A Flexible Inference Method for an Autoregressive Stochastic Volatility Model with an Application to Risk Management

by

Yijun Xie

B.Sc., University of Notre Dame du Lac, 2015

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

Master of Science

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES
(Statistics)

The University of British Columbia
(Vancouver)

April 2017

© Yijun Xie, 2017

Abstract

The Autoregressive Stochastic Volatility (ARSV) model is a discrete-time stochastic volatility model that can model the financial returns time series and volatilities. This model is relevant for risk management. However, existing inference methods have various limitations on model assumptions. In this report we discuss a new inference method that allows flexible model assumption for innovation of the ARSV model. We also present the application of ARSV model to risk management, and compare the ARSV model with another commonly used model for financial time series, namely the GARCH model.

Preface

This dissertation is original, unpublished work by the author, Yijun Xie. The dataset used in this thesis is downloaded from Yahoo Finance (https://finance.yahoo.com/).

Table of Contents

List of Tables

List of Figures

List of Algorithms

Acknowledgments

First and foremost, I would like express my sincere gratitude to my supervisor, Professor Natalia Nolde, for her patience, insightfulness, and immense knowledge. Without her guidance I would never be able to finish this thesis. It is my fortune and honor to work under her tutelage.

Besides my advisor, I would like to thank the Department of Statistics at UBC for the supportive environment. Thank you to everyone in my office, for the office dinners and all those wonderful time we spent together. Thank you to Ho Yin Ho and Tingting Zhao for your help and great advice when I was writing the thesis and being my lunch/dinner buddies. Special thanks to my best friend Mengtian Zhao and her cat King for being my emotional support every time when I feel depressed, even though you are literally 2000 miles away.

My sincere thanks also goes to my second reader Professor Harry Joe, who generously spent time reading and critiquing my works.

Last but not least, I would like to thank my parents, Zhaoliang Xie and Huaping Xie, for their unconditional love to me.

Chapter 1

Introduction

Modeling financial returns time series data is an essential component in a wide range of problems in finance, including risk management, portfolio optimization, pricing and hedging of financial risks. In order to develop a realistic model for time series of financial returns, it is important to first identify key features that such data tend to exhibit. These key features are usually referred to as the stylized facts (McNeil, Frey, and Embrechts, 2015a). The most widely acknowledged among these stylized facts are the following:

1. The conditional expectation of financial returns is close to zero (see Fig. 1.1 (left panel));

2. The conditional standard deviation, known as the volatility, varies over time and with large values having a tendency to cluster (see Fig. 1.1 (left panel));

3. The marginal distribution of return series has heavier tail than that of the normal distribution (See Fig. 1.1 (right panel));

4. There is little serial correlation between returns (see Fig. 1.2 (left panel)). However, the squared or absolute values of returns show strong correlation (see Fig. 1.2 (right panel)).

As the conditional expectation is usually close to zero, it is the volatility that has the dominant effect on the dynamics of financial return series. For this reason, the models for financial returns are often called volatility models. Several types

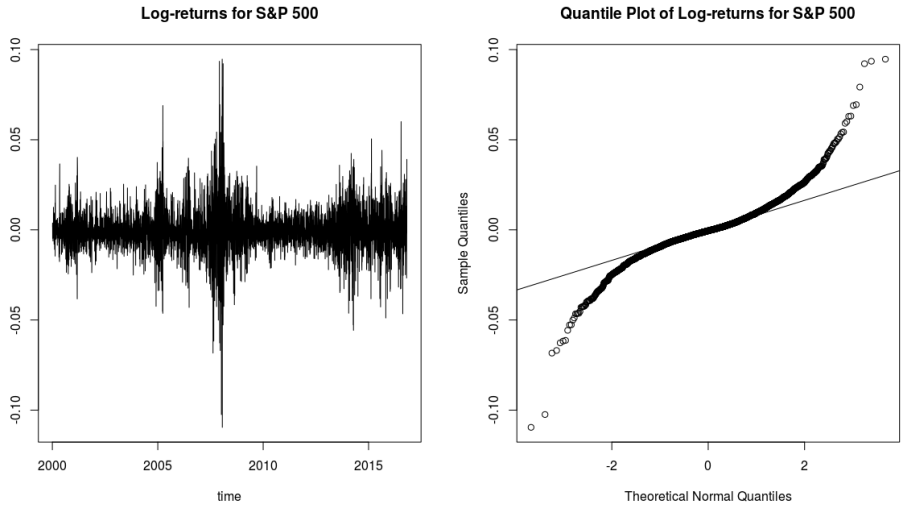

Figure 1.1: The left panel shows the daily log-returns for the S&P 500 Index from January 1, 2000 to October 26, 2016. Large values of log-returns are clustered. The right panel shows the normal quantile plot of this log-return series, which has heavier tails than that of a normal distribution.

of volatility models have been proposed in the literature. Among these models, arguably the most famous and widely used one is the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) process (Engle (1982), Bollerslev (1986)). However, in this report, we discuss another type of a volatility model called Autoregressive Stochastic Volatility (ARSV) process. Both of these processes can capture the above mentioned stylized facts, but they exhibit different extremal dependence properties for consecutive observations. The returns modeled by GARCH process have the property of tail dependence, while those modeled by ARSV process are tail independent. Detailed discussion of tail dependence and independence as well as these two volatility models is provided in Chapter 2.

There is empirical evidence (Drees, Segers, and Warchoł, 2015) suggesting that for some financial time series, the returns over consecutive periods are likely to be tail independent. That is, extreme values at consecutive time points are independent and hence will tend not to co-occur. However, there should be a stronger dependence among large but not extremal observations than the classic ARSV model

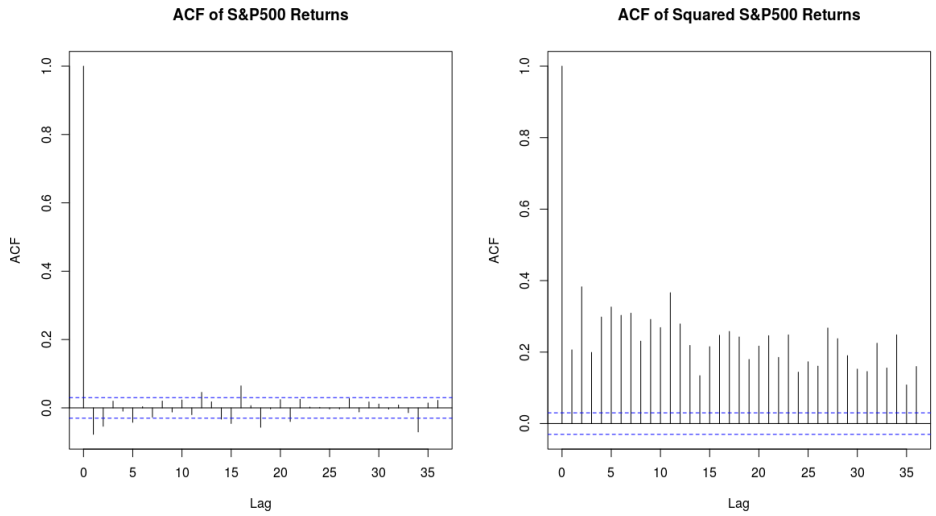

Figure 1.2: The left and right panel show the autocorrelation function (ACF) for the log-return series and squared log-returns of the same dataset as in Figure 1.1.

implies. Janssen and Drees (2016) propose a variation of the classic ARSV model, and they show that this new class of ARSV model which has a special form of the heavy-tailed second innovation, has stronger tail dependence than classic ARSV model, while remains tail independent for extremal observations.

Inspired by their ideas, in this report we consider another extension of the classic ARSV model, which also has a heavy-tailed second innovation and light-tailed first innovation. We conjecture that this model is also tail independent for extremal observations, but has stronger dependence for sub-extremal observations than the classic ARSV model. The conjecture is supported via simulations.

However, the inference for ARSV models is notoriously difficult. Most of the current inference methods for ARSV models require the assumption that both innovations have normal distributions. Therefore, they can only be applied to the classic ARSV models but not to the extended model we want to focus on in this report. We propose a new inference method for ARSV models which allows arbitrary choices of distributions of both innovations. We show that this new inference method works as good as traditional methods for the inference of the classic ARSV

model, while it can accurately estimate parameters when traditional methods fail. This new approach is discussed in Chapter 3.

One of the most important applications for volatility models is to measure risk for risk management purposes. It is also used by financial market regulators to set capital requirements for financial institutes. A correctly specified and flexible volatility model can help both practitioners and regulators to accurately capture features of the market data, and hence to measure risk more precisely. This is paramountly important for the stability of the financial system. In Chapter 4 we discuss two risk measures: Value-at-Risk (VaR) and Conditional Value-at-Risk (CoVaR). The estimation methods for VaR under the GARCH model are already well studied. In this report, we develop estimation methods for CoVaR under the GARCH model as well as for VaR and CoVaR under the proposed ARSV model. We then compare risk measure estimates under the GARCH and ARSV models.

In Chapter 5, we provide discussion of the results and an outlook for future research.

Chapter 2

Background

In this chapter, we first discuss the property of tail dependence. We also briefly introduce two volatility models, namely the GARCH process and ARSV process. Our project is motivated by the different tail dependence properties between these two models.

2.1 Measures of Tail Dependence

A question for the practitioners of risk management would be that given a large loss today, how likely it is to experience a large loss tomorrow or h days into the future.

To answer this question, we need to understand the idea of tail dependence and tail independence. Different models for financial time series would have difference tail dependence properties, and it is important to choose the model with the tail dependence property close to that of the real data.

Consider two random variables Y and Z with continuous distribution functions F_Y and F_Z, respectively. The upper tail dependent coefficient between Y and Z is defined as (Joe, 1997)

$$\begin{aligned}\chi &= \lim_{u\to 1} \mathbb{P}(Z > F_Z^{-1}(u) | Y > F_Y^{-1}(u)) \\ &= \lim_{u\to 1} \frac{\mathbb{P}(Z > F_Z^{-1}(u), Y > F_Y^{-1}(u))}{\mathbb{P}(Y > F_Y^{-1}(u))},\end{aligned} \tag{2.1}$$

provided the limit exits. It is the limit of the ratio between the probability of the occurrence of jointly extremal observations of Y and Z and the probability of the occurrence of an extreme observation in one of the variables. The lower tail dependence coefficient can be defined similarly.

If $\chi = 0$, then Y and Z are said to be tail independent. On the other hand, when $0 < \chi \leq 1$, Y and Z are called (upper) tail dependent. The value of χ measures how strong the tail dependence is.

However, χ only provides us information about how strong the tail dependence is when it is greater than zero. When $\chi = 0$, we cannot tell how fast χ approaches zero as $u \to 1$. We need a more refined measure to tell us the speed of χ approaching zero.

First, we call a function L on $(0,\infty)$ slowly varying at ∞ if

$$\lim_{x\to\infty} \frac{L(tx)}{L(x)} = 1, \qquad t > 0,$$

while a function h on $(0,\infty)$ is called regularly varying at ∞ with index p if

$$\lim_{x\to\infty} \frac{h(tx)}{h(x)} = t^p, \qquad t > 0.$$

Let us define $p(t) = \mathbb{P}(Z > F_Z^{-1}(1-1/t), Y > F_Y^{-1}(1-1/t))$, and suppose $p(t)$ is regularly varying with index $-1/\eta$ for some $\eta \in (0,1]$. Then η is called the residual tail dependence coefficient (Ledford and Tawn, 1996).

Ledford and Tawn (1996) show that if $\eta < 1$, then $\chi = 0$ and hence Y and Z are tail independent, and η measures the speed of convergence to tail independence.

We can further define (Coles, Heffernan, and Tawn, 1999)

$$\bar{\chi} = 2\eta - 1,$$

where $-1 < \bar{\chi} \leq 1$. When $1/2 < \eta \leq 1$ or $0 < \bar{\chi} \leq 1$, Y and Z are non-negatively dependent, and when $\eta = 1/2$ or $\bar{\chi} = 0$, Y and Z are exactly tail independent (Coles, Heffernan, and Tawn, 1999; Ledford and Tawn, 2003).

Now we have the pair $(\chi, \bar{\chi})$ to help us describe the extremal dependence. When $\chi > 0$ and $\bar{\chi} = 1$ the two random variables are tail dependent, and the value

of χ measures the strength of the dependence of this pair. If $\chi = 0$ and $-1 < \bar{\chi} < 1$, the two random variables are tail independent, and the value of $\bar{\chi}$ measures the strength of dependence of this pair.

Understanding the type of tail dependence structure is very important for choosing good models. Let $\{X_t\}$ denote a process of financial log-returns with losses in the upper tail. As discussed in Chapter 1, financial data often suggests tail independence between consecutive returns. In other words, although high volatilities tend to be persistent, extremal levels of returns should be tail independent (Laurini and Tawn, 2008) so that $\mathbb{P}(X_t > F_{X_t}^{-1}(u)|X_{t-1} > F_{X_{t-1}}^{-1}(u)) \to 0$ as $u \to 1$. A tail dependent process may lead to an overestimation of the potential loss. On the other hand, Janssen and Drees (2016) suggest that a good model for returns should be tail independent, while retaining stronger tail dependence at sub-extremal levels than the exactly tail independent case. In other words, we are looking for a stochastic volatility process such that that for the pair (X_{t-1}, X_t), $\chi = 0$ and $0 < \bar{\chi} < 1$.

2.2 Volatility Models

2.2.1 GARCH Process

One of the most commonly discussed volatility models is the generalized autoregressive conditionally heteroscedastic (GARCH) model. GARCH family was first proposed by Engle (1982) and Bollerslev (1986). A GARCH(p,q) process for $\{X_t, t = 1, 2, ..., T\}$ is defined as follows:

$$X_t = \sigma_t \varepsilon_t$$
$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i \sigma_{t-i}^2 + \sum_{j=1}^{q} \beta_j X_{t-j}^2$$

where $\{\varepsilon_t\}$ is a strict white noise process with zero mean and unit variance, $\alpha_0 > 0$, $\alpha_i \geq 0, i = 1, 2, ..., p$, and $\beta_j \geq 0, j = 1, 2, ...q$. In order to achieve stationarity, we also assume $\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1$. In financial statistics, X_t is usually the return on the financial asset observed at time t and σ_t is the conditional standard deviation, or the volatility, at time t, which generally cannot be directly observed.

Let $\mathscr{F}_t$ denote the sigma algebra of all available information up to time t. Then,

for the GARCH model, volatility σ_t is $\mathscr{F}_{t-1}$ measurable. In other words, given the information up to $t-1$, σ_t is known.

The empirical evidence suggests that volatilities are clustered: when high volatility occurs at time *t-1*, it would be more likely to have large volatility again at time *t*. For a GARCH(p,q) model, the conditional squared volatility is based on a linear combination of previous squared returns and squared volatilities, so a large absolute value in $X_{t-h}, h = 1,2,...q$ or a large value of $\sigma_{t-h}, h = 1,2,...p$ lead to a large value of σ_t. Therefore, we can observe the clustered behavior of large volatility from a GARCH model. For simplicity, we only consider the GARCH(1,1) process, which in practice is believed to be sufficient to model the volatility process most of the time.

Mikosch and Starica (2000) and Basrak, Davis, and Mikosch (2002) show that for GARCH models,

$$\lim_{x\to\infty} \mathbb{P}(X_h > x | X_0 > x) > 0,$$

i.e., $\eta = 1$ or (X_h, X_0) are tail dependent. Laurini and Tawn (2008) suggest declustering returns over high threshold to remove the tail dependence. However, we can also look for an alternative model such that the series of $\{X_t\}$ is tail independent while volatility clustering is preserved.

2.2.2 ARSV Process

An alternative to the GARCH model is the autoregressive stochastic volatility model (ARSV). This model is first studied by Harvey, Ruiz, and Shephard (1994) and Jacquier, Polson, and Rossi (1994), among many others. We focus on the simplest ARSV(1) model, which is defined as

$$\begin{aligned} X_t &= \sigma_t \varepsilon_t \\ \log(\sigma_t^2) &= \beta_0 + \beta_1 \log(\sigma_{t-1}^2) + \delta \eta_t \end{aligned}$$

where $\{\varepsilon_t\}$ and $\{\eta_t\}$ are two independent strict white noise processes with zero mean and unit variance, and $\delta > 0$ is a constant to adjust the standard deviation of the second innovation. β_0 and β_1 are coefficients for the log-volatility process. For ARSV(1) model, $0 < \beta_1 < 1$ is required for the process to be stationary.

