Summer 8-9-2019

GARCH Modeling of Value at Risk and Expected Shortfall Using Bayesian Model Averaging

Ismail Kheir
CUNY Hunter College

GARCH Modeling of Value at Risk and Expected Shortfall Using Bayesian Model Averaging

by

Ismail Kheir

Submitted in partial fulfillment
of the requirements for the degree of
Master of Arts in Economics, Hunter College
The City University of New York

2019

Thesis Sponsor:

August 9, 2019	Temisan Agbeyegbe
Date	Signature
August 9, 2019	Matthew Baker
Date	Signature of Second Reader

Acknowledgments:

A great deal of thanks goes to Professor Agbeyegbe and Professor Baker for their guidance and shaping my understanding in this arduous but fulfilling process. I've learned a lot about time series modeling, probability theory, and finance through completing this thesis and hope to build on my understanding for years to come. I would also like to thank my mom, dad, and sisters for their encouragement and support for my pursuit of higher education. Last but not least, I would like to thank my peers for embarking on similar journeys alongside me.

Table of Contents

1) Introduction
2) Literature Review
3) Characteristics of Financial Data
    - High Kurtosis
    - Volatility Clustering
    - Long Memory
    - Leverage Effects
    - Spillover Effects
    - Skewness
4) Definitions
    - Value at Risk
    - Expected Shortfall
    - Prior Probability Distribution
    - Posterior Probability Distribution
    - Bayesian Model Averaging
    - Monte Carlo
    - Maximum Likelihood Estimation
    - NIG-CIR
    - Time Series Models
5) Data
6) The Mixed Model
7) Rolling Window Estimation
    - Parameter Estimation
    - Simulation Paths
8) Model Weights
9) Value at Risk and Expected Shortfall Calculation
10) Estimation Results
11) Model Fit Diagnostics
12) Conclusion

Abstract

This thesis conducts Value at Risk (VaR) and Expected Shortfall (ES) estimation using GARCH modeling and Bayesian Model Averaging. Existing research often estimates Value at Risk and Expected Shortfall using a single model at a time. Afterwards comparison of the performance of different models takes place. Model averaging provides an alternative, by considering multiple models simultaneously weighted by a function of some information criterion. To estimate model parameters, in this thesis, a rolling window estimation is used on a dataset generated from a Normal Inverse Gaussian – Cox Ingersoll Ross distribution. The process provides a great deal of flexibility in the data that can be generated from it. This thesis finds that value at risk and expected shortfall estimates can be improved upon, when compared with single, or non-averaged models.

1 Introduction

One who models some financial data process will quickly run into the difficulty of determining the correct model to use due to model uncertainty. This uncertainty is called model risk, an unavoidable product of using models. It is the risk attributable to a misspecified model or incorrectly assuming the underlying assumption of the model has been met. For a financial firm, model risk can lead to losses, and thus model risk must be minimized. Choosing the correct model is not a simple task as the model space, or the set of all potential models, is large. Regardless of the type of study, picking the correct model to use is a difficult choice. As such, there is risk in choosing a model as a misspecified model may yield suboptimal estimations. Moreover, even choosing one good model is not always enough.

There are a few approaches to dealing with model risk. The first being picking the model with the highest performance, but this approach is problematic for a couple of reasons. As said before, determining the best model in the model space is a difficult task on its own. Secondly, a model that works well for one set of data may not work as well on a different set of data. Another method is to take the estimations of several models and do a simple average between them. The issue with this approach is that it does not

consider the relative performance of each model. If a model's performance is better or worse than another, it should be weighted differently by some criterion. This thesis' method of minimizing model risk is Bayesian Model Averaging (BMA). Bayesian Model Averaging acknowledges that the true model is unknown, therefore it takes multiple models and assigns weights based on some information criterion. The advantage BMA provides is the incorporation of insights from all the models while minimizing uncertainty based on the choice of model.

The unique properties of financial data are another reason why BMA is important to this modeling task. Financial data exhibits behavior such as skewness, autocorrelation, volatility clustering, fat tails return asymmetry, and slowly decaying volatility correlation Cont (2001). These traits not only render simpler modeling methods relatively ineffective but will likely warrant the use of BMA as financial data processes can change their patterns of behavior over time.

This thesis deals with value at risk (VaR) and expected shortfall (ES), two different measures of portfolio risk used in finance. Value at risk seeks to assign a probability that a portfolio will lose a certain percentage in returns. Similarly, expected shortfall estimates the probability that a portfolio will lose a certain percentage or more in returns by averaging out the sampled returns below a certain quantile of data. In this thesis, estimation of VaR and ES occurs by modeling the data generation process, and then calculating estimates for VaR and ES using data generated from the models. The common practice in statistics is to develop one model which is supposed to represent the data generation process. Using a diverse group of models with BMA captures different aspects of the unknown data generation process and yields estimates with a greater degree of certainty.

The plan of this thesis is as follows: Section 2 of this thesis provides a review of the relevant literature in estimating value at risk and expected shortfall and using Bayesian Model Averaging. Section 3 discusses the characteristics of financial data. Section 4 iterates definitions needed to understand the methodologies used in this thesis. Section 5 outlines the data creation process. Section 6 reveals the models used in estimation. Section 7 discusses the rolling window estimation, section 8 is about BMA and the choice of

model weights. Section 9 involves the calculation of VaR and ES using mixed model data, while section 10 displays and analyzes the results. Section 11 is concerned with discussion of model fits including the mixed model.

2 Literature Review

The literature is not short of research illustrating methodology for estimating financial data. In the past, financial concepts have relied upon the assumption of asset returns following a normal distribution. However, this is an erroneous assumption as large losses or gains occur more frequently than what is allowed by the normal distribution. Moreover, the normal distribution is a symmetric distribution, but empiricism shows that the probabilities of equal percentages of losses or gains do not need to be equal Harmantzis (2005).

Time changed Lévy processes were proposed by Carr et al. (2003) and Carr and Wu (2004) as vanilla Lévy processes and stochastic volatility models were insufficient for modeling prices. The reasons for this deal with the stylized facts of financial data. Time changed Lévy processes are Lévy processes where the volatility is governed by a stochastic clock. Time changed Lévy processes are further defined in section 4.3. A popular choice for a time changed Lévy process is the NIG-CIR process. Bottern (2015) and Schoutens, Symens (2002) utilize NIG-CIR processes for their estimations. The NIG-CIR process is an excellent choice due to the flexibility of the data that can be generated from it.

A source of differences between this thesis and those by Bottern (2015) and Schoutens, Symens (2002) lies in the choices made in the data generation process. Parameters must be chosen to generate NIG-CIR data, and these parameters affect the shape of the NIG distribution and the behavior of the CIR process. While this thesis models its parameter set after the parameters used in Bottern (2015), it does not change the parameter set halfway through the data generation process. The decision by Bottern was made to test how the models react to different conditions Bottern (2015), instead a choice is made to fully observe the behavior of each model holding the parameters constant. Schoutens and Symens choose a different set of

parameters for the NIG-CIR data altogether. The choice of parameters is significant in that they control the behavior of the data generated. As a result of BMA estimation, this can have a cascading effect on the model fits, the weights that are assigned, and the estimation results.

The use of Bayesian model averaging in economics and finance papers is ubiquitous, and its merits well documented. Masih (2010) comparatively discerns the quality of short-term and long-term forecasts of asset returns using simple averaging vs Bayesian Model Averaging. Their explanatory paper finds that BMA models outperform simple averaged models and random walk with drift models consistently if longer term forecasts are used Masih (2010).

To analyze the determinants of bond yield spreads Maltritz, Molchanov (2013) use BMA to determine the best predictors through estimating different models. Specifying a model that predicts well across countries was an especially difficult task for a couple of reasons. One reason was the variables that influenced default risk could potentially be significant but dependent on the country. Several variables exist that determine a country's ability and willingness to make bond payments and these variables can vary between countries. Furthermore, the lack of theoretical guidance on the determinants of default risk make choosing components of the model difficult Maltritz, Molchanov (2013). Since BMA fits models to data, the choice of weights depends on the evidence provided by the data and thus important determinants of bond yields were identified Arin, Braunfels (2018).

Following Bottern (2015) this thesis utilizes Bayesian model averaging to incorporate insights from multiple different models. However, it diverges from Bottern (2015) in the choice of models used to estimate value at risk and expected shortfall. Bottern fits data to distributions such as the normal distribution and the student's t distribution which by the author's own omission do not lend themselves well to financial data Bottern (2015). For this reason, this thesis does not fit these distributions to the data generated and have been excluded from the model mix.

Normalized model weights are also varied for each parameter set created by rolling window estimation in Bottern (2015). Instead, this thesis elects to average the information criterion across all parameter sets for every model and create one consistent set of normalized weights for every model across all parameter sets. This decision was made to generalize the behavior of each model that contributes to the mixed model across time.

3 Characteristics of Financial Data

Financial data has unique characteristics relative to other time series data that make it more difficult to model than other processes. In this section those characteristics such as, high kurtosis, volatility clustering, long memory, leverage effects, spillover effects, and skewness will be explained and discussed.

3.1 High Kurtosis

Sampling data from a distribution with high kurtosis will yield more data points further from the mean than a distribution with lower kurtosis. In the past, financial theory assumed normality in the return of financial assets. More recently, this was shown to be inaccurate relative to the more realistic assumption of a Levy distribution Ding (2011). Because of the poor performance of the normal distribution in the context of financial data, other distributions were tried and tested such as the t distribution Bollerslev (1987), the normal mixture distribution of a Poisson Jorion (1988), the power exponent distribution Baillie, Bolleslev (1989), and the expansion of the exponential distribution Nelson (1991).

