricing call options while the ARSV(1) model performs substantially better for put options. Although the implied versions of these two models improve the in-sample pricing errors as expected, they are sensitive to the initial conditions and not robust for the out-of-sample predictions.

The rest of this chapter is organized as follows: In Section 4.2, we discuss the parameter estimation methods for the GARCH(1, 1) and ARSV(1) models under both the physical and risk-neutral measures. Section 4.3 introduces the dataset and discuss the methodologies for the comprehensive comparison of these two models. Section 4.4 investigates the empirical results. In section 4.5, we summarize this chapter and point out directions for the future work. All technical details are included in Appendix B.

4.2 Models and Parameter Estimation Methods

4.2.1 Estimating GARCH(1, 1) Model under the Physical Measure

Considering the historical returns, a volatility model under the physical measure is estimated by fitting the return series as precisely as possible. The GARCH model relates the current conditional variance to the lagged squared residuals and lagged conditional variance estimates. As the most commonly-used GARCH model, the GARCH(1, 1) model under the physical measure

is constructed as in Eq. (4.2.1):

$$\begin{aligned} y_t &= \sigma_t z_t, \quad z_t \overset{\text{i.i.d}}{\sim} N(0,1), \\ \sigma_t^2 &= a_0 + a_1 \sigma_{t-1}^2 z_{t-1}^2 + b_1 \sigma_{t-1}^2. \end{aligned} \tag{4.2.1}$$

where $\{y_t\}_{t\geq 1}$ denotes the log daily return series (can be magnified), σ_t^2 stands for the conditional variance at time t. To ensure the stationarity of the variance, it is required that the parameters satisfy $a_0 > 0, a_1 \geq 0, b_1 \geq 0$ and $a_1 + b_1 < 1$.

Since z_t conditionally follows a normal distribution, the maximum likelihood estimation method can be adopted to estimate the GARCH(1, 1) model. Suppose $\theta_1 = (a_0, a_1, a_2)$ is its constant parameter vector to be estimated under the physical measure. Given a sample of T realized log daily returns, the GARCH(1, 1) model under the physical measure is estimated by maximizing its log likelihood function, denoted by $\ln[p_\theta(y_{1:T})]$, as follows:

$$\begin{aligned} \hat{\theta}_1 = (\hat{a}_0, \hat{a}_1, \hat{b}_1) &= \underset{a_0, a_1, b_1}{\arg\max} \ \ln[p_\theta(y_{1:T})], \\ \ln[p_\theta(y_{1:T})] &= -\frac{1}{2}\left[T\ln(2\pi) + \sum_{t=1}^{T} \ln(\sigma_t^2) + \sum_{t=1}^{T} \frac{y_t^2}{\sigma_t^2} \right]. \end{aligned} \tag{4.2.2}$$

4.2.2 Estimating ARSV(1) Model under the Physical Measure

There have been a lot of studies on Taylor's ARSV model. Actually, in Bayesian time series analysis, the ARSV model is a fundamental example when studying the Markov chain Monte Carlo (MCMC) methods and Particle Filter methods due to its non-linearity that makes the traditional Kalman filter method infeasible. Assuming the latent log variance follows a Gaussian AR(1) process, the ARSV(1) model is generally given by

$$\begin{aligned} y_t &= \exp(\frac{x_t'}{2})\xi_t, \quad \xi_t \overset{\text{i.i.d}}{\sim} N(0,1), \\ x_t' &= \alpha + \phi x_{t-1}' + \gamma \eta_t, \quad \eta_t \overset{\text{i.i.d}}{\sim} N(0,1), \quad \xi_t \perp \eta_t. \end{aligned} \tag{4.2.3}$$

where $\{y_t\}_{t\geq 1}$ is still the log daily return series and $x'_t = \ln \sigma_t^2$ denotes the log conditional variance at time t.

In this chapter, we adopt the following ARSV(1) model since it is easier to estimate and more straightforward to implement in the option pricing model. It can be shown that Eq. (4.2.3) is equivalent to Eq. (4.2.4) as long as we consider $\beta^2 \exp(x_t)$ in Eq. (4.2.4) to be the conditional variance (See Appendix B.1).

$$\begin{aligned} y_t &= \beta \exp(\frac{x_t}{2})\xi_t, \quad \xi_t \overset{\text{i.i.d}}{\sim} N(0,1), \\ x_t &= \phi x_{t-1} + \gamma \eta_t, \quad \eta_t \overset{\text{i.i.d}}{\sim} N(0,1), \quad \xi_t \perp \eta_t. \end{aligned} \tag{4.2.4}$$

Suppose $\theta_2 = (\phi, \gamma^2, \beta^2)$ is the constant parameter vector of the ARSV(1) model under the physical measure. The scale parameter β replaces the constant drift of the original log variance process in Eq. (4.2.3). Sometimes a constant or even a stochastic return drift is added in the first equation of Eq. (4.2.4) to capture the sharp return changes. Here no return drift is added to keep in line with the GARCH(1, 1) model above and this selection also makes it easier to estimate. ϕ can be seen as a persistence parameter in log variance process and $|\phi| < 1$ should be ensured to satisfy the stationarity condition. In addition, $\{\xi_t\}_{t\geq 1}$ and $\{\eta_t\}_{t\geq 1}$ are two independent processes in this chapter though in many cases they are assumed to be correlated in order to capture the leverage effect.

Unlike the GARCH(1, 1) model, the log variance process in the ARSV(1) model has an unexpected noise term, which implies why it is called stochastic volatility. Despite its lack of the analytical likelihood function, many Monte Carlo simulation methods have been proposed to estimate the ARSV(1) model. MCMC algorithms such as Gibbs sampler can be adopted to calculate the posterior distributions of its parameters. Moreover, Particle Filter, which is also referred to as Sequential Monte Carlo, is effective to estimate its likelihood and then the expectation-maximization (EM) or gradient ascent method can be implemented to maximize the estimated likelihood. Particle MCMC, a combination of the Particle Filter and MCMC methods, also works for parameter learning of the ARSV(1) model. In this chapter, we adopt a

forward-only version of the Forward Filter Backward Smoothing (FFBS) algorithm with EM method (Del Moral, Doucet, & Singh, 2010) to maximize the particle-based likelihood of the ARSV(1) model. In what follows we will introduce this algorithm and derive its implementation to the estimation of the ARSV(1) model.

4.2.2.1 Forward Filter Backward Smoothing

As expressed in the form of Eq. (4.2.4), the ARSV(1) model belongs to the family of state space model (SSM) which is constructed by a transition equation for state variables and a measurement equation for observations. SSM assumes the observation and the latent transition processes are both Markovian. For the ease of presentation, hereafter $u_{i:j}$ denotes the sequence $\{u_i, u_{i+1}, \ldots, u_j\}$. Moreover, x_1 follows a given prior distribution. $f_\theta(\cdot|\cdot)$ and $g_\theta(\cdot|\cdot)$ stand for the transition and measurement equations, respectively. In the ARSV(1) model with a parameter vector θ, the unobservable state variable, x_t, is independent of the history of states and observations once x_{t-1} is given:

$$p(x_t|x_{1:t-1}, y_{1:t-1}) = p(x_t|x_{t-1}, x_{1:t-2}, y_{1:t-1}) = f_\theta(x_t|x_{t-1}). \tag{4.2.5}$$

If $\{\xi_t\}_{t\geq 1}$ are independent of $\{\eta_t\}_{t\geq 1}$, we also have

$$p(y_t|x_{1:t}, y_{1:t-1}) = p(y_t|x_t, x_{1:t-1}, y_{1:t-1}) = g_\theta(y_t|x_t). \tag{4.2.6}$$

Given n observations, suppose $s_k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}, k \in \mathbb{N}$, is a sequence of functions and $S_n(x_{1:n}) = \sum_{k=2}^{n} s_k(x_{k-1}, x_k)$ denotes the corresponding sequence of additive functionals built from s_k. Then the *smoothed additive functional* is defined by

$$\mathcal{S}_n^\theta = \mathbb{E}_\theta[S_n(x_{1:n})|y_{1:n}] = \int \left[\sum_{k=2}^{n} s_k(x_{k-1}, x_k)\right] p_\theta(x_{1:n}|y_{1:n}) dx_{1:n}. \tag{4.2.7}$$

The FFBS algorithm can be adopted through two steps to compute $\mathcal{S}_n^\theta$ that is inevitable for a maximum-likelihood method. In the particle filter framework, the first step is to approximate the particle-based $p_\theta(x_k|y_{1:k})$ through a forward particle filtering for $k = 1, 2, \ldots, n$. At any time step k, let $x_k^{(i)}$ denote the i-th particle and $w_k^{(i)}$, an approximation of $p_\theta(x_k^{(i)}|y_{1:k})$, stand for its weight for $i = 1, 2, \ldots, N$ where N is the number of particles at each time step. Using the particles with corresponding weights, the second step is to calculate $p_\theta(x_{k-1}, x_k|y_{1:n})$ by backward recursion. Subsequently, the particle-based approximation of $\mathcal{S}_n^\theta$ is given by

$$\widehat{\mathcal{S}}_n^\theta = \sum_{k=2}^{n} \int s_k(x_{k-1}, x_k) \hat{p}_\theta(dx_{k-1:k}|y_{1:n}). \tag{4.2.8}$$

Though feasible, Eq. (4.2.8) is quite complicated in computation. Fortunately, when $S_n(x_{1:n})$ is in the form of additive functionals, the challenging backward recursion step can be avoided through a forward-only version of the FFBS algorithm. Define an auxiliary function by

$$T_n^\theta(x_n) \equiv \int S_n(x_{1:n}) p_\theta(x_{1:n-1}|y_{1:n-1}, x_n) dx_{1:n-1}. \tag{4.2.9}$$

$T_1^\theta(x_1)$ is generally set to be 0 and $T_n^\theta(x_n)$ can be obtained recursively as follows:

$$\begin{aligned}
T_n^\theta(x_n) &= \int [S_{n-1}(x_{1:n-1}) + s_n(x_{n-1}, x_n)] p_\theta(x_{1:n-1}|y_{1:n-1}, x_n) dx_{1:n-1} \\
&= \int \left[\int S_{n-1}(x_{1:n-1}) p_\theta(x_{1:n-2}|y_{1:n-2}, x_{n-1}) dx_{1:n-2} \right] p_\theta(x_{n-1}|y_{1:n-1}, x_n) dx_{n-1} \\
&\quad + \int s_n(x_{n-1}, x_n) p_\theta(x_{1:n-1}|y_{1:n-1}, x_n) dx_{1:n-1} \\
&= \int T_{n-1}^\theta(x_{n-1}) p_\theta(x_{n-1}|y_{1:n-1}, x_n) dx_{n-1} \\
&\quad + \int s_n(x_{n-1}, x_n) p_\theta(x_{n-1}|y_{1:n-1}, x_n) dx_{n-1} \\
&= \int \left[T_{n-1}^\theta(x_{n-1}) + s_n(x_{n-1}, x_n) \right] p_\theta(x_{n-1}|y_{1:n-1}, x_n) dx_{n-1}.
\end{aligned} \tag{4.2.10}$$

where $p_\theta(x_{n-1}|y_{1:n-1}, x_n)$ is approximated using the forward-filtering weighted particles:

$$\hat{p}_\theta(dx_{n-1}|y_{1:n-1}, x_n) = \frac{\sum_{j=1}^{N} \omega_{n-1}^{(j)} f_\theta(x_n|x_{n-1}^{(j)}) \delta_{x_{n-1}^{(j)}}(dx_{n-1})}{\sum_{l=1}^{N} \omega_{n-1}^{(l)} f_\theta(x_n|x_{n-1}^{(l)})}. \tag{4.2.11}$$

Particle Filter methods such as Bootstrap Filter (BF) and Auxiliary Particle Filter (APF) (see Appendix B.4.1) are feasible for the necessary forward filtering process. Obviously, we can use $T_n^\theta(x_n)$ to express Eq. (4.2.7) as follows:

$$\begin{aligned}
\mathcal{S}_n^\theta &= \int S_n(x_{1:n}) p_\theta(x_{1:n}|y_{1:n}) dx_{1:n} = \int S_n(x_{1:n}) p_\theta(x_{1:n-1}|y_{1:n}, x_n) p_\theta(x_n|y_{1:n}) dx_{1:n} \\
&= \int S_n(x_{1:n}) p_\theta(x_{1:n-1}|y_{1:n-1}, x_n) p_\theta(x_n|y_{1:n}) dx_{1:n} \\
&= \int \int S_n(x_{1:n}) p_\theta(x_{1:n-1}|y_{1:n-1}, x_n) dx_{1:n-1} p_\theta(x_n|y_{1:n}) dx_n \\
&= \int T_n^\theta(x_n) p_\theta(x_n|y_{1:n}) dx_n.
\end{aligned} \tag{4.2.12}$$

where $p_\theta(x_{1:n-1}|y_{1:n}, x_n) = p_\theta(x_{1:n-1}|y_{1:n-1}, x_n)$ because of the Markovian property of SSM. Accordingly, based on Eq. (4.2.10)∼Eq. (4.2.12), an algorithm to approximate $\mathcal{S}_n^\theta$ is derived in Appendix B.2.

4.2.2.2 Off-line EM Method

When maximizing the particle-based likelihood, techniques such as gradient ascent and EM methods can be applied in the on-line or off-line scheme. The off-line scheme means updating parameter estimations after capturing all observations while the on-line scheme updates parameter estimations at each time step when a new observation arrives. Generally, the dataset of a real market problem is not large enough for the on-line scheme so the off-line scheme will be used in this chapter. The EM method consists of two steps: the expectation step (E-step) and the maximization step (M-step). In addition, the EM method, if feasible, is preferred to the gradient ascent method because it has no worries about the step size problem.

In the off-line EM method, assuming θ_s is the parameter vector at s-th iteration, then the

parameter vector at $(s+1)$-th iteration is computed by

$$\theta_{s+1} = \arg\max_{\theta} Q(\theta_s, \theta). \tag{4.2.13}$$

where $Q(\theta_s, \theta) = \int \log p_\theta(x_{1:n}, y_{1:n}) p_{\theta_s}(x_{1:n}|y_{1:n}) dx_{1:n}$ denotes the expectation of the log likelihood in the E-step. When $p_\theta(x_{1:n}, y_{1:n})$ belongs to the exponential family, the M-step in EM method can be simply finished through a function of sufficient statistics calculated by the forward-only FFBS algorithm. Since the ARSV(1) model admits this beautiful property, it is straightforward to update its parameter estimates in Eq. (4.2.13).

Suppose $\{s^h\}_{h=1}^m$ is the collection of m sufficient statistics that are needed in the parameter update function. The summary statistic is calculated by

$$S_{h,n}^{\theta} = \int S_{h,n}(x_{1:n}, y_{1:n}) p_\theta(x_{1:n}|y_{1:n}) dx_{1:n}, \quad h = 1, \ldots, m. \tag{4.2.14}$$

where $S_{h,n}(x_{1:n}, y_{1:n}) = \sum_{k=1}^n s^h(x_{k-1}, x_k, y_k)$ is in the additive form of Eq. (4.2.7). That's why the forward-only FFBS algorithm is feasible to calculate the summary statistics. For a model in the exponential family, we first transform $p_\theta(x_{k+1}, y_{k+1}|x_k)$ to the following form:

$$p_\theta(x_{k+1}, y_{k+1}|x_k) = v(x_{k+1}, y_{k+1}) \exp\big(\langle \psi(\theta), s(x_k, x_{k+1}, y_{k+1})\rangle - A(\theta)\big). \tag{4.2.15}$$

where $s(x_k, x_{k+1}, y_{k+1}) = [s^1(x_k, x_{k+1}, y_{k+1}), \ldots, s^m(x_k, x_{k+1}, y_{k+1})]$ is the vector of sufficient statistics, θ denotes the parameter vector and $\langle\cdot\rangle$ stands for the scalar product. According to Cappé (2011), the maximizing step can be completed through the following update function:

$$\theta_{s+1} = \Lambda(\frac{S_n^{\theta_s}}{n}). \tag{4.2.16}$$

$S_n^{\theta_s}$ is a m-dimensional vector whose h-th element, $S_{h,n}^{\theta_s}$, can be derived by Eq. (4.2.14) using the forward-only FFBS algorithm and $\Lambda(s(x_k, x_{k+1}, y_{k+1}))$ is the unique solution of $\nabla_\theta \psi(\theta) s(x_k, x_{k+1}, y_{k+1}) - \nabla_\theta A(\theta) = 0$.

4.2.2.3 Particle-based Parameter Estimation of ARSV(1) Model

Now let us implement the forward-only FFBS algorithm with the off-line EM method to estimate the ARSV(1) model. According to Eq. (4.2.4), we write down $p_\theta(x_{t+1}, y_{t+1}|x_t)$ into the form of Eq. (4.2.15):

$$\begin{aligned}
& p_\theta(x_{t+1}, y_{t+1}|x_t) \\
&= p_\theta(y_{t+1}|x_{t+1}, x_t) p_\theta(x_{t+1}|x_t) \\
&= \frac{1}{\sqrt{2\pi}\beta \exp(x_{t+1}/2)} \exp[-\frac{y_{t+1}^2}{2\beta^2 \exp(x_{t+1})}] \frac{1}{\sqrt{2\pi}\gamma} \exp[-\frac{(x_{t+1}-\phi x_t)^2}{2\gamma^2}] \\
&= \frac{1}{2\pi \exp(x_{t+1}/2)} \exp[-\frac{\phi^2 x_t^2}{2\gamma^2} + \frac{\phi x_t x_{t+1}}{\gamma^2} - \frac{x_{t+1}^2}{2\gamma^2} - \frac{y_{t+1}^2 \exp(-x_{t+1})}{2\beta^2} - \frac{1}{2}\ln(\beta^2\gamma^2)].
\end{aligned}$$

where $\theta = (\phi, \gamma^2, \beta^2)$, $v(x_{t+1}, y_{t+1}) = \frac{1}{2\pi \exp(x_{t+1}/2)}$, $A(\theta) = \frac{1}{2}\ln(\gamma^2) + \frac{1}{2}\ln(\beta^2)$, $\psi(\theta) = (-\frac{\phi^2}{2\gamma^2}, \frac{\phi}{\gamma^2}, -\frac{1}{2\gamma^2}, -\frac{1}{2\beta^2})$ and $s_{t+1}(x_t, x_{t+1}, y_{t+1}) = (x_t^2, x_t x_{t+1}, x_{t+1}^2, y_{t+1}^2 \exp(-x_{t+1}))$. For ease of presentation, let vector (z_1, z_2, z_3, z_4) denote the vector $(x_t^2, x_t x_{t+1}, x_{t+1}^2, y_{t+1}^2 \exp(-x_{t+1}))$. The unique solution to the maximizing step is $\Lambda(z_1, z_2, z_3, z_4) = (\frac{z_2}{z_1}, z_3 - \frac{z_2^2}{z_1}, z_4)$ (see Appendix B.3). Since there are four sufficient statistics, m is equal to 4 in the off-line EM method. The algorithm to estimate parameters in ARSV(1) under the physical measure is listed as follows:

1 Obtain initial parameter estimates $\theta_0 = (\phi_0, \gamma_0^2, \beta_0^2)$.

2 For iteration $l = 0, 1, \ldots,$ ItN, update parameter estimate by the following steps:

- 2.1 Generate $\{x_1^{(i)}\}_{i=1}^N$ from the prior distribution with θ_l, initialize $\widehat{T}_1^{(i)}(\theta_l) = 0$ and set weighted filtering particles $\{x_1^{(i)}, \omega_1^{(i)} = \frac{1}{N}\}_{i=1}^N$.
- 2.2 Repeat the following steps for time steps $k = 2, \ldots, T$ with θ_l by APF algorithm:
  - 2.2.1 Obtain normalized weighted filtering particles $\{x_k^{(i)}, \omega_k^{(i)}\}$ for $i = 1, \ldots, N$.
  - 2.2.2 $\widehat{T}_k^{(i)}(\theta_l) = \dfrac{\sum_{j=1}^N \omega_{k-1}^{(j)} f_\theta(x_k^{(i)}|x_{k-1}^{(j)})[\widehat{T}_{k-1}^{(j)}(\theta_l) + s_k(x_{k-1}^{(j)}, x_k^{(i)}, y_k)]}{\sum_{j=1}^N \omega_{k-1}^{(j)} f_\theta(x_k^{(i)}|x_{k-1}^{(j)})}, i = 1, \ldots, N.$
- 2.3 Obtain $\widehat{\mathcal{S}}_T = \sum_{i=1}^N \omega_T^{(i)} \widehat{T}_T^{(i)}(\theta_l)$ and update parameter estimates by $\theta_{l+1} = \Lambda(T^{-1}\widehat{\mathcal{S}}_T)$.

