Calculation of Value-at-Risk and Expected Shortfall under model uncertainty

Alexandra Böttern

Master's thesis
2015:E37

Lund University

Faculty of Engineering
Centre for Mathematical Sciences
Mathematical Statistics

Centrum Scientiarum Mathematicarum

Abstract

This thesis studies the concept of calculation of Value-at-Risk and Expected Shortfall when the choice of model is uncertain. The method used for solving the problem is chosen to be Bayesian Model Averaging, using this method will reduce the model risk by taking several models into account. Monte Carlo methods are used to perform the model averaging and the calculation of the risk measurements.

A NIG-CIR process is used to generate the data that is to be considered unknown, for which Value-at-Risk and Expected Shortfall is to be calculated. It is chosen since it have behaviour that often occur in financial data. The model averaging is performed using six different processes of varying levels of complexity. Both a weighted average based on BIC and a equally weighted average is calculated for the two risk measurements. The more complex models that are used in the Bayesian Model Averaging is GARCH processes, an EGARCH process and a stochastic volatility model, namely the Taylor 82 model. The methods used for parameter estimation are Maximum likelihood estimation and Kalman filtering.

The results in this thesis clearly shows that it is advantageous to calculate Value-at-Risk and Expected Shortfall using model averaging. But there is not any clear conclusion on which weights that give the most accurate estimate. But considering the time and effort that goes in to calculating the weights, using equal weight seems preferable.

Preface

Deciding on a topic for my Master thesis was a tricky decision. But after talking to my supervisor assoc. prof. Erik Lindström, and he presented the idea for this thesis, I made my decision. I found this topic interesting for many reasons, one of them was that the concept of averaging over several models to get better a result felt intriguing.

To average over several values is something we are taught from an early age and to use this concept to better model unknown data was something I found interesting. Of course it was not as easy to calculate the model average as it was to calculate the averages one did when growing up, but that is what makes the process interesting. Another thing I found interesting is that although the area of application in this thesis was towards financial risk measurements, the concept of model averaging is possible to apply in many different areas of study.

I have learnt a great deal during this thesis, both knowledge wise and about myself as a person. I would like to begin by thanking my supervisor Erik Lindström for introducing me to this project but also for your support and guidance during this project. Secondly I would like to thank my family and friends for their support during this time, your support have made this process a lot easier.

I do hope you find the contents of this thesis interesting and to get a brief overview of the topic I would recommend to read the abstract.

Contents

Abbrevations

ACF Auto Correlation Function.

ARCH Autoregressive Conditionally Heteroscedastic.

BIC Bayesian Information Criterion.

BMA Bayesian Model Averaging.

CIR Cox-Ingersoll-Ross.

EGARCH Exponential GARCH.

ES Expected Shortfall.

FSA Fishers Scoring Algorithm.

GARCH Generalized Autoregressive Conditionally Heteroscedastic.

MLE Maximum Likelihood Estimation.

NIG Normal Inverse Gaussian.

SV Stochastic Volatility.

SWN Strict White Noise.

VaR Value-at-Risk.

1 Introduction

A problem one never eludes when having unknown data is the process of selecting an appropriate model to describe the data. This decision will affect the future results that the calculations based on the selected model yields. To choose an appropriate model is a task that has proven to be difficult and there are several different model selection processes that are regularly used. Despite this, it still is not always enough to get a good model for the data.

When faced with financial data, the data often demonstrates several complex behaviours such as skewness and volatility clustering [5]. Choosing one single model to represent this data will with a high probability prove to be inadequate to use for some purposes, this is a problem called model risk. There are several risk categories often defined in finance i.e. market risk, credit risk and operational risk. The notation of model risk is one that occurs in almost all areas of risk.

This thesis will focus on the two risk measurements Value-at-Risk and Expected Shortfall. Value-at-Risk is one of the most widely used risk measures and is mentioned in Basel II that is recommended to use for calculation of market risk. But due to that Value-at-Risk have problems capturing behaviours such as tail risk, it is in Basel III recommended that Expected Shortfall is to replace VaR. Expected Shortfall is closely related to Value-at-Risk and is becoming more popular as time passes as it is a coherent risk measure. Nevertheless there are some problems with the risk measurements, one of these is model risk. Model risk is the risk that a financial institution obtain losses, due to that their risk-models are misspecified or the underlying assumptions of the model might not be met. For example, if one tries to model losses using a Normal distribution, the occurrence of volatility clustering might go unrecognized by the model.

The most standard practise in statistics is to choose one model that supposedly is the one that has generated the data. However this approach disregard the fact that the chosen model might not be, or even most certainly, is not the true model. Consequently rendering an over-confident model which in turn might lead to greater losses due to taking riskier decisions. A way to reduce the model uncertainty, therefore reducing the model risk, is to use model averaging. Bayesian Model Averaging is the method considered in this thesis. By averaging over several different models the positive contribution from each model will contribute to the final result, thus yielding a less uncertain measure.

1.1 Problem formulation

Deciding on how to carry out this thesis began by deciding on some main points, i.e. which data should be used to perform the analysis and which method should be used to perform the modelling.

It was decided that a NIG-CIR process should be used as a data source during the simulations and that this data was to be considered unknown. The NIG-CIR process were chosen due to that it has several properties that are common to financial data.

The next problem was to decide on which models that were to be fitted to the data. It was decided to use two simpler models, the Normal distribution and the Student's t distribution, but also three more complex processes. These three processes are a GARCH process, an EGARCH process and a Stochastic Volatilty model, Taylor 82 model. Fitting these distributions to the data is another topic in this thesis.