3.2 Volatility Clustering

As is commonly known, volatility clustering is defined as the tendency for large changes to follow large changes while small changes follow small changes in a time series process. External forces affect the volatility of a process, and this has lasting impact Ding (2011). Modeling processes with volatility clustering involves estimating the volatility at any point in time. For this reason, models which can estimate conditional variance like ARCH and GARCH emerged.

3.3 Long Memory

Long memory refers to the persistence of the effects of a volatility shock, which often have slow decay. One can detect this persistence through autocorrelations in the measures of volatility Ding (2011). It was found by Fama, French (1988) and Poterba, Summers (1987) that in the short term there is positive correlation in stock returns, while in the long term there is negative correlation.

3.4 Leverage Effects

Leverage effects describe the difference in influence that good news and bad news have on the volatility of stock returns. Usually, it is the case that bad news causes more intense fluctuations in volatility as compared to good news. A few models emerged to model this asymmetry, including the EGARCH model Nelson (1991), the GJR-GARCH model Glosten, Jagannathan, Runkle (1993), and the "Asymmetric Power" APARCH model Ding, Granger, Engle (1993).

3.5 Spillover Effects

Spillover effects are effects which come from sources not directly involved in the market of the affected. For instance, financial markets are often intertwined with each other, where one financial market can affect the performance of another. Spillover effects are more apparent between developed economics Ding (2011) as there is a greater degree of connectivity between developed countries than those between developing countries.

3.6 Skewness

Skewness is the degree of distortion from the symmetric normal distribution in a set of data. It is well known that returns in financial data are asymmetric. The probability of losses is usually not equal to the probability of gains. Along with kurtosis, skewness is used to predict the likelihood of events residing within the tails of a probability distribution.

4 Definitions

This section defines terms and equations needed to understand this thesis. It begins by formally defining financial terms such as value at risk and expected shortfall. Then statistical definitions, and methodologies are explained. Afterwards, the data distributions and processes are discussed. Following this, several model definitions are provided, building up from their simpler ancestors to the variants of GARCH models that comprise the mixed model.

4.1 Value at Risk

As previously stated, value at risk (VaR) is a risk measure for financial portfolios. value at risk calculates the smallest value l such that losses are greater than l with probability α. The value l also provides the converse insight such losses will not exceed l with probability $1 - \alpha$. In other words, VaR is the quantile function of the loss distribution.

Let Δ be a fixed time horizon, the loss distribution and the loss distributions distribution function are defined in Bottern (2015) as:

$$L_{[s,s+\Delta]} \coloneqq -(V(s+\Delta) - V(s))$$

$$F_L(l) = P(L \leq l)$$

Value at risk is defined as such:

$$VaR_\alpha = \inf\{\, l \in (R) : P(L > l) \leq 1 - \alpha \,\}$$

$$= \inf\{\, l \in (R) : F_L(l) \geq \alpha \,\} = q_\alpha(F_L)$$

As a measure of risk, VaR does not satisfy all four axioms of coherence. The axioms of coherence for a risk measure ϱ: $M \rightarrow \mathbb{R}$ on the convex cone M, is as such:

1) Translation invariance: For all $L \in M$ and every $l \in \mathbb{R}$, $\varrho(L + l) = \varrho(L) + l$
2) Subadditivity: For all $L_1, L_2 \in M$, $\varrho(L_1 + L_2) \leq \varrho(L_1) + \varrho(L_2)$
3) Positive homogeneity: For all $L \in M$ and every $\lambda > 0$, $\varrho(\lambda L) = \lambda\varrho(L)$

4) Monotonicity: For $L_1, L_2 \in M$ such that $L_1 \leq L_2$ almost surely, $\varrho(L_1) \leq \varrho(L_2)$

VaR fails on the axiom of subadditivity, which means VaR is not a coherent risk measure. Alternatively, Expected Shortfall does meet all axioms to be considered a coherent risk measure.

4.2 Expected Shortfall

Expected shortfall (ES) is another financial risk measure that is like value at risk, but expected shortfall averages out the returns in the tail of the loss distribution that exceed the value at risk. This makes it the more conservative risk measure as it better accounts for tail risk. Tail risk is the risk of the returns of an asset or portfolio of assets moving more than 3 standard deviations away from the mean is greater than what would be expected assuming a normal distribution. Averaging out the returns in the tail of a loss distribution accounts for extreme values and thus provides a more cautious "worst case scenario" risk measure.

Iterating the loss distribution and loss distributions distribution function again:

$$L_{[s,s+\Delta]} \coloneqq -(V(s+\Delta) - V(s))$$

$$F_L(l) = P(L \leq l)$$

Expected shortfall is defined in a few equivalent ways:

$$ES_\alpha = \frac{E(L; L \geq q_\alpha(L))}{1-\alpha} = E(L|L \geq VaR_\alpha)$$

$$ES_\alpha = \frac{1}{1-\alpha}\int_\alpha^1 q_u(F_L)du$$

$$ES_\alpha = \frac{1}{1-\alpha}\int_\alpha^1 VaR_u(L)du$$

Thus, expected shortfall is calculated by averaging VaR over all levels $u \leq \alpha$.

4.3 Prior Probability Distribution

A prior probability distribution of an uncertain quantity expresses the modeler's beliefs about the uncertain quantity or random variable before evidence is taken into consideration. This is also known as the unconditional probability because data is not considered before its assignment.

In further sections, the prior probability distribution will be referred to as $p(\theta)$.

4.4 Posterior Probability Distribution

A posterior probability distribution is the probability distribution of an uncertain quantity or random variable that is assigned after evidence is considered.

The equation for a posterior probability distribution is defined as:

$$p(\theta|x) = \frac{p(x \mid \theta)\, p(\theta)}{p(x)}$$

Where $p(\theta)$ is the prior distribution, x are the observations, $p(x)$ is the prior probability of the observation, and $p(x \mid \theta)$ is the likelihood of data given the prior.

4.5 Bayesian Model Averaging (BMA)

Bayesian model averaging is an ensemble learning methodology that recognizes that the true model is unknown and instead relies on an information criterion to weight models. A situation where models are equally weighted emphasizes the importance of BMA as this indicates that each model is equally proficient in modeling the data. Moreover, BMA can also indicate which models are more proficient than others as indicated by the weight assigned to each model. A larger weight relative to others indicates greater relative importance to the modeling. For an in-depth explanation of BMA, this section follows closely the BMA section of Bottern (2015).

BMA finds $\pi(\mu|y)$, which is the posterior density of μ given y, nonconditional on any of the models. First, BMA specifies the prior probabilities and prior densities of each model, $P(M_j)$ and $\pi\left(\theta_j \middle| M_j\right)$

respectively, where θ_j is a vector of parameters for model M_j. The likelihood of model M_j can then be represented as:

$$P(M_j|y) \propto \lambda_{n,j}(y) = \int \mathcal{L}_{n,j}(y, \theta_j)\pi(\theta_j|M_j)d\theta_j$$

Where $\mathcal{L}_{n,j}$ is the likelihood function for model M_j and $\lambda_{n,j}(y)$ is the marginal density of the unobserved data. The posterior density of model M_j can then be derived using Bayes theorem:

$$P(M_j|y) = \frac{P(M_j)\lambda_{n,j}(y)}{\sum_{i=1}^{k} P(M_i)\lambda_{n,i}(y)}$$

The posterior density of μ can then be calculated as:

$$\pi(\mu|y) = \sum_{j=1}^{k} P(M_j|y)\pi(\mu|M_j, y)$$

The posterior density of μ, or $\pi(\mu|y)$ does not assume one model to be true, but rather is a weighted average of the conditional posterior densities of all k models being considered. Conditioning on other models yields the posterior mean equation:

$$E(\mu, y) = \sum_{j=1}^{k} P(M_j \mid y)E(\mu \mid M_j, y)$$

And the posterior variance:

$$V(\mu|y) = \sum_{j=1}^{k} P(M_j \mid y)\left[V(\mu \mid M_j, y) + (E(\mu|M_j, y) - E(\mu|y))^2\right]$$

However, currently the weights for BMA are simply the posterior probabilities of each model. To normalize the weights so that they sum to one the equation is as follows:

$$w_i = e^{-\frac{1}{2} * \frac{\sum_{j=1}^{K} BIC_{i,j}}{K}}$$

Where K is the number of parameter sets. This leaves the definition for the information criterion, BIC.

4.5.1 Bayesian Information Criterion (BIC)

In this thesis, Bayesian information criterion is the measurement of fit that is used to gauge how well a model fits data. BIC uses the number of parameters as a proxy for model complexity and punishes complex models with higher BIC values. A low value for BIC is indicative of a good model fit.