In the algorithm above, T is the number of observations, ItN is the number of iterations,

$\widehat{T}_k^{(i)}(\theta_l)$ is a 4-dimensional vector and the APF algorithm adopted in Step 2.2 is listed in Appendix B.4.1. In addition, $s_k(x_{k-1}, x_k, y_k)$ and $\Lambda(\cdot) : \mathbb{R}^4 \to \mathbb{R}^3$ have been derived as above.

4.2.3 Estimating Models under the Risk-Neutral Measure

Instead of maximizing likelihood, parameter estimation under the risk-neutral measure is related to option pricing. Duan (1995) shows the risk-neutral GARCH(1, 1) model is given by

$$\begin{aligned} \ln \frac{S_t}{S_{t-1}} &= r_1 - \frac{1}{2}\tilde{\sigma}_t^2 + \tilde{\sigma}_t z_t, \quad z_t \overset{\text{i.i.d}}{\sim} N(0,1), \\ \tilde{\sigma}_t^2 &= \tilde{a}_0 + \tilde{a}_1 \tilde{\sigma}_{t-1}^2 z_{t-1}^2 + \tilde{b}_1 \tilde{\sigma}_{t-1}^2. \end{aligned} \tag{4.2.17}$$

where r_1 denotes the daily risk-free interest rate and S_t denotes the asset spot price at time t. Similarly, the risk-neutral ARSV(1) model can be derived as follows:

$$\begin{aligned} \ln \frac{S_t}{S_{t-1}} &= r_1 - \frac{1}{2}\tilde{\beta}^2 \exp(\tilde{x}_t) + \tilde{\beta}\exp(\frac{\tilde{x}_t}{2})\xi_t, \quad \xi_t \overset{\text{i.i.d}}{\sim} N(0,1), \\ \tilde{x}_t &= \tilde{\phi}\tilde{x}_{t-1} + \tilde{\gamma}\eta_t, \quad \eta_t \overset{\text{i.i.d}}{\sim} N(0,1), \quad \xi_t \perp \eta_t. \end{aligned} \tag{4.2.18}$$

As for a certain European call/put option at time t which expires at time T, its price is calculated as the discounted average pay-off at maturity under the risk-neutral probability measure Q:

$$\begin{aligned} C_t &= e^{-r_1(T-t)}\mathbb{E}^Q[\max(S_T - K, 0)|\mathcal{F}(t)], \\ P_t &= e^{-r_1(T-t)}\mathbb{E}^Q[\max(K - S_T, 0)|\mathcal{F}(t)]. \end{aligned} \tag{4.2.19}$$

where K is its strike price, S_T is the asset price at maturity T, $\mathcal{F}(t)$ is a filtration and r_1 is still the daily risk-free interest rate. The unit of time in Eq. (4.2.19) is 'trading day'. Without a analytical solution for the option price for both the risk-neutral GARCH(1, 1) and ARSV(1) models, the option needs to be priced by Monte Carlo simulation. Suppose the initial volatility estimate at time t is σ_t. Based on Eq. (4.2.17), the asset price at maturity T under risk-neutral GARCH(1, 1) model is computed by a simulation of the return process step by step (the log

variance at each time step is also determined) as follows:

$$S_T^G = S_t \exp\left[(T-t)r_1 - \frac{1}{2}\sum_{k=t+1}^{T}\tilde{\sigma}_k^2 + \sum_{k=t+1}^{T}\tilde{\sigma}_k z_k\right]. \tag{4.2.20}$$

where the superscript G stands for the GARCH(1, 1) model. Repeating the simulation m times, then the simulated European call/put option price at time t of the risk-neutral GARCH(1, 1) model is calculated by

$$\begin{aligned}\hat{C}_t^G(\theta) &\approx \frac{1}{m}e^{-r_1(T-t)}\sum_{j=1}^{m}\max(S_T^G(j)-K,0),\\ \hat{P}_t^G(\theta) &\approx \frac{1}{m}e^{-r_1(T-t)}\sum_{j=1}^{m}\max(K-S_T^G(j),0).\end{aligned} \tag{4.2.21}$$

where θ is the parameter vector of the risk-neutral GARCH(1, 1) model and j is the simulation index.

Unlike the GARCH(1, 1) option pricing model, the risk-neutral ARSV(1) model has a volatility process that is independent of the return dynamics. Since σ_t is the initial volatility estimate at time t, then the initial x_t is set to be $2\ln(\sigma_t/\beta)$. Conditional on one simulated $\{x_k\}_{k=t+1}^{T}$ sequence, the asset call/put option price at time t of the risk-neutral ARSV(1) model can be calculated with the assistance of the famous Black-Scholes (B-S) formula as follows:

$$\mathbb{E}^Q[C_t^A|\exp(\tilde{x}_{t+1}),\exp(\tilde{x}_{t+2}),\ldots,\exp(\tilde{x}_T)] = BS_{\text{call}}(T-t,S_t,K,r_1,\hat{\sigma}), \tag{4.2.22}$$

$$\mathbb{E}^Q[P_t^A|\exp(\tilde{x}_{t+1}),\exp(\tilde{x}_{t+2}),\ldots,\exp(\tilde{x}_T)] = BS_{\text{put}}(T-t,S_t,K,r_1,\hat{\sigma}), \tag{4.2.23}$$

$$\hat{\sigma}^2 = \frac{1}{T-t}\sum_{k=t+1}^{T}\tilde{\beta}^2\exp(\tilde{x}_k). \tag{4.2.24}$$

We still use time unit 'trading day' for parameters of the B-S formula in Eq. (4.2.22), Eq. (4.2.23) and superscript A stands for the ARSV(1) model. Despite the commonly-used yearly-expressed parameters in the B-S formula, the daily-expressed ones actually work in the same way. In other words, $BS_{\text{call}}(\frac{T-t}{Y},S_t,K,Yr_1,\sqrt{Y}\hat{\sigma}) = BS_{\text{call}}(T-t,S_t,K,r_1,\hat{\sigma})$ given T, t, r_1 and $\hat{\sigma}$ are

all expressed in 'day' and Y ($Y = 252$ in this chapter) is the amount of trading days in a year. The proof of Eq. (4.2.22) and Eq. (4.2.23) is straightforward.

Proof. Suppose one simulated sequence $\{\tilde{x}_k\}_{k=t+1}^{T}$ has been obtained. Then we have

$$\ln S_T^A|\{\exp(\tilde{x}_k)\}_{k=t+1}^{T} \overset{D}{=} \ln S_t + r_1(T-t) - \frac{1}{2}\tilde{\beta}^2 \sum_{k=t+1}^{T} \exp(\tilde{x}_k) + \sum_{k=t+1}^{T} \tilde{\beta}\exp(\frac{\tilde{x}_k}{2})\xi_k, \quad (4.2.25)$$

where $\overset{D}{=}$ stands for the operator for 'equivalence in distribution'. As $\xi_k \overset{\text{i.i.d}}{\sim} N(0,1)$ for $k = t+1, \ldots, T$, we have $\sum_{k=t+1}^{T} \tilde{\beta}\exp\left(\frac{\tilde{x}_k}{2}\right)\xi_k \sim N\big(0, \tilde{\beta}^2 \sum_{k=t+1}^{T} \exp(\tilde{x}_k)\big)$ by properties of the normal distribution. Hence, using Eq. (4.2.24), we can simplify Eq. (4.2.25) by

$$\begin{aligned}
&\ln S_T^A|\{\exp(\tilde{x}_k)\}_{k=t+1}^{T} \\
\overset{D}{=}\ & \ln S_t + r_1(T-t) - \frac{1}{2}\tilde{\beta}^2 \sum_{k=t+1}^{T} \exp(\tilde{x}_k) + \Big(\sqrt{\tilde{\beta}^2 \sum_{k=t+1}^{T} \exp(\tilde{x}_k)}\Big) Z \\
\overset{D}{=}\ & \ln S_t + r_1(T-t) - \frac{1}{2}\hat{\sigma}^2(T-t) + \hat{\sigma}\sqrt{T-t}\, Z, \quad Z \sim N(0,1), \qquad (4.2.26)
\end{aligned}$$

It is obvious that Eq. (4.2.26) can been seen as a model in which the asset price follows a geometric Brownian motion. In fact, the B-S formula is derived from such model. Firstly, we transform Eq. (4.2.26) by

$$S_T^A|\{\exp(\tilde{x}_k)\}_{k=t+1}^{T} \overset{D}{=} S_t e^{(r_1 - \frac{1}{2}\hat{\sigma}^2)(T-t) + \hat{\sigma}\sqrt{T-t}\, Z}, \quad Z \sim N(0,1). \qquad (4.2.27)$$

Based on the first equation of Eq. (4.2.19) and Eq. (4.2.27), we can derive the conditional European call option price at time t of the ARSV(1) model by

$$\begin{aligned}
&\big(e^{r_1(T-t)} C_t^A\big)|\{\exp(\tilde{x}_k)\}_{k=t+1}^{T} \\
=\ & \mathbb{E}\big[\max\big((S_T^A|\{\exp(\tilde{x}_k)\}_{k=t+1}^{T}) - K, 0\big)\big] \\
=\ & \int_{-\infty}^{\infty} \max(S_t e^{(r_1 - \frac{1}{2}\hat{\sigma}^2)(T-t) + \hat{\sigma}\sqrt{T-t}\, x} - K, 0) \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx
\end{aligned}$$

$$
\begin{aligned}
&= \int_a^{\infty} (S_t e^{(r_1 - \frac{1}{2}\hat{\sigma}^2)(T-t) + \hat{\sigma}\sqrt{T-t}\,x} - K) \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \\
&= S_t e^{(r_1 - \frac{1}{2}\hat{\sigma}^2)(T-t)} \int_a^{\infty} \frac{1}{\sqrt{2\pi}} e^{\hat{\sigma}\sqrt{T-t}\,x - \frac{x^2}{2}} dx - K(1 - \Phi(a)) \\
&= S_t e^{(r_1 - \frac{1}{2}\hat{\sigma}^2)(T-t) + \frac{\hat{\sigma}^2(T-t)}{2}} \int_a^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x - \hat{\sigma}\sqrt{T-t})^2}{2}} dx - K(1 - \Phi(a)) \\
&= S_t e^{(r_1 - \frac{1}{2}\hat{\sigma}^2)(T-t) + \frac{\hat{\sigma}^2(T-t)}{2}} \int_{a - \hat{\sigma}\sqrt{(T-t)}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x'^2}{2}} dx' - K(1 - \Phi(a)) \\
&= S_t e^{r_1(T-t)} (1 - \Phi(a - \hat{\sigma}\sqrt{T-t})) - K(1 - \Phi(a)), \qquad (4.2.28)
\end{aligned}
$$

where $\Phi(\cdot)$ denotes the CDF of the standard normal distribution, $x' \equiv x - \hat{\sigma}\sqrt{T-t}$, and $a \equiv \frac{\ln(\frac{K}{S_t}) - (r_1 - \frac{1}{2}\hat{\sigma}^2)(T-t)}{\hat{\sigma}\sqrt{T-t}}$. Further simplifying Eq. (4.2.28) using $1 - \Phi(x) = \Phi(-x)$, we have

$$
C_t^A | \{\exp(\tilde{x}_k)\}_{k=t+1}^T = S_t \Phi(\hat{\sigma}\sqrt{T-t} - a) - e^{-r_1(T-t)} K \Phi(-a). \qquad (4.2.29)
$$

Eq. (4.2.29) is equivalent to the B-S formula for a call option with a constant daily volatility $\hat{\sigma}$, a constant daily risk-free rate r_1, time to maturity $T - t$ (in trading day), a spot price S_t and a strike price K. Therefore, we have

$$
\mathbb{E}^Q[C_t^A | \{\exp(\tilde{x}_k)\}_{k=t+1}^T] = BS_{\text{call}}(T-t, S_t, K, r_1, \hat{\sigma}). \qquad \textbf{Q.E.D}
$$

Then the simulated European call option price of the risk-neutral ARSV(1) model is derived by

$$
\hat{C}_t^A(\theta) = \mathbb{E}^Q\left[\mathbb{E}^Q[C_t^A | \{\exp(\tilde{x}_k)\}_{k=t+1}^T]\right] \approx \frac{1}{m} \sum_{j=1}^{m} BS_{\text{call}}(T-t, S_t, K, r_1, \hat{\sigma}(j)). \qquad (4.2.30)
$$

Similarly, the European put option can be valued by

$$
\hat{P}_t^A(\theta) = \mathbb{E}^Q\left[\mathbb{E}^Q[P_t^A | \{\exp(\tilde{x}_k)\}_{k=t+1}^T]\right] \approx \frac{1}{m} \sum_{j=1}^{m} BS_{\text{put}}(T-t, S_t, K, r_1, \hat{\sigma}(j)). \qquad (4.2.31)
$$

where $\hat{\sigma}(j)$ is computed from the j-th simulation path of the ARSV(1) log variance process with a given parameter vector θ. In this chapter, the simulation number m is set to be 10^4 and the

common random number technique is applied for options with different strike prices.

As Eq. (4.2.30) and Eq. (4.2.31) show, all we need to simulate is the log variance dynamics when pricing options with the risk-neutral ARSV(1) model even though it originally has two innovation processes. On the other hand, the conditional variance process in the risk-neutral GARCH(1, 1) model is deterministic and fully depends on the previous return. That's why only the simulation of the return dynamics is needed in the GARCH(1, 1) option pricing model as indicated by Eq. (4.2.20) and Eq. (4.2.21).

Suppose we have a collection of options with their observed market prices. As for both the risk-neutral GARCH(1, 1) and ARSV(1) models, the nonlinear-least-squares parameter estimator is obtained by minimizing the mean squared pricing error (MSPE):

$$\theta^* = \arg\min_{\theta} \frac{1}{n}\sum_{i=1}^{n}(v_i - \bar{v}_i(\theta))^2, \qquad \text{MSPE} = \frac{1}{n}\sum_{i=1}^{n}(v_i - \bar{v}_i(\theta))^2. \tag{4.2.32}$$

where v_i is the observed market price of the i-th option, n is the number of options in the collection and $\bar{v}_i(\theta)$ is the theoretical price of the i-th option derived from the corresponding volatility model with the parameter vector θ.

Another popular error loss function is the mean squared relative pricing errors (MSRPE), as defined in Eq. (4.2.33).

$$\text{MSRPE} = \frac{1}{n}\sum_{i=1}^{n}\Big(\frac{v_i - \bar{v}_i(\theta)}{v_i}\Big)^2. \tag{4.2.33}$$

This error loss function focuses on the valuation error in proportion to the market option price. Consequently, an option with a larger market price bears a relatively larger pricing error. For MSPE, however, only the absolute pricing error matters.

There is no doubt that different error loss functions will lead to different parameter estimates and rank the examined models in a different way. The choice of the loss functions is a worthwhile topic but it is outside the scope of this chapter. As Christoffersen, Jacobs, et al. (2001) points out, once a loss function is used in the parameter estimation, the model evaluation should also

depend on it. If we estimate in-sample parameters with one loss function and subsequently evaluate the out-of-sample performance with another one, the comparison among different models will become unfair and cannot reveal the authentic results. Therefore, in this chapter, only MSPE is adopted as the loss function in the parameter estimation and model evaluation under the risk-neutral measure.

4.3 Methodology and Data

4.3.1 Comparison under the Physical Measure

Using historical return series, we estimate a volatility model by maximizing its log-likelihood. For the GARCH(1, 1) model, its log-likelihood is straightforwardly calculated while for the ARSV(1) model, it needs to be computed by Particle Filter. Since both of the two models have three parameters, the maximum likelihood is valuable for the in-sample fitting comparison under the physical measure. Besides, some other statistics can be compared using the estimated volatility at each time step.

Suppose the size of the in-sample dataset is T. As for the GARCH(1, 1) model, given the parameter estimates under physical measure and the observed return series, we can fully determine the series of conditional variance σ_t^2. On the other hand, for the ARSV(1) model, its conditional variance is still a latent state variable even though its parameters have been determined. Suppose $(\hat{\phi}, \hat{\gamma}^2, \hat{\beta}^2)$ is the estimated parameter vector of the ARSV(1) model. Using particle smoothing algorithms(see Appendix B.4.2), the particle-based volatility estimate, $\mathbb{E}[\hat{\beta}\exp(x_t/2)|y_{1:T}]$, is approximated for $t = 1, 2, \ldots, T$ by Eq. (B.4.4). Subsequently, through normality tests such as the Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests, we can investigate how good the error sequence $(\left\{\frac{y_t}{\sigma_t}\right\}_{t=1}^{T}$ for GARCH(1, 1) model and $\left\{\frac{y_t}{\mathbb{E}[\hat{\beta}\exp(x_t/2)|y_{1:T}]}\right\}_{t=1}^{T}$ for ARSV(1) model) is to fit the assumed standard normal distribution.

As for any volatility model, fitting in-sample observations is one thing while forecasting volatility in the future is an entirely different challenge. A preferred model for in-sample

comparison does not necessarily guarantee a better out-sample forecast. Patton (2011) studies the properties of well-documented loss functions developed for volatility forecast evaluation and shows only MSE and QLIKE, defined in Eq. (4.3.1) and Eq. (4.3.2), are robust to noisy volatility proxy. Therefore, in this chapter we refer to these two loss functions for the out-of-sample one-step-ahead volatility forecast comparison between the two volatility models.

$$\text{MSE} \equiv \frac{1}{n}\sum_{t=1}^{n}(\hat{\sigma}_t^2 - h_{t|t-1})^2, \tag{4.3.1}$$

$$\text{QLIKE} \equiv \frac{1}{n}\sum_{t=1}^{n}(\ln h_{t|t-1} + \frac{\hat{\sigma}_t^2}{h_{t|t-1}}). \tag{4.3.2}$$

where n is the number of out-of-sample observations, $h_{t|t-1}$ is the conditional variance forecast for time t given the information set till time $t-1$, $\hat{\sigma}_t^2$ is the true conditional variance or one conditionally unbiased variance proxy at time t. In practice, true conditional variances are unobservable and the realized volatility, computed by the sum of intra-daily returns, is often considered to be a good proxy. Though such computation is not complicated, the high-frequency intra-daily return data is not always accessible. We adopt a variance proxy suggested by Awartani & Corradi (2005). This variance proxy adopts the squared filtered daily return and ensures a correct ranking of volatility forecast models; that is, in Eq. (4.3.1) and Eq. (4.3.2), $\hat{\sigma}_t^2$ is replaced by $(y_t' - \bar{y})^2$ where y_t' is the out-of-sample log daily return and $\bar{y} = \frac{1}{n}\sum_{t=1}^{n} y_t'$ denotes the mean of the out-of-sample log daily returns.

We can directly use the maximum-likelihood estimators of a volatility model to price a collection of options and then evaluate the pricing errors. Such evaluation, however, is not meaningful enough because MLE only considers historical return series while option pricing allows for the information in the future. Christoffersen, Jacobs, et al. (2002) also points out the maximum-likelihood estimators are neither suitable for option valuation nor reliable to rank models' capabilities of pricing options. Therefore, in this chapter, the performances of the two models regarding option pricing with MLE parameters will not be provided.

4.3.2 Comparison under the Risk-Neutral Measure

The risk-neutral versions of the GARCH(1, 1) and ARSV(1) models will be evaluated based on the in-sample price fitting and out-of-sample price prediction. Instead of handling a collection of options over a long time span, we analyze options in a single day, a choice adopted by Bakshi, Cao, & Chen (1997), for the following reasons. Firstly, though the option sample limited in a single day might not be sufficient to obtain a robust parameter estimation, it indeed eases the computation burden when minimizing the in-sample pricing error. Moreover, the one-day collection makes more sense because parameter estimation is updated as soon as the new information reaches while a long-term sample has to assume the parameters stay unchanged for a long time. This choice also avoids mixing up known information with unknown conditions. For example, if the collection includes options in two different days, we have to ignore the asset price after the first day when pricing first-day options even though that information is provided.

Another issue is on the selection of out-of-sample options. After estimating parameters by minimizing the in-sample option pricing error, we adopt them to calculate the pricing error of the out-of-sample options, which indicates a model's capability of predicting option prices in the future. Indeed, the out-of-sample performance draws much more attention of participants in the derivative market. A smaller in-sample error implies a model is better to fit the observed option prices based on current information set, but it is the out-of-sample prediction that instructs participants' subsequent behaviors. Following the single day in-sample collection, we also select the out-of-sample options in a single day.