The difference between a GARCH model and an ARSV model lies in the volatility process. For a GARCH model, σ_t is $\mathscr{F}_{t-1}$ measurable as mentioned above. However, the volatility process of an ARSV model contains a second innovation term η_t. Thus, after conditioning on all information up to $t-1$, σ_t is still a random variable.

The tail dependence properties of ARSV(1) model are studied by Breidt and Davis (1998) and Hill (2011). They show that for either normally distributed or heavy-tailed ε_t, the extremes of X_t are independent. Liu and Tawn (2013) suggest the difference comes from the source of clustering. For the GARCH model, the components of previous volatilities have negligible effect on the current volatility when α_i's $(i \geq 1)$ are small. However, the return process and volatility process are interconnected. Therefore, extremal return observed at time $t-1$ will lead to large volatility value at time t, and hence a large probability of observing an extremal return at time t. This is illustrated below.

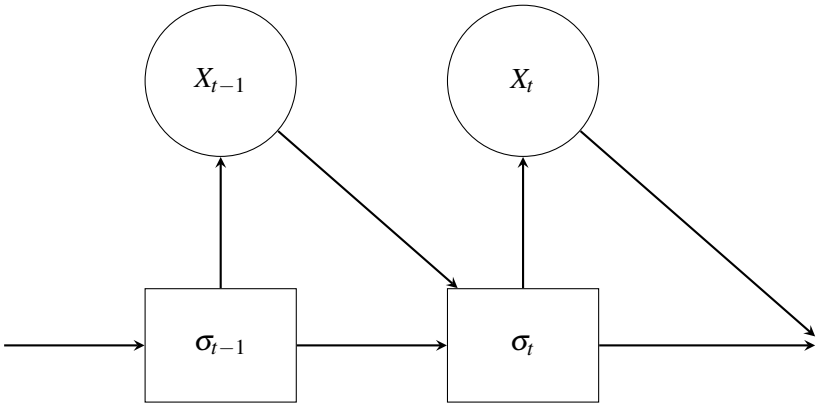

Figure 2.1: Structure of GARCH process.

However, for the ARSV model, the process of the log-volatility is independent from the observed values of return process, while the return is the realization of current volatility times a noise term. Therefore, X_{t-1} and σ_t are independent given σ_{t-1}, and a large volatility value at time $t-1$ does not necessarily lead to a large return value at time t.

The tail behaviors of GARCH process and ARSV process can also be illustrated with the residual tail dependence coefficient η, or measures of tail depen-

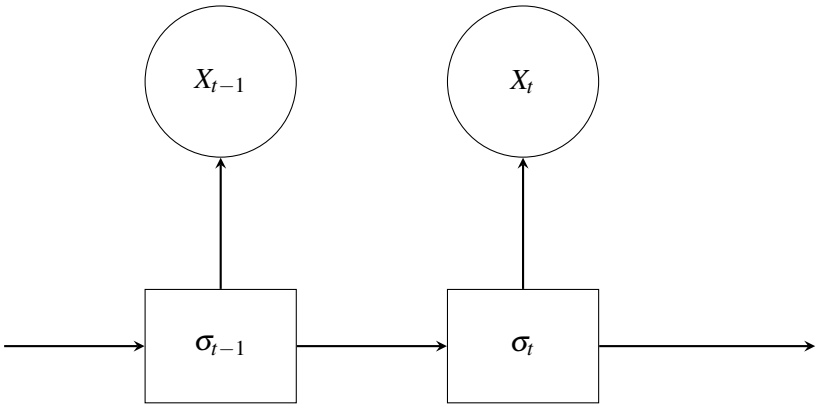

Figure 2.2: Structure of ARSV process.

dence χ and $\bar{\chi}$ discussed in Section 2.1. In Figure 2.3 we present the estimate of residual tail dependence coefficient, η, for the GARCH(1,1) process with $\alpha_0 = 1 \times 10^{-6}$, $\alpha_1 = 0.8$, and $\beta_1 = 0.1$; classic ARSV(1) process with $\beta_0 = -0.5$, $\beta_1 = 0.95$, and $\delta = 0.35$; and an extension of the classic ARSV(1) process with $\beta_0 = -0.5$, $\beta_1 = 0.95$, $\delta = 0.35$, and a standardized second innovation η_t which is Student's t distributed with 5 degress of freedom. The corresponding values of the $(\chi, \bar{\chi})$ pairs [1] are presented in Figure 2.4. Janssen and Drees (2016) suggest that an ARSV(1) process with heavy-tailed second innovation distribution has a stronger tail dependence at sub-extremal levels than classic ARSV model while remains tail independent. Figure 2.3 and Figure 2.4 both support this suggestion. As we can see from these two plots, when the quantile approaches 1 we have evidence to support the claim $\eta = 1$ (*i.e.* tail dependent), $\eta = 0.5$ (*i.e.* exactly tail independent), and $0.5 < \eta < 1$ (*i.e.* tail independent but not exactly tail independent) respectively for the three scenarios. Similarly in Figure 2.4, for the GARCH(1,1) process, when the quantile approaches 1 the empirical value of χ is significantly larger than 0, which implies that (X_{t-1}, X_t) are tail dependent. For the classic ARSV(1) process, both empirical values of χ and $\bar{\chi}$ are not significantly larger than 0, which implies that (X_{t-1}, X_t) are exactly tail independent. For the extended ARSV(1) processes, the empirical value of χ is not significantly larger than 0 but the empirical value

[1] Here the empirical values of χ and $\bar{\chi}$ are estimated using the `evd` package, which calculates both values using approximation method.

of $\bar{\chi}$ is significantly larger than 0, which implies that (X_{t-1}, X_t) have stronger tail dependence than product of margins.

We further compare the simulated data with an actual financial time series. We estimate volatilities of the S&P 500 Index from 2000-01-03 to 2016-10-26 with some unbiased estimators, and then fit the log squared volatilities to an AR(1) process. The top left panel of Figure 2.5 illustrates the quantile plot of residuals of this AR(1) process against a normal distribution, and the top right panel illustrates the quantile plot against a Student's t distribution with 3 degrees of freedom. These two plots hint that a heavy-tailed second innovation η_t might be a more suitable choice. Also, we present the $\chi/\bar{\chi}$ plots of the negative returns of this dataset. Comparing the bottom two panels of Figure 2.5 with all panels in Figure 2.4, we can also conjecture that the $\chi/\bar{\chi}$ plots from extended ARSV(1) process with the second innovation Student's t distributed are the closest ones to those from the real data. Thus we might prefer to model the financial return series using the extended ARSV(1) model.

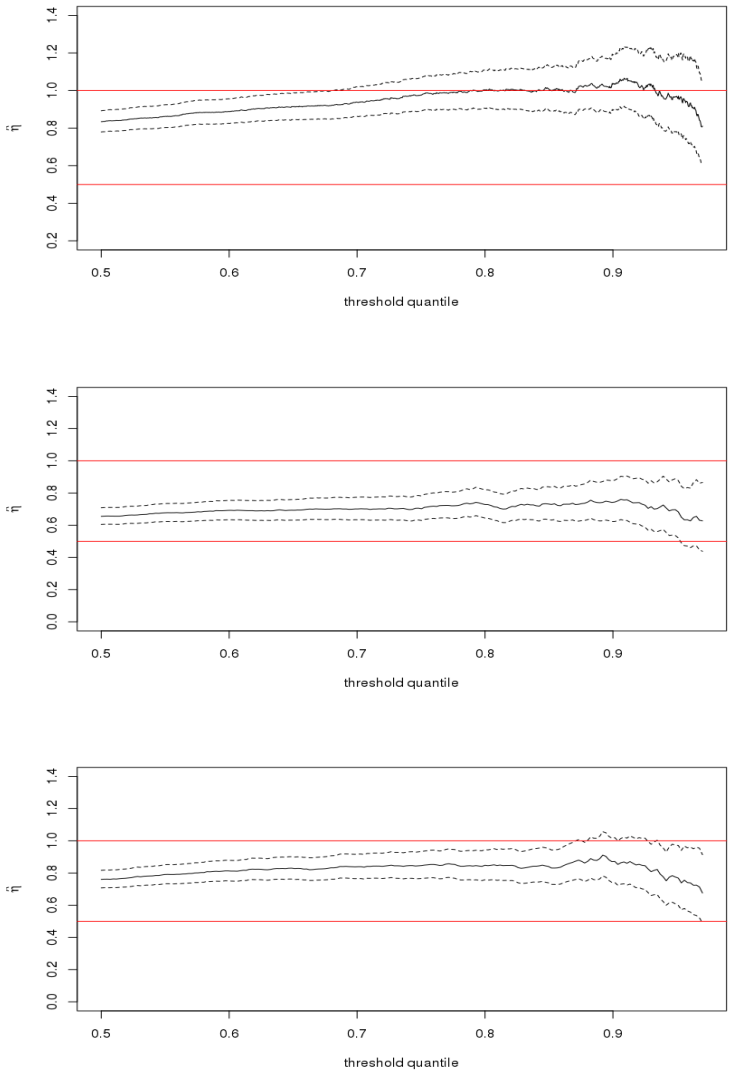

Figure 2.3: Plot of η for a simulated GARCH(1,1) process (top panel), a simulated classic ARSV(1) process (middle panel), and a simulated extended ARSV(1) process with first innovation distribution normal and second innovation distribution Student's t (bottom panel).

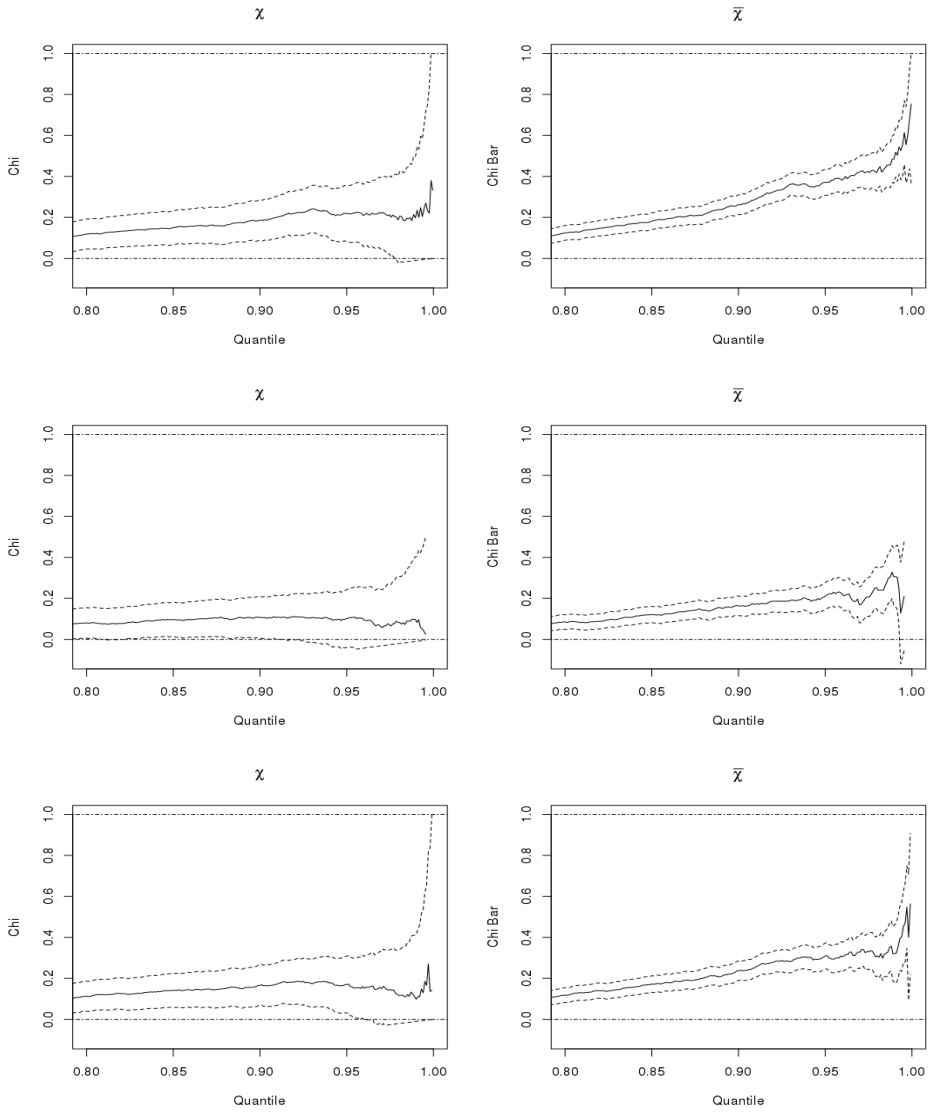

Figure 2.4: $\chi/\bar{\chi}$-plot for a simulated GARCH(1,1) process (top panels), a simulated classic ARSV(1) process (middle panels), and a simulated extended ARSV(1) process with first innovation distribution normal and second innovation distribution Student's t (bottom panels).

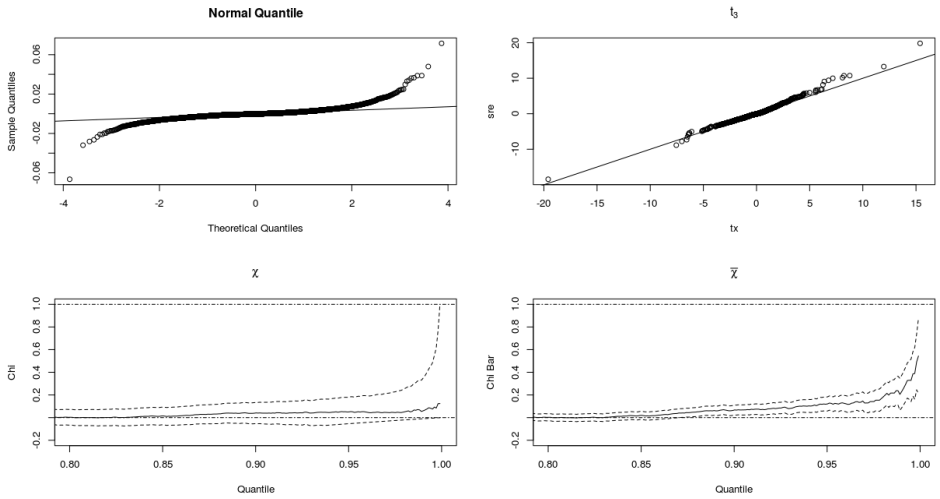

Figure 2.5: Quantile plots and $\chi/\bar{\chi}$ plots of the negative daily log-returns of S&P 500 Index from 2000-01-03 to 2016-10-26.

Chapter 3

Inference for Autoregressive Stochastic Volatility Model

Many efforts have been made to develop inference methods for the traditional ARSV model with both innovation distributions as normal. In this section, we first briefly review the methods that inspire our new inference methods.

However, these existing methods cannot be used to estimate parameters for the extension of the ARSV model discussed at the end of last section. Therefore, we propose a new approach that allows flexible choices of both innovation distributions of the ARSV model. This new method can work as well as the existing methods for the traditional ARSV model, and can provide good parameter estimates when existing methods fail. We describe the details of the new method in this section. We also compare the results from this new method with those from the existing inference method.

3.1 Review of Existing Methods

In this section we review inference methods implementing the full Bayesian approach with pre-specified prior distributions. Jacquier, Polson, and Rossi (1994) propose a cyclic MCMC approach for the classic ARSV(1) model with both inno-

vations normally distributed. Consider the model

$$\begin{aligned} X_t &= \sigma_t \varepsilon_t \\ \log \sigma_t^2 &= \beta_0 + \beta_1 \log \sigma_{t-1}^2 + \delta \eta_t \\ (\varepsilon_t, \eta_t) &\sim \mathrm{N}(0, I_2), \end{aligned}$$

where I_2 is a two-dimensional identity matrix. Let ω denote the set of parameters $(\beta_0, \beta_1, \delta)$. An inverse gamma distribution $p(\delta) \propto \exp(-\nu_0 s_0^2/2\delta^2)/\delta^{\nu_0+1}$ is chosen as the prior distribution for δ, where (ν_0, s_0) are two hyperparameters. The prior distributions of $\beta_0 \sim \mathrm{N}(0, 100)$ and $\beta_1 \sim \mathrm{N}(0, 10)$ are independent and essentially flat. The algorithm proposed by Jacquier, Polson, and Rossi (1994) includes two stages:

1. Sample parameters ω: $p(\omega | \log \sigma_t^2)$ is the posterior from a linear regression, and therefore a direct draws can be made.