The equation for the BIC of a model m is as follows

$$BIC_m = \log(N) * K - 2 * LL$$

Where N refers to the size of the data fit to the model, K refers to the number of parameters in the model, and LL refers to the maximized log-likelihood value for the model. As observed, BIC penalizes models with a high number of parameters. In the Bayesian approach, model selection is done by picking the model with the best fit, posteriori. To explain BIC further we recall Bottern (2015) in this section. The posterior probabilities of the set of models $M_1, \dots, M_k$ are derived using Bayes theorem:

$$P(M_j|y) = \frac{P(M_j)}{f(y)} \int_{\Theta_j} f(y|M_j, \theta_j)\, \pi(\theta_j|M_j) d\theta_j$$

Where Θ_j is the parameter space, or the set of all possible parameters, and $\theta_j \in \Theta_j$. $y_1, \dots, y_n$ is the data, $P(M_j)$ is the prior probability of model M_j, $f(y)$ is the unconditional likelihood of the data y, $f(y\,|M_j, \theta_j) = \mathcal{L}_{n,j}(\theta_j)$ represents the likelihood function for the data, and $\pi(\theta_j|M_j)$ is the prior density of θ_j given the data. The unconditional likelihood of the data is specified as:

$$f(y) = \sum_{j=1}^{k} M_j \lambda_{n,j}(y)$$

Where $\lambda_{n,j}$ is the marginal likelihood of model j with θ_j integrated with respect to the prior

$$\lambda_{n,j} = \int_{\Theta_j} \mathcal{L}_{n,j}(\theta_j)\pi(\theta_j|M_j)d\theta_j$$

Since $f(y)$ is a summation across all the models, it is a constant value and unimportant to comparing the posterior probabilities $P(M_j|y)$ with respect to the different models. However, its component $\lambda_{n,j}$ is important to calculating an exact BIC, and sequentially, calculating the posterior probabilities $P(M_j|y)$.

$$BIC_{n,j}^{exact} = 2\log \lambda_{n,j}(y)$$

$$P(M_j|y) = \frac{P(M_j)e^{(\frac{1}{2}BIC_{n,j}^{exact})}}{\sum_{i=1}^{k} P(M_j)e^{\frac{1}{2}BIC_i^{exact}}}$$

The trouble with calculating $BIC_{n,j}^{exact}$ is that the numeric computations are very difficult, instead an approximation is used. This calculation starts with using a Laplace approximation for $\lambda_{n,j}(y)$

$$\lambda_{n,j} = \int_{\Theta} e^{nh_{n,j}(\theta)}\pi(\Theta|M_j)d\theta$$

Where $h_{n,j} = \frac{\ell_{n,j}(\theta)}{n}$ and ℓ is the log likelihood function $log\mathcal{L}$. Using the Laplace approximation gives:

$$\int_{\Theta} e^{nh(\theta)}g(\theta)d\theta = \left(\frac{2\pi}{n}\right)^{\frac{p}{2}} e^{nh(\theta_u)}\left(g(\theta_0)\,|J(\theta_0)|^{-\frac{1}{2}} + O(n^{-1})\right)$$

Where p is the length of θ, θ_0 is the value that maximizes h and therefore the log likelihood, and J is the Hessian matrix. The approximation is exact if h has a negative quadratic form and g is a constant. The maximum likelihood estimator $\widehat{\theta_j}$ of model M_j is the maximizer of $h_{n,j} = \frac{\ell_{n,j}(\theta)}{n}$ and $J_{n,j}(\hat{\theta}_j)$ defines the Fisher information matrix used in the marginal likelihood of M_j equation that follows:

$$\lambda_{n,j}(y) \approx \mathcal{L}_{n,j}(\hat{\theta})\left(\frac{2\pi}{n}\right)^{\frac{p}{2}} |\, J_{n,j}(\hat{\theta}_j)\,|^{-1/2}\pi(\hat{\theta}_j|M_j)$$

Doing a logarithmic transformation of $\lambda_{n,j}(y)$ and multiplying with 2, yields:

$$BIC^*_{n,j} \approx 2\log\lambda_{n,j}(y) = 2l_{n,j}\left(\hat{\theta}_j\right) - p_j \log(n) + p_j \log(2\pi) - log\left|J_{n,j}\left(\hat{\theta}_j\right)\right| + 2log\pi_j(\hat{\theta}_j)$$

Since the first two terms are dominant, the $BIC^*_{n,j}$ equation can be simplified to:

$$2log\lambda_{n,j}(y) \approx BIC_{n,j} = 2\ell_{n,j,max} - p_j\log(n)$$

$$= 2\ell(M) - \log(n)\, p$$

$$= 2\ell(M) - \log(n)\dim(M)$$

$$= 2\ell(M) - \log(n)\,\mathrm{K}$$

In this form, the BIC with the highest value has the best model fit. As defined previously, the BIC equation can also take the form:

$$BIC(M) = -2log\mathcal{L}(M) + \log(n)\dim(M)$$

$$BIC(M) = -2log\mathcal{L}(M) + \log(n)\,\mathrm{K}$$

$$BIC(M) = \log(N) * K - 2 * LL$$

4.6 Monte Carlo Methods

Monte Carlo methodology is a broad class of computational algorithms that rely on repeated random sampling to obtain numerical results. Monte Carlo was inspired by problems that are too difficult to calculate deterministically, and the underlying concept is that randomness can be used to approximate solutions to these deterministic problems. For the mathematical explanation of Monte Carlo methodology, this section is guided by the section found in Bottern (2015).

Monte Carlo integration is a common application of the methodology. Consider the integral.

$$\tau = E\big(\phi(X)\big) = \int \phi(x) f(x) dx$$

Where ϕ: $\mathbb{R}^d \mapsto \mathbb{R}, X \in \mathbb{R}^d$ and f is the probability density of X. Simply put, the integral finds the expected value of $\phi(X)$, through integrating the values of $\phi(X)$ with their respective probabilities. The probabilities correspond to ϕ being the indicator function or as it is also known the characteristic function.

$$P(X \in A) = \int 1\{x \in A\} f(x) dx$$

An approximation to τ is reached according to the following equation by the law of large numbers:

$$t_n = t(x_1, \dots, x_N) = \frac{1}{N} \sum_{i=1}^{N} \phi(x_i)$$

Where $x_1, \dots, x_N$ are independently drawn from f. In this thesis, Monte Carlo is used to calculate value at risk and expected shortfall by generating random numbers from the distributions that comprise the model mix. Value at risk is simply the α quantile of the random numbers. Expected shortfall is calculated as the average of the numbers that exceed the VaR-limit.

$$ES_\alpha = E(L|L \geq VaR_\alpha) = \frac{1}{N_\alpha} \sum_{i=1}^{N} 1\{L_i \geq VaR_\alpha\} L_i$$

Where N_α is the number of samples that exceeds the VaR limit. With expected shortfall, it is important to calculate an appropriate sample size that will ensure enough data points in the tail of the distribution such that the calculation may be precise. The binomial distribution can be used to calculate that sample size. Let $\hat{p}$ be equal to:

$$\hat{p} = \frac{Bin(N, p)}{N}$$

Where p is the confidence level, N is the sample size, and $Bin(N, p)$ is a binomial distribution with N trials, each trial with a probability of success p. To find N_α the following equations can be used:

$$E(\hat{p}) = \frac{Np}{N} \qquad V(\hat{p}) = \frac{Np(1-p)}{N^2} \approx \frac{p}{N}$$

$$D(\hat{p}) = \sqrt{V(\hat{p})} = CE(\hat{p})$$

Substituting $E(\hat{p})$ into $D(\hat{p})$ yields the equation for sample size:

$$Cp = \sqrt{\frac{p}{N}}$$

$$N = \frac{1}{pC^2}$$

Putting the equation in terms of α:

$$N_\alpha = \frac{1}{\alpha C^2}$$

Where C is a constant and α is the confidence level for the risk measure.

4.7 Maximum Likelihood Estimation

Maximum Likelihood Estimation (MLE), is the algorithm that is used in this thesis to maximize log likelihood. When the data is known, MLE chooses parameter values such that the likelihood function is maximized given the observed results.

$$\hat{\theta}_{ML} = argmax_\theta f_x(x_1, \dots, x_n \mid \theta) = argmax_\theta \mathcal{L}(\theta, x)$$

Where x is the data, $f_x(x_1, \dots, x_n|\theta)$ is the pdf of the data and $\mathcal{L}(\theta, x)$ is the likelihood function.

4.8 NIG-CIR

The Normal Inverse Gaussian (NIG) - Cox Ingersoll Ross (CIR) process subordinates sample draws from an NIG distribution, by a stochastic time component determined by the CIR process. Data generated from this process will exhibit stochastic changes in volatility due to the changes in the stochastic clock modeled by the CIR process.

4.8.1 Normal Inverse Gaussian Process

The Normal Inverse Gaussian (NIG) distribution with parameters $\alpha > 0, -\alpha < \beta < \alpha$ and $\delta > 0$ has a characteristic function given defined as:

$$\phi NIG(u; \alpha, \beta, \delta) = \exp(-\delta(\sqrt{\alpha^2 - (\beta + iu)^2} - \sqrt{\alpha^2 - \beta^2}))$$

Where α controls the tail heaviness, β is the asymmetry parameter, and δ is the scale parameter.

And the NIG process is defined as:

$$I_k \sim IG(1, \delta\sqrt{\alpha^2 + \beta^2})$$

$$n_k = \delta^2 \beta I_k + \delta\sqrt{I_k} u_k$$

$$X_0 = 0, X_t = X_{t_{k-1}} + n_k, k \geq 1$$

Where I_k are Inverse Gaussian distributed numbers, n_k is the NIG process at time k, and u_k are normally distributed random numbers. The NIG process starts with a value 0, and at any points afterwards are determined by the sum of the previous point and n_k.

4.8.2 The Cox Ingersoll Ross Process

The Cox Ingersoll Ross (CIR) process combines innovations and mean reversion. Given y_0, the characteristic equation of Y_t is written as:

$$\phi_{CIR}(u, t; \kappa, \eta, \lambda, y_0) = E[\exp(iuY_t)[y_0] = \frac{\exp\left(\frac{\kappa^2 \eta t}{\lambda^2}\right) \exp(\frac{2y_0 iu}{\kappa + \gamma \coth\left(\frac{\gamma t}{2}\right)})}{(\cosh\left(\frac{\gamma t}{2}\right) + \kappa \sinh(\gamma t/2)/\gamma)^{2\kappa\eta/\lambda^2}}$$

Where $\gamma = \sqrt{\kappa^2 - 2\lambda^2 iu}$. As a stochastic differential equation, the following represents the process in terms of small incremental changes, this is used as the rate of time change:

$$dy_t = \kappa(\eta - y_t)dt + \lambda y_t^{\frac{1}{2}} dW_t, \qquad y_0 \geq 0$$

Where κ is the rate of mean reversion, η is the long run mean, y_t is the previous value of the process, dt is the amount of time that has passed, λ is coefficient of variance, and dW_t is a strict white noise process with distribution $\sim N(0,1)$.