Christoffersen, Jacobs, et al. (2002) value options in the next Wednesday with parameters estimated from the current Wednesday when examining different GARCH models. This is a favorable selection because it leaves five days to update parameters. However, such scheme is not suitable for the ARSV(1) model. As previously demonstrated, the new observed asset prices will directly affect the deterministic conditional variance in the GARCH(1, 1) model. However, leaving some days between in-sample and out-of-the sample collections is trivial for the ARSV(1) model because in that model, the new price information will not get involved in the log variance

dynamics which is stochastic and independent of the return process. Therefore, in this chapter, the out-of-sample options come from the day following the in-sample single day.

Moreover, an indispensable precondition for the risk-neutral parameter estimation is the initial conditional variance. Suppose the in-sample options are selected from Day t and the out-of-sample options come from the next day—Day $t+1$. In this chapter, the conditional variance of Day t, denoted by σ_t,[1] is set as the unconditional sample variance of the 180-day log return series (original, not magnified) before Day t and the conditional variance estimates after Day t are updated based on the corresponding volatility model. We also investigate the pricing error after parameterizing the initial volatility. When the risk-neutral parameters for Day t are estimated by minimizing the in-sample MSPE using Eq. (4.2.32), we assume they stay unchanged overnight. Therefore, the out-of-sample options on Day $t+1$ are valued under the corresponding volatility model using its in-sample parameter estimates, the initial conditional variance of Day t and the new information set on Day $t+1$. Then the out-of-sample pricing errors are thus determined. Moreover, the yearly risk-free interest rate is assumed to constantly equal 2.5%, which approximately sets the interest rate for each trading day as $\frac{2.5\%}{252}$.

4.3.3 Data

In this chapter, we focus on the daily close value and options of S&P 500 Index. For comparison under the physical measure, the in-sample observations, Sample A, are from the log daily return sequence during a ten-year period starting on January 2nd, 1996, and ending on December 30th, 2005. When estimating the in-sample MLE parameters analytically (GARCH) or numerically (ARSV), we consider the magnified return series, $y_t = 100 \times \ln \frac{S_t}{S_{t-1}}, t = 1, \ldots, T$, which are computed by multiplying the original log daily returns by 100. On the other hand, 250 log daily returns, magnified in the same way, follow Sample A construct the out-of-sample dataset—Sample B. The summary of Sample A and Sample B is reported in in Table 4.1.

For the comparison under the risk-neutral measure, as previously mentioned, we adopt the

[1]The risk-neutral GARCH(1, 1) model adopts σ_t directly while the ARSV(1) model use it to determine the initial x_t by solving $\beta \exp(x_t/2) = \sigma_t$ where β is from the present parameter estimate.

'one-day in-sample with second-day out-of-sample' rule. During the period of October 30th, 2017 to February 1st, 2018, all in-sample and out-of-sample pairs are selected from every two continuous days in each week; that is, we will have four such pairs every week. For example, every Monday will only be selected as the in-sample day and the following Tuesday is its out-of-sample day. That Tuesday itself also offers the second in-sample options in that week while its matched out-of-sample day is the Wednesday, and so on. We skip holidays in the weekdays and avoid selecting Friday as in-sample day so that there is no weekend gap between the in-sample and out-of-sample days. At each day, around thirty-five options with highest volumes whose ratio of index to strike is also located in $[0.9, 1.1]$ will be selected and the market price of each chosen option is set to be the mean value of the last bid and ask prices.

Table 4.1: Characteristics of the magnified in-sample (Sample A) and out-of-sample (Sample B) datasets under the physical measure.

Sample	Size	Mean	Max	Min	Std. Dev.	Skewness	Kurtosis
A	$T = 2518$	2.775×10^{-2}	5.574	-7.113	1.154	-9.077×10^{-2}	5.956
B	$T' = 250$	5.288×10^{-2}	2.134	-1.850	0.631	9.547×10^{-2}	4.157

4.4 Empirical Study

4.4.1 Results under the Physical Measure

4.4.1.1 In-Sample Comparison under the Physical Measure

The parameter estimation results of the GARCH(1, 1) and ARSV(1) models for Sample A under the physical measure are listed in Table 4.2 in which 'LL' is short for log-likelihood and the standard errors are reported in brackets. The parameter estimates and their standard errors of the ARSV(1) model are sample mean and sample standard deviations of the estimates of the last 250 iterations in its off-line EM method. It shows the ARSV(1) model has a larger

log-likelihood.

Table 4.2: Parameter estimates under the physical measure, Sample A.

Volatility Model	Parameter Estimates			LL
GARCH(1, 1)	$\hat{a}_0 = 1.26345 \times 10^{-2}$ (4.9×10^{-3})	$\hat{a}_1 = 7.76129 \times 10^{-2}$ (1.2×10^{-2})	$\hat{b}_1 = 9.15091 \times 10^{-1}$ (1.4×10^{-2})	-3682.529
ARSV(1)	$\hat{\phi} = 9.86795 \times 10^{-1}$ (1.7×10^{-4})	$\hat{\gamma}^2 = 1.50959 \times 10^{-2}$ (7.0×10^{-5})	$\hat{\beta}^2 = 1.02930$ (2.7×10^{-2})	-3656.791

When calculating the maximum-likelihood estimators of the GARCH(1, 1) model, the initial volatility for Day 1 (the first in-sample day), denoted by $\hat{\sigma}_1$, is set as the standard deviation of Sample A[2]. In the s-th EM off-line iteration for estimating the ARSV(1) model, we also set a prior normal distribution for particles of Day 1 such that the expected volatility on that day is same as $\hat{\sigma}_1$. Hence, the prior distribution can be derived as follows.

Suppose $x_1^i \sim N(\mu_s, \sigma_s^2)$ in the s-th iteration for $i = 1, \ldots, N$ where N is the number of particles. The parameter estimate $(\phi_s, \gamma_s^2, \beta_s^2)$ used in s-th iteration are updated from the maximum step of the $(s-1)$-th iteration. One popular choice for σ_s^2 is $\frac{\gamma_s^2}{1-\phi_s^2}$. Here we let $\sigma_s^2 = \max(\frac{\gamma_s^2}{1-\phi_s^2}, 1.35)$ to ensure the diversity of the particles. μ_s can be calculated by solving $\beta_s \mathbb{E}[\exp(\frac{x_1}{2})] = \hat{\sigma}_1$, as in Eq. (4.4.1):

$$\mathbb{E}\left[\exp(\frac{x_1}{2})\right] = \exp(\frac{1}{2}\mu_s + \frac{1}{8}\sigma_s^2) = \frac{\hat{\sigma}_1}{\beta_s}, \qquad \mu_s = 2\ln(\frac{\hat{\sigma}_1}{\beta_s}) - \frac{1}{4}\sigma_s^2. \tag{4.4.1}$$

The maximum-likelihood parameter estimates of the GARCH(1, 1) model under the physical measure, whose p-values are all less than 0.01, are significantly different from 0. Those estimates are obtained with the maximum log-likelihood -3682.529. Many empirical studies have a positive log-likelihood because their returns are not subject to the 100-times magnification. Revisiting the second equation in Eq. (4.2.2), when magnifying y_t and σ_t at the same time, the first term

[2]When implementing the MLE method for the GARCH(1, 1) model, the subsequent conditional variances are updated by the model while the initial volatility or conditional variance for the first day should be provided.

$-\frac{1}{2}\ln(2\pi)$ and the third term $-\frac{1}{2}\sum_{t=1}^{T}\frac{y_t^2}{\sigma_t^2}$ stay unchanged while the second term $-\frac{1}{2}\sum_{t=1}^{T}\ln(\sigma_t^2)$ is indeed affected. A larger σ_t^2 series may decrease the log-likelihood to a negative value.

Furthermore, as illuminated in section 4.2.2.3, the ARSV(1) model is estimated by the off-line EM method with a forward-only FFBS algorithm and the estimates of all iterations are shown in Figure 4.1. We run 2730 off-line iterations in total of which the last 200 iterations have 1250 particles and the remaining ones use 800 particles. For each parameter estimate, the arithmetic mean of the last 250 iterations is depicted by a blue dash line with the mean value marked on the right side while the red line sketches the estimate of each iteration. According to Figure 4.1, estimates of ϕ and γ^2 quickly reach their convergent values, and estimates of β^2 fluctuates in a considerably small range around the blue dashed line.

Once the parameter estimates of the ARSV(1) model are determined, its particle-based computations of both the log-likelihood and $\mathbb{E}[\hat{\beta}\exp(\frac{x_t}{2})|y_{1:T}]$ $(t \leq T)$ can be implemented together by particle filtering and smoothing using Bootstrap Filter algorithm with 10^5 particles as demonstrated in Appendix B.4.2. Its numerical log-likelihood estimation, as approximated by Eq. (B.4.10) in Section B.4.3, does not ask for such a large number of particles while the subsequent particle smoothing does since it suffers the degeneracy problem resulted from a large observation number ($T = 2518$).

Now let us investigate the distribution of error terms in the return dynamics. Given the maximum-likelihood parameter estimates and the initial volatility, the volatility at time step t $(1 \leq t \leq T)$ of the GARCH(1, 1) model is fully determined so its error sequence is thus $\left\{\frac{y_t}{\sigma_t}\right\}_{t=1}^{T}$. The ARSV(1) counterpart needs to be estimated through the particle smoothing as indicated in Appendix B.4.2 and subsequently we can obtain the error sequence $\left\{\frac{y_t}{\mathbb{E}[\hat{\beta}\exp(x_t/2)|y_{1:T}]}\right\}_{t=1}^{T}$, of which the denominator (in-sample volatility estimate) is computed by Eq. (B.4.4). Both models assume the errors in the return process follow the standard normal distribution.

The Q-Q plots versus the assumed standard normal distribution for the error sequences of the two models are shown in Figure 4.2 and the corresponding normality test results are reported in Table 4.3 in which the preferred value in each row is underlined. The Q-Q plots

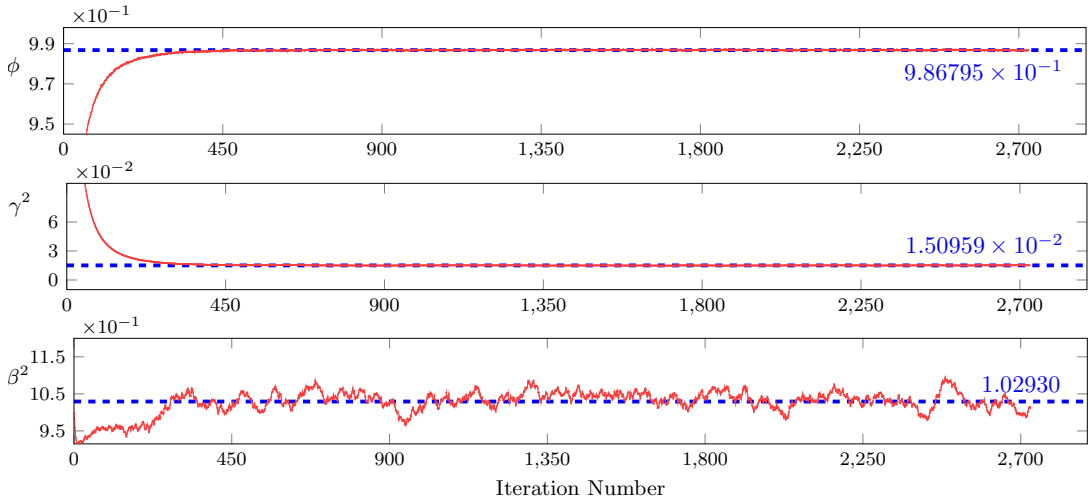

Figure 4.1: Particle-based parameter estimation for the ARSV(1) model under the physical measure, Sample A.

indicate the error sequences of both models have slightly lighter right tails than the standard normal distribution. As for the left tail, the error sequence of the GARCH(1, 1) model is much heavier than the standard normal distribution while the ARSV(1) counterpart is identical to the standard normal distribution. In addition, considering the result of each normality test whose null hypothesis assumes the tested sample follows the standard normal distribution, the ARSV(1) always has a larger p-value. Therefore, when it comes to fitting historical return series, the normality assumption for the error sequence in the ARSV(1) model is more adequate than the one in the GARCH(1, 1) model. This conclusion is in accordance with the finding of M. A. Carnero, Peña, & Ruiz (2004).

Table 4.3: Normality tests results for error sequences, Sample A.

normality test	p-value, GARCH(1, 1)	p-value, ARSV(1)
Kolmogorov-Smirnov	$< 10^{-3}$	0.013
Lilliefors	$< 10^{-3}$	0.472
Anderson-Darling	0.006	0.015

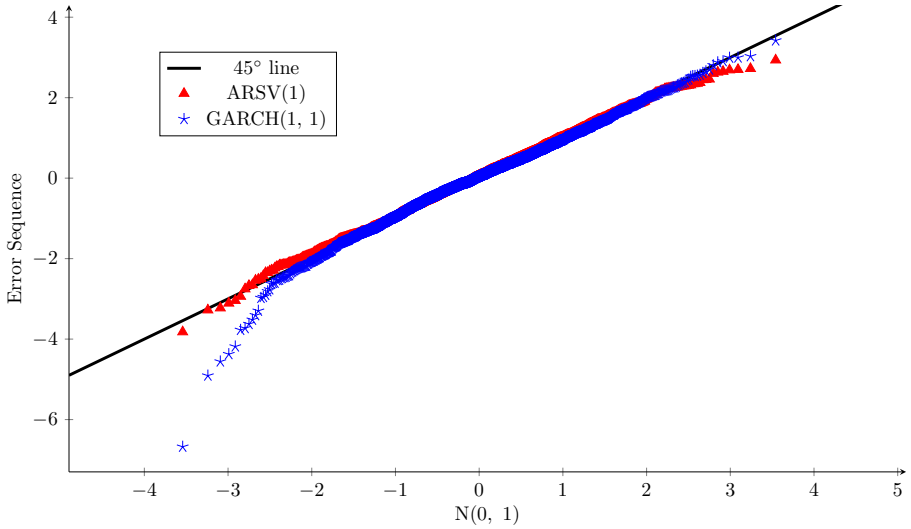

Figure 4.2: Q-Q Plots: error sequences of GARCH(1, 1) and ARSV(1) models vs. standard normal distribution, Sample A.

As a whole, the ARSV(1) model has a larger likelihood when fitting historical return series and is better to satisfy the assumption that the error sequence in the return process belongs to the standard normal distribution. Therefore, the ARSV(1) model outperforms the GARCH(1, 1) model in terms of the in-sample comparison under the physical measure.

4.4.1.2 Out-of-Sample Comparison under the Physical Measure

As previously mentioned, Sample B, the out-of-sample return dataset, follows Sample A, the in-sample dataset, without any gap. Therefore, the last in-sample volatility estimate can be directly used to generate the first out-of-sample volatility estimate. Such one-step-ahead prediction is natural for the GARCH(1, 1) model since actually, it is how that model works. Given the last in-sample conditional variance estimate and the out-of-sample magnified return series, then all the out-of-sample conditional variances are determined through the specification of conditional variance in the GARCH(1, 1) model step by step as follows:

$$h^G_{t+1|t} = \hat{a}_0 + \hat{a}_1 y_t^2 + \hat{b}_1 h^G_{t|t-1}, \quad t = T, T+1, \ldots, T+T'-1. \tag{4.4.2}$$

where the superscript G stands for the GARCH(1, 1) model, $(\hat{a}_0, \hat{a}_1, \hat{b}_1)$ denote the in-sample maximum-likelihood parameter estimates, $h^G_{T|T-1}$ is the last in-sample conditional variance estimate, T is the size of the in-sample dataset and T' is the size of the out-of-sample one.

In light of the ARSV(1) model, the out-of-sample volatility forecasts are still particle-based. Fortunately, as demonstrated in Appendix B.4.4, the one-step-ahead prediction can be connected seamlessly with the particle filtering; that is, the ARSV(1) out-of-sample conditional variance estimate is given, as in Eq. (4.4.3):

$$h^A_{t+1|t} = \mathbb{E}[\hat{\beta}^2 \exp(x_{t+1})|y_{1:t}], \quad t = T, T+1, \ldots, T+T'-1. \tag{4.4.3}$$

where the superscript A stands for the ARSV(1) model. Its parameter estimates are determined based on the in-sample dataset and stay unchanged for out-of-sample forecasts. Moreover, both the in-sample and out-of-sample volatility estimates are sketched in Figure 4.3, which shows the two models' volatility estimates have similar patterns but the magnitudes are different.

The out-of-sample volatility forecast results of the two models are summarized in Table 4.4 in which the preferred value in each row is underlined. As it shows, the ARSV(1) model has smaller values for the MSE and QLIKE loss functions so it is also superior to the GARCH(1, 1) model in terms of the out-of-sample volatility forecast under the physical measure.

Table 4.4: Out-of-sample volatility forecast results of the GARCH(1, 1) and ARSV(1) models, Sample B

loss function	GARCH(1, 1)	ARSV(1)
MSE	5.0372×10^{-1}	$\underline{5.0367 \times 10^{-1}}$
QLIKE	6.8369×10^{-2}	$\underline{6.7339 \times 10^{-2}}$

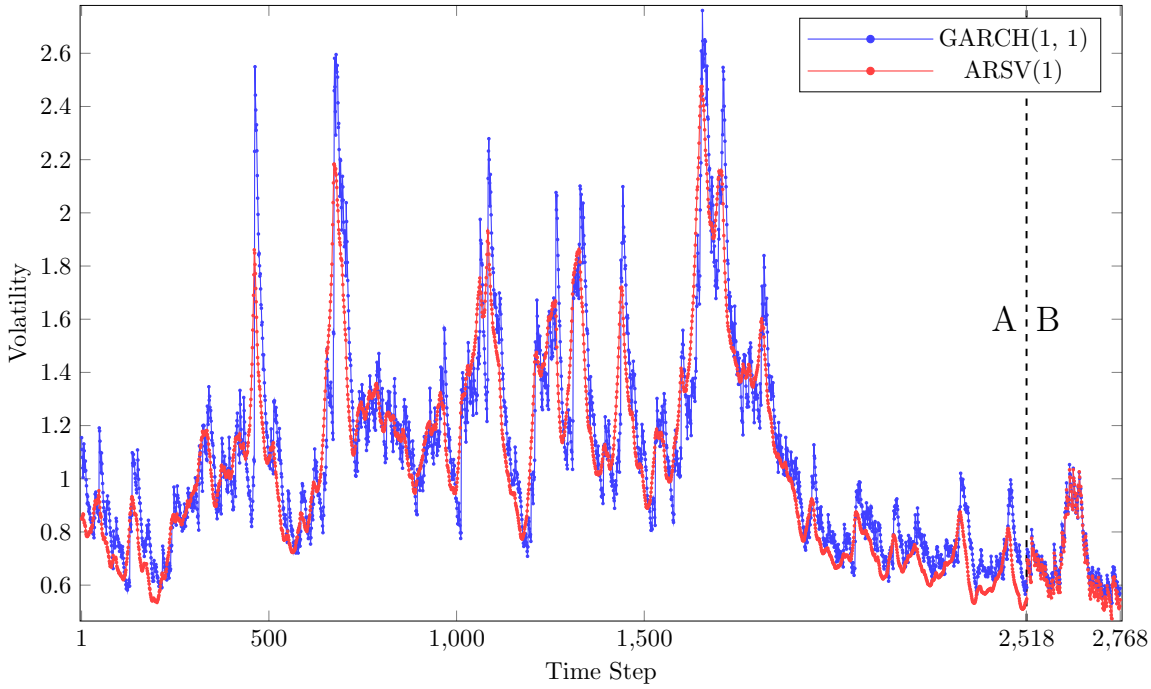

Figure 4.3: In-sample (A) and out-of-sample (B) volatility estimates.

4.4.2 Results under the Risk-Neutral Measure

The S&P 500 Index European call and put options are investigated separately. For each kind of option, we split them into two groups: one expires in about 30 calendar days and the other expires in about 50 calendar days.