The final step is to use these processes to generate the Bayesian Model Averaging estimates of the Value-at-Risk and Expected Shortfall estimates. The calculation of the estimates will be performed using Monte Carlo methods. The idea is to show that by averaging over several different models instead of choosing one single model, the estimate will be more accurate. There will be a comparison between an equally weighted average and a weighted average, where it was decided that the weights were to depend on Bayes Information Criterion, more details on the weights can be seen in 2.3.

2 Theory

2.1 Stylized facts of financial data

Financial data have some special properties that is unusual in other data types. These so called stylized facts are based on empirical observations of financial data [1]. There are several different ones, some of them are mentioned below

- Little to no autocorrelation in the return series
- Conditional expected returns close to zero
- The return series has heavy tails and/or is leptokurtic
- The return often shows signs of skewness
- Volatility clustering often exists
- The absolute value or square of the returns have a clear serial correlation. This is the so called Taylor effect.

Heavy tails & leptokurtosis
The Gaussian distribution is often a poor fit for financial data, the data contains much more extreme events than the Gaussian distribution can predict. In general, financial return data seems to have a higher kurtosis than the Gaussian distribution, it is hence said to be leptokurtic. A distribution that is leptokurtic has a higher peak in the middle and heavier and longer tails [1]. This behavior of the data make for a problem when calculating Value at Risk and Expected Shortfall using a Gaussian distribution. When calculating VaR and ES using a Gaussian distribution fitted to financial data, the estimates will be overestimated on low confidence levels and underestimated on high confidence levels[15]. This is because the financial data have heavier tails than the Gaussian distribution have the ability to predict.

Volatility clustering
Volatility clustering is a phenomenon often occurring in financial data. This means that one extreme return often is followed by one or more extreme returns, not necessarily with the same sign [1]. That a process is heteroscedastic means that the volatility of the process changes over time. Most heteroscedastic processes contains volatility clusters, examples of these are ARCH and GARCH processes which will be presented in 2.6.2.

2.2 Value at Risk & Expected Shortfall

Value-at-Risk (VaR) [1] is a commonly used risk measure in financial institutions. And as mentioned before, it is the risk measurment that Basel II recommend shall be used for calculating market risk.

Consider a fixed time horizon Δ the loss distribution and the loss distributions distribution function is then defined according to

$$\begin{aligned} L_{[s,s+\Delta]} &:= -(V(s+\Delta) - V(s)) \\ F_L(l) &= P(L \leq l) \end{aligned} \tag{1}$$

The VaR-measure is calculated as the quantile function of the loss distribution, given a confidence level α, $\alpha \in [0,1]$. In other words it is the smallest value l such that the loss L is not greater than l with a probability $(1-\alpha)$. The expression of the VaR_α is shown in equation (2)

$$\begin{aligned} \text{VaR}_\alpha &= \inf\{l \in (R) : P(L > l) \leq 1-\alpha\} \\ &= \inf\{l \in (R) : F_L(l) \geq \alpha\} = q_\alpha(F_L) \end{aligned} \tag{2}$$

where $q_\alpha(F_L)$ denotes the quantile function. VaR is not a coherent risk measurement, meaning it does not behave in the way we would have wanted. In order for a risk measure ϱ to be coherent it needs to satisfy the four axioms of coherence. The axioms of coherence for a risk measure $\varrho : \mathcal{M} \to \mathbb{R}$ on the convex cone $\mathcal{M}$, follows below

1. Translation invariance.
   For all $L \in \mathcal{M}$ and every $l \in \mathbb{R}$ we have $\varrho(L + l) = \varrho(L) + l$

2. Subadditivity.
   For all $L_1, L_2 \in \mathcal{M}$ we have $\varrho(L_1 + L_2) \leq \varrho(L_1) + \varrho(L_2)$

3. Positive homogeneity.
   For all $L \in \mathcal{M}$ and every $\lambda > 0$ we have $\varrho(\lambda L) = \lambda\varrho(L)$

4. Monotonicity.
   For $L_1, L_2 \in \mathcal{M}$ such that $L_1 \leq L_2$ almost surely we have $\varrho(L_1) \leq \varrho(L_2)$

VaR does not satisfy all the axioms of coherence, it does not satisfy the axiom of subaddidivity, and thus is not a coherent risk measure. Expected shortfall on the other hand does satisfy all criteria to be coherent. In Basel III it is recommended that ES shall replace VaR for calculation of market risk since ES is able to capture some behaviours VaR can not, i.e. tail risk. Due to this to study both Expected Shortfall and Value-at-Risk in this thesis.

Using the definitions of the loss distribution as in (1) where $E(|L|) < \infty$ the expected shortfall at confidence level $\alpha \in (0, 1)$ is defined as in (3) if the loss distribution is continuous

$$\mathrm{ES}_\alpha = \frac{E(L; L \geq q_\alpha(L))}{1 - \alpha} = E(L | L \geq \mathrm{VaR}_\alpha) \tag{3}$$

That is, the expected loss given that VaR is exceeded. Another definition that often is used is

$$\begin{aligned} \mathrm{ES}_\alpha &= \frac{1}{1-\alpha} \int_\alpha^1 q_u(F_L) du \\ &= \frac{1}{1-\alpha} \int_\alpha^1 \mathrm{VaR}_u(L) du \end{aligned} \tag{4}$$

It can be interpreted as that expected shortfall is calculated by averaging VaR over all levels $u \leq \alpha$.