And the process definition is:

$$y_{t_n} = y_{t_{n-1}} + \kappa(\eta - y_{t_{n-1}})\Delta t + \lambda y_{t_{n-1}}^{\frac{1}{2}}\sqrt{\Delta t}W_t$$

Where Δt is the change in time. Since creating a CIR process must be discretized for computational reasons, it is important to pick a small value for Δt to accurately produce one.

4.8.3 The NIG-CIR Process

The data used in this thesis was generated with a MATLAB script taken from Kienitz, Wetterau (2012). A set of instructions that can be used to generate this process will be described here for completeness. It is worth noting this is not the only way to generate an NIG-CIR process.

To create an NIG-CIR process one creates an NIG process sampled from the NIG distribution, using a time changed delta. The time changed delta $\delta\Delta t$, is calculated by multiplying the NIG parameter δ, with the change in the CIR process which is represented as Δt in the following equation.

$$n_k \sim NIG(\alpha, \beta, \delta\Delta t)$$

4.9 Time Series Models

Despite the unique properties of financial data, there are a multitude of options in modeling them. Namely multiple variations of GARCH models and stochastic volatility models. Before announcing the models that comprised the mixed model, definitions of their simpler predecessors will be provided as well as relevant statistical definitions.

Strict Stationarity

Let $(X_t)_{t\in\mathbb{Z}}$ be a time series process. $(X_t)_{t\in\mathbb{Z}}$ is strictly stationary if the following condition is met:

$$(X_{t_1}, \dots, X_{t_n}) = (X_{t_{1+k}}, \dots, X_{t_{n+k}})$$

For all $t_1, \dots, t_n, k \in \mathbb{Z}$ and $n \in \mathbb{N}$.

4.9.1 Covariance Stationarity

A common assumption among many time series models is covariance stationarity. Covariance stationarity defines a process whose unconditional joint probability distribution does not change as time goes by. Furthermore, because of this trait, mean and variance also do not change over time.

As is well known, a time series $\{y_t\}$ is covariance stationary if:

$E[y_t] = \mu$ for all t

$cov(y_t, y_{t-j}) = E[(y_t - \mu)(y_{t-j} - \mu)] = \gamma_j$ for all t and any j

Notice that the covariance function computes the covariance between values of y_t at different points in the time series process this is called an autocorrelation function (ACF). As compared to strict stationarity, covariance stationarity is the more relaxed assumption that still allows the same type of analysis to be carried out. For the sake of brevity, covariance stationarity is often called stationarity. Nonstationary data can be transformed to become stationary by differencing the data.

4.9.2 Autocorrelation Function

As the name implies, an autocorrelation function is one which describes correlation between current and past values of the same time series process, i.e. the lag. The autocorrelation function (ACF) for a covariance stationary process is defined as:

$$p(h) = p(X_h, X_0) = \frac{\gamma(h)}{\gamma(0)}, \forall h \in \mathbb{Z}$$

Where $(X_t)_{t \in \mathbb{Z}}$ is a covariance stationary process, and h is the lag considered.

4.9.3 Strict White Noise

A process $(X_t)_{t\in\mathbb{Z}}$, is a strict white noise process if the following conditions are met.

- $E[(X_t)] = 0\ \forall t$
- $(X_t)_{t\in\mathbb{Z}}\ is\ iid$
- $\sigma_X^2 = E({X_t}^2)$

4.9.4 White Noise

A process $(X_t)_{t\in\mathbb{Z}}$ is a white noise process if it has an autocorrelation function of:

$$p(h) = \begin{cases} 1 & h = 0 \\ 0 & h \neq 0 \end{cases}$$

Simply put, a process is white noise if it has no autocorrelation with any value except the present value.

4.9.5 Martingales

Martingales are commonly used in studying GARCH processes because martingales comprise the ARCH process. Let y_t be a sequence of random variables and let $I_t = \{y_t, y_{t-1}, \dots\}$ denote an information set of the past values of y_t. The sequence of $\{y_t, I_t\}$ is called martingale if:

- $I_{t-1} \subset I_t\ (I_t\ is\ a\ filtration)$
- $E[|y_t|] < \infty$
- $E[y_t \mid I_{t-1}] = y_{t-1} (martingale\ property)$

A common example of a martingale is the random walk $y_t = y_{t-1} + \epsilon_t, \epsilon_t \sim WN(0, \sigma^2)$, where y_0 is a initial fixed value. Letting $I_t = \{y_t, \dots, y_0\}$ implies $E[y_t\ |I_{t-1}] = y_{t-1}$ since $E[\epsilon_t\ |I_{t-1}] = 0$ Zivot, Wang (2002).

4.9.6 Martingale Difference Sequence

A martingale difference sequence is a stochastic series in which its expectation with respect to the past is zero.

$$X_t = Y_t - Y_{t-1}$$

Martingale difference sequences are useful because although they are an uncorrelated process, they are not required to be an independent process. This allows dependency in higher order moments of ϵ_t Zivot, Wang (2002). Because of this fact, they can be used to build models that have predictive power for variance with respect to past values of variance, which is essential for GARCH modeling.

4.9.7 AR Model

An Autoregressive Model uses past values of a process along with a strict white noise component to determine the next step at time t. An $AR(1)$ model looks at only 1 past value while an $AR(n)$ model looks at n past values. Written in mean-adjusted form, an $AR(n)$ model is defined as:

$$y_t - \mu = \phi_1(y_{t-1} - \mu) + \cdots + \phi_n(y_{t-n} - \mu) + \epsilon_t$$

Or equivalently:

$$y_t = \mu + \sum_{i=1}^{n} \phi_i y_{t-i} + \epsilon_t$$

Where μ is the long run mean, u_t is the summation of n previous values demeaned by the coefficients ϕ_i where $0 < \phi < 1$, and ϵ_t is a strict white noise process with distribution $\sim N(0,1)$.

4.9.8 MA Model

A moving average model uses the mean of the process at time t, and both current and past innovations to change the path of a model. Moving average models take the form:

$$y_t = \mu + \epsilon_t + \theta_1\epsilon_{t-1} + \cdots + \theta_n\epsilon_{t-n}, \epsilon_t \sim WN(0, \sigma^2)$$

Where μ is the mean value of the process, $\epsilon_t$is a strict white noise innovation, and ϕ_i is a coefficient that controls the degree to which past innovations affect the current value of y_t.

4.9.9 ARMA Model

Auto Regressive Moving Average (ARMA) models incorporate the predictive abilities of both AR and MA models. An ARMA model has the mean adjusted form of:

$$y_t = \mathrm{c} + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \epsilon_t + \theta\epsilon_{t-1} + \cdots + \theta\epsilon_{t-q}$$

The mean of an ARMA process can be calculated as:

$$\mu = \frac{c}{1 - \phi_1 - \cdots - \phi_p}$$

However, AR, MA, and ARMA models have no capability to account for conditional variance, which makes them a poor suit for financial data.

4.9.10 ARCH Process

ARCH stands for Autoregressive Conditionally Heteroskedastic. An ARCH model allows the previous innovations to affect the conditional variance at time t.

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i \epsilon_{t-i}^2$$

4.9.11 GARCH Models

Generalized Auto Regressive Conditional Heteroskedasticity Models allow dependency on past values of the time series process, while simultaneously allowing the conditional variance of the process to change over time. This allows modeling of volatility clusters. Additionally, GARCH models do not assume homoskedasticity in the residuals. GARCH models are based on their simpler ARCH models.

This thesis considers GARCH with normal distribution, GARCH with students t distribution, EGARCH with students t distribution, and APGARCH with generalized error distribution. Each of these models exhibit different behavior which when averaged results in a model which allows mixed consideration for all the models.

4.9.12 GARCH Process

GARCH stands for Generalized Autoregressive Conditionally Heteroskedastic. The generalized ARCH model allows the conditional volatility of the process to depend on the past squared volatilities. The ability to do this allows modeling of volatility clustering, an import feature of financial data. A GARCH model takes the form:

$$X_t = \sigma_t \epsilon_t$$

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

Where ϵ_t is a strict white noise process, and $(\epsilon_t)_{t \in Z} \sim SWN(0, 1)$. When $q = 0$, there is no difference between an ARCH model and a GARCH model.

As is common in the literature, GARCH (1, 1) model are considered, where $p = 1$ and $q = 1$.

4.9.13 GARCH with Normal Distribution

A GARCH model with normal distribution allows normally distributed innovations in the estimated variance of the process.

4.9.14 GARCH with Student's t Distribution

A GARCH model with student's t distribution allows innovations following a student's t distribution in the estimated variance of the process. A student's t distribution allows more volatility in the process as the kurtosis of a student's t distribution is higher, allowing for fatter tails. This means that movements will more often be larger as compared to the normal distribution.