In addition to the GARCH(1, 1) and ARSV(1) models, we also investigate the traditional B-S model and the B-S model with the implied volatility (BS-IV). Both of the traditional BS model and the BS-IV model consider a constant volatility across the time to maturity when pricing options. The initial volatility in the traditional B-S model is set to be same as the initial one of the GARCH(1, 1) and the ARSV(1) models which is the unconditional sample standard deviation of the 180-day log return series before the in-sample day. On the other hand, the BS-IV model parameterize its initial volatility when minimizing the in-sample MSPE. Assuming

there are n call options in the in-sample day t, then the daily implied volatility is calculated by

$$\tilde{\sigma}_{\text{iv}} = \arg\min_{\sigma} \frac{1}{n} \sum_{i=1}^{n} [v_i - BS_{\text{call}}(T_i - t, S_t, K_i, r_1, \sigma)]^2. \tag{4.4.4}$$

where S_t is the close price of day t, r_1 is the constant daily risk-free rate ($r_1 = 2.5\%/252$), $T_i - t$, K_i and v_i are the time to maturity[3] (in trading days), strike price and market price of the i-th option, respectively. The risk-neutral GARCH(1, 1) and ARSV(1) models for the n in-sample call options at t are estimated as follows:

$$\text{GARCH}(1,\,1) : \tilde{\theta}^G = \arg\min_{\theta^G} \frac{1}{n} \sum_{i=1}^{n} [v_i - \hat{C}_t^G(\theta^G, \sigma_t, i)]^2, \qquad \theta^G = (a_0, a_1, b_1), \tag{4.4.5}$$

$$\text{ARSV}(1) : \tilde{\theta}^A = \arg\min_{\theta^A} \frac{1}{n} \sum_{i=1}^{n} [v_i - \hat{C}_t^A(\theta^A, \sigma_t, i)]^2, \qquad \theta^A = (\phi, \gamma, \beta). \tag{4.4.6}$$

where theoretical option prices $\hat{C}_t^G(\theta^G, \sigma_t, i)$ and $\hat{C}_t^A(\theta^A, \sigma_t, i)$ are respectively calculated by Eq. (4.2.21) and Eq. (4.2.30) using the given initial volatility estimate—σ_t, along with the identifications (strike price K_i and time to maturity $T_i - t$) of the i-th option. All the three risk-neutral models (BS-IV, GARCH(1, 1) and ARSV(1)) for put options are estimated similarly.

It is worth noting that the risk-neutral parameter estimates stay unchanged for the out-of-sample option pricing. Moreover, the initial volatility estimate of the risk-neutral B-S, GARCH(1, 1) and ARSV(1) models comes from the historical daily returns (unconditional sample standard deviation of the 180-day log returns before the in-sample day) while it is considered to be the only parameter in the BS-IV model. Actually, the B-S model has no parameter and its in-sample and out-of-sample performances are fully determined by the initial volatility estimate and option selections. The average results are reported in Table 4.6 and Table 4.7 in which the preferred value in each column is underlined and the corresponding sample standard deviations are included in the brackets. In addition, Table 4.5 summarizes the average risk-neutral parameter estimates with the corresponding sample standard deviations included in the brackets.

[3]Though our in-sample or out-of-sample options are chosen in a single day and expire in about 30 or 50 days, they can have a little different times to maturity such as 31 days or 33 days.

Table 4.5: Average in-sample risk-neutral parameter estimates.

	GARCH(1, 1)			ARSV(1)			BS-IV
	$\tilde{a}_0$	$\tilde{a}_1$	$\tilde{b}_1$	$\tilde{\phi}$	$\tilde{\gamma}$	$\tilde{\beta}$	$\tilde{\sigma}$
call, 30 days	6.949×10^{-6}	0.144	0.571	0.442	0.525	6.787×10^{-3}	4.557×10^{-3}
	(8.10×10^{-6})	(0.20)	(0.35)	(0.28)	(0.52)	(1.55×10^{-2})	(7.87×10^{-4})
call, 50 days	4.477×10^{-6}	0.204	0.605	0.367	0.234	4.589×10^{-3}	4.817×10^{-3}
	(3.05×10^{-6})	(0.21)	(0.33)	(0.12)	(0.35)	(9.42×10^{-4})	(6.85×10^{-4})
put, 30 days	6.727×10^{-6}	0.939	0.061	0.886	1.228	1.038×10^{-3}	6.207×10^{-3}
	(2.4×10^{-6})	(0.08)	(0.08)	(0.08)	(0.40)	(8.36×10^{-4})	(8.36×10^{-4})
put, 50 days	6.134×10^{-6}	0.863	0.137	0.873	1.417	9.888×10^{-4}	6.924×10^{-3}
	(1.66×10^{-6})	(0.11)	(0.11)	(0.05)	(0.32)	(4.90×10^{-4})	(8.19×10^{-4})

According to Table 4.5, the risk-neutral parameters of the GARCH(1, 1) and ARSV(1) models are quite different from their physical counterparts[4]. For example, the put options have $\tilde{b}_1$ in the GARCH(1, 1) model less than 0.2 while its physical counterpart is close to one. By contrast, $\tilde{\gamma}$ in the ARSV(1) model is much larger than its physical counterpart. The risk-neutral estimates also depend on the kind of option. $\tilde{a}_1$ in the GARCH(1, 1) is such an example. Interestingly, it seems the put options have larger volatilities as the corresponding $\tilde{\sigma}_{\text{iv}}$ apparently grows. That's why pricing errors of the traditional B-S model are exacerbated for put options because the initial volatility is always around 4.2×10^{-3} with tiny fluctuations.

Table 4.6: Average in-sample and out-of-sample **call** option pricing errors with a given initial volatility.

	Call, 30 calendar days		Call, 50 calendar days	
MSPE	In-Sample	Out-of-Sample	In-Sample	Out-of-Sample
GARCH(1, 1)	2.1912	4.0073	3.4537	5.5005
	(2.14)	(3.54)	(2.59)	(3.85)
ARSV(1)	2.2057	4.0540	3.6125	5.6605
	(2.19)	(3.58)	(2.66)	(3.92)
BS-IV	2.2675	3.9935	3.6264	5.6303
	(2.23)	(3.59)	(2.68)	(3.92)
B-S	11.3419	11.2559	21.3919	20.7546
	(15.56)	(15.39)	(33.34)	(33.21)

[4]The return sequence is 100-times magnified. Without the magnification, the previous $\hat{a}_0$ in GARCH(1, 1) and $\hat{\beta}$ in ARSV(1) would divide by 100 while other parameter estimates are unchanged.

Table 4.7: Average in-sample and out-of-sample **put** option pricing errors with a given initial volatility.

MSPE	Put, 30 calendar days In-Sample	Put, 30 calendar days Out-of-Sample	Put, 50 calendar days In-Sample	Put, 50 calendar days Out-of-Sample
GARCH(1, 1)	2.9376 (2.19)	4.0737 (3.40)	5.6874 (2.84)	7.5993 (5.00)
ARSV(1)	0.6949 (1.06)	2.0679 (2.21)	1.1797 (1.47)	2.7540 (2.86)
BS-IV	9.8680 (3.55)	10.8460 (4.88)	21.3281 (6.32)	23.7384 (8.46)
B-S	46.4342 (41.65)	44.1500 (39.52)	143.4384 (91.28)	137.2699 (91.59)

Furthermore, though estimating the ARSV(1) model is much more complicated than the GARCH(1, 1) model under the physical measure, estimating their risk-neutral versions asks for similar computation burden. The reason is that as previously demonstrated, only one process needs to be simulated in the risk-neutral ARSV(1) model.

4.4.2.1 In-Sample Comparison under the Risk-Neutral Measure

As for call options, the GARCH(1, 1) model is slightly superior to the three rivals in terms of the average in-sample MSPE and the standard deviation regardless of the time to maturity. Therefore, the GARCH(1, 1) model fits the observed call option prices better with less dispersion. However, its superiority over the ARSV(1) and BS-IV models, however, are not obvious.

The in-sample performance of the put options is another thing. The ARSV(1) model remarkably dominates the others for both the 30-day and 50-day put options. In addition, the GARCH(1, 1) model also substantially outperforms the BS-IV model.

Moreover, options with longer time to maturity tend to have a larger average in-sample MSPE. Not surprisingly, the traditional B-S model is always inferior to the GARCH(1, 1) and BS-IV model regarding the in-sample pricing error. The reason is that with the same initial volatility estimate, the B-S model refers to a special case of the GARCH(1, 1) model ($a_0 = \hat{\sigma}_1^2$, $a_1 = b_1 = 0$). On the other hand, the BS-IV model, whose initial volatility is parameterized, is

an optimal version of the B-S model with respect to the in-sample MSPE.

4.4.2.2 Out-of-Sample Comparison under the Risk-Neutral Measure

As expected, for all models, the out-of-sample average MSPE is larger than the in-sample counterpart. In addition, the longer the time to maturity is, the harder it is to predict the out-of-sample option prices. In light of call options, the BS-IV model performs better than others for 30-day options while the GARCH(1, 1) model is preferred for 50-day options. Like the in-sample call options, the out-of-sample call pricing errors are very similar among the models except the traditional B-S model.

The obvious superiority of the ARSV(1) model over others within in-sample put options is kept for the out-of-sample pricing performances. As a whole, the ARSV(1) model is indeed preferable when pricing put options. By contrast, put options are less suitable for the GARCH(1, 1) model than call options. This finding is similar to the result of Heston & Nandi (1997).

One thing we need to pay attention to is the initial volatility estimate of each in-sample trading day. Admittedly, we set the initial volatility casually without examining other strategies. Undoubtedly, some refined initial values will improve the GARCH(1, 1) and ARSV(1) models in terms of the option pricing performance. For example, the initial volatility for a risk-neutral model can be estimated by the physical version of itself. On the other hand, the BS-IV model adopts an implied volatility but it is still inferior to the two models with variant volatilities in most of the examined scenarios. Therefore, dynamic volatility models such as the ARSV(1) and GARCH(1, 1) models are more accurate for option pricing than a constant volatility.

Instead of using historical returns, we can also derive the initial volatility directly from preceding option prices. That adjustment brings about the implied versions of the GARCH(1, 1) and ARSV(1) models whose pricing performances are explored in the following subsection.

4.4.3 Risk-Neutral GARCH(1, 1) and ARSV(1) Models using Implied Volatilities

Rather than set a casual value, we can also parameterize the initial volatility estimate in the risk-neutral GARCH(1, 1) and ARSV(1) models. For call options, the implied versions of the two models are estimated by minimizing the in-sample MSPE of the n options as follows:

$$\text{GARCH-IV}(1,\,1): \tilde{\theta}^G_{\text{iv}} = \arg\min_{\theta^G_{\text{iv}}} \frac{1}{n}\sum_{i=1}^{n}[v_i - \hat{C}^G_t(\theta^G_{\text{iv}}, i)]^2, \qquad \theta^G_{\text{iv}} = (a_0, a_1, b_1, \sigma_t), \quad (4.4.7)$$

$$\text{ARSV-IV}(1): \tilde{\theta}^A_{\text{iv}} = \arg\min_{\theta^A_{\text{iv}}} \frac{1}{n}\sum_{i=1}^{n}[v_i - \hat{C}^A_t(\theta^A_{\text{iv}}, i)]^2, \qquad \theta^A_{\text{iv}} = (\phi, \gamma, \beta, \sigma_t). \quad (4.4.8)$$

where the suffix '-IV' or subscript 'iv' stands for the implied version that parameterizes the initial volatility estimate σ_t. In addition, $\hat{C}^G_t(\theta^G_{\text{iv}}, i)$ and $\hat{C}^A_t(\theta^A_{\text{iv}}, i)$, which denote the theoretical prices of the i-th call option with the risk-neutral GARCH(1, 1) and ARSV(1) models, are calculated by Eq. (4.2.21) and Eq. (4.2.30), respectively. Here σ_t is an extra parameter. The counterparts for put options can be estimated in a similar way. The in-sample and out-of-sample average MSPE of the implied versions of the two risk-neutral models are presented in Table 4.8, in which the standard deviations are included in brackets.

Table 4.8: Average in-sample and out-of-sample option pricing errors of the implied GARCH(1, 1) and ARSV(1) models.

	GARCH-IV(1, 1)		ARSV-IV(1)	
MSPE	In-Sample	Out-of-Sample	In-Sample	Out-of-Sample
call, 30 days	1.9666 (2.03)	12.3583 (11.47)	2.2000 (2.20)	4.5315 (3.87)
call, 50 days	3.3549 (2.45)	5.7784 (4.62)	3.4535 (2.50)	6.4111 (5.16)
put, 30 days	0.6459 (1.00)	7.3234 (5.97)	0.6787 (1.05)	2.2551 (2.56)
put, 50 days	1.1828 (1.35)	9.5250 (6.28)	1.1271 (1.48)	2.9827 (3.42)

It can be seen from Table 4.8 that parameterizing the initial volatility leads to a smaller average in-sample MSPE for both the risk-neutral GARCH(1, 1) and ARSV(1) models. Most corresponding standard deviations are also decreased. Such improvements are within our expectation since the original non-implied versions are just special cases of the implied versions when minimizing the in-sample MSPE.

However, an extra volatility parameter also brings about more uncertainties in the out-of-sample results. In terms of the average values and standard deviations of the out-of-sample pricing errors, the implied versions of both models are inferior to their original non-implied counterparts that casually choose the predetermined initial volatility estimates from historical returns. Therefore, a more sophisticated model does not necessarily imply a more robust model for the out-of-sample prediction.

One reason is that the in-sample nonlinear-least-squares parameter estimators are to some extent sensitive to the input option identifications such as the spot price and time to maturity while parameterizing the volatility would further add to such sensitivity. Moreover, it is likely that the implied models will have an abnormal initial volatility estimate; that is, when minimizing the in-sample MSPE, we may obtain an extremely large initial volatility estimate in the implied models. Then the subsequent out-of-sample prediction error tends to become out of control. In theory, volatility is a positive real number without an upper bound. To avoid the abnormal cases, within our implementation, the daily initial volatility is constrained to be less than 0.023 (the yearly equivalent upper bound is $\sqrt{252} \times 0.023 = 36.5\%$). When the in-sample initial volatility estimate reaches that upper bound, there is a high possibility that their out-of-sample prediction error will be substantially large. Allowing for a smaller upper bound, the out-of-sample results may be better but it will make the extra volatility parameter less meaningful. In addition, the upper bound should be connected with the market situation. For example, in a bear market that upper bound can be relatively larger. How to set a reasonable upper bound for the initial volatility parameter is worth investigating in the future. As a whole, for our option samples, the implied versions of the GARCH(1, 1) and ARSV(1) models are not recommended. After all, the

out-of-sample prediction error matters more than the in-sample counterpart.

4.5 Conclusion and Future Work

This chapter conducts a comprehensive comparison of the GARCH(1, 1) and ARSV(1) models under both the physical and risk-neutral measures. Under the physical measure, we investigate their log-likelihoods after fitting historical returns and test the normality assumption for their error terms in the return process. Moreover, two robust loss functions, MSE and QLIKE, are adopted for the one-step-ahead volatility forecast comparison. The results indicate the ARSV(1) model outperforms the GARCH(1, 1) model in terms of the in-sample fitting and out-of-sample prediction performances under the physical measure.

On the other hand, under the risk-neutral measure, the in-sample and out-of-sample option pricing errors of the two models are explored. We show that only the volatility process of the ARSV(1) model needs to be simulated for option pricing. In addition, the original and implied versions of the two risk-neutral models are taken into consideration. The traditional and implied B-S models are also examined as benchmarks. We find the performances of the two models are considerably similar when pricing call options while the ARSV(1) model is remarkably superior to the GARCH(1, 1) model for put options. However, their implied versions are not robust for the out-of-sample prediction.

For the original non-implied version of the risk-neutral GARCH(1, 1) and ARSV(1) model, we adopt a causal initial volatility when investigating their in-sample and out-of-sample pricing errors. It is much likely that some other selections will lead to a better performance. For example, the initial volatility for a risk-neutral model can be estimated by the physical version of itself. Moreover, when parameterizing the initial volatility in the implied version, how to set a reasonable upper bound for that parameter is also worth studying in the future.

CHAPTER 5

A Comparison of Option Pricing Models with Leverage Effect

5.1 Introduction

Stylized facts, mostly used in economics, refer to common statistical properties in the financial series that are widely found in empirical studies. Only when a volatility model is capable of capturing some stylized facts can it precisely estimate or forecast volatilities. A lot of dynamic volatility models have been proposed to exhibit certain stylized facts. In this chapter, we will concentrate our attention on two widely-used volatility models—the GARCH(1, 1) and ARSV(1) models, which have been studied in Chapter 4. Both of the models can explain two prevalent stylized facts: excess kurtosis and volatility clustering (M. A. Carnero, Peña, & Ruiz, 2004). They, however, fail to directly capture another stylized fact—leverage effect, which refers to a negative correlation between the return and the subsequent volatility. Black (1976) and Christie (1982) attempt to explain this common characteristic by the financial leverage (equivalently, the ratio of debt to equity). Figlewski & Wang (2000), by contrast, conclude the leverage effect is more related to the falling stock prices than the financial leverage. Bouchaud, Matacz, & Potters

(2001) also rationalize the leverage effect using a retarded volatility model.

To capture the leverage effect, the ARSV(1) model just needs to introduce a negative correlation between the return and volatility innovation processes but the GARCH(1, 1) model has to replace its volatility specification with other forms such as the Exponential GARCH (EGARCH) (Nelson, 1991) and Nonlinear Asymmetric GARCH (NGARCH) (Engle & Ng, 1993) models. M. A. Carnero, Peña, & Ruiz (2004) show that introducing the leverage effect will not change the conclusions about the persistence and kurtosis captured in the GARCH(1, 1) and ARSV(1) models. Until now, there have been no empirical studies on the comparison of these two volatility models under the risk-neutral measure after incorporating the leverage effect.

The objective of this chapter is to compare the capacities of the GARCH(1, 1) and ARSV(1) models with the leverage effect to fit the observed option prices and to forecast option prices in the future. In addition, we propose a continuous-time option pricing model that captures the leverage effect by straightforwardly relating the volatility to the exponential decayed weighted average (EDWA) cumulative asset return. We show that incorporating the leverage effect brings about substantial improvements to the GARCH(1, 1) and ARSV(1) models in terms of the in-sample and out-of-sample option pricing performances. Moreover, our Cumulative Return model with only two parameters dominates other sophisticated models for predicting call options and has as accurate results as others when it comes to pricing put options. The performance of our model in an outlier day also further highlights its robustness.

The rest of this chapter is composed as follows: Section 5.2 discusses the GARCH(1, 1) and ARSV(1) models with the leverage effect under the risk-neutral measure. In Section 5.3, we put forward the new continuous-time Cumulative Return option pricing model. Section 5.4 investigates the empirical results and Section 5.5 summarizes this chapter and provides suggestions for the future work.

5.2 GARCH(1, 1) and ARSV(1) models with Leverage Effect

5.2.1 Risk-Neutral ARSV(1) Model with Leverage Effect

The standard ARSV(1) option pricing model, as shown by Eq. (4.2.18), has independent observation and state processes with uncorrelated $\{\xi_t\}$ and $\{\eta_t\}$ sequences. Taking the leverage effect into consideration, the ARSV(1) option pricing model has the following form:

$$\begin{aligned} \ln \frac{S_t}{S_{t-1}} &= r_1 - \frac{1}{2}\tilde{\beta}^2 \exp(\tilde{x}_t) + \tilde{\beta} \exp(\frac{\tilde{x}_t}{2})\xi_t, \\ \tilde{x}_{t+1} &= \tilde{\phi}\tilde{x}_t + \tilde{\gamma}\eta_{t+1}. \end{aligned} \tag{5.2.1}$$

where $\begin{pmatrix} \xi_t \\ \eta_{t+1} \end{pmatrix} \sim N(0, \Sigma)$, $\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ and the parameter ρ measures the leverage effect.