2. Sample σ_t^2 from

$$\begin{aligned} p(\sigma_t^2 | \sigma_{t-1}^2, \sigma_{t+1}^2, \omega, x_t) &\propto f_X(x_t | \sigma_t^2) f_a(\sigma_t^2 | \sigma_{t-1}^2) f_b(\sigma_{t+1}^2 | \sigma_t^2) \\ &\propto \frac{1}{\sigma_t} \exp\Big(\frac{-x_t^2}{2\sigma_t^2}\Big) \times \frac{1}{\sigma_t^2} \exp\Big(\frac{-(\log \sigma_t^2 - \mu_t)^2}{2\delta^2}\Big), \end{aligned}$$

   where $\mu_t = (\beta_0(1-\beta_1) + \beta_1(\log \sigma_{t+1} + \log \sigma_{t-1}^2))/(1+\beta_1^2)$, and $\delta^2 = u_\eta^2/(1+\beta_1^2)$; f_X is the probability density function of X_t conditioned on σ_t^2; f_a and f_b are probability density functions of σ_t^2 conditioned on σ_{t-1}^2 and σ_{t-1}^2 conditioned on σ_t^2 respectively; π is the posterior density function of σ_t.

Jacquier, Polson, and Rossi (2004) extend the method above to deal with the case where the first innovation ε_t follows a Student t distribution with ν degrees of freedom. They treat the heavy-tailed ε_t as a scale mixture of the inverse gamma distribution and normal distribution. This allows us to write X_t as $X_t = \sigma_t \sqrt{\lambda_t} Z_t$, where Z_t follows the standard normal distribution and λ_t follows an inverse gamma distribution $IG(\nu/2, 2/\nu)$, or $\nu/\lambda_t \sim \chi_\nu^2$. Let us denote $X_t^* = X_t/\sqrt{\lambda_t}$. Then we can replace $\{x_t, t = 1, 2, ...T\}$ with $\{x_t^*, t = 1, 2, ...T\}$, and sample σ^2 and ω in the similar way as described above. Jacquier, Polson, and Rossi (2004) choose a

uniform discrete prior on $[3,40]$ for ν, and prior distributions of other parameters are the same as above. The algorithm includes three stages:

1. Sample from the posterior distribution $p(\sigma^2, \omega|\lambda, \nu, x^*)$ using the algorithm in Jacquier, Polson, and Rossi (1994).

2. Sample λ from the posterior distribution

$$p(\lambda_t|x_t, \sigma_t^2, \nu) \sim IG\Big(\frac{\nu+1}{2}, \frac{2}{(x_t^{*2}/\log\sigma_t^2)+\nu}\Big).$$

3. Sample ν from the posterior distribution

$$p(\nu|\sigma^2, x, \omega) = p(\nu)\prod_{t=1}^{T}\frac{\nu^{\nu/2}\Gamma(\nu+1/2)}{\Gamma(\nu/2)\Gamma(1/2)}(\nu+x_t^2/\sigma_t^2)^{-(\nu+1)/2},$$

   where $p(\nu)$ is the prior distribution of ν.

Kastner and Frühwirth-Schnatter (2014) propose another method based on $\log X_t^2$ instead of X_t. By taking square and then logarithm on X_t, we have

$$\log X_t^2 = \log\sigma_t^2 + \log\varepsilon_t^2,$$

where $\log\varepsilon_t^2$ can be approximated by a mixture of normal distributions. They further implement the ancillary-sufficiency interweaving strategy to sample in the $[\log X^2, \log\sigma^2, \omega]$ space. This method is implemented in the **stochvol** package.

Broto and Ruiz (2004) provide a more comprehensive and detailed review on other approaches for the classic ARSV model.

3.2 New Inference Method

The methods discussed in the preceding section assume that the second innovation distribution in the ARSV model is normal, and the first innovation distribution can only be normal or Student t. However, when the second innovation in the ARSV model is non-Gaussian, it would be very difficult to derive the corresponding posterior distribution, and hence very difficult to sample from the posterior distribution. In order to make model inference for the extension of the classic ARSV model we

discussed earlier, we propose a new inference method which does not require presepcified prior distributions, and hence does not rely on the normality of innovation distributions.

For the rest of the report, we consider the model

$$X_t = \sigma_t \varepsilon_t$$
$$\log \sigma_t^2 = \beta_0 + \beta_1 \log \sigma_{t-1}^2 + \delta \eta_t,$$

where ε_t follows a standard normal distribution and η_t follows a standardized Student t distribution with zero mean, unit variance, and ν_η degrees of freedom.

To start, we first need to make an initial guess of the volatilities and parameters. Popular model-free methods for volatilities include a simple moving average of squared returns, or the Exponentially Weighted Moving Average (EWMA) among many others. Let us define $h_t := \log \sigma_t^2$, $t = 1, 2, ...T$. With estimated volatilities $\{\sigma_0^{(0)}, ..., \sigma_T^{(0)}\}$, we can calculate h_t, $t = 1, ..., T$ and have an initial guess of parameters $\theta^{(1)}$ by fitting an AR process as described in Section 3.2.4. In this report, we choose the 5-day moving average of squared returns as our initial guess of volatility.

Assume that after the $(i-1)^{th}$ iteration, we have sampled the sequence of $h_t^{(i-1)}$, $t = 1, ..., T$ and have updated parameters $\theta^{(i)} = [\beta_0^{(i)}, \beta_1^{(i)}, \delta^{(i)}, \nu_\eta^{(i)}]$ in the way discussed in Section 3.2.4. Then in the i^{th} iteration, we sample $\{h_t^{(i)}, t = 2, ..., T-1\}$ in the way described in the following sections.

3.2.1 Proposing Step

The sequence $\{h_t^{(i)}, t = 2, ..., T-1\}$ is sampled sequentially. Suppose that after the $(t-1)^{th}$ step in the i^{th} iteration, we have already sampled $h_{t-1}^{(i)}$. Since the volatility process is independent from the observed returns, the distribution of h_t will only depend on the value of h_{t-1} and the parameters $\theta = [\beta_0, \beta_1, \delta, \nu_\eta]$. Therefore, by plugging in the estimated parameters $\theta^{(i)} = [\beta_0^{(i)}, \beta_1^{(i)}, \delta^{(i)}, \nu_\eta^{(i)}]$ and the sampled $h_{t-1}^{(i)}$ from the previous step, we can sample $h_t^{'}$ from the Student t distribution with mean $\beta_0^{(i)} + \beta_1^{(i)} h_{t-1}^{(i)}$, standard deviation $\delta^{(i)}$, and degrees of freedom $\nu_\eta^{(i)}$. We can continue this sequence by sampling $h_{t+1}^{(i)}$ based on $h_t^{(i)}$ and $\theta^{(i)}$ for all $t = 2, 3, ...T-1$.

This step is similar to the commonly used technique in MCMC called Gibbs sampler. However, Gibbs sampler is applied more often in the scenario when we are only dealing with a smaller sampling space. Our case is different from the regular setting as we want to sample a sequence of $T-2$ random variables, where the T can be as large as 1000 to 2000. Also, h_t only depends on h_{t-1} and this process can be analogous to a random walk process. Therefore, the classic Gibbs sampling method would be very inefficient in this case, as it would take a very long time to explore all regions with high probability. In Figure 3.1 we show the values of the estimated parameter $\hat{\beta}_0$ in each iteration when we only use the classic Gibbs sampler with the parameters estimation method discussed in Section 3.2.4. The true value is -0.5. However, we can see that the estimated values do not approach the true value even after nearly 2000 iterations.

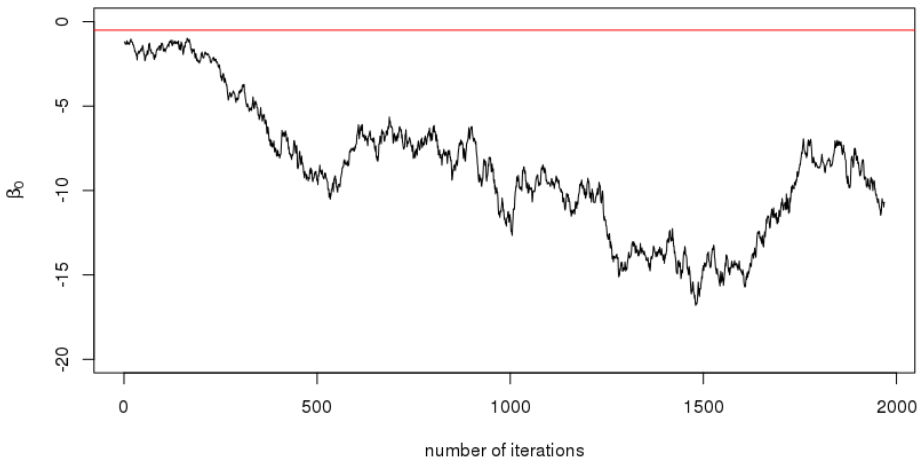

Figure 3.1: Path of β_0 estimates when only implementing Gibbs sampler. The red line represents the true value.

3.2.2 Classic Metropolis-Hastings Algorithm

To overcome the problem of slow convergence, we need to regulate the acceptance rate of each new sample. Instead of accepting all newly proposed samples, we can

adopt the Metropolis-Hastings' algorithm to adjust our acceptance rate. For the rest of this section, we consider sampling a generic sequence $x^{(i)}$ from distribution π, where generally the functional form of π is known except for a normalizing constant. The classic Metropolis-Hastings' algorithm contains the following three steps ((Murphy, 2012)):

First, in the i^{th} iteration, we sample a new sample x' from a proposal distribution $q(x^{(i)}|x^{(i-1)}, \mathscr{D})$, where $x^{(i-1)}$ is the value we obtained in the previous iteration, and $\mathscr{D}$ is the set of parameters.

Second, we calculate the acceptance rate, r, which is found by

$$\begin{aligned} r &= \min(1, \alpha), \\ \text{with } \alpha &= \frac{q(x^{(i-1)}|x')\pi(x')}{q(x'|x^{(i-1)})\pi(x^{(i-1)})}, \end{aligned} \tag{3.1}$$

where $\pi(\cdot)$ is known up to a normalizing constant.

The idea behind the acceptance rate is that in order to reveal the true distribution of x, we want our algorithm to be able to explore the whole space without stucking at one point, and to visit the regions with higher probability more often. The $\frac{q(x^{(i-1)}|x')}{q(x'|x^{(i-1)})}$ part ensures that the possibility of the sampler to revisit the previous point is reserved. At the same time, the $\frac{\pi(x')}{\pi(x^{(i-1)})}$ part ensures that regions with higher probability would be more likely to be visited.

The final step is to reject or accept the proposal we make in the first step. We accept this newly sampled proposal with the probability equals r. This is often done by generating a uniformly distributed random variable u between 0 and 1. Then

$$x^{(i)} = \begin{cases} x', u < r \\ x^{(i-1)}, u \geq r \end{cases}$$

3.2.3 Metropolis-within-Gibbs

However, the classic Metropolis-Hastings algorithm cannot be directly applied to our problem. We need to know the full joint probability density function of the

newly sampled $h_t^{'}$ defined in Section 3.2.1 and that of $h_t^{(i-1)}$ from the previous iteration, which could be very difficult to find. Instead, we could implement an algorithm with the similar idea as the Metropolis-within-Gibbs algorithm (Roberts and Rosenthal, 2006). The details of our algorithm are described as below.

In the first step, when $t \geq 2$, we propose the new $h_t^{'}$ conditioned on $h_{t-1}^{(i)}$ rather than $h_t^{(i-1)}$ [1]. As discussed in Section 3.2.1, the distribution of $h_t^{(i)}$ is only determined by $h_{t-1}^{(i)}$ and $\theta^{(i)}$. Therefore, our proposal distribution would be a Student t distribution with mean $\beta_0^{(i)} + \beta_1^{(i)} h_{t-1}^{(i)}$, standard deviation $\delta^{(i)}$, and degrees of freedom $\nu_\eta^{(i)}$. We denote this distribution as $G_t^{(i)}$.

In the second step, since $h_t^{'}$ and $h_t^{(i-1)}$ are independent fixing the parameter vector $\theta^{(i)}$, the first part of the acceptance probability is

$$\frac{q(h_t^{(i-1)}|h_t^{'}, \theta^{(i)})}{q(h_t^{'}|h_t^{(i-1)}, \theta^{(i)})} = \frac{G_t^{(i)}(h_t^{(i-1)})}{G_t^{(i)}(h_t^{'})}.$$

The full joint density function of h_t's cannot be easily derived. However, the Markovian structure of the process of $\{h_t\}$ suggests that we only need to focus on the joint density functions in the Markov blankets of $h_t^{'}$ and $h_t^{(i-1)}$. A Markov blanket is the smallest set of data that can grant h_t conditional independence from all other variables. The structure of ARSV process suggests that the Markov blanket of h_t includes $\{h_{t-1}, h_{t+1}, x_t\}$. That is, conditioned on the observed return at time t, x_t, the log-volatility sampled from the previous step, h_{t-1}, and the log-volatility of the next step, h_{t+1}, the h_t is independent from all other observed return values or volatility values. In our case, the Markov blanket of $h_t^{'}$ includes $\{h_{t-1}^{(i)}, h_{t+1}^{(i-1)}, x_t\}$. These are the log-volatility sampled in the previous step of this iteration, the log-volatility of the next step from the previous iteration, and the known return value.

[1] Note that in the Gibb's sampling scheme, a new value is drawn by conditioning on the value from the previous iteration. However, here we condition on the value from the previous time step in the same iteration.

Therefore,

$$\begin{aligned}
\pi(h_t^{'}) &= f_{h_t}(h_t^{'}|x_t, h_{t-1}^{(i)}, h_{t+1}^{(i-1)}, \theta^{(i)}) \\
&\propto f_1(h_t^{'}, x_t, h_{t-1}^{(i)}, h_{t+1}^{(i-1)}, \theta^{(i)}) \\
&= f_X(x_t|h_{t-1}^{(i)}, h_t^{'}, h_{t+1}^{(i-1)}, \theta^{(i)}) f_2(h_{t-1}^{(i)}, h_t^{'}, h_{t+1}^{(i-1)}, \theta^{(i)}) \\
&= f_X(x_t|h_t^{'}) f_{h_{t+1}}(h_{t+1}^{(i-1)}|h_t^{'}, h_{t-1}^{(i)}, \theta^{(i)}) f_3(h_t^{'}, h_{t-1}^{(i)}, \theta^{(i)}) \\
&= f_X(x_t|h_t^{'}) f_{h_{t+1}}(h_{t+1}^{(i-1)}|h_t^{'}, \theta^{(i)}) f_{h_t}(h_t^{'}|h_{t-1}^{(i)}, \theta^{(i)}) f_{h_{t-1}}(h_{t-1}^{(i)}|\theta^{(i)}) f_p(\theta^{(i)})
\end{aligned}$$

where f_X is the conditional probability density function of returns [2]; f_{h_t} and $f_{h_{t+1}}$ are the conditional probability density functions of h_t and h_{t+1} respectively; $f_{h_{t-1}}$ and f_p are the marginal probability density functions of h_{t-1} and θ respectively; f_1, f_2, and f_3 are joint probability density functions of $\{h_t, X_t, h_{t-1}, h_{t+1}, \theta\}$, $\{h_t, h_{t-1}, h_{t+1}, \theta\}$, and $\{h_{t-1}, h_t, \theta\}$ respectively.