4.9.15 EGARCH

An EGARCH model is like the GARCH normal model but the key difference lies in that the behavior of the model is defined by a product and exponential transformation as opposed to a GARCH normal model.

$$X_t = \sigma_t \epsilon_t$$

$$\log(\sigma_t^2) = \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i} + \sum_{j=1}^{q} b_i \log(\sigma_{t-j}^2)$$

$$\sigma_t^2 = e^{\alpha_0} \prod_{i=1}^{p} e^{\alpha_i X_{t-i}} + \sum_{j=1}^{q} b_i \sigma_{t-j}^2$$

4.9.16 APGARCH

An APGARCH model accounts well for fat tails, excess kurtosis, and leverage effects Ding (2011)

$$X_t = \sigma_t \epsilon_t$$

$$\sigma_t^{\delta} = c + \sum_{i=1}^{p} \beta_i \ (\sigma_{t-i})^{\delta} + \sum_{j=1}^{q} a_j \ (|X_{t-j}| - \gamma_j X_{t-j})^{\delta}$$

Where $c, \alpha_j, \gamma_j, \beta_i, and\ \delta$ are the parameters of the APGARCH model. A positive γ_j entails that negative information has a stronger effect on volatility. δ represents the leverage effect. Ding (2011)

5 Data

It is desired to create data that mimics the properties held by financial data. Generating data as opposed to using empirical data allows generalization of the relationship to some unknown data and control its characteristics. To achieve this flexibility of choice, a decision is made to use the NIG-CIR process. As noted in the literature review, this thesis' parameter set is modeled after the first parameter set used in Bottern (2015).

This thesis utilizes this parameter set which describes the NIG distribution.

NIG Parameters

$$\alpha = 21.1975$$

$$\beta = -1.1804$$

$$\delta = 7.0867$$

And these parameters which describe the CIR process.

CIR Parameters

$$\kappa = 5.7101$$

$$\eta = 5.5507$$

$$\lambda = 2.7864$$

$$y_0 = 1$$

As one can imply from the CIR definition, the CIR process which governs the stochastic time change was a high degree of mean reversion κ, relative to its coefficient on innovation λ. This means that movements away from the long run mean will be relatively short lived and soon be brought back towards the center of the data. This can be observed in Figure #1**.**

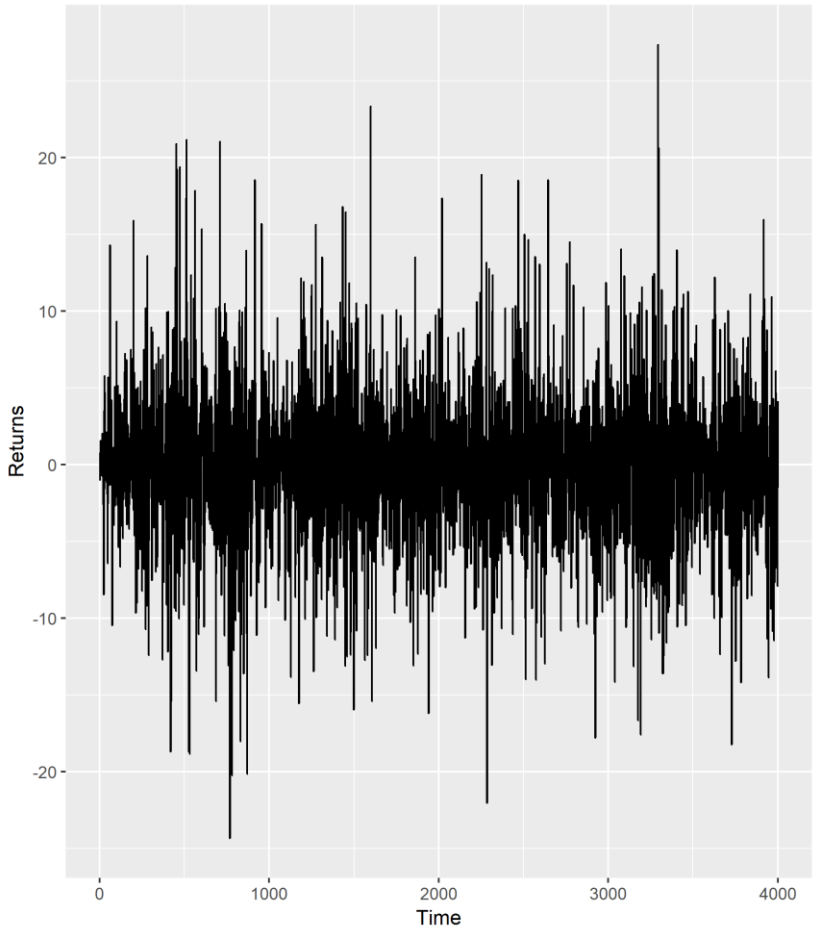

Figure 1: An NIG-CIR process of length 4000. This process models the returns of a hypothetical investment at different points in time

6 The Mixed Model

The mixed model is comprised of the following GARCH models.

- GARCH with normal distribution
- GARCH with student's t distribution
- EGARCH with students t distribution
- APGARCH with generalized error distribution

Earlier stochastic volatility models are mentioned as a viable method to estimate financial data. However, due to severe estimation error compared to the data generated, they are omitted from the mixed model. Nonetheless what is left are a rich mix of models, each describing behaviors of financial data differently.

7 Rolling Window Estimation

Rolling window estimation is employed to subsample the 4000 data points in such a way that each subsample describes a different window of data. The algorithm is described here.

Taking the NIG-CIR process of length 4000, rolling window estimation is used to capture 1000 consecutive data points at a time, estimating parameters for each of the models which contribute to the mixed model, and simulating a model path using the estimated parameters. After this is accomplished, a step size of 50 data points forward is used from where the window was previously started to begin the new window. If this process is repeated across all the NIG-CIR data, 60 parameter sets are generated, each describing the model parameters at a different window of the data.

A couple of tasks must be executed at every window generated. Firstly, model parameters are estimated by fitting the models to the data windows which is described in section 7.1. Secondly, a model path for each model estimated is created which is described in section 7.2.

7.1 Parameter Estimation

Estimating model parameters involves fitting each model specification to the window of data. The rugarch package in R was used to fit the models to the data windows.

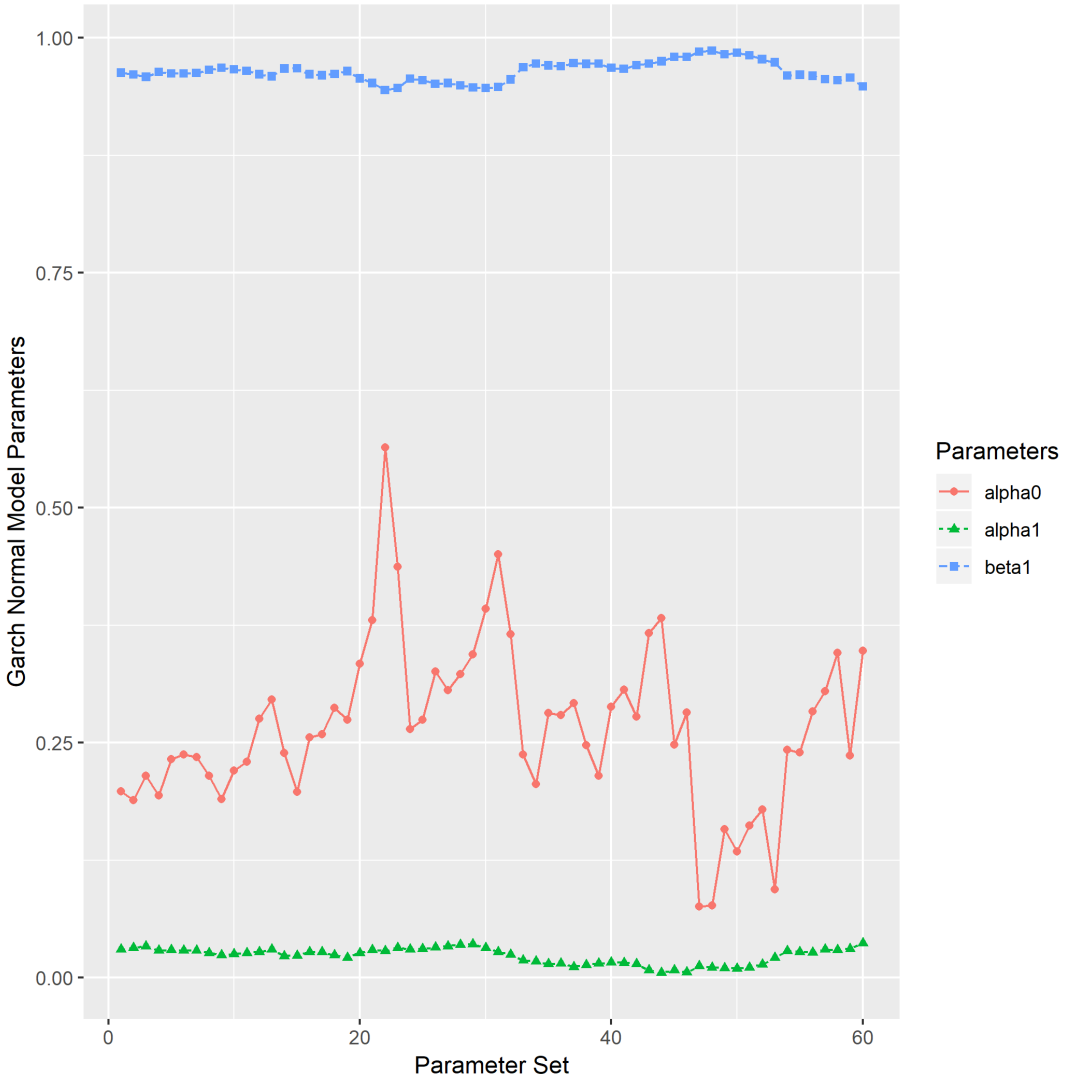

Figure 2: The parameters of the GARCH with Normal Distribution Model plotted at every parameter set