Hereafter, the ARSV(1) option pricing model incorporating the leverage effect is denoted by $\rho-$ARSV(1). Instead of generating two random variables from a multivariate normal distribution, we determine η_{t+1} at first and subsequently calculate ξ_t; that is, we have

$$\xi_t = \rho\eta_{t+1} + \sqrt{1-\rho^2}\zeta_t, \quad \eta_{t+1} \sim N(0,1), \quad \zeta_t \sim N(0,1), \quad \eta_{t+1} \perp \zeta_t. \tag{5.2.2}$$

This substitution avoids the correlated innovation processes which are troublesome in price simulation. Plugging Eq. (5.2.2) into Eq. (5.2.1) implies the equivalent form of the $\rho-$ARSV(1) model as follows:

$$\begin{aligned} \ln \frac{S_t}{S_{t-1}} &= r_1 - \frac{1}{2}\tilde{\beta}^2 \exp(\tilde{x}_t) + \rho\tilde{\beta} \exp(\frac{\tilde{x}_t}{2})\eta_{t+1} + \sqrt{1-\rho^2}\tilde{\beta} \exp(\frac{\tilde{x}_t}{2})\zeta_t, \\ \tilde{x}_{t+1} &= \tilde{\phi}\tilde{x}_t + \tilde{\gamma}\eta_{t+1}, \qquad \eta_{t+1} \overset{\text{i.i.d}}{\sim} N(0,1), \quad \zeta_t \overset{\text{i.i.d}}{\sim} N(0,1), \quad \eta_{t+1} \perp \zeta_t. \end{aligned} \tag{5.2.3}$$

At first sight, the model above is more complicated than its counterpart without the leverage effect. Fortunately, like the ARSV(1) model, only the volatility process in the $\rho-$ARSV(1) model needs to be simulated when pricing options. Starting at time t, suppose the sequence $\{\tilde{x}_k\}_{k=t+1}^{T}$ has been simulated using a corresponding random number sequence $\{\eta_k\}_{k=t+1}^{T+1}$ and

an initial value $\tilde{x}_t$. Given $\tilde{x}_t$ and $\{\eta_k\}_{k=t+1}^{T+1}$ from one simulation path, the conditional asset price at maturity T, denoted by S_T^{AL}, can be derived as follows[1]:

$$\begin{aligned}\ln S_T^{AL}|\tilde{x}_t, \{\eta_k\}_{k=t+1}^{T+1} \overset{D}{=} \ln S_t + r_1(T-t) - \frac{1}{2}\tilde{\beta}^2 \sum_{k=t+1}^{T} \exp(\tilde{x}_k) \\ + \rho\tilde{\beta} \sum_{k=t+1}^{T} \exp(\frac{\tilde{x}_k}{2})\eta_{k+1} + \sqrt{1-\rho^2}\tilde{\beta} \sum_{k=t+1}^{T} \exp(\frac{\tilde{x}_k}{2})\zeta_k,\end{aligned} \tag{5.2.4}$$

where r_1 still denotes the risk-free rate and the superscript AL stands for the ARSV(1) model with the leverage effect. It is worth noting that η_{T+1} is needed in Eq. (5.2.4) even though it is beyond the maturity date.

Moreover, since the independent and identically distributed standard normal error sequence $\{\zeta_k\}_{k=t+1}^{T}$ in Eq. (5.2.3) is independent of $\{\eta_k\}_{k=t+1}^{T+1}$, we have $\tilde{\beta}\sum_{k=t+1}^{T}\exp(\frac{\tilde{x}_k}{2})\zeta_k \sim N\big(0, \tilde{\beta}^2\sum_{k=t+1}^{T}\exp(\tilde{x}_k)\big)$. Hence, we can expand Eq. (5.2.4) through

$$\begin{aligned}\ln S_T^{AL}|\tilde{x}_t, \{\eta_k\}_{k=t+1}^{T+1} &\overset{D}{=} \ln S_t + r_1(T-t) - \frac{1}{2}\tilde{\beta}^2 \sum_{k=t+1}^{T} \exp(\tilde{x}_k) \\ &\quad + \rho\tilde{\beta} \sum_{k=t+1}^{T} \exp(\frac{\tilde{x}_k}{2})\eta_{k+1} + \sqrt{1-\rho^2}\Big(\sqrt{\tilde{\beta}^2 \sum_{k=t+1}^{T} \exp(\tilde{x}_k)}\Big)\ Z \\ &\overset{D}{=} \ln S_t + r_1(T-t) - \frac{1}{2}(1-\rho^2+\rho^2)\hat{\sigma}^2(T-t) \\ &\quad + \rho\tilde{\beta} \sum_{k=t+1}^{T} \exp(\frac{\tilde{x}_k}{2})\eta_{k+1} + \sqrt{1-\rho^2}\hat{\sigma}\sqrt{T-t}\ Z, \quad Z \sim N(0,1),\end{aligned}$$

where $\hat{\sigma}^2(T-t) = \sum_{k=t+1}^{T}\tilde{\beta}^2\exp(\tilde{x}_k)$ as defined in Eq. (4.2.24). Furthermore, we can write the conditional log asset price at T into the following form:

$$\begin{aligned}\ln S_T^{AL}|\tilde{x}_t, \{\eta_k\}_{k=t+1}^{T+1} \overset{D}{=} \ln S_t + r_1(T-t) - \frac{1}{2}\rho^2\hat{\sigma}^2(T-t) + \rho\tilde{\beta} \sum_{k=t+1}^{T} \exp(\frac{\tilde{x}_k}{2})\eta_{k+1} \\ - \frac{1}{2}(1-\rho^2)\hat{\sigma}^2(T-t) + \sqrt{1-\rho^2}\hat{\sigma}\sqrt{T-t}\ Z\end{aligned}$$

[1]According to the second equation in Eq. (5.2.3) that describes the dynamics of $\tilde{x}_t$, $\{\tilde{x}_k\}_{k=t+1}^{T}$ is determined given the initial $\tilde{x}_t$ and corresponding error term sequence $\{\eta_k\}_{k=t+1}^{T}$.

$$\stackrel{D}{=} \ln S_t + \big(r_1 - \frac{1}{2}\rho^2\hat{\sigma}^2 + \frac{\rho\tilde{\beta}\sum_{k=t+1}^{T}\exp(\frac{\tilde{x}_k}{2})\eta_{k+1}}{T-t}\big)(T-t)$$
$$- \frac{1}{2}(1-\rho^2)\hat{\sigma}^2(T-t) + \sqrt{1-\rho^2}\hat{\sigma}\sqrt{T-t}\ Z. \tag{5.2.5}$$

Let $r^L \equiv r_1 - \frac{1}{2}\rho^2\hat{\sigma}^2 + \frac{\rho\tilde{\beta}\sum_{k=t+1}^{T}\exp(\frac{\tilde{x}_k}{2})\eta_{k+1}}{T-t}$ and $\hat{\sigma}^L \equiv \sqrt{1-\rho^2}\hat{\sigma}$. Based on Eq. (5.2.5), we find the dynamics of $S_T^{AL}|\tilde{x}_t, \{\eta_k\}_{k=t+1}^{T+1}$ thus belongs to a geometric Brownian motion with a constant daily volatility $\hat{\sigma}^L$ and a constant daily risk-free rate r^L. With the assistance of the Black-Scholes option pricing formula, the conditional European call/put option prices can be derived by

$$\mathbb{E}^Q[C_t^{AL}|\tilde{x}_t, \{\eta_k\}_{k=t+1}^{T+1}] = BS_{\text{call}}(T-t, S_t, K, r^L, \hat{\sigma}^L), \tag{5.2.6}$$
$$\mathbb{E}^Q[P_t^{AL}|\tilde{x}_t, \{\eta_k\}_{k=t+1}^{T+1}] = BS_{\text{put}}(T-t, S_t, K, r^L, \hat{\sigma}^L). \tag{5.2.7}$$

where C_t^{AL} and P_t^{AL} respectively denote the European call and put option prices at time t (expire at time T) derived by the $\rho-$ARSV(1) model. The proof is similar to the one in Section 4.2.3. Then the corresponding unconditional European call/put option prices are simulated by

$$\hat{C}_t^{AL}(\theta^L) = \mathbb{E}^Q\big[\mathbb{E}^Q[C_t^L|\tilde{x}_t, \{\eta_k\}_{k=t+1}^{T+1}]\big] \approx \frac{1}{m}\sum_{j=1}^{m} BS_{\text{call}}(T-t, S_t, K, r^L(j), \hat{\sigma}^L(j)), \tag{5.2.8}$$
$$\hat{P}_t^{AL}(\theta^L) = \mathbb{E}^Q\big[\mathbb{E}^Q[P_t^L|\tilde{x}_t, \{\eta_k\}_{k=t+1}^{T+1}\big] \approx \frac{1}{m}\sum_{j=1}^{m} BS_{\text{put}}(T-t, S_t, K, r^L(j), \hat{\sigma}^L(j)). \tag{5.2.9}$$

where j in the parentheses stands for the j-th simulated path under the risk-neutral $\rho-$ARSV(1) model using the parameter vector $\theta^L = (\rho, \phi, \gamma, \beta)$. Similar to its counterpart without leverage effect, when pricing European options, only the volatility dynamics needs to be simulated though there are two innovation processes in that model. In this chapter, the simulation number m is still set to be 10^4.

5.2.2 Risk-Neutral GARCH(1, 1) Model with Leverage Effect

When incorporating the leverage effect, the GARCH(1, 1) model is not as natural as the ARSV(1) model because the variance process in the framework of the GARCH model is deterministic. Rather than easily adjust the correlation between two innovation processes, the GARCH(1, 1) model has to modify the specification of its conditional variance to capture the leverage effect.

5.2.2.1 Exponential GARCH(1, 1) Model

One commonly-used candidate is the EGARCH(1, 1) model (Nelson, 1991), which puts the log variance into a linear regression equation such that

$$\ln(\tilde{\sigma}_t^2) = \tilde{a}_0 + \tilde{a}_1 z_{t-1} + \tilde{\gamma}_1[|z_{t-1}| - \mathbb{E}(|z_{t-1}|)] + \tilde{b}_1 \ln(\sigma_{t-1}^2). \tag{5.2.10}$$

where $\mathbb{E}(|z_{t-1}|) = \sqrt{\frac{2}{\pi}}$ for any standard normal error term z_{t-1}. The logarithmic transformation of variance removes the positivity restrictions on parameters in Eq. (5.2.10) and $|\tilde{b}_1| < 1$ guarantees the stationarity. Parameters $\tilde{a}_1$ and $\tilde{\gamma}_1$ allow for asymmetry and leverage effect. The responses of log variance are $(\tilde{a}_1 + \tilde{\gamma}_1)z_{t-1}$ and $(\tilde{a}_1 - \tilde{\gamma}_1)z_{t-1}$ for positive and negative z_{t-1}, respectively. If $\tilde{\gamma}_1$ is zero, then negative and positive return have same effect on volatility. That's why the asymmetry is counted unless $\tilde{\gamma}_1$ is negative ($\tilde{a}_1 - \tilde{\gamma}_1 > a_1 + \tilde{\gamma}_1$). Moreover, the leverage effect, which refers to the negative correlation between the return and the subsequent volatility, asks for two extra constraints: $\tilde{a}_1 + \tilde{\gamma}_1 < 0$ and $\tilde{a}_1 - \gamma_1 > 0$ (equivalently, $\tilde{\gamma}_1 < \tilde{a}_1 < -\tilde{\gamma}_1$).

The second popular asymmetric model within the ARCH/GARCH family is the GJR-GARCH(1, 1) model (Glosten, Jagannathan, & Runkle, 1993), which adds an indicator function into the variance process to judge whether the previous return is positive. It has the following variance process:

$$\tilde{\sigma}_t^2 = \tilde{a}_0 + \big(\tilde{a}_1 + \tilde{\gamma}_1 \mathbb{1}_{\{z_{t-1}<0\}}\big)\tilde{\sigma}_{t-1}^2 z_{t-1}^2 + \tilde{b}_1 \tilde{\sigma}_{t-1}^2. \tag{5.2.11}$$

To ensure the conditional variance is positive, $a_1 \geq 0$ has to be satisfied. However, with a

positive a_1, a larger positive return cannot subsequently lead to a smaller volatility. Therefore, the GJR-GARCH(1, 1) model is unable to capture the leverage effect.

5.2.2.2 Nonlinear Asymmetric GARCH(1, 1) Model

Another GARCH model with leverage effect that has been adopted by many researchers is the NGARCH(1, 1) model. This model is proposed by Engle & Ng (1993), as in Eq. (5.2.12):

$$\tilde{\sigma}_t^2 = \tilde{a}_0 + \tilde{a}_1(z_{t-1} - \tilde{\gamma}_1)^2\tilde{\sigma}_{t-1}^2 + \tilde{b}_1\tilde{\sigma}_{t-1}^2. \tag{5.2.12}$$

where $\tilde{a}_0 > 0$, $\tilde{a}_1 \geq 0$, $\tilde{b}_1 \geq 0$. To derive its covariance stationarity constraint, we need to have a look at the corresponding persistence that equals to $\tilde{b}_1+\tilde{a}_1\mathbb{E}(z_{t-1}-\tilde{\gamma}_1)^2$. As $z_{t-1}-\tilde{\gamma}_1 \sim N(-\tilde{\gamma}_1, 1)$, we have $\mathbb{E}(z_{t-1} - \tilde{\gamma}_1)^2 = 1 + \tilde{\gamma}_1^2$. Therefore, $\tilde{b}_1 + \tilde{a}_1(1 + \tilde{\gamma}_1^2) < 1$ ensures the stationarity of the NGARCH(1, 1) model.

When it comes to pricing the European call/put option price by the GARCH(1, 1) model with the leverage effect, we can refer to Eq. (4.2.20) and Eq. (4.2.21) in which the volatility at each step is updated by Eq. (5.2.10) or Eq. (5.2.12) for the EGARCH(1, 1) or NGARCH(1, 1) model, respectively. Hereafter, the European call/put options prices simulated by the EGARCH(1, 1) model are denoted by $\hat{C}^E(\theta^E)$ and $\hat{P}^E(\theta^E)$ where $\theta^E = (a_0^E, a_1^E, \gamma_1^E, b_1^E)$ stands for the EGARCH(1, 1) parameter vector. Similarly, the European call/put option prices simulated by the NGARCH(1, 1) model are denoted by $\hat{C}^N(\theta^N)$ and $\hat{P}^N(\theta^N)$ where $\theta^N = (a_0^N, a_1^N, \gamma_1^N, b_1^N)$ stands for its parameter vector.

5.3 A New Option Pricing Model with the Leverage Effect

Enlightened by the constant elasticity of variance (CEV) model put forward by Cox (1975) which directly relates the current asset price with the subsequent volatility, we propose a Cumulative Return option pricing model. Given a invariant base time point t_0 which is before the option

trading day, our model satisfies the following stochastic differential equations:

$$dR_t = \big(r_1 - \frac{1}{2}f^2(Y_t)\sigma^2\big)dt + f(Y_t)\sigma dW_t, \tag{5.3.1}$$

$$Y_t = \frac{1}{M}\int_{-h}^{0} e^{\lambda s} R_{t+s} ds, \quad M = \int_{-h}^{0} e^{\lambda s} ds = \frac{1 - e^{-\lambda h}}{\lambda}. \tag{5.3.2}$$

where r_1 is the constant daily risk-free interest rate, σ is the constant daily volatility factor, $R_t = \ln S_t - \ln S_{t_0}$ denotes the cumulative return from time t_0 till current time t, Y_t is the EDWA cumulative return of previous h days with a decay rate λ, W_t is a Brownian motion and the form of $f(Y_t)$ will be given later. It is worth noting that σ alone in Eq. (5.3.1) does not stand for a daily volatility estimate. It is just a constant factor of volatility and $f(Y_t)\sigma$ together can be seen as the volatility estimate at time $t + \Delta t$. We first transform Y_t linearly into the following form:

$$\frac{1}{M}\int_{-h}^{0} e^{\lambda s} R_{t+s} ds \overset{\mu=t+s}{=} \frac{1}{M}\int_{t-h}^{t} e^{\lambda(\mu - t)} R_\mu d\mu, \tag{5.3.3}$$

Then by Leibniz integral rule, the differential of Y_t is given by

$$\begin{aligned}
\frac{dY_t}{dt} &= \frac{1}{M}\Big[\frac{d\int_{t-h}^{t} e^{\lambda(\mu-t)} R_\mu d\mu}{dt}\Big] = \frac{1}{M}\Big[R_t - e^{-\lambda h} R_{t-h} + \int_{t-h}^{t} -\lambda e^{\lambda(\mu-t)} R_\mu d\mu\Big] \\
&= -\lambda Y_t + \frac{1}{M}(R_t - e^{-\lambda h} R_{t-h}), \\
dY_t &= [-\lambda Y_t + \frac{1}{M}(R_t - e^{-\lambda h} R_{t-h})]dt.
\end{aligned} \tag{5.3.4}$$

The day on which the option needs to be valued is set as **Day** 0 and the spot price of the underlying asset is S_0. Our base time point t_0 is defined to be **Day** $-h$, which means h days before **Day** 0.[2] The option will be exercised at **Day** T so the time to maturity is T trading days.[3] In addition, the strike price of the option is denoted by K.

[2]We only consider trading days and assume one year have 252 trading days. When k is negative, **Day** k denotes the day that is k days before **Day** 0.

[3]Generally, time to maturity is denoted by τ. Since the option has a valid period from **Day** 0 to **Day** T, its time to maturity also equals T days.

Suppose the volatility estimate for for **Day** 1, denoted by σ_1 is given at **Day** 0. To indicate the leverage effect, $f(\cdot)$ should be a nonnegative and decreasing function of the EDWA cumulative return Y_t. In this chapter, we set $f(Y_t) = e^{\alpha(Y_0+rt-Y_t)}$ with a constant parameter α that is constrained to be positive. The reason why we add $r_1 t$ in the function is that this term can offset the time effect. Considered to be the EDWA of h previous cumulative returns with a fixed base time, Y_t is expected to increase as time t grows. Moreover, $f(Y_0)\sigma$ should be ensured to equal σ_1 so that the estimated daily volatility for **Day** 1 is in line with our initialization; that is, $\sigma = \sigma_1$ since $f(Y_0) = 1$ here.

5.3.1 Partial Differential Equations for Option Pricing

To calculate the option price, one method is to numerically solve the corresponding PDE equations that are related to the stochastic differential equations of our option pricing model by the Discounted Feynman-Kac Theorem.

Hereafter, the time unit is a trading day while other units can be applied if the intra-daily data are available. Assuming V is the value of an option on the underlying asset whose pay-off at maturity time T is $g(X(T))$. Under the risk-neutral measure Q, the value of the option at time t $(0 \leq t \leq T)$ is given by

$$V(t, X(t), Y(t)) = \mathbb{E}^Q\big(e^{-r_1(T-t)} g(X(T))|\mathcal{F}(t)\big), \tag{5.3.5}$$

where $\mathcal{F}(t)$ is a filtration. It is worth noting that $V(t, X(t), Y(t))$ itself is not a Q-martingale. To build a martingale process, we need to add an extra discount term such that the subsequent expected value will not depend on the time t. Therefore, we have

$$e^{-r_1 t} V(t, X(t), Y(t)) = \mathbb{E}^{\mathbb{Q}}\big(e^{-r_1 T} g(X(T))|\mathcal{F}(t)\big). \tag{5.3.6}$$

It can be proved that $e^{-r_1 t} V(t, X(t), Y(t))$ is a Q-martingale:

Proof. For $0 \le s \le t \le T$, then

$$\begin{aligned}
\mathbb{E}^Q[e^{-r_1 t}V(t,X(t),Y(t))|\mathcal{F}(s)] &= \mathbb{E}^Q[\mathbb{E}^Q[e^{-r_1 T}g(X(T))|\mathcal{F}(t)]|\mathcal{F}(s)] \\
&= \mathbb{E}^Q[e^{-r_1 T}g(X(T))|\mathcal{F}(s)] \\
&= e^{-r_1 s}\mathbb{E}^{\mathbb{Q}}[e^{-r_1(T-s)}g(X(T))|\mathcal{F}(s)] \\
&= e^{-r_1 s}V(s,X(s),Y(s)). \qquad \textbf{Q.E.D}
\end{aligned}$$

Then the differential of this Q-martingale is

$$\begin{aligned}
&d\big(e^{-r_1 t}V(t,X(t),Y(t))\big) \\
&= e^{-r_1 t}[-r_1 V dt + V_t dt + V_x dX + V_y dY + \frac{1}{2}V_{xx}dXdX + \frac{1}{2}V_{yy}dYdY + V_{xy}dXdY] \\
&= e^{-r_1 t}\big[- r_1 V dt + V_t dt + V_x(r_1 - \frac{1}{2}f^2(Y(t))\sigma^2 dt + f(Y(t))\sigma dW(t)) \\
&\quad + V_y\big(- \lambda Y(t) + \frac{1}{M}(X(t) - e^{-\lambda h}X(t-h))\big)dt + \frac{1}{2}V_{xx}f^2(Y(t))\sigma^2 dt\big] \\
&= e^{-r_1 t}f(Y(t))\sigma dW(t) + e^{-r_1 t}\big[- r_1 V + V_t + V_x\big(r_1 - \frac{1}{2}f^2(Y(t))\sigma^2\big) + \frac{1}{2}V_{xx}f^2(Y(t))\sigma^2 \\
&\quad + V_y\big(- \lambda Y(t) + \frac{1}{M}(X(t) - e^{-\lambda h}X(t-h))\big)\big]dt. \qquad (5.3.7)
\end{aligned}$$

Since $e^{-r_1 t}V$ is a Q-martingale, the dt term should be zero. Therefore, $V(t,X(t),Y(t))$ $(0 \le t \le T)$ must satisfy the following partial differential equations:

$$\begin{cases}
\quad V_y[-\lambda Y(t) + \frac{1}{M}\big(X(t) - e^{-\lambda h}X(t-h)\big)] \\
-r_1 V + V_t + V_x[r - \frac{1}{2}f^2(Y(t))\sigma^2] + \frac{1}{2}V_{xx}f^2(Y(t))\sigma^2 = 0, \\
X(s) = \ln S(\lfloor s \rfloor) - \ln S(-h), \quad -h \le s \le 0, \quad Y(0) = \frac{\sum_{j=-h+1}^{0} e^{\lambda j}X(j)}{\sum_{j=-h+1}^{0} e^{\lambda j}}, \\
V(T,X(T),Y(T)) = g(X(T)).
\end{cases} \qquad (5.3.8)$$

where $\lfloor \cdot \rfloor$ denotes the floor function, the sequence of historical close asset prices, denoted by $S(k)$, $k = -h, \ldots, 0$, are known and the form of $g(X(T))$ depends on the option kind.