Similarly,

$$\begin{aligned}
&\pi(h_t^{(i-1)}) \propto \\
&f_X(x_t|h_t^{(i-1)}) f_{h_{t+1}}(h_{t+1}^{(i-1)}|h_t^{(i-1)}, \theta^{(i)}) f_{h_t}(h_t^{(i-1)}|h_{t-1}^{(i)}, \theta^{(i)}) f_{h_t}(h_{t-1}^{(i)}|\theta^{(i)}) f_p(\theta^{(i)})
\end{aligned}$$

We know that $f_{h_t}(h_t^{'}|h_{t-1}^{(i)}, \theta^{(i)}) = G_t^{(i)}(h_t^{'})$, and we can assume that

[2] Note that in this example we are considering a simple case where $\varepsilon_t \sim \mathrm{N}(0,1)$, therefore the distribution of X_t only depends on the value of σ_t. If we have further assumptions about ε_t then $f_X(x_t|h_t^{'})$ should be replaced with $f_X(x_t|h_t^{'}, \theta^{(i)})$.

$f_{h_t}(h_t^{(i-1)}|h_{t-1}^{(i)},\theta^{(i)}) \approx G_t^{(i)}(h_t^{(i-1)})$. Then, α in (3.2) can be written as

$$
\begin{aligned}
\alpha &= \frac{q(h_t^{(i-1)}|h_t^{'})}{q(h_t^{'}|h_t^{(i-1)})} \times \frac{\pi(h_t^{'})}{\pi(h_t^{(i-1)})} \\
&= \frac{G_t^{(i)}(h_t^{(i-1)})}{G_t^{(i)}(h_t^{'})} \times \\
&\frac{f_X(x_t|h_t^{'}) f_{h_{t+1}}(h_{t+1}^{(i-1)}|h_t^{'},\theta^{(i)}) f_{h_t}(h_t^{'}|h_{t-1}^{(i)},\theta^{(i)}) f_{h_{t-1}}(h_{t-1}^{(i)}|\theta^{(i)}) f_p(\theta^{(i)})}{f_X(x_t|h_t^{(i-1)}) f_{h_{t+1}}(h_{t+1}^{(i-1)}|h_t^{(i-1)},\theta^{(i)}) f_{h_t}(h_t^{(i-1)}|h_{t-1}^{(i)},\theta^{(i)}) f_{h_{t-1}}(h_{t-1}^{(i)}|\theta^{(i)}) f_p(\theta^{(i)})} \\
&= \frac{f_X(x_t|h_t^{'}) f_{h_{t+1}}(h_{t+1}^{(i-1)}|h_t^{'},\theta^{(i)}) f_{h_{t-1}}(h_{t-1}^{(i)}|\theta^{(i)}) f_p(\theta^{(i)})}{f_X(x_t|h_t^{(i-1)}) f_{h_{t+1}}(h_{t+1}^{(i-1)}|h_t^{(i-1)},\theta^{(i)}) f_{h_{t-1}}(h_{t-1}^{(i)}|\theta^{(i)}) f_p(\theta^{(i)})} \\
&= \frac{f_X(x_t|h_t^{'}) f_{h_{t+1}}(h_{t+1}^{(i-1)}|h_t^{'},\theta^{(i)})}{f_X(x_t|h_t^{(i-1)}) f_{h_{t+1}}(h_{t+1}^{(i-1)}|h_t^{(i-1)},\theta^{(i)})}.
\end{aligned} \tag{3.2}
$$

Expression (3.3) can be evaluated easily for $t = 2,3,...,(T-1)$ and $i \geq 1$. It can be understood intuitively as the ratio of partial conditional likelihoods between $h_t^{'}$ and $h_t^{(i-1)}$. Instead of comparing the full likelihood times the inverse of proposal kernel like the classic Metropolis-Hastings' algorithm, the acceptance rate here is determined by comparing the likelihood of observing the current return value and the log-volatility value of the next step between the newly proposed $h_t^{'}$ and the value from the last iteration $h_t^{(i-1)}$. The decision of either keeping the new value or retaining the old value is made in the same way as classic Metropolis-Hastings algorithm. However, when $t = 1$, h_0 is unknown and therefore we cannot sample h_1, and need to make an arbitrary guess about it. Similarly, when $t = T$, h_{T+1} is unknown and the method described above cannot be applied to sample h_T. We need to make an arbitrary guess about h_T as well. One possible way to minimize the impact of h_1 and h_T in the i^{th} iteration is to discard a number of $\{h_t^{(i)}\}$ for $t < T_1$ and $t > T_2$, and only retain a subset of $\{h_t^{(i)}\}$ as $\{h_t^{(i)} : T_1 \leq t \leq T_2\}$, where T_1 and T_2 are two arbitrary numbers to be selected prior to running the algorithm.

Note that we do not have theoretical proof for the convergence of this algorithm. However, it is reasonable to believe that our method enjoys the same convergence property as the classic Metropolis-Hastings' algorithm from the empirical results such as in Figure 3.2.

3.2.4 Parameter Estimation

Previous methods implement the fully Bayesian approach to estimate the parameters of the ARSV model. With carefully chosen prior distributions for the parameters and the normality assumption of both innovations one can sample the parameters from their posterior distributions. However, it would be very difficult if the innovation distributions, especially the second innovation distribution of the ARSV(1) model, are not conjugate with each other or the prior distribution chosen for parameters.

In our approach, we estimate the parameters separately. After each iteration we obtain the $\{h_t^{(i)} : T_1 \leq t \leq T_2\}$, and we want to estimate the parameters $(\beta_0, \beta_1, \delta, \nu_\eta)$ of an AR(1) process

$$h_t = \beta_0 + \beta_1 h_{t-1} + \delta \eta_t$$

where $T_1 \leq t \leq T_2$ and η_t is a strict white noise process with zero mean and unit variance. Then the paramter estimation can be simplified to the problem of estimating the parameters of an AR(1) process with non-Gaussian innovation, which is well studied and can be easily applied (for example, (Grunwald, Hyndman, Tedesco, and Tweedie, 2000)). In this study, we use the `arfimafit` function in the **rugarch** package, which provides estimation of parameters for an autoregressive model with Student t distribution with an MLE approach. If there is no available program for the parameter estimation, one can try different methods, such as maximum likelihood method or Bayesian method, to estimate the parameters. Figure 3.2 illustrates the distribution of estimated parameters. It can be seen that with a starting value -1.2, the Markov chain values of β_0 soon move close to the true value -0.5 after a few hundred iterations, and remain around the true value as the number iterations increases. Also notice that the Markov chain values of the degrees of freedom is not very stable, however, most of the iterations will return estimations that are reasonably close to the true value. Therefore, we could discard the first K iterations as the burn-in period, where the K is determined by observation, and use some robust statistic such as the median of Markov chain values of each parameter from remaining Markov chain values as our estimator of the param-

eter[3]. This estimator will help us to obtain a good estimation of each parameter. In Table 3.1 we present an example of the results of our algorithm. Here the length of Markov chain is 3000 iterations, and the length burn-in period is 500 iterations. In Figure 3.2 we also present the Markov values after each iteration.

Table 3.1: An example of estimating parameters of an ARSV(1) process with $\varepsilon_t \sim N(0,1)$ and $\eta_t \sim$ standardized t_5. Median of Markov chain values after discarding the values from the first 1000 iterations. The values in the parentheses are standard error of each estimator.

True Value	$\beta_0 = -0.5$	$\beta_1 = 0.95$	$\delta = 0.35$	$\nu_\eta = 5$
estimation	-0.544(0.080)	0.944(0.008)	0.360(0.039)	5.46(20.7)

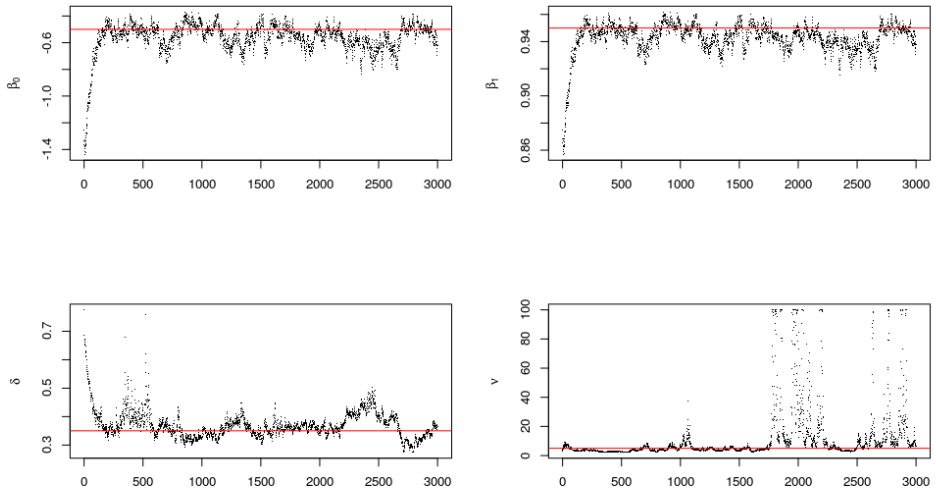

Figure 3.2: Illustration of Markov chain values of parameters after each iteration. Each dot represents the estimated value of the corresponding parameter after each iteration. The red lines represent the true value.

[3]For simplicity we use median as the estimator for all parameters. However, particularly for the estimator for degrees of freedom, mode is also recommended.

3.2.5 Discussion on the Flexibility

During the process of writing this report, we find a paper by Fridman and Harris (1998) which also discusses a flexible inference method for ARSV process based on the maximum likelihood approach. The new method discussed in this report and their method achieve the flexibility of allowing arbitrary choices of innovation distribution in a different way. For the method proposed by Fridman and Harris (1998), inference for ARSV(1) process is simplified to finding a good numerical approximation method to evaluate the integration on a given probability density function. However, our new method simplifies the inference for ARSV(1) process to inference for an AR(1) process with given distribution of error terms. Furthermore, it is easy for the method proposed by Fridman and Harris (1998) to include the ARCH term in the volatility process, while increasing the lags in the volatility process will lead to an exponential growth in the complexity of this method. Our method, on the other hand, can handle the inference of ARSV(p) process with trivial modifications. However, it might be more difficult for our method to estimate parameters for the ARSV model with the ARCH term in the volatility process. Therefore, these two methods can be complementary to each other. The choice should be made depending on the assumption of underlying process.

3.2.6 Algorithm

The algorithm for our new method contains two layers of loops. Before starting our iterations, first we need to get an initial estimation of $\{h_t^{(0)}\}$ using some proxies such as the moving average of squared returns. Then we initialize the parameter set $\theta^{(1)} = [\beta_0^{(1)}, \beta_1^{(1)}, \delta^{(1)}, \nu_\eta^{(1)}]$ based on methods discussed in Section 3.2.4. This algorithm contains two layers of loops. In the i^{th} iteration of the external loop, we follow sampling method discussed in Section 3.2.1 and 3.2.3 to generate the sequence of $\{h_t^{(i)}\}$ for $t = 1, 2, ...N$. After the internal loop ends, we update parameter set based on the newly sampled $\{h_t^{(i)}\}, t = 1, 2, ...N$ with methods discussed in Section 3.2.4. Detailed algorithm is given below, and simulation results are presented in the next section.

Algorithm 1 Flexible Inference for ARSV Model Inference

initialize ite, $\{h_t^{(0)}\}$, $\theta^{(1)}$, T, T_1, and T_2
while i $<$ ite **do**
  for t from 2 to T-1 **do**
    sample $h_t^{'}$ from $f_{h_t}(\cdot|h_{t-1}^{(i)}, \theta^{(i)})$
    set $\alpha = \frac{f_X(x_t|h_t^{'}) f_{h_{t+1}}(h_{t+1}^{(i-1)}|h_t^{'}, \theta^{(i)})}{f_X(x_t|h_t^{(i-1)}) f_{h_{t+1}}(h_{t+1}^{(i-1)}|h_t^{(i-1)}, \theta^{(i-1)})}$
    calculate acceptance rate $A = min\{1, \alpha\}$
    generate u from $unif(0,1)$
    if $u < A$ **then** $h_t^{(i)} = h_t^{(i-1)}$
    else $h_t^{(i)} = h_t^{'}$
    end if
  end for
  keep $\{h_t^{(i)}, T_1 \leq t \leq T_2\}$
  update $\theta^{(i+1)}$ by fitting an AR(1) process to $\{h_t^{(i)}, T_1 \leq t \leq T_2\}$
end while

3.3 Comparison of Inference Methods

In this section we present the results of the model inference. We show that our new method can estimate the parameters of the classic ARSV(1) model as well as previous methods, and can be applied to the extension of classic ARSV(1) when previous methods may fail.

3.3.1 Parameter Estimation for Simulated Data

First we want to show that our new approach works as well as previous methods. The simulated data is generated from the following process:

$$X_t = \sigma_t \varepsilon_t \tag{3.3}$$

$$\log \sigma_t^2 = -0.5 + 0.95 \log \sigma_{t-1}^2 + 0.35 \eta_t, \tag{3.4}$$

where $\varepsilon_t \sim N(0,1)$, $\eta_t \sim N(0,1)$, and η_t and ε_t are independent.

We generate 200 datasets with length $T = 2500$ based on (3.4) and (3.5) with

different random seeds. We then run our new method for 5000 iterations and discard the first 2000 runs. We compare our results with the **stochvol** package (Kastner, 2016). This package is one of the latest packages for ARSV(1) model inference, and we regard it as the representative of existing methods. **stochvol** package assumes the second innovation distribution $\{\eta_t\}$ of ARSV(1) model to be normal, and can work with both normal innovation distributions or a Student t first innovation and a normal second innovation.

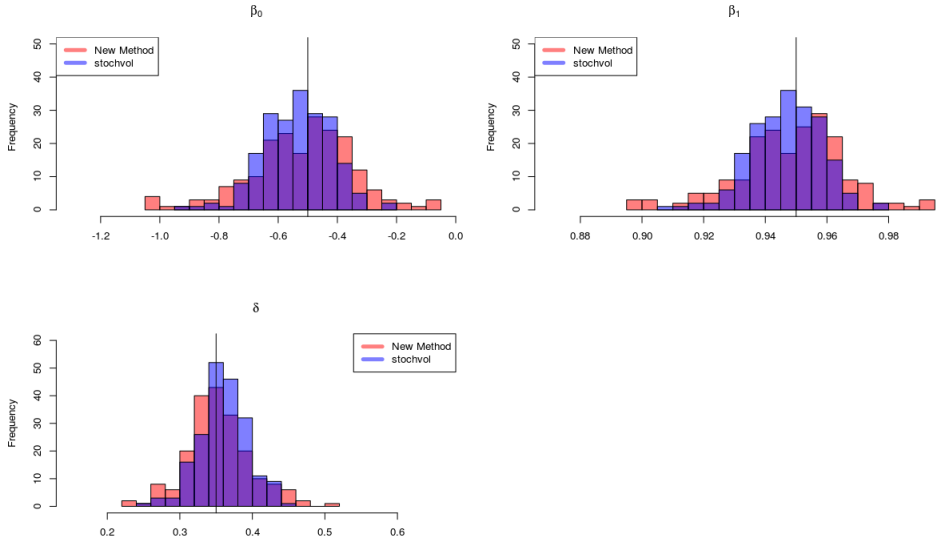

Figure 3.3: Distribution of estimated parameters based on simulated data generated from (3.4) and (3.5) with $\varepsilon_t \sim N(0,1)$ and $\eta_t \sim N(0,1)$. Estimates are from our new method and the **stochvol** package.

Table 3.2: Median of estimated parameters for simulated data in Figure 3.3 using our new method. The values in the parentheses are standard deviations of the 200 estimated values of each parameter.

Method	$\beta_0 = -0.5$	$\beta_1 = 0.95$	$\delta = 0.35$
stochvol	-0.527(0.113)	0.947(0.011)	0.359(0.033)
New Method	-0.495(0.177)	0.950(0.017)	0.348(0.045)

The example above illustrates that our new method works as well as previous methods for the inference of the classic ARSV(1) model. Although the standard deviations of our estimated parameters are slightly larger, in general two methods are very close to each other.

Next we want to show that our method is more flexible than previous methods. We consider two different scenarios: (1) $\varepsilon_t \sim N(0,1)$ and $\eta_t \sim$ standardized t_5; (2) $\varepsilon_t \sim$ standardized t_5 and $\eta_t \sim$ standardized t_5.

In the first scenario, 300 datasets with length $T = 2500$ are generated. We compare the results in the same way as above.