2.3 Bayesian Model Averaging

Bayesian model averaging is a mathematical method, originating from Bayes formula. It assumes that there is a set of different models $M_1, ..., M_k$, that all are a fairly good fit for estimating the quantity μ from the data y [9]. The parameter μ is defined and has an interpretation that is common for all the models. Instead of using just one of these models, Bayesian Model Averaging (BMA) constructs $\pi(\mu|y)$, which is the posterior density of μ given y, unconditioned on any of the models. BMA begins by specifying the prior probabilities, $P(M_j)$, for the different models and by specifying the prior densities, $\pi(\theta_j|M_j)$, where θ_j is the parameters of model M_j. The integrated likelihood of model M_j is then given by (5)

$$P(M_j|y) \propto \lambda_{n,j}(y) = \int \mathcal{L}_{n,j}(y, \theta_j)\pi(\theta_j|M_j)d\theta_j \tag{5}$$

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. Using Bayes theorem yields the posterior density of model M_j as (6)

$$P(M_j|y) = \frac{P(M_j)\lambda_{n,j}(y)}{\sum_{j'=1}^{k} P(M_{j'})\lambda_{n,j'}(y)} \tag{6}$$

The next step is to calculate the posterior density of μ, this can be done as in (7)

$$\pi(\mu|y) = \sum_{j=1}^{k} P(M_j|y)\pi(\mu|M_j, y) \tag{7}$$

This means that instead of assuming one model to be true, the posterior density in (7) is a weighted average of the conditional posterior densities. The weights are the posterior probabilities of each model. Since the method does not condition on any given model, it does not ignore the problem of model uncertainty. The posterior mean and posterior variance can be calculated as in (8) and (9) respectively [9].

$$E(\mu|y) = \sum_{j=1}^{k} P(M_j|y)E(\mu|M_j, y) \tag{8}$$

$$V(\mu|y) = \sum_{j=1}^{k} P(M_j|y)\Big(V(\mu|M_j, y) + \left(E(\mu|M_j, y) - E(\mu|y)\right)^2\Big) \tag{9}$$

In this thesis the quantity μ is VaR and ES and the average is calculated using Monte Carlo methods. The weight for M_j will be calculated as,

$$w_j = e^{-\mathrm{BIC}/2} \tag{10}$$

which is the posterior model probability seen in (6) and the integrated likelihood in (5) is calculated as the approximate BIC seen in (21). Averaging over all considered models has been shown to provide a better predictive ability in average, measured using a logarithmic scoring rule, than using one single model M_j conditional on all considered models [11].

2.4 Bayes Information Criterion

The Bayesian Information Criterion (BIC) is a measurement on how good estimate fit a model is and it is a way of deciding which model is the best fit to the data. The BIC is a measurement which punishes model complexity using a Bayesian framework [9]. The criterion takes the form of a penalised log-likelihood function as seen in (11).

$$\text{BIC}(M) = 2\log\mathcal{L}(M) - \log(n)\dim(M) \tag{11}$$

Where M is the model, dim(M) is the number of parameters that is estimated in the model and n is the sample size of the data set. The model with the largest BIC value is the one that is the best fit. The Bayesian approach to model selection, is to select the model that out of several models is the one which is a posteriori the best fit. The posterior probabilities of models, $M_1, ..., M_k$, derived from Bayes theorem is shown in (12)

$$P(M_j|y) = \frac{P(M_j)}{f(y)} \int_{\Theta_j} f(y|M_j, \theta_j)\pi(\theta_j|M_j)d\theta_j \tag{12}$$

where Θ_j is the parameter space and $\theta_j \in \Theta_j$ and $y_1, ..., y_n$ is the data. $P(M_j)$ is the prior probability of model M_j, $f(y)$ is the unconditional likelihood of the data y, $f(y|M_j, \theta_j) = \mathcal{L}_{n,j}(\theta_j)$ is 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, $f(y)$, is computed from (13)

$$f(y) = \sum_{j=1}^{k} P(M_j)\lambda_{n,j}(y) \tag{13}$$

$\lambda_{n,j}$ is the marginal likelihood of model j with θ_j integrated out with respect to the prior, as seen in (14) [9]

$$\lambda_{n,j} = \int_{\Theta_j} \mathcal{L}_{n,j}(\theta_j)\pi(\theta_j|M_j)d\theta_j \tag{14}$$

When comparing the posterior probabilities $P(M_j|y)$ with respect to the different models, the unconditional density $f(y)$ is irrelevant since it is constant across the models, it is $\lambda_{n,j}$ that is important to evaluate. Define the exact BIC estimates, $\text{BIC}_{n,j}^{\text{exact}}$, as (15), yielding the posterior probabilities in (16) [9]

$$\text{BIC}_{n,j}^{\text{exact}} = 2\log\lambda_{n,j}(y) \tag{15}$$

$$P(M_j|y) = \frac{P(M_j)e^{(\frac{1}{2}\text{BIC}_{n,j}^{\text{exact}})}}{\sum_{j'=1}^{k} P(M_{j'})e^{(\frac{1}{2}\text{BIC}_{n,j'}^{\text{exact}})}} \tag{16}$$

these exact BIC measures as seen in (15) are seldom used in practice, since they are very hard to compute numerically [9]. In order to get an expression that is more useful in practice, the approximation seen in (11), one begin with using the Laplace approximation [9].

Begin by writing (14) as (17), the integral is then of the kind where the basic Laplace approximation works [9].

$$\lambda_{n,j}(y) = \int_{\Theta} e^{nh_{n,j}(\theta)}\pi(\theta|M_j)d\theta \tag{17}$$

where $h_{n,j} = \frac{\ell_{n,j}(\theta)}{n}$ and ℓ is the log likelihood function, $\log\mathcal{L}$. The using the Laplace approximation yields (18)

$$\int_{\Theta} e^{nh(\theta)} g(\theta) d\theta = \Big(\frac{2\pi}{n}\Big)^{(p/2)} e^{nh(\theta_0)} \Big(g(\theta_0)|J(\theta_0)|^{-\frac{1}{2}} + O(n^{-1})\Big) \tag{18}$$

where p is the length of θ, θ_0 is the value that maximises h and J is the Hessian matrix. The approximation is an exact solution if h has a negative quadratic form and g is constant. The maximiser of $h_{n,j} = \frac{l_{n,j}(\theta)}{n}$ is the maximum likelihood estimator $\hat{\theta}_j$ of model M_j and $J_{n,j}(\hat{\theta_j})$ is the Fisher information matrix, which is seen in (19).