For call options, $g(X(T)) = \max(e^{X(T)+\ln S(-h)} - K, 0)$ while for put options, $g(X(T)) = \max(K - e^{X(T)+\ln S(-h)}, 0)$. When λh is large enough, M approaches $\frac{1}{\lambda}$ and we can approximately simplify dY_t by

$$dY_t = \lambda(R_t - Y_t)dt. \tag{5.3.9}$$

5.3.2 Monte Carlo Simulation for Option Pricing

In addition to solving partial differential equations, our Cumulative Return model can also value options by Monte Carlo simulations. Firstly, we initialize R_0 and Y_0 by

$$R_0 = \ln S_0 - \ln S_{-h}, \qquad Y_0 = \frac{\sum_{j=-h+1}^{0} e^{\lambda j} R_j}{M'}, \qquad M' = \sum_{j=-h+1}^{0} e^{\lambda j}. \tag{5.3.10}$$

Subsequently, let us define the discrete-time versions of EDWA cumulative returns at **Day** k and $k+1$ as follows:

$$Y_k = \frac{1}{M'} \sum_{s=-h+1}^{0} e^{\lambda s} R_{k+s}, \qquad Y_{k+1} = \frac{1}{M'} \sum_{s=-h+1}^{0} e^{\lambda s} R_{k+1+s}. \tag{5.3.11}$$

Then Y_{k+1} can be thus expressed using Y_k, as in Eq. (5.3.12).

$$Y_{k+1} = e^{-\lambda}\big(Y_k - \frac{1}{M'} e^{\lambda(-h+1)} R_{k-h+1}\big) + \frac{1}{M'} R_{k+1}. \tag{5.3.12}$$

where R_k for $k = -h+1, \ldots -1$ should also be provided ($R_k \equiv \ln S_k - \ln S_{-h}$). Now the underlying asset price at **Day** T can be simulated step by step by

$$\begin{aligned} R_{k+1} &= R_k + r_1 - \frac{1}{2}(f(Y_k))^2 \sigma_1^2 + f(Y_k)\sigma_1 z_k, \\ Y_{k+1} &= e^{-\lambda}\big(Y_k - \frac{1}{M'} e^{\lambda(-h+1)} R_{k-h+1}\big) + \frac{1}{M'} R_{k+1}. \end{aligned} \tag{5.3.13}$$

where $z_k \overset{\text{i.i.d}}{\sim} N(0,1)$ for $k = 0, 1, \ldots, T-1$. Therefore, the cumulative return R_T^C at maturity is calculated by

$$R_T^C = R_0 + r_1 T - \frac{1}{2}\sigma_1^2 \sum_{k=0}^{T-1} (f(Y_k))^2 + \sigma_1 \sum_{k=0}^{T-1} f(Y_k) z_k. \tag{5.3.14}$$

where the superscript C stands for the Cumulative Return option pricing model. As for this simulation, the asset price at the end of **Day** T is calculated by $S_T^C = e^{R_T^C + \ln S_{-h}}$ because $R_T^C = \ln S_T^C - \ln S_{-h}$. Then the call/put option price at **Day** 0 can be derived by

$$\text{Call}: \quad \hat{C}_0^C \approx \frac{1}{m} e^{-r_1 T} \sum_{j=1}^{m} \max(S_T^C(j) - K, 0), \tag{5.3.15}$$

$$\text{Put}: \quad \hat{P}_0^C \approx \frac{1}{m} e^{-r_1 T} \sum_{j=1}^{m} \max(K - S_T^C(j), 0). \tag{5.3.16}$$

where m is the total simulation number and $S_T^C(j)$ is the asset price at **Day** T calculated by the j-th simulated path under our Cumulative Return option pricing model.

5.4 Methodology and Empirical Study

In this section, we will introduce the methodology for the comparison among the risk-neutral models incorporating the leverage effect and investigate the empirical results. In the empirical study, our Cumulative Return option pricing model will value options by the Monte Carlo simulation method.

5.4.1 Methodology

In what follows we will conduct a comparison among the GARCH(1, 1) and ARSV(1) models with the leverage effect and our Cumulative Return option pricing model under the risk-neutral measure. Except for our model, both the implied and non-implied versions of the examined models are explored. As for the non-implied versions, we adopt the initial volatility choice of

Chapter 4, which is the unconditional standard deviation of the 180-day log returns before the option trading day. Regarding the implied version, the upper bound for the initial volatility parameter is also set in line with Chapter 4. Moreover, we keep the 'one-day in-sample with second-day out-of-sample' rule and deal with the same dataset of options as the one used in the comparison between the risk-neutral GARCH(1) and ARSV(1) models without leverage effect in Chapter 4.

In light of our Cumulative Return option pricing model, we only report the results of its implied version for the following reasons. Firstly, compared to the GARCH(1, 1) or ARSV(1) model with the leverage effect which has four parameters, the non-implied version of our option pricing model has only one parameter—α if initial volatility is not parameterized. Though a model with fewer parameters is easier to estimate, it is less flexible when fitting in-sample observations. In fact, we examine the performance of the non-implied version of our option pricing model and the results are significantly inferior to its non-implied rivals with the leverage effect: the EGARCH(1, 1), NGARCH(1, 1) and $\rho-$ARSV(1) models. Secondly, our model relates the volatility to a function of the EDWA cumulative return Y_t so that the new asset price of the out-of-sample day (**Day** 1) has little effect on Y_1. If the initial volatility is parameterized, it is expected to be robust for the out-of-sample prediction. Therefore, only through the implied version can we take advantage of this feature.

In this chapter, the risk-neutral models are still estimated by minimizing the in-sample mean squared pricing error (MSPE) which is defined in (4.2.32). If it is not parameterized, σ_0 is set to be the unconditional standard deviation of the 180-day log returns before **Day** 0.[4] Assuming n in-sample options are selected on **Day** 0, then all the examined models with the leverage effect are estimated as follows:

$$\rho - \text{ARSV}(1) : \tilde{\theta}^L = \arg\min_{\theta^L} \frac{1}{n} \sum_{i=1}^{n} [v_i - \hat{C}_0^{AL}(\theta^L, \sigma_0, i)]^2, \quad \theta^L = (\rho, \phi, \gamma, \beta),$$

[4]As illustrated in Chapter 4, the ρ-ARSV(1) and ρ-ARSV-IV(1) models use σ_0 to determine $\tilde{x}_0$ by solving $\beta \exp(\tilde{x}/2) = \sigma_0$ where β is from the present parameter estimates.

$$\rho - \text{ARSV-IV}(1) : \tilde{\theta}^L_{\text{iv}} = \arg\min_{\theta^L_{\text{iv}}} \frac{1}{n} \sum_{i=1}^{n} [v_i - \hat{C}^{AL}_0(\theta^L_{\text{iv}}, i)]^2, \quad \theta^L_{\text{iv}} = (\rho, \phi, \gamma, \beta, \sigma_0),$$

$$\text{EGARCH}(1,\,1) : \tilde{\theta}^E = \arg\min_{\theta^E} \frac{1}{n} \sum_{i=1}^{n} [v_i - \hat{C}^E_0(\theta^E, \sigma_0, i)]^2, \quad \theta^E = (a^E_0, a^E_1, \gamma^E_1, b^E_1),$$

$$\text{EGARCH-IV}(1,\,1) : \tilde{\theta}^E_{\text{iv}} = \arg\min_{\theta^E_{\text{iv}}} \frac{1}{n} \sum_{i=1}^{n} [v_i - \hat{C}^E_0(\theta^E_{\text{iv}}, i)]^2, \quad \theta^E_{\text{iv}} = (a^E_0, a^E_1, \gamma^E_1, b^E_1, \sigma_0),$$

$$\text{NGARCH}(1,\,1) : \tilde{\theta}^N = \arg\min_{\theta^N} \frac{1}{n} \sum_{i=1}^{n} [v_i - \hat{C}^N_0(\theta^N, \sigma_0, i)]^2, \quad \theta^N = (a^N_0, a^N_1, \gamma^N_1, b^N_1),$$

$$\text{NGARCH-IV}(1,\,1) : \tilde{\theta}^N_{\text{iv}} = \arg\min_{\theta^N_{\text{iv}}} \frac{1}{n} \sum_{i=1}^{n} [v_i - \hat{C}^N_0(\theta^N_{\text{iv}}, i)]^2, \quad \theta^N_{\text{iv}} = (a^N_0, a^N_1, \gamma^N_1, b^N_1, \sigma_0),$$

$$\text{CR} : \tilde{\theta}^C_{\text{iv}} = \arg\min_{\theta^C_{\text{iv}}} \frac{1}{n} \sum_{i=1}^{n} [v_i - \hat{C}^C_0(\theta^C_{\text{iv}}, , h, \lambda, i)]^2, \quad \theta^C_{\text{iv}} = (\alpha, \sigma_1).$$

where v_i is the market price of the i-th call option, σ_0 is the initial volatility, $\hat{C}^{AL}_0(\cdot, i)$, $\hat{C}^E_0(\cdot, i)$, $\hat{C}^N_0(\cdot, i)$ and $\hat{C}^C_0(\cdot, i)$ are the theoretical prices of the corresponding volatility models for the i-th call option with the identifications (strike price K_i and time to maturity $T_i - t$). In addition, 'CR' refers to our Cumulative Return option pricing model and the suffix '-IV' stands for the implied versions that parameterize the initial volatility. As previously mentioned, $\hat{C}^{AL}_0(\cdot, i)$ and $\hat{C}^C_0(\cdot, i)$ are calculated by Eq. (5.2.8) and Eq. (5.3.15), respectively. $\hat{C}^E_0(\cdot, i)$ and $\hat{C}^N_0(\cdot, i)$ can be calculated by what we have described at the end of Section 5.2.2. The put option counterparts are estimated in a similar way.

5.4.2 Empirical Results

The in-sample and out-of-sample results of risk-neutral volatility models with leverage effect are presented in Table 5.1 and Table 5.2. The preferred value in each column is underlined. Regarding our cumulative return model, h is set to be 20 while there are three choices for the decay rate λ: 0.2, 0.5 and 0.9.

First of all, compared to the original risk-neutral GARCH(1, 1) and ARSV(1) models in Chapter 4, both of the two models are subject to substantial improvements in terms of the average

in-sample and out-of-sample pricing errors after incorporating the leverage effect, indicating the leverage effect should be considered when pricing options. The subsequent in-sample and out-of-sample comprehensive comparisons will be divided into two groups by call and put options. Moreover, the results of an outlier day (February 1st, 2018) are also investigated.

Table 5.1: Average in-sample and out-of-sample **call** option pricing errors of risk-neutral models with leverage effect.

	Call, 30 calendar days		Call, 50 calendar days	
MSPE	In-Sample	Out-of-Sample	In-Sample	Out-of-Sample
$\rho-$ARSV(1)	0.0460 (0.09)	2.3828 (3.29)	0.0171 (0.05)	2.5883 (3.68)
$\rho-$ARSV-IV(1)	0.0454 (0.07)	3.6186 (6.21)	0.0126 (0.01)	2.2966 (3.72)
EGARCH(1, 1)	0.0750 (0.15)	2.5348 (3.36)	0.0184 (0.02)	2.6446 (3.60)
EGARCH-IV(1, 1)	0.0668 (0.15)	7.3703 (8.88)	0.0157 (0.02)	3.8715 (5.03)
NGARCH(1, 1)	0.1157 (0.20)	2.5401 (3.34)	0.1051 (0.10)	2.7322 (3.79)
NGARCH-IV(1, 1)	0.1144 (0.20)	2.7526 (3.69)	0.0794 (0.10)	2.6770 (3.80)
CR, $h = 20$, $\lambda = 0.2$	0.1952 (0.22)	1.7967 (2.54)	0.1865 (0.13)	1.9500 (2.62)
CR, $h = 20$, $\lambda = 0.5$	0.1731 (0.21)	1.8395 (2.61)	0.1836 (0.13)	1.8498 (2.30)
CR, $h = 20$, $\lambda = 0.9$	0.1678 (0.21)	1.7787 (2.53)	0.1902 (0.13)	1.9836 (2.52)

5.4.2.1 In-Sample Comparison

As for call options, the implied $\rho-$ARSV(1) model dominates other models in terms of the average value and standard deviation of the in-sample MSPE no matter whether the time to maturity is 30 or 50 calendar days. With respect to the 30-day put options, the implied EGARCH(1, 1) model is preferred in terms of the average value and standard deviation of the

Table 5.2: Average in-sample and out-of-sample **put** option pricing errors of risk-neutral models with leverage effect.

MSPE	Put, 30 calendar days In-Sample	Put, 30 calendar days Out-of-Sample	Put, 50 calendar days In-Sample	Put, 50 calendar days Out-of-Sample
$\rho-$ARSV(1)	0.0198 (0.07)	1.5919 (2.16)	0.0057 (0.0104)	1.8927 (2.59)
$\rho-$ARSV-IV(1)	0.0092 (0.014)	2.4377 (5.64)	0.0038 (0.0064)	1.7785 (2.38)
EGARCH(1, 1)	0.0108 (0.02)	1.6437 (2.32)	0.0062 (0.0078)	1.8895 (2.79)
EGARCH-IV(1, 1)	0.0024 (0.004)	2.2565 (2.23)	0.0036 (0.0075)	2.8590 (3.26)
NGARCH(1, 1)	0.0629 (0.06)	1.7850 (2.42)	0.1360 (0.24)	1.9173 (2.43)
NGARCH-IV(1, 1)	0.0613 (0.10)	1.9754 (2.49)	0.0934 (0.18)	2.0558 (2.59)
CR, $h = 20$, $\lambda = 0.2$	0.3030 (0.24)	2.6692 (3.95)	0.3706 (0.63)	2.5786 (2.93)
CR, $h = 20$, $\lambda = 0.5$	0.2656 (0.23)	2.2479 (3.37)	0.4180 (0.73)	2.2924 (2.55)
CR, $h = 20$, $\lambda = 0.9$	0.2804 (0.26)	2.0178 (2.57)	0.5158 (0.80)	2.3455 (1.94)

in-sample MSPE. Regarding the 50-day put options, the implied $\rho-$ARSV(1) model has as good results as the implied EGARCH(1, 1) model and these two models are superior to the others.

Not surprisingly, all the implied versions have better in-sample performances than their non-implied counterparts while the improvements are considerably insignificant. Moreover, all the examined risk-neutral models including our Cumulative Return option pricing model have remarkably satisfying in-sample results. Consequently, models with leverage effects are capable of precisely fitting the observed in-sample option prices. Interestingly, all the ARSV(1) and GARCH(1, 1) models with leverage effects have a smaller average in-sample MSPE for 50-day options than their 30-day counterparts. Options with a longer time to maturity are usually believed to be more difficult to fit. The reason of the uncommon case here is worth studying in the future but at least it shows incorporating the leverage effect makes risk-neutral models

more promising to fit observed option prices that will be exercised in a long time.

Now let us have a look at our Cumulative Return option pricing model. Its average in-sample MSPE, though a little inferior to the other models with leverage effect, are still acceptable. After all, as previously mentioned, it is the capability of out-of-sample prediction that decides whether a model can be adopted to price options in the market. Next, putting aside the in-sample results, we will focus on the out-of-sample option pricing errors.

5.4.2.2 Out-of-Sample Comparison

For both the 30-day and 50-day call options, our Cumulative Return option pricing model is superior to other models in terms of the average value and standard deviation of the out-of-sample option pricing errors. Such results are surprising since our option pricing model has only two parameters and seems to be much easier than the other examined models. In addition, the decay rate in our option pricing model can be further refined to provide a better out-of-sample prediction performance.

As for put options, the $\rho-$ARSV(1) model dominates other models regarding the 30-day out-of-sample prediction performance while the implied $\rho-$ARSV(1) has the smallest average MSPE for the 50-day options. Our cumulative return model with $\lambda = 0.9$ is preferred to other models in terms of the standard deviation of the out-sample MSPE for the 50-day put options.

As expected, the average out-of-sample MSPE is larger than the in-sample counterpart for each risk-neutral model. Furthermore, the 50-day out-of-sample results are not always inferior to their 30-day counterparts, indicating the potential of such models to price options with a long time to maturity.

Moreover, except for the $\rho-$ARSV(1) model for the 50-day call and put options and the NGARCH(1, 1) model for the 50-day call options, the implied models especially the implied EGARCH(1, 1) model are inferior to their non-implied versions. Although parameterizing the initial volatility can better fit the in-sample observations, it undermines a model's robustness for the out-of-sample prediction. However, compared to the results shown in Table 4.8 of Chapter

4, the impact of the extra volatility parameter on the models that capture the leverage effect is not as strong as the one on the original GARCH(1, 1) and ARSV(1) models without the leverage effect. Therefore, the introduction of the leverage effect makes the implied versions of the GARCH(1, 1) and ARSV(1) models more robust for the out-of-sample pricing performance.

Revisiting the initial setting of our Cumulative Return option pricing model, we assume h is equal to 20. Whether other values of h will lead to a better performance is out of the scope of this chapter. Moreover, for a given h, how can we derive the optimal decay rate λ? Such topics are worth investigating in the future.

5.4.2.3 Comparison On An Outlier Day

Table 5.3: In-sample and out-of-sample put option pricing error of risk-neutral models with the leverage effect on the outlier day.

Feb 1, 2018	Put, 30 calendar days		Put, 50 calendar days	
MSPE	In-Sample	Out-of-Sample	In-Sample	Out-of-Sample
$\rho-$ARSV(1)	0.0024	58.5681	0.0012	64.9115
$\rho-$ARSV-IV(1)	0.0019	55.6732	0.0012	62.6181
EGARCH(1, 1)	0.0089	57.9871	0.0018	62.9838
EGARCH-IV(1, 1)	0.0015	74.0426	0.0010	86.2504
NGARCH(1, 1)	0.0273	55.9743	0.0169	61.6368
NGARCH-IV(1, 1)	0.0995	55.7715	0.0011	56.1960
CR, $h = 20$, $\lambda = 0.2$	0.1885	<u>0.4663</u>	0.0095	<u>0.6730</u>
CR, $h = 20$, $\lambda = 0.5$	0.1515	2.8194	0.0128	8.4193
CR, $h = 20$, $\lambda = 0.9$	0.1226	17.8401	0.0161	22.3763

For the results shown in Table 5.1 and Table 5.2, the last out-of-sample day in our collection is Thursday, February 1st, 2018. The U.S. stock market experienced a series plunge on the next Monday, February 5th. In fact, if we minimize the in-sample pricing error for the put options on Thursday, February 1st and the corresponding out-of-sample results on Friday, February 2nd will

reveal some clues of the plunge.[5] That day can be seen an outlier day because the out-of-sample errors of all examined option pricing models except our proposed model are exceptionally large. The results are reported in Table 5.3, which illustrates our Cumulative Return model substantially outperforms other models on that outlier day. Therefore, our model is capable of tracking the true volatility trend at the extreme situation by combining the in-sample parameter estimates and the slight change in EDWA cumulative Y_t when the new out-of-sample asset price arrives. In addition, regarding our model, the performance when $\lambda = 0.2$ dominates the one when $\lambda = 0.5$ or 0.9. As shown in Table 5.4, a smaller λ leads to a larger α and a smaller initial volatility estimates on that outlier day. The reason of the superiority of that case might be: the smaller initial volatility makes the model less sensitive to the new asset price while the larger α ensures its capability of identifying the volatility trend from a insignificant change in the EDWA cumulative return.

Table 5.4: In-sample nonlinear-least-square parameter estimates of our model for put options on the outlier day.