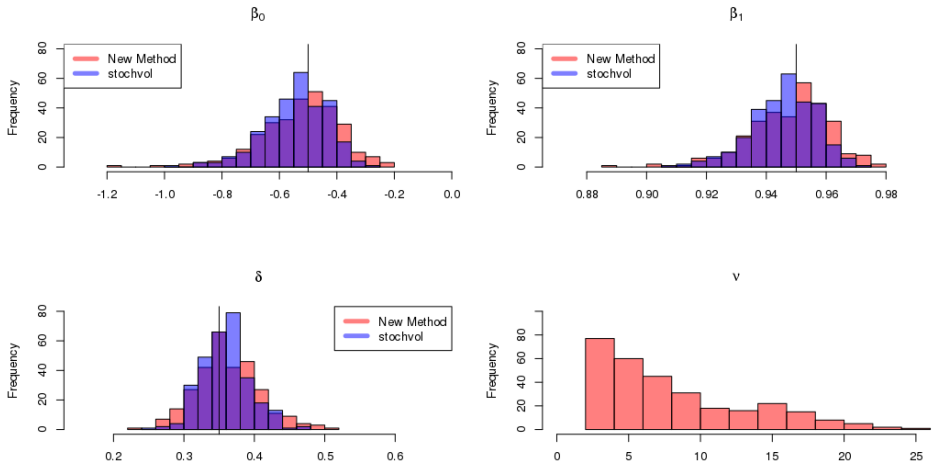

Figure 3.4: Distribution of estimated parameters based on simulated data generated from (3.4) and (3.5) with $\varepsilon_t \sim N(0,1)$ and $\eta_t \sim$ standardized t_5. Estimates are from our new method and the **stochvol** package.

Table 3.3: Median of estimated parameters for simulated data in Figure 3.4. The values in the parentheses are standard deviations of the 300 estimated values of each parameter.

Method	$\beta_0 = -0.5$	$\beta_1 = 0.95$	$\delta = 0.35$	$\nu_\eta = 5$
stochvol	-0.531(0.110)	0.947(0.011)	0.359(0.033)	NA
New Method	-0.506(0.138)	0.950(0.014)	0.357(0.045)	6.54(5.15)

From Figure 3.4 and Table 3.3 we can see that estimated parameters from both our new method and the **stochvol** package are very close. However, our new method estimates the degrees of freedom of the second innovation distribution reasonably well, while the **stochvol** package has to assume that the second innovation is normally distributed.

We do similar simulation study for the second case, and the results are presented below. Note that we estimate the parameters of ARSV model using the **stochvol** package with two different model specifications, namely assuming $\varepsilon_t \sim N(0,1)$ or $\varepsilon_t \sim$ standardized t_{ν_ε}, to show how bad the performance can be when we specify the model incorrectly. The results are presented in Table 3.4.

When both innovations follow Student t distribution, we can observe a much better performance for our new method in comparison to the **stochvol** package. The difference is significant especially for estimating the shape parameter of ε_t, as the result from our method is less biased with a much smaller standard deviation. We can also see that when the model specification is incorrect, the performance of the method implemented in **stochvol** package could be very poor.

Next, we want to show that our new method is robust against the incorrect model specification. First, let us consider the case when the true model has both innovations normally distributed. We estimate the parameters by specifying both innovations as Student t distributed. The results are presented in Table 3.5.

We also consider the case when the first innovation of the true model is normally distributed and the second innovation distribution is a skewed Student t distribution. The skewing parameter $\gamma = 1.5$, and to amplify the impact of the incorrect model specification we assume that the standard deviation of η_t equals 0.5. We estimate the parameters by assuming that the first innovation distribution is normal while the second innovation distribution is Student t. The results are presented in Table 3.6.

Table 3.4: Median of estimated parameters for simulated data with $\varepsilon_t \sim$ standardized t_5 and $\eta_t \sim$ standardized t_5. The values in the parentheses are standard deviations of the estimated values of each parameter.

Method	$\beta_0 = -0.5$	$\beta_1 = 0.95$	$\delta = 0.35$	$\nu_\varepsilon = 5$	$\nu_\eta = 5$
stochvol(t-Normal)	-0.561(0.156)	0.946(0.015)	0.363(0.054)	5.44(4.58)	NA
stochvol(Normal-Normal)	-1.098(0.312)	0.893(0.031)	0.575(0.064)	NA	NA
New Method	-0.502(0.154)	0.950(0.015)	0.357(0.059)	4.98(1.02)	5.37(4.78)

Table 3.5: Median of estimated parameters for simulated data with $\varepsilon_t \sim N(0,1)$ and $\eta_t \sim N(0,1)$. However, we estimate the parameters by assuming that $\varepsilon_t \sim$ standardized t_{ν_ε} and $\eta_t \sim$ standardized t_{ν_η}. The values in the parentheses are standard deviations of the estimated values of each parameter.

True Value	$\beta_0 = -0.5$	$\beta_1 = 0.95$	$\delta = 0.35$	$\nu_\varepsilon = \infty$	$\nu_\eta = \infty$
Estimated Value	-0.493(0.139)	0.951(0.014)	0.344(0.042)	9.39(5.96)	92.22(44.6)

Table 3.6: Median of estimated parameters for simulated data with $\varepsilon_t \sim N(0,1)$ and $\eta_t \sim skewed-t(0,1,5,1.5)$. However, we estimate the parameters by assuming that $\varepsilon_t \sim \mathrm{N}(0,1)$ and $\eta_t \sim$ standardized t_{ν_η}. The values in the parentheses are standard deviations of the estimated values of each parameter.

True Value	$\beta_0 = -0.5$	$\beta_1 = 0.95$	$\delta = 0.5$	$\nu_\varepsilon = 5$	$\gamma = 1.5$
Estimated Value	-0.545(0.129)	0.946(0.013)	0.521(0.051)	4.93(5.19)	NA

The two examples above suggest that our new method is robust against the incorrect model specification. Note that our method can work for arbitrary choices of innovation distributions. Therefore, we can always "over-specify" our model with more general innovation distributions to achieve robustness. For example, in the example illustrated in Table 3.4, we estimate the degrees of freedom for η_t as 9.39 with standard deviation 5.96 and the degrees of freedom for ε_t as 92.22 with standard deviation 44.6 while these two innovations are actually normally distributed. However, a Student t distribution with such a large degrees of freedom would practically behave very similar to a normal distribution. Therefore, our method allows a more generalized model specification which brings robustness, while the previous methods do not enjoy this flexibility.

3.3.2 Parameter Estimation for the S&P 500 Index

In Fridman and Harris (1998), results for the model inference are compared across several methods utilizing a Bayesian approach, semi-maximum likelihood approach and maximum likelihood approach. Here we compare the estimated parameters from the three methods mentioned above with results from our proposed method and results from the **stochvol** package.

We fit an ARSV(1) model to the daily log-return of the S&P 500 Index from 1980 to 1987. In total, 2022 observations are used for model inference. We estimate the parameters following both the traditional model assumptions that the two innovation distributions are Gaussian. The results are presented in Table 3.7. Due to the difference approach in achieving flexible model assumption, in Fridman and Harris (1998) the authors do not fit an ARSV(1) process with light-tailed first innovation and heavy tailed second innovation. They consider the scenario that the first innovation is Student t distributed while the second innovation is normally distributed, which can also be fitted using the **stochvol** package. We fit the data with the same model assumption as well as a more generalized model assumption that both innovations are Student t distributed. The results are presented in Table 3.8.

Table 3.7: Summaries statistics of estimated parameters with the model assumption that both innovations are normally distributed. The "n" in column names stands for "normal". The values in the parentheses are asymptotic standard deviations for SML and ML methods, posterior standard deviations for Bayes method and the method implemented in **stochvol** package, and interquartile range (IQR) for the new approach.

	Bayes	SML	ML(n-n)	stochvol	New Method(n-n)
β_0	-.002(.004)	-.002(.004)	-.002(.0004)	-.270(.007)	-.010(.022)
β_1	.970(.008)	.958(.014)	.959(.005)	.971(.001)	.989(.003)
δ	.150(.017)	.161(.026)	.159(.009)	.153(.002)	.071(.014)

Table 3.8: Summaries statistics of estimated parameters with the model assumption that the first innovation distribution follows a Student t distribution. The values in the parentheses are asymptotic standard deviations for ML methods, posterior standard deviations for the method implemented in **stochvol** package, and interquartile range (IQR) for the new approach.

	ML(t-n)	stochvol	New Method(t-n)	New Method(t-t)
β_0	-.0038(.0013)	-.1384(.0108)	-.1020(.0473)	-.1718(.2979)
β_1	.9813(.0056)	.9855(.0011)	.9897(.0054)	.9821(.0312)
δ	.0942(.0199)	.1016(.0042)	.0799(.01853)	.1153 (.0958)
ν_ε	10.39(5.88)	11.42(1.76)	10.11(1.07)	10.47(1.75)
ν_η	NA	NA	NA	2.50(1.96)

Note that our new approach achieves the flexibility of arbitrary choices of innovations at the cost of stability and efficiency. Therefore, outliers appear a few times when estimating the parameters. More robust summary statistics such as median/IQR are preferred over mean/standard deviation. The results presented in Table 3.7 and Table 3.8 show that the results from our method are close to those from existing method, except that the estimated δ using our method in Table 3.7 is significantly lower than those using other methods. The reason requires further study.

Chapter 4

Conditional Risk Measurement with the ARSV Model

One of the applications for volatility models is to estimate the potential risk given known information. In general, there are two approaches to risk measures, namely conditional risk measures and unconditional risk measures. When we are discussing GARCH process and ARSV process in Chapter 2, we assume that $\mathscr{F}_{t-1}$, the sigma algebra of all available information up to time $t-1$, is known. If we further assume that the distribution of the return X_t is conditioned on $\mathscr{F}_{t-1}$, then our measure of risk at time t should be conditioned on $\mathscr{F}_{t-1}$. On the other hand, we can also measure the unconditional risk from the unconditional distribution of X_t. One can choose whether to use the conditional or unconditional distribution for risk measure forecasting. In this chapter we adopt the conditional estimation approach, and focus on the Value-at-Risk and Conditional Value-at-Risk risk measures under the ARSV process.

4.1 Value-at-Risk forecasting under the ARSV Model

The Value-at-Risk (VaR) is arguably one of the most widely used risk measures, and for a confidence level $\alpha \in (0,1)$ it is defined as:

$$\mathrm{VaR}_\alpha(X) = \inf\{x : \mathbb{P}(X \geq x) \leq 1-\alpha\}, \tag{4.1}$$

and typically α is close to 1 (*e.g.*, 0.9 or 0.95). For an ARSV(1) model, $X_t = \sigma_t \varepsilon_t$, where σ_t and ε_t are independent random variables. Since $\sigma_t > 0$ almost surely, we have

$$\begin{aligned}\mathbb{P}(X_t \geq x|\mathscr{F}_{t-1}) &= \mathbb{P}(\sigma_t \varepsilon_t \geq x|\mathscr{F}_{t-1}) \\ &= \int_0^\infty \mathbb{P}(\varepsilon_t \geq x/s) f_{\sigma_t|\mathscr{F}_{t-1}}(s|\mathscr{F}_{t-1}) ds, \end{aligned} \tag{4.2}$$

where f_{σ_t} is the conditional probability density function of σ_t given $\mathscr{F}_{t-1}$ is the information set at time $t-1$.

Since $\log \sigma_t^2 = \beta_0 + \beta_1 \log \sigma_{t-1}^2 + \eta_t$,

$$\begin{aligned}\mathbb{P}(\sigma_t \leq s|\mathscr{F}_{t-1}) &= \mathbb{P}(\log \sigma_t^2 \leq \log s^2|\mathscr{F}_{t-1}) \\ &= \mathbb{P}(\beta_0 + \beta_1 \log \sigma_{t-1}^2 + \delta \eta_t \leq \log s^2|\mathscr{F}_{t-1}) \\ &= \mathbb{P}(\eta_t \leq (\log s^2 - \beta_0 - \beta_1 \log \sigma_{t-1}^2)/\delta|\mathscr{F}_{t-1}) \\ &= F_\eta((\log s^2 - \beta_0 - \beta_1 \log \sigma_{t-1}^2)/\delta),\end{aligned}$$

where F_η is the cumulative distribution function of a standardized Student's t distribution with zero mean, unit variance and ν_η degrees of freedom and assuming that $\mathscr{F}_{t-1}$ contains σ_{t-1}. Hence, the conditional density of σ_t is given by

$$f_{\sigma_t|\mathscr{F}_{t-1}}(s|\mathscr{F}_{t-1}) = f_\eta((\log s^2 - \beta_0 - \beta_1 \log \sigma_{t-1}^2)/\delta)\frac{2}{\delta s}, \qquad s > 0..$$

Therefore, (4.2) can be written as

$$\begin{aligned}&\mathbb{P}(X_t \geq x|\mathscr{F}_{t-1}) \\ &= \frac{2}{\delta}\int_0^\infty \mathbb{P}(\varepsilon_t \geq x/s) f_\eta((\log s^2 - \beta_0 - \beta_1 \log \sigma_{t-1}^2)/\delta)\frac{1}{s} ds.\end{aligned} \tag{4.3}$$

With estimated parameters $\hat{\theta} = [\hat{\beta}_0, \hat{\beta}_1, \hat{\delta}, \hat{\nu}_\eta]$ and estimated volatility $\hat{\sigma}_{t-1}$, we can plug (4.3) into (4.1) and solve for $\mathrm{VaR}_\alpha(X_t)$ using numerical methods. Parameters $\hat{\theta}$ can be estimated by the method we described in Chapter 3. The volatility $\hat{\sigma}_{t-1}$ can be estimated separately using unbiased volatility estimators such as average squared returns or EWMA, or using inference methods for volatility models

that provide volatility estimation.

4.2 CoVaR forecasting: GARCH Model vs ARSV Model

Adrian and Brunnermeier (2011) were first to define Conditional Value-at-Risk (CoVaR) as:

$$\text{CoVaR}_{\alpha}^{(1)}(X_{t+1}) = \inf\{x : \mathbb{P}(X_{t+1} \geq x) \leq 1-\alpha | X_t = \text{VaR}_{\alpha'}(X_t), \mathscr{F}_{t-1}\}, \quad (4.4)$$

where both α and α' are two constants that are close to 1 and VaR_α is defined in Section 4.1. For simplicity, we assume $\alpha = \alpha'$.

Girardi and Ergün (2013) modify the definition of CoVaR to:

$$\text{CoVaR}_{\alpha}^{(2)}(X_{t+1}) = \inf\{x : \mathbb{P}(X_{t+1} \geq x) \leq 1-\alpha | X_t \geq \text{VaR}_{\alpha}(X_t), \mathscr{F}_{t-1}\}. \quad (4.5)$$

The one-step ahead forecasts of VaR_α are based on the estimate of α-quantile of the distribution of X_t give the history $\mathscr{F}_{t-1}$. CoVaR_α forecasting, on the other hand, makes a two-step forward estimation of the conditional quantile of the distribution of X_{t+1} conditioned on X_t and history $\mathscr{F}_{t-1}$. So instead of measuring the potential large loss or gain on the next day as VaR does, CoVaR measures the potential consecutive large gains for two days. In this section we discuss the CoVaR estimation under both GARCH and ARSV models for both CoVaR definitions.

4.2.1 First Definition of CoVaR

For the GARCH(1,1) process, given all information up to time $t-1$ and σ_{t-1}, the volatility at time t is fixed as $\sqrt{\alpha_0 + \alpha\sigma_{t-1}^2 + \beta X_{t-1}^2}$. Suppose that $X_t = \text{VaR}_\alpha(X_t) =:$

v, then σ_{t+1} is also fixed and equals $\sqrt{\alpha_0+\alpha\sigma_t^2+\beta v^2}$. Then

$$\begin{aligned}
&\mathbb{P}(X_{t+1}\geq x|X_t=v,\mathscr{F}_{t-1})\\
&=\mathbb{P}(\sigma_{t+1}\varepsilon_{t+1}\geq x|X_t=v,\mathscr{F}_{t-1})\\
&=\mathbb{P}(\varepsilon_{t+1}\geq x/\sigma_{t+1}|X_t=v,\mathscr{F}_{t-1}) \qquad (4.6)\\
&=\mathbb{P}(\varepsilon_{t+1}\geq x/\sqrt{\alpha_0+\alpha\sigma_t^2+\beta v^2}|\mathscr{F}_{t-1})\\
&=P\left(\varepsilon_{t+1}\geq \frac{x}{\sqrt{\alpha_0(\alpha+1)+\alpha^2\sigma_{t-1}^2+\alpha\beta X_{t-1}^2+\beta v^2}}\right). \qquad (4.7)
\end{aligned}$$

Substituting (4.7) into (4.4), we can solve for $\text{CoVaR}_\alpha^{(1)}(X_{t+1})$ numerically when $\{X_t\}$ follows a GARCH(1,1) process.