$$\lambda_{n,j}(y) \approx \mathcal{L}_{n,j}(\hat{\theta})\Big(\frac{2\pi}{n}\Big)^{p/2} |J_{n,j}(\hat{\theta}_j)|^{-1/2} \pi(\hat{\theta}_j|M_j) \tag{19}$$

Taking the logarithm and multiplying with 2, $2\log\lambda_{n,j}(y)$, yields (20)

$$\mathrm{BIC}^*_{n,j} = 2l_{n,j}(\hat{\theta}_j) - p_j\log(n) + p_j\log(2\pi) - \log|J_{n,j}(\hat{\theta}_j)| + 2\log\pi_j(\hat{\theta}_j) \tag{20}$$

where the first two terms are dominant with size $O_P(n)$ and $\log(n)$ respectively, yielding the BIC (21) as it is usually recognized.

$$\begin{aligned} 2\log\lambda_{n,j}(y) \approx \mathrm{BIC}_{n,j} &= 2\ell_{n,j,max} - p_j\log(n) \\ &= 2\ell(M) - \log(n)p \\ &= 2\ell(M) - \log(n)\dim(M) \end{aligned} \tag{21}$$

As can be seen in [10], the BIC can also be defined as (22), which mean that instead of multiplying with a factor 2, we multiply by -2.

$$\mathrm{BIC}(M) = -2\log\mathcal{L}(M) + \log(n)\dim(M) \tag{22}$$

With this definition of BIC, the model with the smallest BIC value is the best fit.

2.5 Mathematical methods

In the following section the theory behind the mathematical methods used in this thesis are introduced. There will be a brief introduction to Monte Carlo methods, Maximum Likelihood Estimation and Kalman filtering.

2.5.1 Monte Carlo Methods

Monte Carlo methods is a class of computational algorithms where you repeatedly generate random numbers to obtain the wanted numerical result.

A common application of Monte Carlo methods is Monte Carlo integration [2]. Consider the integral

$$\tau = E(\phi(X)) = \int \phi(x) f(x) dx \tag{23}$$

Where $\phi : \mathbb{R}^d \mapsto \mathbb{R}$, $X \in \mathbb{R}^d$ and f is the probability density of X. The probabilities correspond to ϕ being the indicator function.

$$P(X \in A) = \int \mathbb{1}\{x \in A\} f(x) dx \tag{24}$$

By using the law of large numbers, an approximation to τ is achieved according to the equation below.

$$t_N = t(x_1, \ldots, x_N) = \frac{1}{N} \sum_{i=1}^{N} \phi(x_i) \tag{25}$$

where $x_1, \ldots, x_N$ are independently drawn from f.

In this thesis VaR is calculated by generating multiple random numbers from the selected distribution and thereafter calculated as the α-quantile of the random numbers. The expected shortfall is calculated in a similar way by generating random numbers and averaging over the numbers that exceed the VaR-limit, as shown in (26).

$$\mathrm{ES}_\alpha = E(L|L \geq \mathrm{VaR}_\alpha) = \frac{1}{N_\alpha} \sum_{i=1}^{N} \mathbb{1}_{\{L_i \geq \mathrm{VaR}_\alpha\}} L_i \tag{26}$$

where N_α is the number of random numbers that exceeds VaR. To get a sample, N_α, of sufficient size one can use the Binomial distribution. Define $\hat{p}$ as

$$\hat{p} = \frac{\mathrm{Bin}(N, p)}{N} \tag{27}$$

where p is the probability level and N is the sample size. By designing N such that it will be possible to calculate what size the original sample N need to be for N_α to be of required size, the expression for this can be seen in below in (29)

$$E(\hat{p}) = \frac{Np}{N} \quad V(\hat{p}) = \frac{Np(1-p)}{N^2} \approx \frac{p}{N} \tag{28}$$

$$D(\hat{p}) = \sqrt{V(\hat{p})} = CE(\hat{p}) \tag{29}$$

Using the calculations in (28) and inserting into (29) yields the expression for the sample size (30)

$$\begin{aligned} C\,p &= \sqrt{\frac{p}{N}} \\ \Rightarrow\ N &= \frac{1}{p\,C^2} \end{aligned} \tag{30}$$

By applying the notation from this section, the sample size needed to obtain a certain amount (N_α) of samples once the VaR have been calculated can be seen in (31)

$$N = \frac{1}{C^2 \alpha} \tag{31}$$

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

2.5.2 Maximum likelihood estimation

Maximum Likelihood Estimation (MLE) is a method commonly used to estimate the unknown parameters of a model when the data is known. By maximizing the likelihood function, and thereby the pdf of the model, it is possible to calculate the estimates of the unknown parameters that are the most likely to be true given the data [4].

$$\hat{\theta}_{ML} = \arg\max_{\theta} f_x(x_1, ..., x_n|\theta) = \arg\max_{\theta} \mathcal{L}(\theta, x) \tag{32}$$

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

Fishers Scoring Algorithm
A commonly used method to calculate the MLE estimates is Fishers Scoring Algorithm (FSA). This algorithm is a version of the Newton-Raphson algorithm [13]. Given a set of estimates θ each step in the iteration, using the Newton-Raphson method is calculated as

$$\Delta\theta = -\left(\frac{\partial^2 \ell}{\partial\theta\partial\theta}\right)^{-1} \frac{\partial \ell}{\partial\theta} \tag{33}$$

where ℓ is the log-likelihood function. The matrix of second partial derivatives is called the Hessian matrix. This yields the algorithm (34)

$$\theta_{i+1} = \theta_i + \Delta\theta_i \tag{34}$$

where $\Delta\theta_i$ is defined as in (33). By replacing the matrix of second partial derivatives in (33) with their expectation, and thereby yielding the Fisher information matrix, one get the Fisher Scoring Algorithm [13].