Feb 1, 2018	Put, 30 calendar days		Put, 50 calendar days	
Parameter Estimates	$\tilde{\sigma}_1$	$\tilde{\alpha}$	$\tilde{\sigma}_1$	$\tilde{\alpha}$
CR, $h = 20$, $\lambda = 0.2$	7.49412×10^{-3}	14.74616	7.464591×10^{-3}	11.13344
CR, $h = 20$, $\lambda = 0.5$	7.49939×10^{-3}	13.09271	7.66398×10^{-3}	10.34252
CR, $h = 20$, $\lambda = 0.9$	7.588897×10^{-3}	11.96045	7.906737×10^{-3}	9.84929

The value of the results on the outlier day is limited for the following reasons. Firstly, regarding that extreme market situation, our predetermined initial volatility for the non-implied models is too small. Secondly, our assumption that the in-sample nonlinear-least-squares parameters will stay unchanged overnight is not suitable for such an abnormal out-of-sample day. Moreover, we observe only one outlier day and other extreme situations may lead to different performances. However, the results at least highlight the robustness of our Cumulative Return option pricing

[5] Only the results of the put options on the outlier day are discussed here since the call options seem to be relatively normal.

model especially for a small λ.

5.5 Conclusion and Future Work

This paper evaluates the in-sample and out-of-sample performances of four option pricing models incorporating the leverage effect: the ρ-ARSV(1), EGARCH(1, 1), NGARCH(1, 1) models and our Cumulative Return option pricing model. The implied versions of the first three models are also investigated. Our first conclusion is that compared to the original risk-neutral GARCH(1, 1) and ARSV(1) models in Chapter 4, both of the two models exhibit substantial improvements in terms of the average in-sample and out-of-sample pricing errors after incorporating the leverage effect. Therefore, the leverage effect plays an indispensable role in option pricing. Secondly, our Cumulative Return option pricing model that has only two parameters are superior to the other sophisticated models regarding the out-of-sample pricing error of call options. As for the put option pricing, the results of most models are considerably close. The comparison of the outlier day also further adds to the robustness of our proposed model for the out-of-sample price prediction.

For the future work, we can focus on the selection of the initial volatility for the non-implied option pricing models. In our proposed model, the choices of h and λ can be further adjusted to have a better performance. Moreover, we only explore the results with $f(Y_t) = e^{\alpha(Y_0+rt-Y_t)}$ and other decreasing functions are worth exploring. We can also investigate the pricing errors of the in-the-money and out-of-the-money options separately.

REFERENCES

Acerbi, C., & Szekely, B. (2014). Back-testing expected shortfall. *Risk*, 76.

Acerbi, C., & Tasche, D. (2002a). Expected shortfall: a natural coherent alternative to value at risk. *Economic Notes*, *31*(2), 379–388.

Acerbi, C., & Tasche, D. (2002b). On the coherence of expected shortfall. *Journal of Banking & Finance*, *26*(7), 1487–1503.

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Ebens, H. (2001). The distribution of realized stock return volatility. *Journal of financial economics*, *61*(1), 43–76.

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2001). The distribution of realized exchange rate volatility. *Journal of the American statistical association*, *96*(453), 42–55.

Areal, N. M., & Taylor, S. J. (2002). The realized volatility of ftse-100 futures prices. *Journal of Futures Markets*, *22*(7), 627–648.

Artzner, P., Delbaen, F., Eber, J.-M., & Heath, D. (1997). Thinking coherently. *Risk*, *10*(11), 68–71.

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

Awartani, B. M., & Corradi, V. (2005). Predicting the volatility of the s&p-500 stock index via garch models: the role of asymmetries. *International Journal of Forecasting*, *21*(1), 167–183.

Bae, T., & Iscoe, I. (2012). Large-sample confidence intervals for risk measures of location–scale families. *Journal of Statistical Planning and Inference*, *142*(7), 2032–2046.

Bakshi, G., Cao, C., & Chen, Z. (1997). Empirical performance of alternative option pricing models. *The Journal of Finance*, *52*(5), 2003–2049.

Baysal, R. E., & Staum, J. (2008). Empirical likelihood for value-at-risk and expected shortfall. *The Journal of Risk*, *11*(1), 3.

Bernardi, M. (2013). Risk measures for skew normal mixtures. *Statistics & Probability Letters*, *83*(8), 1819–1824.

Black, F. (1976). Stuedies of stock price volatility changes. proceedings of the 1976 meetings of the american statistical association, business and economics section. , 177–181.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, *31*(3), 307–327.

Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. *The review of economics and statistics*, 542–547.

Bouchaud, J.-P., Matacz, A., & Potters, M. (2001). Leverage effect in financial markets: The retarded volatility model. *Physical review letters*, *87*(22), 228701.

Briers, M., Doucet, A., & Maskell, S. (2010). Smoothing algorithms for state–space models. *Annals of the Institute of Statistical Mathematics*, *62*(1), 61–89.

Broda, S. A., & Paolella, M. S. (2011). Expected shortfall for distributions in finance. *Statistical Tools for Finance and Insurance*, 57–99.

Cappé, O. (2011). Online em algorithm for hidden markov models. *Journal of Computational and Graphical Statistics*, *20*(3), 728–749.

Carnero, M. Á., Peña, D., & Ruiz, E. (2001). Is stochastic volatility more flexible than garch?

Carnero, M. A., Peña, D., & Ruiz, E. (2004). Persistence and kurtosis in garch and stochastic volatility models. *Journal of Financial Econometrics*, *2*(2), 319–342.

Carver, L. (2013). Mooted var substitute cannot be back-tested, says top quant. *Risk, March*, *8*.

Chen, Q., Gerlach, R., & Lu, Z. (2012). Bayesian Value-at-Risk and expected shortfall forecasting via the asymmetric Laplace distribution. *Computational Statistics & Data Analysis*, *56*(11), 3498–3516.

Chen, S. X. (2007). Nonparametric estimation of expected shortfall. *Journal of Financial Econometrics*, *6*(1), 87–107.

Chen, S. X., & Tang, C. Y. (2005). Nonparametric inference of value-at-risk for dependent financial returns. *Journal of Financial Econometrics*, *3*(2), 227–255.

Chernov, M., Gallant, A. R., Ghysels, E., & Tauchen, G. (2003). Alternative models for stock price dynamics. *Journal of Econometrics*, *116*(1), 225–257.

Christie, A. A. (1982). The stochastic behavior of common stock variances: Value, leverage and interest rate effects. *Journal of financial Economics*, *10*(4), 407–432.

Christoffersen, P., Jacobs, K., et al. (2001). *The importance of the loss function in option pricing*. CIRANO.

Christoffersen, P., Jacobs, K., et al. (2002). *Which volatility model for option valuation?* CIRANO.

Comte, F., Coutin, L., & Renault, É. (2012). Affine fractional stochastic volatility models. *Annals of Finance*, *8*(2), 337–378.

Costanzino, N., & Curran, M. (2015). Backtesting general spectral risk measures with application to expected shortfall.

Cox, J. (1975). Notes on option pricing i: Constant elasticity of variance diffusions. *Unpublished note, Stanford University, Graduate School of Business*.

Davison, A. C., & Smith, R. L. (1990). Models for exceedances over high thresholds. *Journal of the Royal Statistical Society. Series B (Methodological)*, 393–442.

Del Moral, P., Doucet, A., & Singh, S. (2010). Forward smoothing using sequential monte carlo. *arXiv preprint arXiv:1012.5390*.

Doucet, A., Godsill, S., & Andrieu, C. (2000). On sequential monte carlo sampling methods for bayesian filtering. *Statistics and computing*, *10*(3), 197–208.

Doucet, A., Godsill, S. J., & West, M. (2000). Monte carlo filtering and smoothing with application to time-varying spectral estimation. , *2*, II701–II704.

Duan, J.-C. (1995). The garch option pricing model. *Mathematical finance*, *5*(1), 13–32.

Embrechts, P., Kaufmann, R., & Patie, P. (2005). Strategic long-term financial risks: Single risk factors. *Computational Optimization and Applications*, *32*(1-2), 61–90.

Emmer, S., Kratz, M., & Tasche, D. (2015). What is the best risk measure in practice? a comparison of standard measures.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the Econometric Society*, 987–1007.

Engle, R. F., & Ng, V. K. (1993). Measuring and testing the impact of news on volatility. *The journal of finance*, *48*(5), 1749–1778.

Figlewski, S., & Wang, X. (2000). Is the 'leverage effect' a leverage effect?

Fissler, T., Ziegel, J. F., et al. (2016). Higher order elicitability and osbands principle. *The Annals of Statistics*, *44*(4), 1680–1707.

Glosten, L. R., Jagannathan, R., & Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The journal of finance*, *48*(5), 1779–1801.

Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une serie aleatoire. *Annals of Mathematics*, *44*(3), 423–453.

Gneiting, T. (2011). Making and evaluating point forecasts. *Journal of the American Statistical Association*, *106*(494), 746–762.

Gordon, N. J., Salmond, D. J., & Smith, A. F. (1993). Novel approach to nonlinear/non-gaussian bayesian state estimation. In *Iee proceedings f (radar and signal processing)* (Vol. 140, pp. 107–113).

Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The review of financial studies*, *6*(2), 327–343.

Heston, S. L., & Nandi, S. (1997). A closed-form garch option pricing model.

Hill, J. B. (2013). Expected shortfall estimation and gaussian inference for infinite variance time series. *Journal of Financial Econometrics*, *13*(1), 1–44.

Hull, J. (2012). *Risk Management and Financial Institutions* (Vol. 733). John Wiley & Sons.

Hull, J., & White, A. (1987). The pricing of options on assets with stochastic volatilities. *The journal of finance*, *42*(2), 281–300.

Inui, K., & Kijima, M. (2005). On the significance of expected shortfall as a coherent risk measure. *Journal of Banking & Finance*, *29*(4), 853–864.

Ionides, E. L., Bhadra, A., Atchadé, Y., King, A., et al. (2011). Iterated filtering. *The Annals of Statistics*, *39*(3), 1776–1802.

Jiménez, J. A., & Arunachalam, V. (2011). Using Tukey's g and h family of distributions to calculate value-at-risk and conditional value-at-risk. *The Journal of Risk*, *13*(4), 95-116.

Johannes, M., & Polson, N. (2009). Particle filtering. *Handbook of Financial Time Series*, 1015–1029.

Kitagawa, G. (1996). Monte carlo filter and smoother for non-gaussian nonlinear state space models. *Journal of computational and graphical statistics*, *5*(1), 1–25.

Kong, A., Liu, J. S., & Wong, W. H. (1994). Sequential imputations and bayesian missing data problems. *Journal of the American statistical association*, *89*(425), 278–288.

Lehar, A., Scheicher, M., & Schittenkopf, C. (2002). Garch vs. stochastic volatility: Option pricing and risk management. *Journal of Banking & Finance*, *26*(2), 323–345.

Liu, J., & West, M. (2001). Combined parameter and state estimation in simulation-based filtering. In *Sequential monte carlo methods in practice* (pp. 197–223). Springer.

Liu, J. S., & Chen, R. (1995). Blind deconvolution via sequential imputations. *Journal of the american statistical association*, *90*(430), 567–576.

Malmsten, H., Teräsvirta, T., et al. (2004). Stylized facts of financial time series and three popular models of volatility. *SSE/EFI Working Paper Series in Economics and Finance*, *563*.

McNeil, A. J. (1997). Estimating the tails of loss severity distributions using extreme value theory. *ASTIN Bulletin: The Journal of the IAA*, *27*(1), 117–137.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the Econometric Society*, 347–370.

Pang, T., & Yang, S. (2015). Copulas and GARCH Models for Market Risks of Relative Returns. *Journal of Finance & Management Research*, *1*(1), 19-38.

Patton, A. J. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics*, *160*(1), 246–256.

Pitt, M. K., & Shephard, N. (1999). Filtering via simulation: Auxiliary particle filters. *Journal of the American statistical association*, *94*(446), 590–599.

Poyiadjis, G., Doucet, A., & Singh, S. S. (2011). Particle approximations of the score and observed information matrix in state space models with application to parameter estimation. *Biometrika*, *98*(1), 65–80.

Scaillet, O. (2004). Nonparametric estimation and sensitivity analysis of expected shortfall. *Mathematical Finance*, *14*(1), 115–129.

Tauchen, G. E., & Pitts, M. (1983). The price variability-volume relationship on speculative markets. *Econometrica: Journal of the Econometric Society*, 485–505.

Taylor, J. W. (2007). Using exponentially weighted quantile regression to estimate value at risk and expected shortfall. *Journal of Financial Econometrics*, *6*(3), 382–406.

Taylor, S. (1982). Financial returns modelled by the product of two stochastic processes, a study of daily sugar prices 1961-79. *Time Series Analysis: Theory and Practice*, *1*, 203–226.

Wang, K.-L., Fawson, C., Barrett, C. B., & McDonald, J. B. (2001). A flexible parametric garch model with an application to exchange rates. *Journal of Applied Econometrics*, *16*(4), 521–536.

Wong, W. K. (2008). Backtesting trading risk of commercial banks using expected shortfall. *Journal of Banking & Finance*, *32*(7), 1404–1415.

Yamai, Y., & Yoshiba, T. (2005). Value-at-risk versus expected shortfall: A practical perspective. *Journal of Banking & Finance*, *29*(4), 997–1015.

Yu, J. (2002). Forecasting volatility in the new zealand stock market. *Applied Financial Economics*, *12*(3), 193–202.

Zhu, D., & Galbraith, J. W. (2011). Modeling and forecasting expected shortfall with the generalized asymmetric Student-t and asymmetric exponential power distributions. *Journal of Empirical Finance*, *18*(4), 765–778.

APPENDICES

APPENDIX A

Appendix for Chapter 3

A.1 Derivations of Eq. (3.2.8) and Eq. (3.2.9)

Suppose $X \sim N(\mu, \sigma^2)$ and $Pr(X \leq A_\alpha) = \alpha$. Let $Z \equiv \frac{X-\mu}{\sigma}$ and $a \equiv \frac{A_\alpha - \mu}{\sigma}$. Then $Z \sim N(0,1)$ and $\Phi(a) = \Phi\left(\frac{A_\alpha-\mu}{\sigma}\right) = Pr\left(Z \leq \frac{A_\alpha-\mu}{\sigma}\right) = Pr(X \leq A_\alpha) = \alpha$, where $\Phi(\cdot)$ is the CDF of the standard normal distribution. Hence, we have

$$\begin{aligned}
\mathbb{E}[X|X > A_\alpha] &= \sigma\mathbb{E}\left[Z|Z > a\right] + \mu = \mu + \frac{\sigma}{1-\Phi(a)}\int_a^\infty z\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}dz \\
&= \mu + \frac{\sigma}{(1-\alpha)\sqrt{2\pi}}e^{-\frac{a^2}{2}} = \mu + \frac{\sigma}{(1-\alpha)\sqrt{2\pi}}e^{-\frac{(A_\alpha-\mu)^2}{2\sigma^2}}, \\
\mathbb{E}[X^2|X > A_\alpha] &= \frac{1}{Pr(X > A_\alpha)}\int_{A_\alpha}^\infty \frac{x^2}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx = \frac{1}{1-\alpha}\int_a^\infty \frac{(z\sigma+\mu)^2}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}dz \\
&= \frac{1}{(1-\alpha)\sqrt{2\pi}}\int_a^\infty (z^2\sigma^2 + 2\mu\sigma z + \mu^2)e^{-\frac{z^2}{2}}dz \\
&= \mu^2 + \sigma^2 + \frac{\sigma(A_\alpha+\mu)}{(1-\alpha)\sqrt{2\pi}}e^{-\frac{(A_\alpha-\mu)^2}{2\sigma^2}},
\end{aligned}$$

where the form $\int_a^\infty z^2 e^{-\frac{z^2}{2}} dz$ is evaluated using integration by parts as follows:

$$\int_a^\infty z^2 e^{-\frac{z^2}{2}} dz = \int_a^\infty -z de^{-\frac{z^2}{2}} = \left(-ze^{-\frac{z^2}{2}}\right)\Big|_a^\infty - \int_a^\infty -e^{-\frac{z^2}{2}} dz = ae^{-\frac{a^2}{2}} + \sqrt{2\pi}(1-\Phi(a)).$$

$\mathbb{E}[X^3|X > A_\alpha]$ can be derived using integration by parts in a similar way.

A.2 EVT estimators for VaR and ES

The EVT estimator for VaR at β level is calculated by solving $Pr(W \leq \text{VaR}_\beta(W)) = \beta$, where the unconditional probability $Pr(W \leq \text{VaR}_\beta(W))$ is given in (3.5.6). When $\xi \neq 0$, we have

$$(1-F(v))\left(1-\left[1+\frac{\xi(\text{VaR}_\beta(W)-v)}{\sigma}\right]^{-1/\xi}\right)+F(v)=\beta. \tag{A.2.1}$$

Hence, $\text{VaR}_\beta(W) = v + \frac{\sigma}{\xi}[(\frac{1-\beta}{1-F(v)})^{-\xi} - 1]$. Since $W - v \sim GPD(\xi, \sigma)$, its ES estimate is

$$\begin{aligned}\text{ES}_\beta(W) &= \mathbb{E}[W|W \geq \text{VaR}_\beta(W)] = v + \mathbb{E}[W-v|W-v \geq \text{VaR}_\beta(W)-v] \\ &= v + \text{VaR}_\beta(W) - v + \frac{\sigma + \xi\text{VaR}_\beta(W) - \xi v}{1-\xi} = \frac{\text{VaR}_\beta(W) + \sigma - \xi v}{1-\xi},\end{aligned} \tag{A.2.2}$$

where (A.2.2) is obtained by GPD's ES closed-form formula in Table A.1. When $\xi = 0$, we have

$$(1-F(v))\left(1-e^{-\frac{1}{\sigma}(\text{VaR}_\beta(W)-v)}\right)+F(v)=\beta. \tag{A.2.3}$$

Thus $\text{VaR}_\beta(W) = v - \sigma\ln(\frac{1-\beta}{1-F(v)})$. Now $W - v$ follow an exponential distribution with the parameter σ and we have the following derivation based on PDF of the exponential distribution:

$$\begin{aligned}&\mathbb{E}[W-v|W-v \geq \text{VaR}_\beta(W)-v] = \frac{1}{1-G_\sigma(b)}\int_b^\infty \frac{y}{\sigma}e^{-\frac{y}{\sigma}}dy = \text{VaR}_\beta(W) - v + \sigma, \\ &\text{ES}_\beta(W) = v + \mathbb{E}[W-v|W-v \geq \text{VaR}_\beta(W)-v] = \text{VaR}_\beta(W) + \sigma.\end{aligned} \tag{A.2.4}$$

where $b = \text{VaR}_\beta(W) - v$ and $G_\sigma(b) = Pr(W - v \leq b) = 1 - e^{-b/\sigma}$.

A.3 Closed-form Formulas for Some Heavy-Tailed Distributions

In what follows we summarize closed-form expressions of some necessary statistics used in this paper. Results are reported in Table A.1, in which A is a constant, $\Gamma(\cdot)$ is the gamma function, $\gamma(\alpha;x)=\int_0^x z^{\alpha-1}e^{-z}dz$ is the lower incomplete gamma function and $\hat{\Gamma}(\alpha;x)=\int_x^\infty z^{\alpha-1}e^{-z}dz$ is the upper incomplete gamma function. Except t and GPD, the domains of PDF and CDF are $w>0$. In GPD, $\xi\neq 0,\ \sigma>0$ always hold and $1+\xi w/\sigma>0$ should be guaranteed when ξ is negative. Some non-analytical expressions need to be computed by numerical methods or software packages. For the ease of presentation, long equations are shown below separately:

$$
\begin{aligned}
m_1^t &= \frac{1}{1-F_W(A)}\frac{\Gamma(\frac{v+1}{2})v}{\sqrt{v\pi}\Gamma(\frac{v}{2})(v-1)}\left(\frac{A^2}{v}+1\right)^{-\frac{v-1}{2}}, \quad v>1;\\
m_2^{GPD} &= A^2+2\left(1+\frac{\xi}{\sigma}A\right)\frac{\sigma(A+\sigma-A\xi)}{(\xi-1)(2\xi-1)}, \quad \xi<\frac{1}{2};\\
m_3^{GPD} &= A^3-\frac{3\sigma\left(1+\xi\frac{A}{\sigma}\right)}{\xi-1}A^2+\frac{6\sigma^2\left(1+\xi\frac{A}{\sigma}\right)^2}{(\xi-1)(2\xi-1)}A-\frac{6\sigma^3\left(1+\xi\frac{A}{\sigma}\right)^3}{(\xi-1)(2\xi-1)(3\xi-1)}, \quad \xi<\frac{1}{3}.
\end{aligned}
$$

Table A.1: Closed-form formulas of 1st, 2nd & 3rd conditional moments.