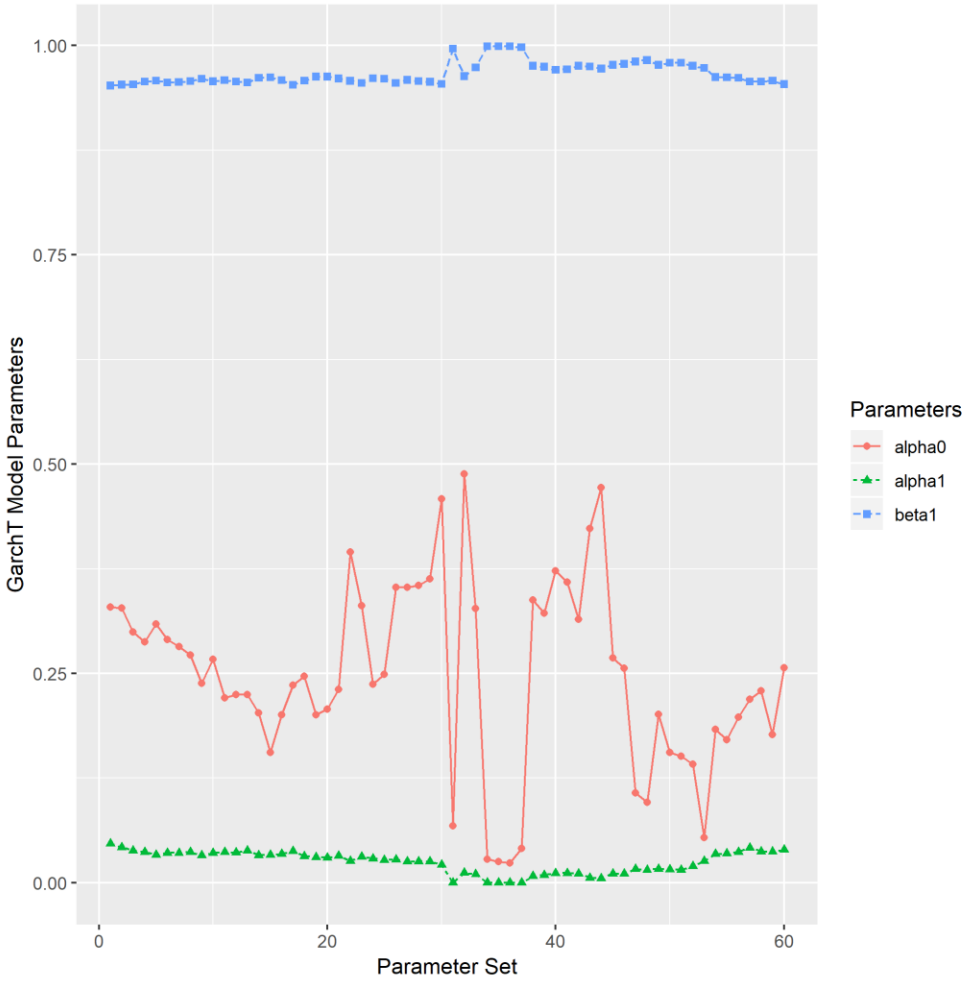

Figure 3: The parameters of the GARCH with Student t Distribution Model plotted at every parameter set

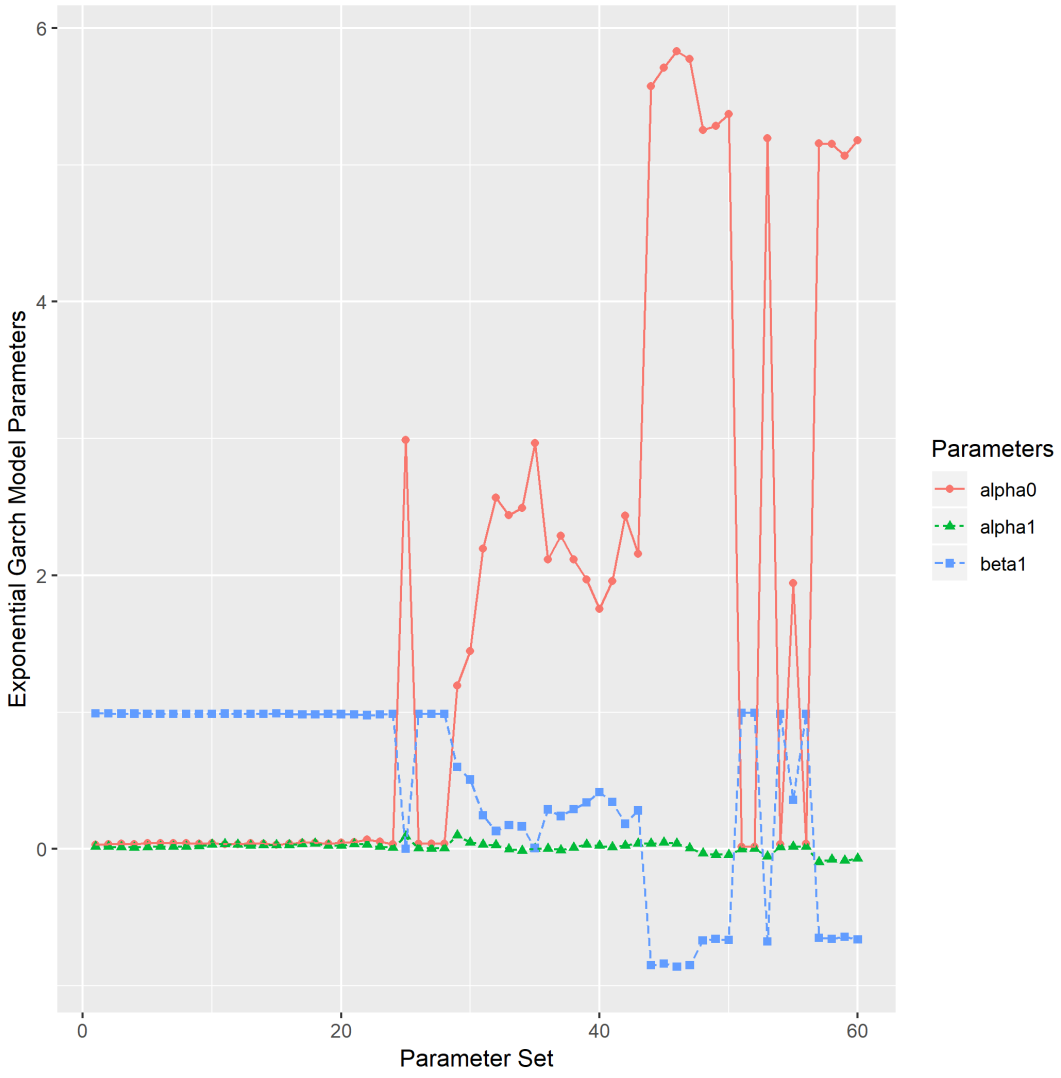

Figure 4: The parameters of the EGARCH Model with student's t distribution plotted at every parameter set.

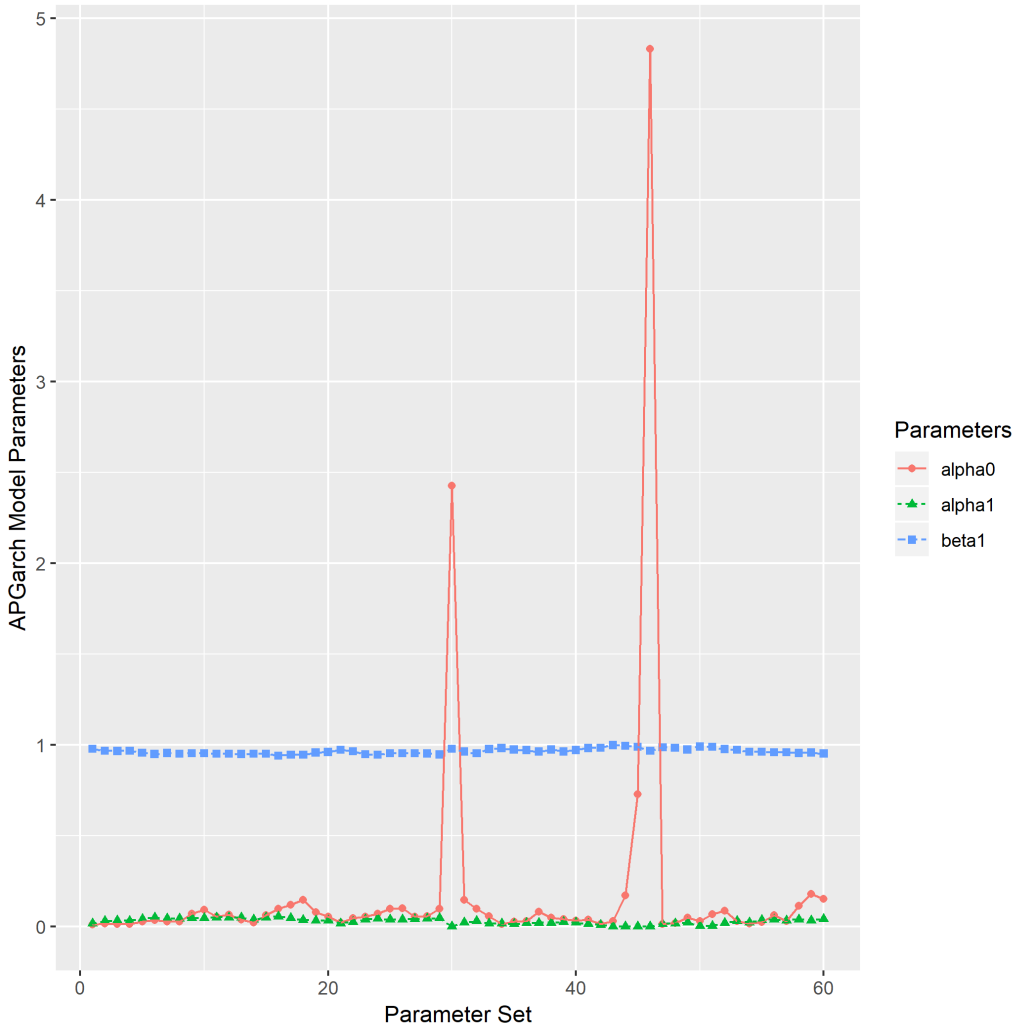

Figure 5: The parameters of the APGARCH Model with Generalized Error Distribution plotted at every parameter set.