2.5.3 Kalman Filter

A Kalman filter produces the optimal linear estimators of the unknown parameters of the underlying state system (35). It is a recursive algorithm which uses the noisy input data to produce the desired estimates.

$$\begin{aligned} x_{t+1} &= Ax_t + Bu_t + e_t \\ y_t &= Cx_t + w_t \end{aligned} \tag{35}$$

Seen above is the state space representation, where y_t is the measured data at t, A,B and C is known matrices. u_t is a known input to the system, e_t is the process, which includes model uncertainties, and w_t is the measurement noise process, describing the noise that disturbs the observed measurements [4].

The algorithm works in two steps, one prediction step and one estimation step. In the prediction step the filter calculates estimates of the current state and their variances. Thereafter estimates are produced using the predictions

and the data. For the interested reader I reference to [4] with more detailed information on how the Kalman filter works.

The equations used for the different steps of the filter is presented below, for a more detailed derivation of the equations I refer once again to [4]. The prediction of x and the variance of the prediction is seen in (36) and (37) respectively.

$$\hat{x}_{t+1|t} = A\hat{x}_{t|t} + Bu_t \tag{36}$$

$$V(\hat{x}_{t+1|t}) = R^{x,x}_{t+1|t} = AR^{x,x}_{t|t}A^T + R_e \tag{37}$$

$$V(\hat{y}_{t+1|t}) = R^{y,y}_{t+1|t} = CR^{x,x}_{t+1|t}C^T + R_w \tag{38}$$

Where the initial estimates is set to (39) and (40) respectively.

$$\hat{x}_{1|0} = E(x_1) = m_0 \tag{39}$$

$$R^{x,x}_{1|0} = V(x_1) = V_0 \tag{40}$$

In (41) it is shown how to calculate the Kalman gain and in (42) and (43) it is shown how to update the reconstruction of x and the variance of x respectively.

$$K_t = R^{x,x}_{t|t-1}C^T[R^{y,y}_{t|t-1}]^{-1} \tag{41}$$

$$\hat{x}_{t|t} = \hat{x}_{t|t-1} + K_t(y_t - C\hat{x}_{t|t-1}) \tag{42}$$

$$R^{x,x}_{t|t} = (I - K_tC)R^{x,x}_{t|t-1} \tag{43}$$

2.6 Different mathematical models

In this thesis several different mathematical models have been used. Both for the purpose of generating the data that were to be studied and the models that were to be fitted to this unknown data. Following is an introduction to the models used in this thesis.

2.6.1 NIG-CIR

The data studied in this thesis is considered to be unknown. The generated data is a Lévy model with stochastic time, to be more specific, it is a NIG-CIR process [3]. Making the time stochastic will result in stochastic volatility effects in the generated data. The NIG-CIR process in this thesis is generated out of two separate independent stochastic processes, the Normal Inverse Gaussian (NIG) process and the Cox-Ingersoll-Ross (CIR) process.

The characteristic function of the NIG Lévy process, NIG(α, β, δ), is seen in (44).

$$\phi_{NIG}(u;\alpha,\beta,\delta) = \exp(-\delta(\sqrt{\alpha^2-(\beta+\mathrm{i}u)^2}-\sqrt{\alpha^2-\beta^2})) \tag{44}$$

Where $\alpha > 0$, $-\alpha < \beta < \alpha$ and $\delta > 0$. Due to that the characteristic function is infinitely divisible the NIG process can be defined as $X^{(NIG)} = \{X_t^{(NIG)}, t \geq 0\}$ with $X_0^{(NIG)} = 0$. Resulting in a process which has stationary independent increments that are NIG distributed. The process can be related to an Inverse Gaussian time changed Brownian motion.

In order to simulate the NIG process, one begin with simulating Normal Inverse Gaussian random numbers. This can be obtained by first simulating Inverse Gaussian (IG) random numbers, $I_k \sim \mathrm{IG}(1, \delta\sqrt{\alpha^2-\beta^2})$. Thereafter one proceeds by sampling Normal random numbers u_k in order to generate the NIG process random numbers n_k

$$n_k = \delta^2\beta I_k + \delta\sqrt{I_k}u_k \tag{45}$$

The final sample path for the NIG process is then obtained according to (46), using the random numbers $n_k \sim \mathrm{NIG}(\alpha, \beta, \delta\Delta t)$, where Δt are the time points.

$$X_0 = 0, X_t = X_{t_{k-1}} + n_k, k \geq 1 \tag{46}$$

The second part of the process is the stochastic clock, in this case a CIR stochastic clock. It is the CIR process that gives the data its stochastic volatility behaviour, by time changing the Lévy process above.