For the ARSV(1) process, given the same information and assuming $X_t=v$, the value of σ_{t+1} is no longer fixed. Therefore, expression (4.6) becomes

$$\begin{aligned}
&\mathbb{P}(X_{t+1}\geq x|X_t=v,\mathscr{F}_{t-1})\\
&=\int_0^\infty \mathbb{P}(\varepsilon_{t+1}\geq x/s|X_t=v,\mathscr{F}_{t-1},\sigma_{t+1}=s)f_{\sigma_{t+1}|X_t=v,\mathscr{F}_{t-1}}(s|X_t=v,\mathscr{F}_{t-1})ds\\
&=\int_0^\infty \mathbb{P}(\varepsilon_{t+1}\geq x/s)f_{\sigma_{t+1}|X_t=v,\mathscr{F}_{t-1}}(s|X_t=v,\mathscr{F}_{t-1})ds, \qquad (4.8)
\end{aligned}$$

where $f_{\sigma_{t+1}|X_t=v,\mathscr{F}_{t-1}}$ is the conditional probability density function of σ_{t+1} given X_t and $\mathscr{F}_{t-1}$. In order to evaluate (4.8), we need to find this conditional probability density function. Since for $s>0$,

$$\mathbb{P}(\sigma_{t+1}\leq s|X_t=v,\mathscr{F}_{t-1})=\mathbb{P}(\log\sigma_{t+1}^2\leq\log s^2|X_t=v,\mathscr{F}_{t-1}),$$

we have

$$f_{\sigma_{t+1}|X_t=v,\mathscr{F}_{t-1}}(s|X_t=v,\mathscr{F}_{t-1})=f_{h_{t+1}|X_t=v,\mathscr{F}_{t-1}}(\log s^2|X_t=v,\mathscr{F}_{t-1})\frac{2}{s},\ s>0.$$

The conditional cumulative distribution function of h_{t+1} is

$$\begin{aligned} F_{h_{t+1}|X_t=v,\mathscr{F}_{t-1}}(u|X_t=v,\mathscr{F}_{t-1}) &= \mathbb{P}(h_{t+1}\le u|X_t=v,\mathscr{F}_{t-1}) \\ &= \mathbb{P}(\beta_0+2\beta_1\log\frac{X_t}{\varepsilon_t}+\delta\eta_{t+1}\le u|X_t=v,\mathscr{F}_{t-1}) \\ &= \mathbb{P}(\eta_{t+1}\le(u-\beta_0-2\beta_1\log\frac{v}{\varepsilon_t})\delta) \\ &= \int_{-\infty}^{\infty}\mathbb{P}(\eta_{t+1}\le(u-\beta_0-2\beta_1\log\frac{v}{z})/\delta)f_\varepsilon(z)dz. \end{aligned} \tag{4.9}$$

Therefore,

$$f_{h_{t+1}|X_t=v,\mathscr{F}_{t-1}}(\log s^2|X_t=v,\mathscr{F}_{t-1})=\frac{1}{\delta}\int_{-\infty}^{\infty}f_\eta((\log s^2-\beta_0-2\beta_1\log\frac{v}{z})/\delta)f_\varepsilon(z)dz,$$

where f_η is the probability density function of η_t.

Then, (4.8) can be expressed as

$$\begin{aligned} &\mathbb{P}(X_{t+1}\ge x|X_t=v,\mathscr{F}_{t-1}) \\ &=\frac{2}{\delta}\int_0^{\infty}\int_{-\infty}^{\infty}\mathbb{P}(\varepsilon_{t+1}\ge x/s)f_\eta((\log s^2-\beta_0-2\beta_1\log\frac{v}{z})\delta)f_\varepsilon(z)\frac{1}{s}dzds. \end{aligned} \tag{4.10}$$

Substituting (4.10) into (4.4), we can solve for $\text{CoVaR}_\alpha^{(1)}(X_{t+1})$ numerically when $\{X_t\}$ follows an ARSV(1) process.

4.2.2 Second Definition of CoVaR

We can also find the CoVaR under the second definition modified by Girardi and Ergün (2013) (see eq. (4.5)) in a similar way as discussed in Section 4.2.1. However, we have a different conditioning event that $X_t \ge v$ instead of $X_t = v$.

For the GARCH(1,1) process, σ_t is still fixed as $\sqrt{\alpha_0+\alpha_1\sigma_{t-1}^2+\beta_1X_{t-1}^2}$. How-

ever, σ_{t+1} now is a random variable $\sqrt{\alpha_0+\alpha_1\sigma_t^2+\beta_1 X_t^2}$. We have

$$\begin{aligned}
&\mathbb{P}(X_{t+1}\geq x|X_t\geq v,\mathscr{F}_{t-1})\\
&=\mathbb{P}(\varepsilon_{t+1}\geq x/\sqrt{\alpha_0+\alpha_1\sigma_t^2+\beta_1 X_t^2}|X_t\geq v,\mathscr{F}_{t-1})\\
&=\int_v^\infty \mathbb{P}(\varepsilon_{t+1}\geq x/\sqrt{\alpha_0+\alpha_1\sigma_t^2+\beta_1 w^2})f_{X_t|X_t\geq v,\mathscr{F}_{t-1}}(w|X_t\geq v,\mathscr{F}_{t-1})dw. \quad (4.11)
\end{aligned}$$

Notice that for $w\geq v$

$$\begin{aligned}
\mathbb{P}(X_t\leq w|X_t\geq v,\mathscr{F}_{t-1}) &= \frac{\mathbb{P}(X_t\leq w, X_t\geq v|\mathscr{F}_{t-1})}{\mathbb{P}(X_t\geq v|\mathscr{F}_{t-1})}\\
&= \frac{1}{\alpha}\mathbb{P}(v\leq X_t\leq w|\mathscr{F}_{t-1}) \text{ by the definition of VaR}_\alpha\\
&= \frac{1}{\alpha}\mathbb{P}(v\leq \varepsilon_t\sigma_t\leq w|\mathscr{F}_{t-1})\\
&= \frac{1}{\alpha}\mathbb{P}(v/\sigma_t\leq \varepsilon_t\leq w/\sigma_t|\mathscr{F}_{t-1})\\
&= \frac{1}{\alpha}(F_{\varepsilon_t}(w/\sigma_t)-F_{\varepsilon_t}(v/\sigma_t)). \quad (4.12)
\end{aligned}$$

From (4.12) we obtain

$$f_{X_t|X_t\geq v,\mathscr{F}_{t-1}}(w|X_t\geq v,\mathscr{F}_{t-1})=\frac{1}{\alpha\sigma_t}f_\varepsilon(w/\sigma_t),\ w\geq v. \quad (4.13)$$

With (4.13), (4.11) can be expressed as

$$\begin{aligned}
&\mathbb{P}(X_{t+1}\geq x|X_t\geq v,\mathscr{F}_{t-1})\\
&=\frac{1}{\alpha\sigma_t}\int_v^\infty P\left(\varepsilon_{t+1}\geq x/\sqrt{\alpha_0+\alpha_1\sigma_t^2+\beta_1 w^2}\right)f_\varepsilon(w/\sigma_t)dw. \quad (4.14)
\end{aligned}$$

Substituting (4.14) into (4.5), we can solve for $\text{CoVaR}_\alpha^{(2)}(X_{t+1})$ numerically when $\{X_t\}$ follows a GARCH(1,1) process.

In order to find the $\text{CoVaR}_\alpha^{(2)}(X_{t+1})$ under an ARSV(1) process, we can start

from (4.9). Under the condition that $X_t \geq v$, (4.9) now becomes

$$\begin{aligned}
&\mathbb{P}(h_{t+1} \leq u | X_t \geq v, \mathscr{F}_{t-1}) \\
&= \mathbb{P}(\beta_0 + 2\beta_1 \log \frac{X_t}{\varepsilon_t} + \delta \eta_{t+1} \leq u | X_t \geq v, \mathscr{F}_{t-1}) \\
&= \mathbb{P}(\eta_{t+1} \leq (u - \beta_0 - 2\beta_1 \log \frac{X_t}{\varepsilon_t})/\delta | X_t \geq v, \mathscr{F}_{t-1}) \\
&= \int_v^\infty \mathbb{P}(\eta_{t+1} \leq (u - \beta_0 - 2\beta_1 \log \frac{w}{\varepsilon_t})/\delta) f_{X_t | X_t \geq v, \mathscr{F}_{t-1}}(w | X_t \geq v, \mathscr{F}_{t-1}) dw.
\end{aligned} \tag{4.15}$$

In (4.15), the conditional probability density function of X_t, $f_{X_t|X_t \geq v, \mathscr{F}_{t-1}}(w|X_t \geq v, \mathscr{F}_{t-1})$, is unknown. However,

$$f_{X_t|X_t \geq v, \mathscr{F}_{t-1}}(w|X_t \geq v, \mathscr{F}_{t-1}) = \frac{1}{\alpha} f_{X_t|\mathscr{F}_{t-1}}(w|\mathscr{F}_{t-1}),\ w \geq v.$$

Then, (4.15) becomes

$$\begin{aligned}
&\mathbb{P}(h_{t+1} \leq u | X_t \geq v, \mathscr{F}_{t-1}) \\
&= \int_v^\infty \mathbb{P}((\eta_{t+1} \leq u - \beta_0 - 2\beta_1 \log \frac{w}{\varepsilon_t})/\delta) \frac{1}{\alpha} f_{X_t|\mathscr{F}_{t-1}}(w|\mathscr{F}_{t-1}) dw \\
&= \frac{1}{\alpha} \int_{-\infty}^\infty \int_v^\infty \mathbb{P}(\eta_{t+1} \leq (u - \beta_0 - 2\beta_1 \log \frac{w}{z})/\delta) f_{X_t|\mathscr{F}_{t-1}}(w|\mathscr{F}_{t-1}) f_\varepsilon(z) dw dz.
\end{aligned}$$

Therefore,

$$\begin{aligned}
&f_{h_{t+1}|X_{t+1} \geq v, \mathscr{F}_{t-1}}(u | X_{t+1} \geq v, \mathscr{F}_{t-1}) \\
&= \frac{1}{\alpha\delta} \int_{-\infty}^\infty \int_v^\infty f_\eta((u - \beta_0 - 2\beta_1 \log \frac{w}{z})/\delta) f_{X_t|\mathscr{F}_{t-1}}(w|\mathscr{F}_{t-1}) f_\varepsilon(z) dw dz.
\end{aligned} \tag{4.16}$$

The conditional cumulative distribution function of X_t is

$$\begin{aligned}
\mathbb{P}(X_t \leq w | \mathscr{F}_{t-1}) &= \mathbb{P}(\sigma_t \varepsilon_t \leq w | \mathscr{F}_{t-1}) \\
&= \int_0^\infty \mathbb{P}(\varepsilon_t \leq w/r) f_{\sigma_t|\mathscr{F}_{t-1}}(r|\mathscr{F}_{t-1}) dr,
\end{aligned}$$

where $f_{\sigma_t|\mathscr{F}_{t-1}}(r|\mathscr{F}_{t-1}) = \frac{2}{r} f_\eta(\log r^2 - \beta_0 - \beta_1 \log \sigma_{t-1}^2)$ since for $r > 0$,

$$\begin{aligned}\mathbb{P}(\sigma_t \le r|\mathscr{F}_{t-1}) &= \mathbb{P}(\log \sigma_t^2 \le \log r^2|\mathscr{F}_{t-1}) \\ &= \mathbb{P}(\beta_0 + \beta_1 \log \sigma_{t-1}^2 + \delta\eta_t \le \log r^2|\mathscr{F}_{t-1}) \\ &= \mathbb{P}(\eta_t \le (\log r^2 - \beta_0 - \beta_1 \log \sigma_{t-1}^2)/\delta).\end{aligned}$$

So

$$f_{X_t|\mathscr{F}_{t-1}}(w|\mathscr{F}_{t-1}) = \frac{1}{\delta}\int_0^\infty f_\varepsilon(w/r) f_\eta(\log r^2 - \beta_0 - \beta_1 \log \sigma_{t-1}^2)\frac{2}{r^2}dr,$$

and (4.16) can be expressed as

$$\begin{aligned}&f_{h_{t+1}|X_{t+1}\ge v,\mathscr{F}_{t-1}}(u|X_{t+1} \ge v, \mathscr{F}_{t-1}) \\ &= \frac{1}{\alpha\delta^2}\int_{-\infty}^\infty\int_v^\infty\int_0^\infty f_\eta((u - \beta_0 - 2\beta_1 \log\frac{w}{z})/\delta) f_\varepsilon(w/r) \\ &f_\eta((\log r^2 - \beta_0 - \beta_1 \log \sigma_{t-1}^2)/\delta) f_\varepsilon(z)\frac{2}{r^2}drdwdz.\end{aligned}$$

Now (4.8) becomes

$$\begin{aligned}&\mathbb{P}(X_{t+1} \ge x|X_t \ge v, \mathscr{F}_{t-1}) \\ &= \frac{4}{\alpha\delta^2}\int_0^\infty\int_{-\infty}^\infty\int_v^\infty\int_0^\infty \mathbb{P}(\varepsilon_{t+1} \ge x/s) f_\eta((\log s^2 - \beta_0 - 2\beta_1 \log\frac{w}{z})/\delta) f_\varepsilon(w/r) \\ &f_\eta((\log r^2 - \beta_0 - \beta_1 \log \sigma_{t-1}^2)/\delta) f_\varepsilon(z)\frac{1}{r^2 s}drdwdzds. \qquad (4.17)\end{aligned}$$

Substituting (4.17) into (4.5), we can solve for $\text{CoVaR}_\alpha^{(2)}(X_{t+1})$ numerically when $\{X_t\}$ follows an ARSV(1) process.

However, given the numerical complexity of the analytic expression, which involves a 4-dimensional integral, we do not follow this approach. Instead, we propose a simulation based computation that is introduced in the next section.

4.3 Simulation Methods to Find CoVaR

The tail dependence properties of different models have a strong impact on the joint distribution of the (X_t, X_{t+1}) pair. To illustrate the difference in the estimation

of CoVaR under the GARCH model and ARSV model, we will only focus on the second definition of CoVaR modified by Girardi and Ergün (2013). However, it could be difficult to find the CoVaR using numerical method, especially under the second definition of CoVaR. We introduce a simulation-based methods as an alternative for find CoVaR. Since CoVaR can be seen as the conditional quantile of X_{t+1} given $\mathscr{F}_{t-1}$, our goal here is to sample $\hat{X}_{t+1}$ with information up to time $t-1$. Then it will be easy to find the empirical quantile of $\{\hat{X}_{t+1}\}$ as our estimated CoVaR.

For GARCH process, given $\mathscr{F}_{t-1}$ and estimated volatility at time $t-1$ $\hat{\sigma}_{t-1}$, $\hat{\sigma}_t^2 = \alpha_0 + \alpha_1 \hat{\sigma}_{t-1}^2 + \beta_1 X_{t-1}^2$ and $\hat{X}_t = \hat{\sigma}_t \varepsilon_t \sim N(0, \hat{\sigma}_t^2)$. Therefore, the condition that $\hat{X}_t \geq \text{VaR}_\alpha(\hat{X}_t)$ suggests that $\varepsilon_t \geq \text{VaR}_\alpha(\varepsilon_t)$. So we can sample ε_t first, and for those ε_t's that are greater than $\text{VaR}_\alpha(\varepsilon_t)$ we further calculate $\hat{\sigma}_{t+1}$ as

$$\begin{aligned}\hat{\sigma}_{t+1} &= \sqrt{\alpha_0 + \alpha_1 \hat{\sigma}_t^2 + \beta_1 \hat{X}_t^2} \\ &= \sqrt{\alpha_0 + \alpha_1 \hat{\sigma}_t^2 + (\beta_1 \hat{\sigma}_t^2)\varepsilon_t^2}.\end{aligned}$$

Then $\hat{X}_{t+1} = \hat{\sigma}_{t+1}\varepsilon_{t+1}$ can be calculated by sampling another ε_{t+1} from the distribution of $\{\varepsilon_t\}$. Details about estimating the CoVaR under the GARCH process as described in Algorithm 2.