Additionally, for each parameter set and each model an N length model path is created which is used as data for the estimation of VaR and ES for each model. The VaR and ES estimate for the mixed model is then a weighted average of the estimates of each model.

7.2 Simulation Paths

In order to generate enough data to precisely calculate value at risk and expected shortfall, a Monte Carlo simulation procedure is employed. At every parameter set, a simulation of an N_α length model path is created for each model. The calculation for the length N_α is as follows.

$$N = \frac{1}{\alpha * C^2}$$

Where α is the confidence level and C is a constant. For this thesis, this specific calculation is done.

$$N_\alpha = \frac{1}{\alpha * (10^{-2})^2}$$

Where α is the quantile in which value at risk and expected shortfall is estimated. Calculating expected shortfall requires averaging returns below a certain quantile of the data. To get reliable estimates, enough data below the quantile must be present. The larger α is, the lower N_α needs be because more data will be present below larger quantiles. For this thesis, VaR and ES is measured using an $\alpha = 0.1$.

8 Model Weights

Weights are assigned to each component of the mixed model using Bayesian Model Averaging (BMA). The weight for a model is calculated as a function of the mean of the BICs of all parameter sets.

$$w_i = e^{-\frac{1}{2} * \frac{\sum_{j=1}^{K} BIC_{i,j}}{K}}$$

Where K is the number of parameter sets.

Because weights are calculated as an average of all BICs for a model across all parameter sets, model weights are consistent throughout the parameter sets. This amounts to a more wholistic interpretation of the entire data, as opposed to weights calculated at every parameter set which is limited by the view of the window.

The weights for the mixed model following this procedure are detailed in Table 1.

Model	Weight
GARCH with normal distribution	0.242
GARCH with student's t distribution	0.253
EGARCH with student's t distribution	0.253
APGARCH with generalized error distribution	0.252

Table 1: The model weights for the mixed model.

9 Value at Risk and Expected Shortfall Calculation

Using the simulated paths at every parameter set for every model allows calculation of the value at risk and expected shortfall estimates of the models. If the data generated by a model path is sorted, it arranges itself in a distribution in which the quantile can be taken.

Less formally, the VaR estimate of a model at a parameter set can be calculated as such:

$$VaR = quantile(\alpha_n) * standard\ deviation(D) + mean(D)$$

While the expected shortfall is calculated as:

$$ES = \frac{\sum_{i=1}^{N_\alpha * \alpha} D_i}{N_\alpha * \alpha}$$

Where D is the sorted process generated or the loss distribution at a parameter set and α_n is the confidence level specified in section 7.2.

10 Estimation Results

Plotting each model against the true estimates reveals a few things about their performance.

1) Certain models fit the true estimates better than others
2) No one model can fully account for all behaviors of the true data.
3) Certain models often underestimated VaR and ES, while others consistently overestimated them