The CIR process solving the stochastic differential equation (SDE) (47) is used as a the rate of time change.

$$dy_t = \kappa(\eta - y_t)dt + \lambda y_t^{1/2}dW_t, \qquad y_0 \geq 0 \tag{47}$$

where $W = \{W_t, t \geq 0\}$ is a Brownian motion. Given y_0, the characteristic equation of Y_t is known and can be seen in (48)

$$\begin{aligned}\phi_{CIR}(u,t;\kappa,\eta,\lambda,y_0) = E[\exp(iuY_t)|y_0] = \\ = \frac{\exp(\kappa^2\eta t/\lambda^2)\exp(2y_0iu/(\kappa+\gamma\coth(\gamma t/2)))}{(\cosh(\gamma t/2)+\kappa\sinh(\gamma t/2)/\gamma)^{2\kappa\eta/\lambda^2}}\end{aligned} \tag{48}$$

where $\gamma = \sqrt{\kappa^2 - 2\lambda^2 iu}$. To simulate a CIR process $y = \{y_t, t \geq 0\}$ the SDE (47) is discretized. If a first order accurate explicit differencing scheme is used in time, the sample path $y = \{y_t, t \geq 0\}$ in the time points $t = n\Delta t, n = 0, 1, 2...$, becomes

$$y_{t_n} = y_{t_{n-1}} + \kappa(\eta - y_{t_{n-1}})\Delta t + \lambda y_{t_{n-1}}^{1/2}\sqrt{\Delta t}\upsilon_n \tag{49}$$

where $\{\upsilon_n, n = 1, 2, ...\}$ are independent standard normally distributed random numbers.

To create the NIG-CIR process one generates the NIG process, using the generated CIR process as Δt in (46).

2.6.2 GARCH & EGARCH

Generalized Autoregressive Conditionally Heteroscedastic (GARCH) process is a type of time series process. As the name says it is an auto-regressive process, meaning that it is dependent on the past values of the process. That the process is conditionally heteroscedastic means that the conditional variance of the process is changing over time, creating volatility clusters. Introduction to the GARCH process will begin by some background concepts proceeded by introducing the ARCH process, followed by an introduction to the GARCH and EGARCH.

Basic definitions

The different processes in this thesis are all univariate stationary processes which is a foundation for the following definitions. In time series analysis the first two moments are commonly used. These are the mean function $\mu(t)$ and autocovariance function $\gamma(s,t)$, as defined in (50) and (51) respectively [1].

$$\mu(t) = E(X_t), \quad t \in \mathbb{Z} \tag{50}$$

$$\gamma(s,t) = E(X_t - \mu(t))E(X_s - \mu(s)), \quad s,t \in \mathbb{Z} \tag{51}$$

where $(X_t)_{t\in\mathbb{Z}}$ is a stochastic process. For the autocovariance function it follows that $\gamma(t,s) = \gamma(s,t)$ for all s,t and that $\gamma(t,t) = V(X_t)$.

Next, strict stationarity (2.1) and covariance stationarity (2.2) will be defined. Most often processes are stationary in both these senses or at least in one of them.

Definition 2.1. (**Strict stationarity**) A process, $(X_t)_{t\in\mathbb{Z}}$, is strictly stationary if,

$$(X_{t_1}, ..., X_{t_n}) \stackrel{\text{def}}{=} (X_{t_{1+k}}, ..., X_{t_{n+k}})$$

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

Definition 2.2. (**Covariance stationarity**)

A process, $(X_t)_{t\in\mathbb{Z}}$, is covariance stationary if both the first two moments exist and satisfy,

$$\mu(t) = \mu, \quad t \in \mathbb{Z}$$

$$\gamma(t,s) = \gamma(t+k, s+k), \quad s,t,k \in \mathbb{Z}$$

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

Definition 2.3. (**Auto Correlation Function**) For a covariance stationary process the Auto Correlation Function (ACF),is defined as

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

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

This can be seen from the definition of covariance stationarity $\gamma(t-s,0) = \gamma(t,s) = \gamma(s,t) = \gamma(s-t,0)$, showing that the covariance between X_t and X_s is only dependent on $|t-s| = h$ which is called the lag.

An important concept in time series analysis is white noise. White noise is a stationary process without serial correlation, the definition follows below.

Definition 2.4. (**White noise**)
A process, $(X_t)_{t\in\mathbb{Z}}$, is a white noise process if it is covariance stationary with ACF

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

Definition 2.5. (**Strict white noise**)
A process, $(X_t)_{t\in\mathbb{Z}}$, is called a Strict White Noise (SWN) if it is a series of iid random variables with finite variance

Martingale difference
The martingale difference property is another noise concept that is commonly used in the study of ARCH and GARCH processes [1]. A martingale difference sequence is a generalized white noise process and it is a martingale difference process if the following holds

Definition 2.6. (**Martingale**) A time series $(X_t)_{t\in\mathbb{Z}}$ is called a martingale with respect to the filtration $(\mathcal{F}_t)_{t\in\mathbb{Z}}$ if the following properties hold

I. $E|X_t| < \infty$

II. X_t is $\mathcal{F}_t - measurable$

III. $E(X_t|\mathcal{F}_{t-1}) = 0, \quad \forall t \in \mathbb{Z}$

The property saying that the expectation of the next value always is zero, makes it appropriate to apply for financial data [1].

ARCH process
The ARCH process is as the name suggests an Autoregressive Conditionally Heteroscedastic process. Let $(Z_t)_{t\in\mathbb{Z}}$ be a Strict White Noise (SWN) process, such that $(Z_t)_{t\in\mathbb{Z}} \sim \text{SWN}(0,1)$. The process $(X_t)_{t\in\mathbb{Z}}$ is defined as an ARCH(p) process if it is strictly stationary and if it for all $t \in \mathbb{Z}$ and for the strictly positive process $(\sigma_t)_{t\in\mathbb{Z}}$ satisfies equation (52) [1]

$$\begin{aligned} X_t &= \sigma_t Z_t \\ \sigma_t^2 &= \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i}^2 \end{aligned} \tag{52}$$

where $\alpha_0 > 0$ and $\alpha_i \geq 0, i = 1, ..., p$.