W	t, df $=v$	Gamma(α,β)	Lognormal(μ,σ^2)	GPD(ξ,σ)	Weibull(k,λ)
PDF, $f_W(w)$	$\frac{\Gamma(\frac{v+1}{2})}{\sqrt{v\pi}\Gamma(\frac{v}{2})}(1+\frac{w^2}{v})^{-\frac{v+1}{2}}$	$\frac{w^{\alpha-1}e^{-\frac{w}{\beta}}}{\Gamma(\alpha)\beta^{\alpha}}$	$\frac{1}{\sqrt{2\pi}\sigma w}e^{-\frac{(\ln w-\mu)^2}{2\sigma^2}}$	$\frac{1}{\sigma}\left(1+\xi\frac{w}{\sigma}\right)^{-\frac{1}{\xi}-1}$	$\frac{k}{\lambda}(\frac{w}{\lambda})^{k-1}e^{-(\frac{w}{\lambda})^k}$
CDF, $F_W(w)$	$\int_{-\infty}^{w} f_W(x)dx$	$\frac{\gamma\left(\alpha;\frac{w}{\beta}\right)}{\Gamma(\alpha)}$	$\Phi(\frac{\ln w-\mu}{\sigma})$	$1-\left(1+\xi\frac{w}{\sigma}\right)^{-\frac{1}{\xi}}$	$1-e^{-(\frac{w}{\lambda})^k}$
$\mathbb{E}[W\|W\geq A]$	m_1^t	$\frac{\beta\hat{\Gamma}\left(\alpha+1;\frac{A}{\beta}\right)}{(1-F_W(A))\Gamma(\alpha)}$	$\frac{\Phi(\frac{\mu+\sigma^2-\ln A}{\sigma})}{1-F_W(A)}e^{\mu+\frac{\sigma^2}{2}}$	$A+\frac{\xi A+\sigma}{1-\xi}, \xi<1$	$\frac{\lambda\hat{\Gamma}\left(\frac{1}{k}+1;\left(\frac{A}{\lambda}\right)^k\right)}{1-F_W(A)}$
$\mathbb{E}[W^2\|W\geq A]$	$\frac{\int_A^{\infty}x^2f_W(x)dx}{1-F_W(A)}, v>2$	$\frac{\beta^2\hat{\Gamma}\left(\alpha+2;\frac{A}{\beta}\right)}{(1-F_W(A))\Gamma(\alpha)}$	$\frac{\Phi(\frac{\mu+2\sigma^2-\ln A}{\sigma})}{1-F_W(A)}e^{2\mu+2\sigma^2}$	m_2^{GPD}	$\frac{\lambda^2\hat{\Gamma}\left(\frac{2}{k}+1;\left(\frac{A}{\lambda}\right)^k\right)}{1-F_W(A)}$
$\mathbb{E}[W^3\|W\geq A]$	$\frac{\int_A^{\infty}x^3f_W(x)dx}{1-F_W(A)}, v>3$	$\frac{\beta^3\hat{\Gamma}\left(\alpha+3;\frac{A}{\beta}\right)}{(1-F_W(A))\Gamma(\alpha)}$	$\frac{\Phi(\frac{\mu+3\sigma^2-\ln A}{\sigma})}{1-F_W(A)}e^{3\mu+\frac{9}{2}\sigma^2}$	m_3^{GPD}	$\frac{\lambda^3\hat{\Gamma}\left(\frac{3}{k}+1;\left(\frac{A}{\lambda}\right)^k\right)}{1-F_W(A)}$

APPENDIX B

Appendix for Chapter 4

B.1 Two Forms of ARSV(1) Model

The first version of the ARSV(1) model is

$$y_t = \exp(\frac{x'_t}{2})\xi_t, \quad \xi_t \overset{\text{i.i.d}}{\sim} N(0,1),$$
$$x'_t = \alpha + \phi x'_{t-1} + \gamma\eta_t, \quad \eta_t \overset{\text{i.i.d}}{\sim} N(0,1), \quad \xi_t \perp \eta_t.$$

The second equation can be transformed by

$$x'_t - \mu = \phi(x'_{t-1} - \mu) + \gamma\eta_t,$$

where $\mu = \frac{\alpha}{1-\phi}$. Let $x_t \equiv x'_t - \mu$. Then we have

$$y_t = \exp(\frac{x_t + \mu}{2})\xi_t = \exp(\frac{\mu}{2})\exp(\frac{x_t}{2})\xi_t.$$

Therefore, the former ARSV(1) model is transformed as follows:

$$y_t = \beta \exp(\frac{x_t}{2})\xi_t, \quad \xi_t \overset{\text{i.i.d}}{\sim} N(0,1),$$
$$x_t = \phi x_{t-1} + \gamma\eta_t, \quad \eta_t \overset{\text{i.i.d}}{\sim} N(0,1), \quad \xi_t \perp \eta_t.$$

where $\beta \equiv \exp(\frac{\mu}{2})$. One change we need to pay attention to is the conditional variance in the former version is $\exp(x_t')$ while it becomes $\beta^2 \exp(x_t)$ in the new version.

B.2 Forward-Only FFBS Algorithm for estimating $\mathcal{S}_n^\theta$

1 Initialize $\widehat{T}_0^\theta(x_0^{(i)}) = 0$ and obtain the weighted filtering particles $\{x_0^{(i)}, \omega_0^{(i)}\}_{i=1}^N$.

2 Repeat the following steps for time steps $k = 1, 2, \ldots, n$.

2.1 Obtain weighted filtering particles $\{x_k^{(i)}, \omega_k^{(i)}\}$ for $i = 1, \ldots, N$.

2.2 $\widehat{T}_k^\theta(x_k^{(i)}) = \dfrac{\sum_{j=1}^N \omega_{k-1}^{(j)} f_\theta(x_k^{(i)}|x_{k-1}^{(j)})[\widehat{T}_{k-1}^\theta(x_{k-1}^{(j)}) + s_k(x_{k-1}^{(j)}, x_k^{(i)})]}{\sum_{j=1}^N \omega_{k-1}^{(j)} f_\theta(x_k^{(i)}|x_{k-1}^{(j)})}, i = 1, \ldots, N.$

2.3 $\widehat{\mathcal{S}}_k^\theta = \sum_{i=1}^N \omega_k^{(i)} \widehat{T}_k^\theta(x_k^{(i)})$.

B.3 Solution to the Maximizing Step

We have $A(\theta) = \frac{1}{2}\ln(\gamma^2) + \frac{1}{2}\ln(\beta^2)$ and $\psi(\theta) = (-\frac{\phi^2}{2\gamma^2}, \frac{\phi}{\gamma^2}, -\frac{1}{2\gamma^2}, -\frac{1}{2\beta^2})$. The unique solution of the complete-data maximum likelihood equation $\nabla_\theta \psi(\theta) s - \nabla_\theta A(\theta) = 0$ is derived as follows:

$$\nabla_\theta \psi(\theta) = \begin{bmatrix} -\frac{\phi}{\gamma^2} & \frac{1}{\gamma^2} & 0 & 0 \\ \frac{\phi^2}{2\gamma^4} & -\frac{\phi}{\gamma^4} & \frac{1}{2\gamma^4} & 0 \\ 0 & 0 & 0 & \frac{1}{2\beta^4} \end{bmatrix}, \quad s = \begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{bmatrix}, \quad \nabla_\theta A(\theta) = \begin{bmatrix} 0 \\ \frac{1}{2\gamma^2} \\ \frac{1}{2\beta^2} \end{bmatrix}.$$

Then we have

$$\nabla_\theta \psi(\theta) s = \begin{bmatrix} -\frac{\phi}{\gamma^2} z_1 + \frac{1}{\gamma^2} z_2 \\ \frac{\phi^2}{2\gamma^4} z_1 - \frac{\phi}{\gamma^4} z_2 + \frac{1}{2\gamma^4} z_3 \\ \frac{1}{2\beta^4} z_4 \end{bmatrix} = \begin{bmatrix} 0 \\ \frac{1}{2\gamma^2} \\ \frac{1}{2\beta^2} \end{bmatrix}.$$

$\hat{\phi}$ and $\hat{\beta}^2$ can be solved from the first and third equations in the linear system, respectively. Plugging $\hat{\phi}$ into the second equation, $\hat{\gamma}^2$ is also solved. We have $\phi = \frac{z_2}{z_1}$, $\gamma^2 = z_3 - \frac{z_2^2}{z_1}$, $\beta^2 = z_4$. Therefore, the unique solution is $\theta(s) = \Lambda(z_1, z_2, z_3, z_4) = (\frac{z_2}{z_1}, z_3 - \frac{z_2^2}{z_1}, z_4)$.

B.4 Particle Filter and Smoothing

B.4.1 Auxiliary Particle Filter

Auxiliary Particle Filter (APF) algorithm with the given parameter vector θ is listed as follows:

step 0 Draw N samples $x_0^{(i)}$ from the initial particle distribution and set $\omega_0^{(i)} = \frac{1}{N}$, for $i = 1, \ldots, N$.

step 1 For each time step, given the new observation y_{t+1} and recent particles $\{x_t^{(i)}, \omega_t^{(i)}\}_{i=1}^N$, $t = 0, 1, \ldots, T-1$, we need to

- Calculate the conditional expected value $\mu_{t+1}^{(i)} = E[x_{t+1}|x_t^{(i)}]$ for $i = 1, \ldots, N$.
- Calculate probabilities for each auxiliary index $p(k^{(i)}) \propto \omega_t^{(i)} g_\theta(y_{t+1}|\mu_{t+1}^{(i)})$ for $i = 1, \ldots, N$ and then normalize them to unity.
- (Re-sampling) Sample auxiliary indices $k^{(i)}$ according to $\{p(k^{(j)})\}_{j=1}^N$ and set $x_t^{(i)} = x_t^{k^{(i)}}, \mu_{t+1}^{(i)} = \mu_{t+1}^{k^{(i)}}$ for $i = 1, \ldots, N$.
- (Propagating) Draw $\{x_{t+1}^{(i)}\}$ by $f_\theta(\cdot|x_t^{(i)})$ for $i = 1, \ldots, N$.
- Update weights as $\omega_{t+1}^{(i)} \propto \dfrac{g(y_{t+1}|x_{t+1}^{(i)})}{g(y_{t+1}|\mu_{t+1}^{(i)})}$ and then normalize them to unity.

B.4.2 Particle Filtering together with Smoothing by Bootstrap Filter

The particle filtering along with smoothing algorithm using Bootstrap Filter with the given parameter vector θ is run as follows:

step 1 At time $t = 1$,

- Draw N samples $x_1^{(i)}$ from the initial particle distribution, $\mu(x_1)$, for $i = 1, \ldots, N$.
- Calculate and normalize particle weights,

  unnormalized weights: $\omega_1'(x_1^{(i)}) = g_\theta(y_1|x_1^{(i)})$, $i = 1, \ldots, N$,

  weight sum at time 1: $\text{ws}(1) = \sum_{i=1}^{N} \omega_1'(x_1^{(i)})$,

  normalized weights: $\omega_1^{(i)} = \frac{\omega_1'(x_1^{(i)})}{\text{ws}(1)}$, $i = 1, \ldots, N$.

step 2 At times $t = 2, \ldots, T$,

- Sample index $A_{t-1}^{(i)}$ based on normalized weights $\{\omega_{t-1}^{(k)}\}_{k=1}^{N}$ using multinomial or stratified re-sampling schemes for $i = 1, \ldots, N$.
- Sample $x_t^{(i)} \sim f_\theta(\cdot|x_{t-1}^{(A_{t-1}^{(i)})})$ and set $x_{1:t}^{(i)} = (x_{1:t-1}^{(A_{t-1}^{(i)})}, x_t^{(i)})$ for $i = 1, \ldots, N$.
- Calculate and normalize particle weights,

  unnormalized weights: $\omega_t'(x_{1:t}^{(i)}) = g_\theta(y_t|x_t^{(i)})$, $i = 1, \ldots, N$,

  weight sum at time t: $\text{ws}(t) = \sum_{i=1}^{N} \omega_t'(x_{1:t}^{(i)})$,

  normalized weights: $\omega_t^{(i)} = \dfrac{\omega_t'(x_{1:t}^{(i)})}{\text{ws}(t)}$, $i = 1, \ldots, N$.

This algorithm re-sample particles with their ancestors so that the smoothing process is also included; that is, the re-sampling step at each time t is conducted for the whole particle path $x_{1:t}$. At each time step t ($t < T$), $p_\theta(x_{1:t}|y_{1:t})$ is approximated by the particles with normalized weights as follows:

$$\hat{p}_\theta(x_t|y_{1:t}) = \hat{p}_\theta(x_{1:t}|y_{1:t}) = \sum_{i=1}^{N} \omega_t^{(i)} \delta_{x_{1:t}^{(i)}}(dx_{1:t}), \tag{B.4.1}$$

Therefore, at final time step T, the joint posterior density $p_\theta(x_{1:T}|y_{1:T})$ can be approximated by

$$\hat{p}_\theta(dx_{1:T}|y_{1:T}) = \sum_{i=1}^{N} \omega_T^{(i)} \delta_{x_{1:T}^{(i)}}(dx_{1:T}), \tag{B.4.2}$$

Actually, this algorithm also solves the smoothing problem. It is straightforward to approximate $p_\theta(x_s|y_{1:T}), 1 \leq s \leq T$ by marginalizing $\hat{p}_\theta(dx_{1:T}|y_{1:T})$, as in (B.4.3).

$$\hat{p}_\theta(dx_s|y_{1:T}) = \sum_{i=1}^{N} \omega_T^{(i)} \delta_{x_s^{(i)}}(dx_s), \tag{B.4.3}$$

where $x_s^{(i)}$ is the s-th element of the vector (or path) $x_{1:T}^{(i)}$. In addition, the volatility at time s ($s \leq T$) is approximated by

$$\mathbb{E}[\beta \exp(x_s/2)|y_{1:T}] = \beta \sum_{i=1}^{N} \omega_T^{(i)} \exp(\frac{x_s^{(i)}}{2}) \delta_{x_s^{(i)}}(dx_s). \tag{B.4.4}$$

However, when $T - s$ is very large, only few original particles at time step s will be kept at final time step. This is called the degeneracy problem. To alleviate such problem, we increase the number of particles to 10^5.

B.4.3 Particle-Based Log-Likelihood Computation

Given the approximated $p_\theta(x_{1:t}|y_{1:t})$ and $p_\theta(x_t)$ $(t \leq T)$, particle filtering provides a numerical solution for the likelihood estimation. Firstly, the marginal likelihood is approximated by

$$\hat{p}_\theta(y_{1:T}) = \hat{p}_\theta(y_1) \prod_{t=2}^{T} \hat{p}_\theta(y_t|y_{1:t-1}), \tag{B.4.5}$$

where $\hat{p}_\theta(y_1)$ and $\hat{p}_\theta(y_t|y_{1:t-1})$ are derived by the Bootstrap Filter algorithm as follows:

$$p_\theta(y_1) = \int p_\theta(y_1|x_1)\mu(x_1)dx_1$$

$$\approx \frac{1}{N}\sum_{i=1}^{N} g_\theta(y_1|x_1^{(i)}) = \frac{1}{N}\sum_{i=1}^{N} \omega_1'(x_1^{(i)}) = \frac{1}{N}\mathrm{ws}(1), \tag{B.4.6}$$

$$\begin{aligned} p_\theta(y_t|y_{1:t-1}) &= \int p_\theta(y_t, x_{1:t-1}, x_t|y_{1:t-1})dx_{1:t} \\ &= \int p_\theta(y_t|x_{1:t-1}, x_t, y_{1:t-1})p_\theta(x_t|x_{1:t-1}, y_{1:t-1})p_\theta(x_{1:t-1}|y_{1:t-1})dx_{1:t} \qquad \text{(B.4.7)} \\ &= \int g_\theta(y_t|x_t)f_\theta(x_t|x_{t-1})p_\theta(x_{1:t-1}|y_{1:t-1})dx_{1:t}, \qquad \text{(B.4.8)} \end{aligned}$$

The simplification from (B.4.7) to (B.4.8) results from the Markovian property of observation and transition processes in the state space model including the ARSV(1) model. Within the Bootstrap Filter algorithm, we generate $x_t \sim f_\theta(x_t|x_{t-1})$ while the previous re-sampling step is conducted based on $p_\theta(x_{1:t-1}|y_{1:t-1})$, then we have $x_{1:t}|y_{1:t-1} \sim f_\theta(x_t|x_{t-1})p_\theta(x_{1:t-1}|y_{1:t-1})$. Consequently, (B.4.8) can be approximated by

$$p_\theta(y_t|y_{1:t-1}) \approx \frac{1}{N}\sum_{i=1}^{N} g_\theta(y_t|x_t^{(i)}) = \frac{1}{N}\sum_{i=1}^{N} \omega_t'(x_{1:t}^{(i)}) = \frac{1}{N}\mathrm{ws}(t), \quad 1 < t \le T. \tag{B.4.9}$$

As a whole, the particle-based computations of the marginal likelihood and log likelihood are approximated by

$$\begin{aligned} \hat{p}_\theta(y_{1:T}) &= \frac{1}{N}\mathrm{ws}(1)\prod_{t=2}^{T}\frac{1}{N}\mathrm{ws}(t), \\ \ln\hat{p}_\theta(y_{1:T}) &= -T\ln N + \sum_{t=1}^{T}\ln\mathrm{ws}(t). \end{aligned} \tag{B.4.10}$$

B.4.4 Particle-Based One-Step Ahead Prediction by Bootstrap Filter

Under the framework of the state space model, the one-step-ahead prediction density $p_\theta(x_{t+1}|y_{1:t})$ is given by

$$\begin{aligned} p_\theta(x_{t+1}|y_{1:t}) &= \int p_\theta(x_{t+1}, x_{1:t}|y_{1:t})dx_{1:t} = \int p_\theta(x_{t+1}|x_{1:t}, y_{1:t})p_\theta(x_{1:t}|y_{1:t})dx_{1:t} \\ &= \int f_\theta(x_{t+1}|x_t)p_\theta(x_{1:t}|y_{1:t})dx_{1:t}, \end{aligned} \tag{B.4.11}$$

Revisiting the particle filtering along with smoothing algorithm using Bootstrap Filter in B.4.2, if the weighted particles $\{x_{1:t}^{(i)}, \omega_t^{(i)}\}_{i=1}^N$ are obtained at time t, we have

$$p_\theta(x_{t+1}|y_{1:t}) \approx \sum_{i=1}^{N} f_\theta(x_{t+1}|x_t^{(i)})\omega_t^{(i)} = \sum_{i=1}^{N} f_\theta(x_{t+1}|x_t^{(A_t^{(i)})}). \tag{B.4.12}$$

where $A_t^{(i)}$ is the re-sampling index at $t+1$. Therefore, the particle-based one-step-ahead prediction by Bootstrap Filter can be summarized as follows:

step 1 Obtain $\{x_{1:T}^{(i)}, \omega_T^{(i)}\}_{i=1}^N$ from the particle filtering together with smoothing by bootstrap filter at the last in-sample step T.

step 2 At times $t = T+1, \ldots, T+T'$,

- Sample index $A_{t-1}^{(i)}$ based on normalized weights $\{\omega_{t-1}^{(k)}\}_{k=1}^N$ using multinomial or stratified re-sampling schemes for $i = 1, \ldots, N$.
- Sample $x_t^{(i)} \sim f_\theta(\cdot|x_{t-1}^{(A_{t-1}^{(i)})})$ and set $x_{1:t}^{(i)} = (x_{1:t-1}^{(A_{t-1}^{(i)})}, x_t^{(i)})$ for $i = 1, \ldots, N$.
- One-step-ahead prediction, conditional variance forecast at t: $\frac{\beta^2}{N}\sum_{i=1}^N \exp(x_t^{(i)})$, volatility forecast at t (if needed): $\frac{\beta}{N}\sum_{i=1}^N \exp(x_t^{(i)}/2)$.
- Capture the new observation y_t'. Calculate and normalize particle weights,

  unnormalized weights: $\omega_t'(x_{1:t}^{(i)}) = g_\theta(y_t'|x_t^{(i)})$, $i = 1, \ldots, N$,

  weight sum at time t: $\text{ws}(t) = \sum_{i=1}^N \omega_t'(x_{1:t}^{(i)})$,

  normalized weights: $\omega_t^{(i)} = \dfrac{\omega_t'(x_{1:t}^{(i)})}{\text{ws}(t)}$, $i = 1, \ldots, N$.

The size of the out-of-sample dataset is T'. The first step above is to obtain the weighted particles at the last in-sample step. Since the out-of-sample dataset follows the in-sample one without any gap, the first step of the out-of-sample observations is on $T+1$. If there is a gap between the in-sample and out-of-sample datasets, the algorithm above should be adjusted.