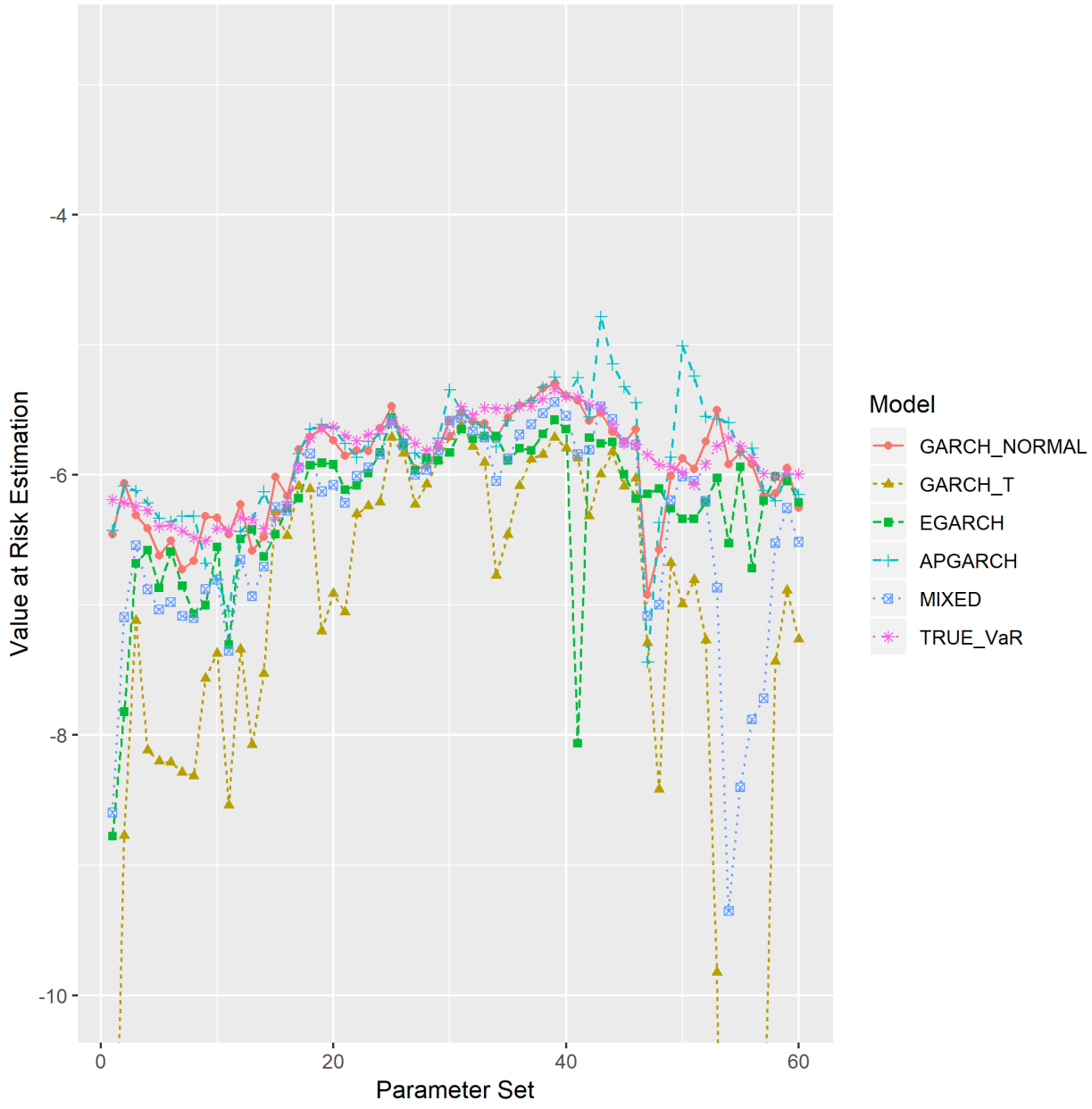

Figure 6: Model Estimations of Value at Risk plotted at every parameter set

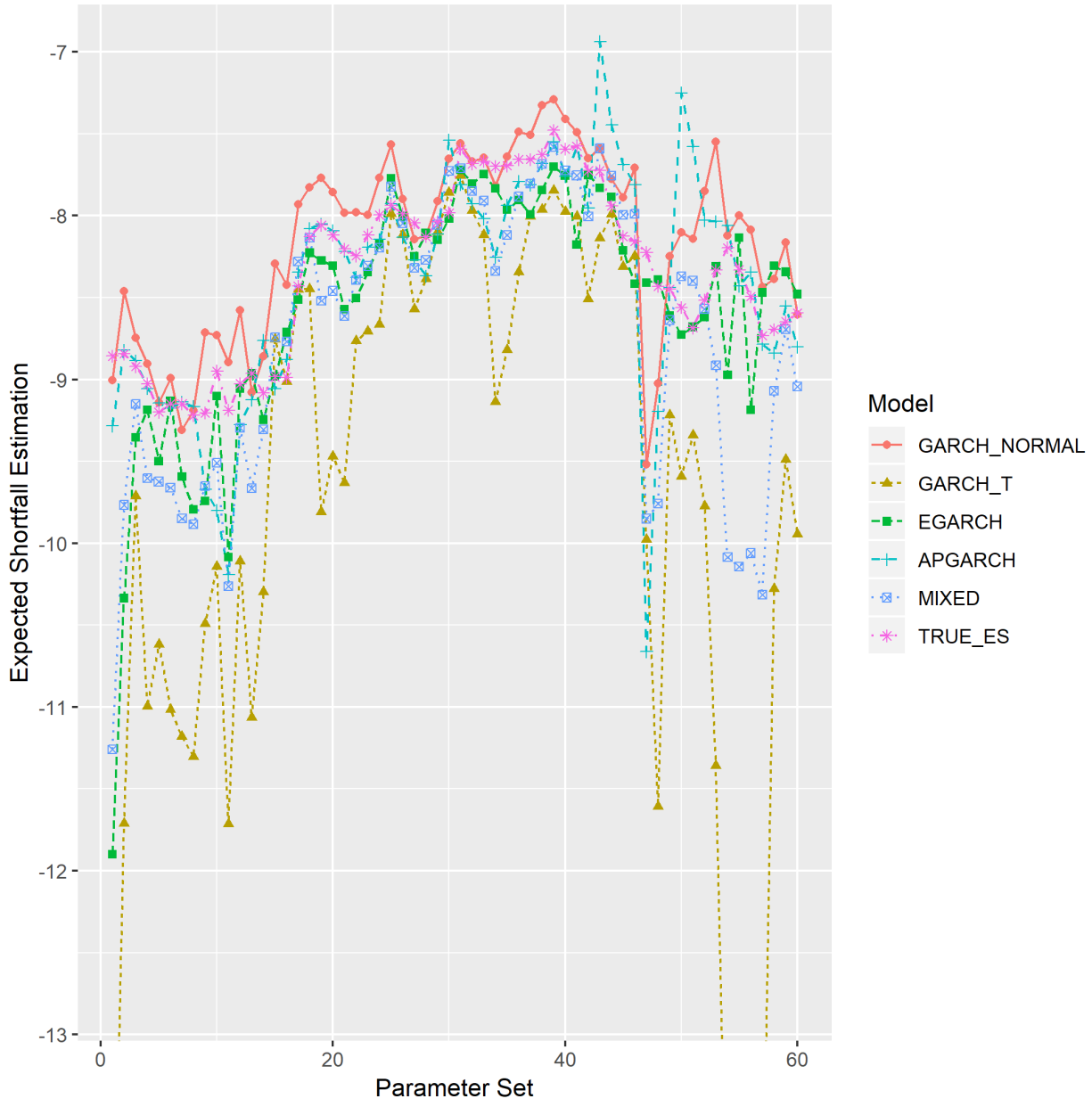

Figure 7: Model Estimations of Expected Shortfall plotted at every parameter set

Considering this data, the GARCH with normal distribution model most closely followed the true VaR and ES measurements, however it often underestimated VaR. Because the true data generated lacked large jumps or innovations between the values of VaR and ES across parameter sets, it would be expected that GARCH with normal distribution performs well. GARCH with students t distribution never underestimates VaR and ES but perhaps too cautiously overestimates it. This is likely due to the fatter

tails of the student's t distribution that generate larger movements more frequently than the normal distribution. In terms of closeness to the true estimates EGARCH with student's t distribution performs somewhere between GARCH with normal distribution and GARCH with student's t distribution. APGARCH was another model to significantly underestimate VaR and ES at some parameter sets, however with most parameter sets the model performed quite well.

The behavior of the mixed model was determined by an almost equal averaging between all these models as seen in Table 1. The models chosen had similar BIC values, and as such they received similar weights. The inclusion of different types of models that capture different behaviors acts as insurance that should the behavior of the process change; the mixed model will have diverse components where at least one will likely be able to closely model real data.

11 Model Fit Diagnostics

To measure model fit more numerically, a simple sum of squared residuals calculation of each model across the 60 parameter sets is used. The results are shown in Table 2 and Table 3 respectively.

Model	Sum of Squared Residuals (VaR)
GARCH with normal distribution	2.255
GARCH with student's t distribution	186.339
EGARCH with student's t distribution	19.158
APGARCH with generalized error distribution	6.261
Mixed Model	21.422

Table 2: The sum of squared residuals for every model calculating VaR

Model	Sum of Squared Residuals (ES)
GARCH with normal distribution	6.895
GARCH with student's t distribution	197.127
EGARCH with student's t distribution	16.648
APGARCH with generalized error distribution	13.410
Mixed Model	20.599

Table 3: The sum of squared residuals for every model calculating ES

In terms of residuals, it is found that the models perform differently. The GARCH with normal distribution model most closely follows both estimates of risk, whereas GARCH with student's t distribution is furthest from the true values of VaR and ES. EGARCH with student's t distribution and APGARCH with generalized error distribution perform similarly to each other when estimating ES.

It is worth noting that had the parameters which generated the data been different, a different model could have performed best in terms of fit instead of GARCH with normal distribution. For example, if the parameters had been changed such that the tails of the data distribution were thicker, GARCH with student's t distribution could have fared better while GARCH with normal distribution could have done worse. However, simply by using BMA, the mixed model is robust to potential changes in conditions because of our diverse model considerations.

12 Conclusion

The unique nature of financial data calls for models capable of capturing those idiosyncrasies. GARCH models are an excellent tool in modeling financial time series data. The ability to estimate conditional variance when the residuals are heteroskedastic is a must when considering the nature of returns data. However, one single GARCH model is not equipped to handle the whole realm of possibilities of movements in financial data. It is important to choose models that capture different types of movements such that each model provides a unique perspective on the data.

Bayesian model averaging allows multiple models to play a role in estimation. The motivation behind Bayesian model averaging is to reduce the risk associated with choosing a model. It is difficult to select the correct model considering the large model space that exists in any area of research. BMA offers a solution to this predicament by avoiding the single model choice, and instead uses a weighted average across many models calculated as a function of an information criterion. By using BMA, the goal is to ensure that insights are diverse enough to gain perspectives from multiple good models while at the same time ranking the authority of each model by a weighted average. Failure to account for different conditions could mean needlessly inaccurate estimations depending on the conditions present at the time. This thesis finds a case where the models selected had almost equal weights, indicating a situation where each model almost equally describes the data generation process. As such, the necessity of Bayesian model averaging is emphasized. BIC was the information criterion used to calculate the model weights, but other information criterion exist that could be tried instead.

This thesis discusses just a few models used to model financial data, but by no means covers the range of possibilities for financial modeling nor the range of questions that can be answered. A collection of GARCH models were used to estimate value at risk and expected shortfall, but of course other options exist. For instance, Extreme Value Theory (EVT) focuses on the tails of asset returns and has been used in insurance, finance, and quantitative risk management. EVT has specifically been used to estimate VaR. Alternatively, the Stable Paretian distribution can be used capture the behavior of the whole return distribution Harmantzis (2005). Even the method of likelihood maximization can be varied. This thesis used maximum likelihood estimation to calculate maximized log likelihood values, but other methods such as maximum a posteriori (MAP) could also be tested. For further research, it is worth exploring other combinations of models and methodologies for the estimation, analysis, and forecasting of value at risk and expected shortfall.

Bibliography

1) Arin, K. Peren, and Elias Braunfels. “The Resource Curse Revisited: A Bayesian Model Averaging Approach.” *Energy Economics*, vol. 70, 2018, pp. 170–178., doi:10.1016/j.eneco.2017.12.033.
2) Baillie, Richard T., and Tim Bollerslev. “The Message in Daily Exchange Rates: A Conditional-Variance Tale.” *Journal of Business & Economic Statistics*, vol. 7, no. 3, 1989, pp. 297–305., doi:10.1080/07350015.1989.10509739.
3) Bollerslev, Tim. “A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return.” *The Review of Economics and Statistics*, vol. 69, no. 3, 1987, p. 542., doi:10.2307/1925546.
4) Bottern, Alexandra. “Calculation of Value-At-Risk and Expected Shortfall Under Model Uncertainty.” 2015.
5) Carr, Peter P., et al. “Stochastic Volatility for Levy Processes.” *SSRN Electronic Journal*, 2003, doi:10.2139/ssrn.314979.
6) Ding, Ding. “Modeling of Market Volatility with APARCH Model.” *U.U.D.M Project Report*, Mar. 2011.
7) Kienitz, Joerg, and Daniel Wetterau. *Financial Modelling - Theory, Implementation and Practice with Matlab*. Wiley Finance Series, 2012.
8) Harmantzis, Fotios, et al. “Empirical Study of Value-at-Risk and Expected Shortfall Models with Heavy Tails.” *SSRN Electronic Journal*, 2005, doi:10.2139/ssrn.788624.
9) Giovannini, Alberto, and Philippe Jorion. “The Time-Variation of Risk and Return in the Foreign Exchange and Stock Markets.” 1988, doi:10.3386/w2573.
10) Glosten, Lawrence R., et al. “On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks.” *The Journal of Finance*, vol. 48, no. 5, 1993, p. 1779., doi:10.2307/2329067.
11) Fama, Eugene F., and Kenneth R. French. “Dividend Yields and Expected Stock Returns.” *Journal of Financial Economics*, vol. 22, no. 1, 1988, pp. 3–25., doi:10.1016/0304-405x(88)90020-7.
12) Maltritz, Dominik, and Alexander Molchanov. “Analyzing Determinants of Bond Yield Spreads with Bayesian Model Averaging.” *Journal of Banking & Finance*, vol. 37, no. 12, 2013, pp. 5275–5284., doi:10.1016/j.jbankfin.2013.07.007.
13) Masih, Rumi, et al. “Model Uncertainty and Asset Return Predictability: an Application of Bayesian Model Averaging.” *Applied Economics*, vol. 42, no. 15, 2010, pp. 1963–1972., doi:10.1080/00036840701736214.
14) Nelson, Daniel B. “Conditional Heteroskedasticity in Asset Returns: A New Approach.” *Econometrica*, vol. 59, no. 2, 1991, p. 347., doi:10.2307/2938260.
15) Poterba, James, and Lawrence Summers. “Mean Reversion in Stock Prices: Evidence and Implications.” 1987, doi:10.3386/w2343.
16) Schoutens, Wim, and Stijn Symens. “The Pricing of Exotic Options By Monte–Carlo Simulations In A Lévy Market With Stochastic Volatility.” *International Journal of Theoretical and Applied Finance*, vol. 06, no. 08, 17 Oct. 2002, pp. 839–864., doi:10.1142/s0219024903002249.
17) Wu, Liuren, and Peter P. Carr. “Time-Changed Levy Process and Option Pricing.” *SSRN Electronic Journal*, 2004, doi:10.2139/ssrn.283999.
18) Zivot, Eric, and Jiahui Wang. *Modeling Financial Time Series with S-Plus*. Springer New York, 2002.