$\mathcal{F}_t = \sigma(X_S : s \leq t)$ is the sigma algebra which contains the information generated by the process until time t, that is $\mathcal{F}_t$ is the natural filtration. From

(52) one can see that σ_t is measurable with respect to $\mathcal{F}_{t-1}$. Provided that $E(|X_t|) < \infty$, one can calculate

$$E(X_t|\mathcal{F}_{t-1}) = E(\sigma_t Z_t|\mathcal{F}_{t-1}) = \sigma_t E(Z_t|\mathcal{F}_{t-1}) = \sigma_t E(Z_t) = 0 \tag{53}$$

in (53) one can see that the ARCH process possesses the martingale difference property (2.6) with respect to $(\mathcal{F}_t)_{t\in\mathbb{Z}}$. If the process is covariance stationary as well, the process is a white noise. Assuming that the process is covariance stationary (54) yields that its conditional volatility is depending on the previously squared values of the process. Hence generating volatility clustering [1].

$$v(X_t|\mathcal{F}_{t-1}) = E(\sigma_t^2 Z_t^2|\mathcal{F}_{t-1}) = \sigma_t^2 v(Z_t^2) = \sigma_t^2 \tag{54}$$

The innovations $(Z_t)_{t\in\mathbb{Z}}$ can in general be any distribution with zero mean and unit variance, i.e. the Gaussian distribution or student's t-distribution.

GARCH process

The GARCH process is a Generalized ARCH process, meaning that the conditional volatility is also allowed to depend on previous squared volatilities.

Let once again Z_t be a strict white noise, $(Z_t)_{t\in\mathbb{Z}} \sim \text{SWN}(0,1)$. The time series process $X_{tt\in\mathbb{Z}}$ is a GARCH(p,q) process if it for all $t \in \mathbb{Z}$ and some strictly positive valued process $(\sigma_t)_{t\in\mathbb{Z}}$ holds for (55) and (56).

$$X_t = \sigma_t Z_t \tag{55}$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-1}^2 \tag{56}$$

Where $\alpha_0 > 0$, $\alpha_i \geq 0, i = 1, ..., p$ and $\beta_j \geq 0, j = 0, ..., q$. In practise mostly lower ordered GARCH processes is used [1]. A GARCH(1,1) process is a covariance stationary white noise process iff $\alpha_1 + \beta_1 < 1$ [1].

EGARCH

The Exponential GARCH (EGARCH) process is an extension of the GARCH model. Assume $(Z_t)_{t\in\mathbb{Z}}$ to be a SWN$(0,1)$ process, $X_{tt\in\mathbb{Z}}$ is an EGARCH(p,q) process, if for all $t \in \mathbb{Z}$ and some strictly positive valued process $(\sigma_t)_{t\in\mathbb{Z}}$, it satisfy the following equations [6].

$$X_t = \sigma_t Z_t \tag{57}$$

$$\log \sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i} + \sum_{j=1}^{q} \beta_j \log \sigma_{t-j}^2 \tag{58}$$

Where $\alpha_i\,, i = 0...p$, $\beta_j\,, j = 1...q$ are real numbers.

$$\sigma_t^2 = e^{\alpha_0} \prod_{i=1}^{p} e^{\alpha_i X_{t-i}} \prod_{j=1}^{q} {\sigma_{j-1}^2}^{\beta_{j-1}} \tag{59}$$

In contrast to the classical GARCH process the volatility in the EGARCH have multiplicative dynamics which can be seen in (59) [6]. Since the logarithm can be of any sign, it's possible to avoid the constraint of strictly positive coefficient, which holds for the GARCH process.

MLE of GARCH and EGARCH parameter estimates
The parameters of a GARCH or EGARCH process can be estimated using Maximum Likelihood Esimation. In this thesis the focus is on GARCH(1,1) and EGARCH(1,1) processes and these will thus set the standard for the rest of this section, but the concept can easily be extend to cover GARCH(p,q) and EGARCH(p,q) processes.

Begin by defining the parameter vector $\theta = (\alpha_0, \alpha_1, \beta_1)$ which is the argument that shall be maximized. The log-likelihood function ℓ can be rewritten as (60)

$$\ell(\theta) = \sum_{t=q+1}^{n} \ell_t(\theta) \tag{60}$$

where $\ell_t(\theta)$ is the log-likelihood function at time t. As written in 2.5.2 the algorithm updates the parameters according to (61), each iteration step use both the first and second order derivative of the log-likelihood function. called ∇L and J respectively [12].

$$\theta_{i+1} = \theta_i + J^{-1}(\theta_j)\nabla L(\theta_j) \tag{61}$$

Where ∇L and J are defined according to (62) and (63) respectively, and J is the Fisher information matrix.

$$\nabla L = \frac{\partial \ell}{\partial \theta} \tag{62}$$

$$J = E\bigg(- \frac{\partial^2 \ell}{\partial \theta \partial \theta} \bigg) \tag{63}$$

The GARCH(1,1) process is defined as in (55) and (56) as

$$\begin{aligned} y_t &= \sigma_t Z_t \\ \sigma^2 &= \alpha_0 + \alpha_1 y_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \end{aligned} \tag{64}$$

and the EGARCH(1,1) process is defined as

$$\begin{aligned} y_t &= \sigma_t Z_t \\ \log \sigma^2 &= \alpha_0 + \alpha_1 y_{t-1} + \beta_1 \log(\sigma_{t-1}^2) \end{aligned} \tag{65}$$

This thesis will consider both the Normal distribution and Student's t Distribution for the SWN process Z_t. The log-likelihood function for the Normal distribution is presented in (66) below [12].

$$\ell_t^N(\theta) = -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma_t^2) - \frac{1}{2}\frac{y^2}{\sigma_t^2} \tag{66}$$

For the Normal distribution the first and second order derivative of (66) are defined as in (67) (68). A more detailed calculation of the derivatives can be found in A.1.