Algorithm 2 Estimation CoVaR Using Simulation under a GARCH(1,1) Process

1: initialize $\hat{\alpha}_0$, $\hat{\alpha}_1$, $\hat{\beta}_1$, $\hat{\sigma}_{t-1}$, X_{t-1}, $N >> \frac{1}{1-\alpha}$, the cumulative distribution function of ε (F_ε), and $\alpha \in (0,1)$.
2: Calculate $\hat{\sigma}_t$ as $\sqrt{\hat{\alpha}_0 + \hat{\alpha}_1 \hat{\sigma}_{t-1}^2 + \hat{\beta}_1 X_{t-1}^2}$
3: **for** i from 1 to N **do**
4: Generate p_t from Uniform$(\alpha, 1)$
5: Find ε_t' such that $F_\varepsilon(\varepsilon_t') = p_t$
6: Let $\hat{X}_t = \hat{\sigma}_t \varepsilon_t'$
7: Let $\hat{\sigma}_{t+1} = \sqrt{\hat{\alpha}_0 + \hat{\alpha}_1 \hat{\sigma}_t^2 + \hat{\beta}_1 \hat{X}_t^2}$
8: Generate p_{t+1} from Uniform$(0,1)$ and find $\varepsilon_{t+1} = F_\varepsilon^{-1}(p_{t+1})$
9: Calculate $\hat{X}_{t+1,i} = \hat{\sigma}_{t+1}\varepsilon_{t+1}$
10: Save $\hat{X}_{t+1,i}$
11: **end for**
12: Find $\text{CoVaR}_\alpha(\hat{X}_{t+1})$ as the α^{th} quantile of $\{\hat{X}_{t+1,i}, i = 1, ..., N\}$

For the ARSV(1) model, since the AR process of $\log \sigma_t^2$'s does not depend on the value of X_t, we need to simulate X_{t+1} in a different way. The idea here is to sample η_t and calculate the corresponding $\hat{\sigma}_t$ following the AR process. Then we sample ε_t to calculate $\hat{X}_t = \hat{\sigma}_t \varepsilon_t$. We repeat the two steps above until this $\hat{X}_t$ is greater than $VaR_\alpha(X_t)$, which can be estimated based on a rolling window of historical data. Then with this $\hat{\sigma}_t$ and a newly sampled η_{t+1} we can calculate the $\hat{\sigma}_{t+1}$ and sample ε_{t+1} to find $\hat{X}_{t+1}$. The details are described in Algorithm 3.

Algorithm 3 Simulating CoVaR under the ARSV Process

initialize $\hat{\beta}_0$, $\hat{\beta}_1$, distribution of ε_t (F_ε), distribution of η_t (F_η), $\hat{\sigma}_{t-1}$, X_{t-1}, $\alpha \in (0,1)$, $\hat{v} = \text{VaR}_\alpha$ from historical data, and $N >> \frac{1}{1-\alpha}$
for i from 1 to N **do**
    Initialize $\hat{X}_t = 2|\hat{v}|$
    while $\hat{X}_t < \hat{v}$ **do**
        Sample $\eta_t \sim F_\eta$
        Let $\hat{\sigma}_t = \sqrt{\exp(\hat{\beta}_0 + \hat{\beta}_1 \log \hat{\sigma}_{t-1}^2 + \eta_t)}$
        Sample $\varepsilon_t \sim F_\varepsilon$
        Calculate $\hat{X}_t = \hat{\sigma}_t \varepsilon_t$
    end while
    Sample $\eta_{t+1} \sim F_\eta$
    Let $\hat{\sigma}_{t+1} = \sqrt{\exp(\hat{\beta}_0 + \hat{\beta}_1 \log \hat{\sigma}_t^2 + \eta_{t+1})}$
    Sample $\varepsilon_{t+1} \sim F_\varepsilon$
    Calculate $\hat{X}_{t+1,i} = \hat{\sigma}_{t+1} \varepsilon_{t+1}$
    Save $\hat{X}_{t+1,i}$
end for
Find $\text{CoVaR}_\alpha(X_{t+1})$ as the α^{th} quantile of $\{\hat{X}_{t+1,i} : i = 1, ..., N\}$

The results of comparisons of estimated VaR and CoVaR under the second definition are shown in the next section.

4.4 Comparison of VaR and CoVaR Forecasts Under GARCH and ARSV Processes

In this section we present the results of risk forecasts from both simulated data and real data. There are two scenarios of data generating processes for the simulation study:

1. An ARSV(1) process:

$$\begin{aligned} X_t &= \sigma_t \varepsilon_t \\ \log \sigma_t^2 &= -0.5 + 0.95 \log \sigma_{t-1}^2 + 0.35 \eta_t. \end{aligned}$$

   where $\varepsilon_t \sim N(0,1)$, $\eta_t \sim t_5$, and η_t and ε_t are independent.

2. A GARCH(1,1) process:

$$\begin{aligned} X_t &= \sigma_t \varepsilon_t \\ \sigma_t^2 &= 5 \times 10^{-6} + 0.85 \sigma_{t-1}^2 + 0.1 X_{t-1}^2, \end{aligned}$$

   where $\varepsilon_t \sim N(0,1)$.

For the data example, we use the daily log-returns of the S&P 500 Index from 1980 to 1987 to estimate parameter to forecast risk measures of the daily log-returns from 1988 to 2003.

4.4.1 Simulation Study

Value-at-Risk

In the simulation study, we generate the data from both scenarios. The length of each generated dataset is 4500 after the burn-in period. We use the first 1000 observations to estimate parameters of the ARSV(1) process with normally distributed first innovation and Student's t distributed second innovation. Then we estimate the VaR for the rest of the data using a rolling window of size 1000. Note that when estimating VaR at time t, we need to know the volatility at time $t-1$ first. However, our inference method cannot estimate σ_{t-1} since we need to discard the last part of

estimated volatilities in our algorithm. Therefore, we need to use some external estimators for σ_{t-1} when estimating VaR. In this project we choose σ_{t-1}'s estimated when fitting the data in the rolling window to the GARCH model. There are two reasons for us to do so: first, it is more convenience since we are also estimating the VaR under the GARCH process. Second, based on some empirical analysis, we found that σ_{t-1}'s estimated by fitting a GARCH(1,1) model are more accurate than those estimated by some unbiased but noisy model-free estimators. For example, in this simulated dataset, the root MSE of estimated σ_{t-1} from GARCH(1,1) model is $\sqrt{3.7\times10^{-5}}$ compared with $\sqrt{5.0\times10^{-5}}$ which is the MSE of estimated σ_{t-1} from the 5-day close-to-close estimator.

We first compare the estimated VaR based on both ARSV(1) model and GARCH(1,1) model. The VaR forecasts under the GARCH(1,1) model are estimated using model-based method (McNeil, Frey, and Embrechts, 2015b), while the VaR forecasts under the ARSV(1) model are estimated by finding the numerical solutions discussed in Section 4.1. The results are shown in Figure 4.1 and Figure 4.2. From the plots it is hard to observe obvious differences between the forecasts estimated under the ARSV(1) model and those under the GARCH(1,1) model. We also present the results of traditional backtests in Table 4.1, and the results of conditional predictive ability tests (Giacomini and White, 2006) in Table 4.2. The scoring function we use here is the piece-wise linear scoring function suggested by Gneiting (2011). From Table 4.1 we can see that at both the 95% level and the 99% level, the VaR forecasts estimated under the ARSV(1) model and those under the GARCH(1,1) model can pass the traditional backtests no matter what the true underlying process is. However, in Table 4.2 we can find that at the 99% level the GARCH(1,1) model has a stronger conditional predictive ability than the ARSV(1) when the true data generating process is a GARCH(1,1) process. Otherwise, there is no significant difference between the forecasts from an ARSV(1) model and those from a GARCH(1,1) model.

CoVaR

The difference is more obvious when we forecast CoVaR with different filters. We first generate a dataset from the ARSV(1) process as specified in Scenario 1. at

the beginning of Section 4.4. We then estimate CoVaR under the second definition modified by Girardi and Ergün (2013) using the simulation methods described in Algorithm 2 and Algorithm 3. The forecasted CoVaR under the GARCH process and the ARSV process are shown in Figure 4.3 and Figure 4.4. It can be observed that the CoVaR forecasts estimated from the GARCH(1,1) model are in general higher that those from the ARSV(1) model. We can also see that the large difference between the two estimated CoVaR is highly correlated with the large return squared.

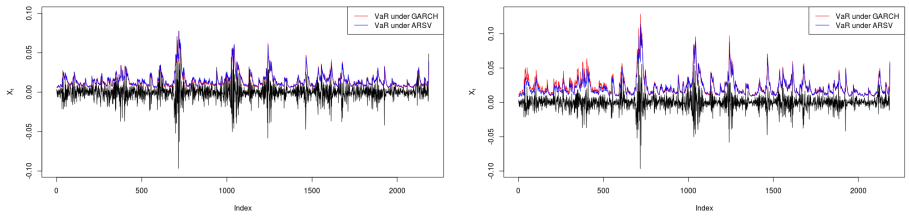

Figure 4.1: Estimated 95% and 99% VaR forecasts for the simulated ARSV(1) process. Black lines represent the simulated daily returns, red lines represent the VaR estimated under the GARCH model, and blue lines represent the VaR under the ARSV model. The left panel illustrates the case of 95% level, while the right panel illustrates the case of 99% level.

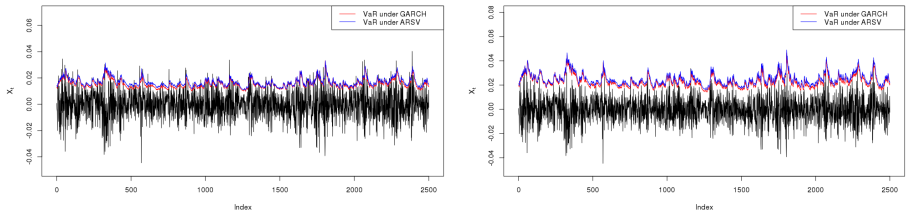

Figure 4.2: Estimated 95% and 99% VaR forecasts for the simulated GARCH(1,1) process. Black lines represent the simulated daily returns, red lines represent the VaR estimated under the GARCH model, and blue lines represent the VaR under the ARSV model. The left panel illustrates the case of 95% level, while the right panel illustrates the case of 99% level.

Table 4.1: Violation rates and corresponding p-values of likelihood-ratio tests for VaR_α forecasts at 95% and 99% levels for simulated Data. Column names are the underlying processes and corresponding risk levels $1-\alpha$, row names are the models used to forecast the risk measures.

	ARSV(1)-95%	ARSV(1)-99%	GARCH(1,1)-95%	GARCH(1,1)-99%
ARSV(1)	4.35% (0.153)	1.19% (0.386)	4.24% (0.074)	0.64% (0.053)
GARCH(1,1)	5.35% (0.452)	1.05% (0.806)	5.24% (0.585)	1.04% (0.842)

Table 4.2: Mean piece-wise linear scores and corresponding p-values of conditional predictive ability tests for VaR_α forecasts at 95% and 99% levels for simulated Data. Column names are the underlying processes and corresponding risk levels $1-\alpha$, row names are the models used to forecast the risk measures.

	ARSV(1)-95%		ARSV(1)-99%		GARCH(1,1)-95%		GARCH(1,1)-99%	
ARSV(1)	9.93×10^{-4}	0.835	2.85×10^{-4}	0.237	1.03×10^{-3}	0.144	2.763×10^{-4}	0.001
GARCH(1,1)	9.97×10^{-4}		2.95×10^{-4}		1.033×10^{-4}		2.756×10^{-4}	

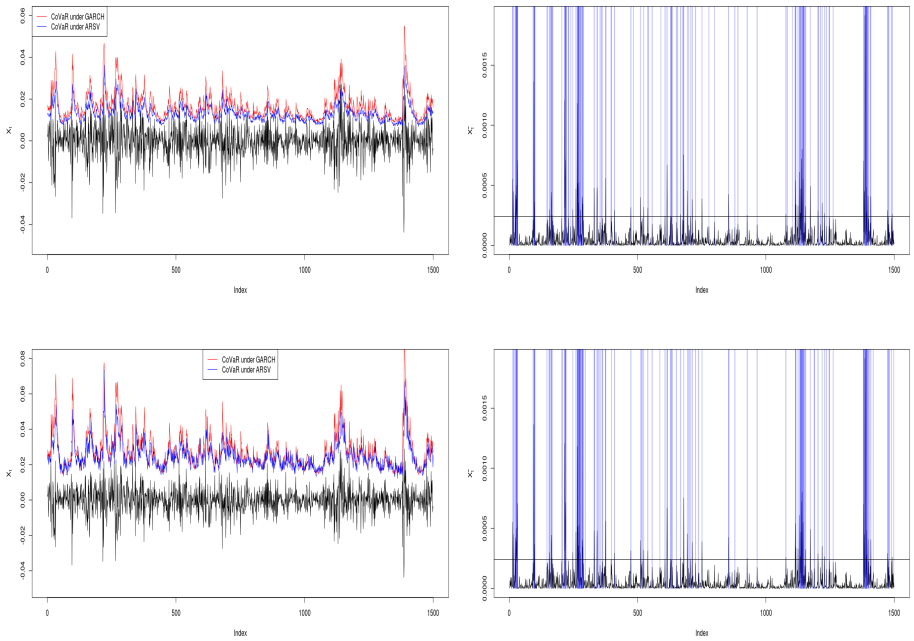

Figure 4.3: Estimated 95% and 99% CoVaR forecasts for the simulated ARSV(1) process. In the left column, black lines represent the simulated daily returns, red lines represent the CoVaR estimated under the GARCH model, and blue lines represent the CoVaR under the ARSV model. In the right column, the black lines represent squared return, and blue beams indicate top 5% largest differences in the estimated CoVaR between GARCH(1,1) model and ARSV(1) model. The top panels illustrate the case of 95% level, while the bottom panels illustrates the case of 99% level.

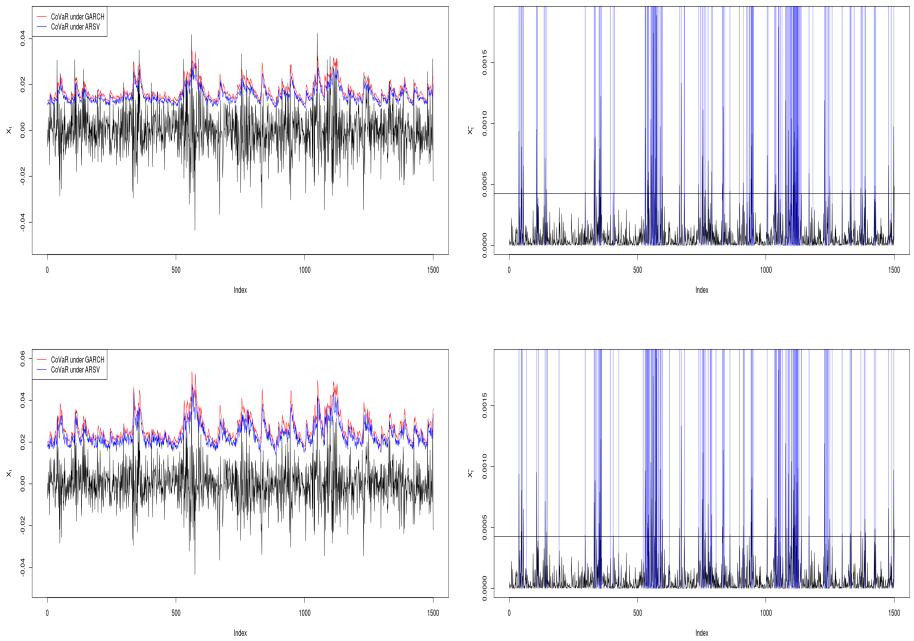

Figure 4.4: Estimated 95% and 99% CoVaR forecasts for the simulated GARCH(1,1) process. In the left column, black lines represent the simulated daily returns, red lines represent the CoVaR estimated under the GARCH model, and blue lines represent the CoVaR under the ARSV model. In the right column, the black lines represent squared return, and blue beams indicate top 5% largest differences in the estimated CoVaR between GARCH(1,1) model and ARSV(1) model. The top panels illustrate the case of 95% level, while the bottom panels illustrates the case of 99% level.

4.4.2 Data Example

To further compare the VaR and CoVaR forecasts under the GARCH model and the ARSV model, we apply the forecasting methods to daily log-returns of S&P 500 Index from 1988 to 2003 with a rolling window of size 1000. The forecasting process is the same as in Section 4.4.1, and the results are presented below.

For the VaR forecasts, we can find that there is no significant difference between the values estimated from the ARSV(1) model and those from the GARCH(1,1) model from Figure 4.5. Both models can provide good forecasts

Table 4.3: Violation rates and corresponding p-values of likelihood-ratio tests for VaR_α forecasts at 95% and 99% levels for daily log-returns of S&P 500 Index. Column names are the risk levels $1-\alpha$, row names are the models used to forecast the risk measures.

	S&P 500-95%	S&P 500-99%
ARSV(1)	4.96% (0.929)	1.31% (0.109)
GARCH(1,1)	5.24% (0.345)	1.34% (0.077)

Table 4.4: Mean piece-wise linear scores and corresponding p-values for comparison of two forecast methods of conditional predictive ability tests for VaR_α forecasts at 95% and 99% levels for daily log-returns of S&P 500 Index. Column names are the risk levels $1-\alpha$, row names are the models used to forecast the risk measures.

	S&P 500-95%		S&P 500-99%	
ARSV(1)	1.06×10^{-3}	0.461	2.85×10^{-4}	0.506
GARCH(1,1)	1.05×10^{-3}		2.79×10^{-4}	

of VaR and pass the traditional backtests (Table 4.3). There is also no significant difference in conditional predictive ability (Table 4.4).

In Figure 4.6 we show the estimated CoVaR forecasts from the ARSV(1) model and GARCH(1,1) model. We can observe a more obvious difference between estimated CoVaR forecasts in the data example, as the values estimated under the GARCH(1,1) model are consistently higher that those under the ARSV(1) model. A possible explanation for this consistent difference could be that the GARCH(1,1) process is asymptotically tail dependent. However, the ARSV(1) process with heavy-tailed second innovation, although has stronger tail dependence than the classic ARSV(1) model, is still asymptotically tail independent. Recall that CoVaR is the conditional quantile of X_{t+1}. Therefore, when conditioned on $X_t \geq \text{VaR}_\alpha(X_t)$ for some α close to 1, X_{t+1} is more likely to also be a large value under the GARCH process than under the ARSV process. So the α^{th} quantile of forecasted X_{t+1} is greater under the GARCH process than that under the ARSV process.

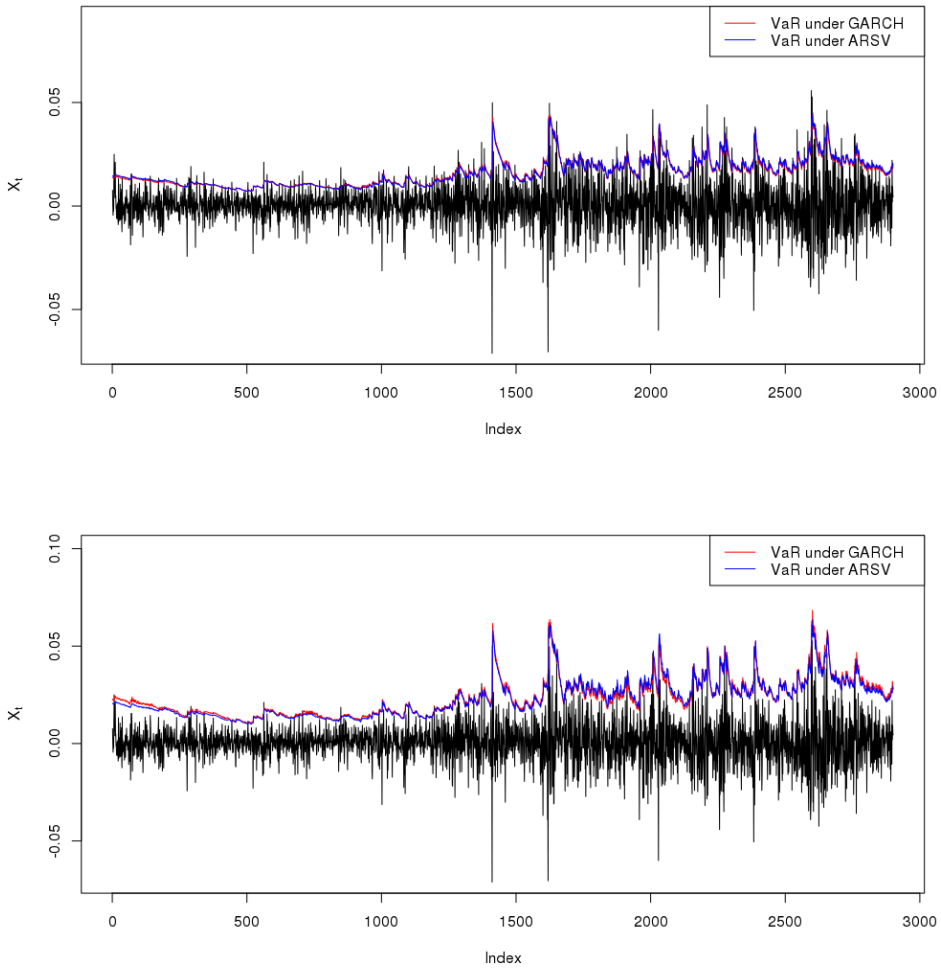

Figure 4.5: Estimated 95% and 99% VaR forecasts for daily log-returns of S&P 500 Index. Black lines represent the simulated daily returns, red lines represent the VaR estimated under the GARCH model, and blue lines represent the VaR under the ARSV model. The top panel illustrates the case of 95% level, while the bottom panel illustrates the case of 99% level.

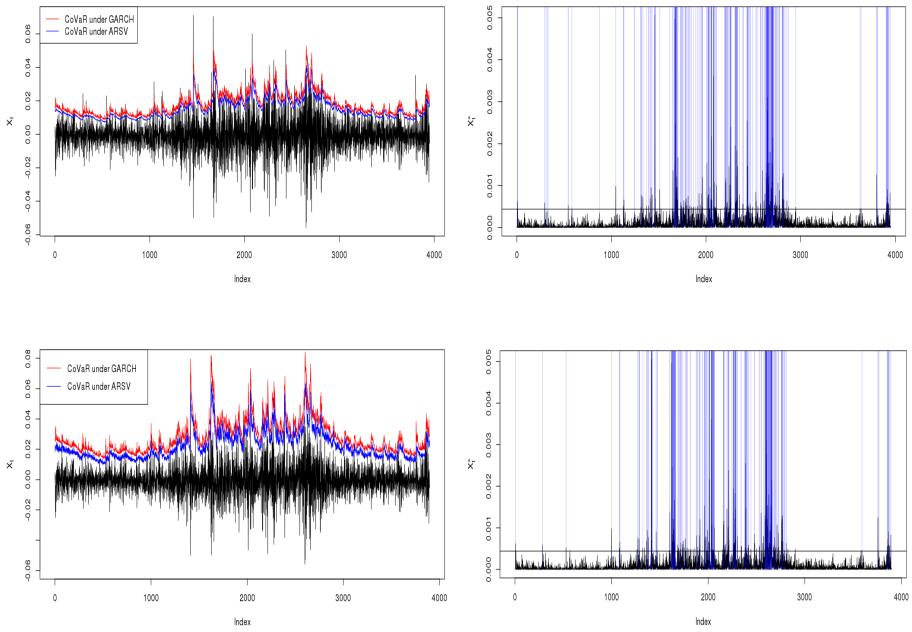

Figure 4.6: Estimated 95% and 99% CoVaR forecasts for daily log-returns of S&P 500 Index. In the left column, black lines represent the simulated daily returns, red lines represent the CoVaR estimated under the GARCH model, and blue lines represent the CoVaR under the ARSV model. In the right column, the black lines represent squared return, and blue beams indicate top 5% largest differences in the estimated CoVaR between GARCH(1,1) model and ARSV(1) model. The top panels illustrate the case of 95% level, while the bottom panels illustrates the case of 99% level.

Chapter 5

Discussion

Empirical evidence suggests that consecutive log-return values may be tail independent. Thus the GARCH process might not be a suitable choice in modeling consecutive losses, as it may over-estimate the potential risk at extremal levels. However, the tail dependence should be stronger at sub-extremal levels than can be modelled by the classic ARSV process with light-tailed second innovation. Therefore, modelling log-returns using the classic ARSV process may lead to under-estimation of the probability of consecutive large losses.

In this thesis report we propose an extension of an ARSV model by taking the second innovation to be Student's t distributed. We conjecture that this model exhibits stronger tail dependence at sub-extremal levels than the classic ARSV model with normally distributed second innovation. Our conjecture is supported via simulation.

However, most existing inference methods for the ARSV process have limited flexibility in the choice of innovation distributions. Most of these methods require the second innovation to be normal, and the first innovation to be either normal or Student's t. There are only few methods allow the second innovation to be heavy-tailed. In this report, we develop a new inference method for the extended ARSV(1) process which also allows flexible distributional assumptions on both innovations. This new method works as well as existing methods in estimating parameters for the classic ARSV(1) model. It can also successfully estimate parameters of the extended ARSV(1) process we consider, which is out of the range

of most existing methods. Furthermore, it has the potential for being adjusted to other extensions of the classic ARSV(1) process. For example, the same scheme can be applied to estimating parameters of an ARSV(p) process for any arbitrary integer $p \geq 1$ and non-Gaussian second innovation, as long as we know how to estimate parameters of the AR(p) process with this desired non-Gaussian error distribution. We hope this new inference method will provide a useful tool for future study of ARSV models.

We also study the VaR and CoVaR risk measures under the ARSV process and compare them with those under the GARCH process. We show that there is no significant difference between estimated VaR under GARCH process and ARSV process. However, we can observe a large difference between the CoVaR estimated under these two processes. This difference reveals the impact of tail dependence properties implied by a chosen model.

For future work, we first need to prove that $\bar{\chi}$ covers the full spectrum of sub-extremal tail dependence; i.e., $0 < \bar{\chi} < 1$, for the extended model. And with the support of a more grounded theory, we also need to improve the computational efficiency. Currently our new inference method is much slower than existing inference methods that similarly require an MCMC scheme. In many cases, the computing speed of the proposed method is up to 100 times slower. Another shortcoming of the new method is that it only provides parameter estimates but volatility estimates can only be obtained for a sub-period of the historical record.

Bibliography

T. Adrian and M. K. Brunnermeier. CoVaR. Technical report, National Bureau of Economic Research, 2011. → pages 36

B. Basrak, R. A. Davis, and T. Mikosch. Regular Variation of GARCH Processes. *Stochastic Processes and Their Applications*, 99:95–115, 2002. → pages 8

T. Bollerslev. Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, 31:307–327, 1986. → pages 2, 7

F. J. Breidt and R. A. Davis. Extremes of Stochastic Volatility Models. *Annals of Applied Probability*, 8:664–675, 1998. → pages 9

C. Broto and E. Ruiz. Estimation Methods for Stochastic Volatility Models: A Survey. *Journal of Economic Surveys*, 18:613–649, 2004. → pages 17

S. Coles, J. Heffernan, and J. A. Tawn. Dependence Measures for Extreme Value Analyses. *Extremes*, 2:339–365, 1999. → pages 6

H. Drees, J. Segers, and M. Warchoł. Statistics for Tail Processes of Markov Chains. *Extremes*, 18:369–402, 2015. → pages 2

R. F. Engle. Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica: Journal of the Econometric Society*, 50:987–1007, 1982. → pages 2, 7

M. Fridman and L. Harris. A Maximum Likelihood Approach for Non-Gaussian Stochastic Volatility Models. *Journal of Business & Economic Statistics*, 16: 284–291, 1998. → pages 26, 32

R. Giacomini and H. White. Tests of Conditional Predictive Ability. *Econometrica*, 74:1545–1578, 2006. → pages 45

G. Girardi and A. T. Ergün. Systemic Risk Measurement: Multivariate GARCH Estimation of CoVaR. *Journal of Banking & Finance*, 37:3169–3180, 2013. → pages 36, 38, 42, 46

T. Gneiting. Making and Evaluating Point Forecasts. *Journal of the American Statistical Association*, 106:746–762, 2011. → pages 45

G. K. Grunwald, R. J. Hyndman, L. Tedesco, and R. L. Tweedie. Theory & Methods: Non-Gaussian Conditional Linear AR(1) Models. *Australian & New Zealand Journal of Statistics*, 42:479–495, 2000. → pages 24

A. Harvey, E. Ruiz, and N. Shephard. Multivariate Stochastic Variance Models. *The Review of Economic Studies*, 61:247–264, 1994. → pages 8

J. B. Hill. Extremal Memory of Stochastic Volatility with an Application to Tail Shape Inference. *Journal of Statistical Planning and Inference*, 141:663–676, 2011. → pages 9

E. Jacquier, N. G. Polson, and P. E. Rossi. Bayesian Analysis of Stochastic Volatility Models. *Journal of Business & Economic Statistics*, 12:69–87, 1994. → pages 8, 15, 16, 17

E. Jacquier, N. G. Polson, and P. E. Rossi. Bayesian Analysis of Stochastic Volatility Models with Fat-tails and Correlated Errors. *Journal of Econometrics*, 122:185–212, 2004. → pages 16

A. Janssen and H. Drees. A Stochastic Volatility Model with Flexible Extremal Dependence structure. *Bernoulli*, 22:1448–1490, 2016. → pages 3, 7, 10

H. Joe. *Multivariate Models and Multivariate Dependence Concepts*. CRC Press, 1997. → pages 5

G. Kastner. *stochvol: Efficient Bayesian Inference for Stochastic Volatility (SV) Models*, 2016. URL https://CRAN.R-project.org/package=stochvol. R package version 1.3.1. → pages 28

G. Kastner and S. Frühwirth-Schnatter. Ancillarity-sufficiency Interweaving Strategy (ASIS) for Boosting MCMC Estimation of Stochastic Volatility Models. *Computational Statistics & Data Analysis*, 76:408–423, 2014. → pages 17

F. Laurini and J. A. Tawn. Regular Variation and Extremal Dependence of GARCH Residuals with Application to Market Risk Measures. *Econometric Reviews*, 28:146–169, 2008. → pages 7, 8

A. W. Ledford and J. A. Tawn. Statistics for Near Independence in Multivariate Extreme Values. *Biometrika*, 83:169–187, 1996. → pages 6

A. W. Ledford and J. A. Tawn. Diagnostics for dependence within time series extremes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(2):521–543, 2003. → pages 6

Y. Liu and J. A. Tawn. Volatility Model Selection for Extremes of Financial Time Series. *Journal of Statistical Planning and Inference*, 143:520–530, 2013. → pages 9

A. J. McNeil, R. Frey, and P. Embrechts. Quantitative Risk Management, 2015a. pg. 133. → pages 1

A. J. McNeil, R. Frey, and P. Embrechts. Quantitative Risk Management, 2015b. pg. 48. → pages 45

T. Mikosch and C. Starica. Limit Theory for the Sample Autocorrelations and Extremes of a GARCH(1, 1) Process. *Annals of Statistics*, 28:1427–1451, 2000. → pages 8

K. P. Murphy. *Machine Learning: A Probabilistic Perspective*. MIT Press, 2012. → pages 20

G. O. Roberts and J. S. Rosenthal. Harris Recurrence of Metropolis-within-Gibbs and Trans-dimensional Markov Chains. *The Annals of Applied Probability*, 16: 2123–2139, 2006. → pages 21