$$\frac{\partial \ell_t^N}{\partial \theta} = \frac{1}{2\sigma_t^2}\frac{\partial \sigma_t^2}{\partial \theta}\left(\frac{y_t^2}{\sigma_t^2} - 1\right) \tag{67}$$

$$\frac{\partial^2 \ell_t^N}{\partial \theta \partial \theta} = \left(\frac{y_t^2}{\sigma_t^2} - 1\right)\frac{\partial}{\partial \theta}\left(\frac{1}{2\sigma_t^2}\frac{\partial \sigma_t^2}{\partial \theta}\right) - \frac{y_t^2}{2(\sigma_t^2)^3}\frac{\partial \sigma_t^2}{\partial \theta}\frac{\partial \sigma_t^2}{\partial \theta} \tag{68}$$

where $\frac{\partial \sigma_t^2}{\partial \theta}$ is updated according to (69) and ∇L and J is calculated as in (70) and (71).

$$\frac{\partial \sigma_t^2}{\partial \theta} = (1, y_{t-1}^2, \sigma_{t-1}^2) + \beta_1 \frac{\partial \sigma_{t-1}^2}{\partial \theta} \tag{69}$$

$$\nabla L = \frac{1}{2}\sum_{t=2}^{n}\frac{1}{\sigma_t^2}\frac{\partial \sigma_t^2}{\partial \theta}\left(\frac{y_t^2}{\sigma_t^2} - 1\right) \tag{70}$$

$$J = \frac{1}{2}\sum_{t=2}^{n}\left(\frac{1}{\sigma_t^4}\frac{\partial \sigma_t^2}{\partial \theta}\frac{\partial \sigma_t^2}{\partial \theta}\right) \tag{71}$$

The way of calculating the iteration step for the Student's t distribution, similar to the way it is calculated for the Normal distribution, but the way one updates the ∇L and J matrices differ. This due to that the first and second order derivatives will differ from the expressions in (67) and (68). The log-likelihood function for the Student's t distribution can be seen in (72)

$$\begin{aligned}\ell_t^T(\theta) &= \log(\Gamma((\nu+1)/2)) - \log(\Gamma(\nu/2)) - \frac{1}{2}\log(\pi) \\ &-\frac{1}{2}\log(\nu-2) - \frac{1}{2}\log(\sigma_t^2) - \frac{\nu+1}{2}\log\left(1 + \frac{y^2}{(\nu-2)\sigma_t^2}\right)\end{aligned} \tag{72}$$

For the t-distribution the following functions will replace the functions valid for the Normal distribution. More detailed calculations of the expression can be found in A.2.

$$\frac{\partial \ell_t^T}{\partial \theta} = \frac{1}{2\sigma_t^2}\frac{\partial \sigma_t^2}{\partial \theta}\left(\frac{y_t^2}{\sigma_t^2}\mathcal{G} - 1\right) \tag{73}$$

$$\frac{\partial^2 \ell_t^T}{\partial \theta \partial \theta} = -\left(\frac{y_t^2}{\sigma_t^2}\mathcal{G} - 1\right)\frac{\partial}{\partial \theta}\left(\frac{1}{\sigma_t^2}\frac{\partial \sigma_t^2}{\partial \theta}\right) - \frac{y_t^2}{2\sigma_t^2}\frac{\partial \sigma_t^2}{\partial \theta}\frac{\partial \sigma_t^2}{\partial \theta}\frac{(\nu+1)(\nu-2)}{(\sigma_t^2(\nu-2)+y^2)^2} \tag{74}$$

where $\mathcal{G}$ is defined as in (75), which yields the expressions seen in (76) and (77) for ∇L and J respectively

$$\mathcal{G} = \frac{\nu + 1}{(\nu - 2) + \frac{y_t^2}{\sigma_t^2}} \tag{75}$$

$$\nabla L = \frac{1}{2} \sum_{t=2}^{n} \frac{1}{\sigma_t^2} \frac{\partial \sigma_t^2}{\partial \theta} \left(\frac{y_t^2}{\sigma_t^2} \mathcal{G} - 1 \right) \tag{76}$$

$$J = \frac{1}{2} \sum_{t=2}^{n} \left(\frac{1}{\sigma_t^4} \frac{(\nu + 1)(\nu - 2)}{((\nu + 2) + 1)^2} \frac{\partial \sigma_t^2}{\partial \theta} \frac{\partial \sigma_t^2}{\partial \theta} \right) \tag{77}$$

Studying the EGARCH(1,1) process shown in the system (65) the SWN process have been chosen to be Normal distributed, implying that most of the equations will remain the same as in the case of GARCH(1,1) with Normal distribution. The difference will be to (69) which instead will be updated according to (78)

$$\frac{\partial \sigma_t^2}{\partial \theta} = (1, y_{t-1}, \log \sigma_{t-1}^2) e^{(\alpha_0 + \alpha_1 y_{t-1} + \beta_1 \log \sigma_{t-1}^2)} + \beta_1 \frac{\partial \sigma_{t-1}^2}{\partial \theta} \tag{78}$$

2.6.3 Stochastic Volatility model - Taylor 82

Stochastic Volatility (SV) models are an alternative to the GARCH model, which also generates volatility clustering, but with SV models the volatility is assumed to be driven by it's own stochastic process. The SV process studied in this thesis is the classical SV model introduced in 1982 by S.J Taylor [7]. This is a standard Gaussian autoregressive SV process in discrete time and is defined according to (79) and (80) [8].

$$y_t = e^{h_t/2} z_t \quad z_t \sim N(0, 1) \tag{79}$$

$$h_t = a_0 + a_1 h_{t-1} + \sigma \eta_t \quad \eta_t \sim N(0, 1) \tag{80}$$

where y_t is the log return at time t and h_t is the log volatility. The error terms z_t and η_t are Gaussian white noise processes. Comparing with the GARCH model the major difference is, conditional on the information set $\mathcal{F}_{t-1}$, h_t^2 is an unobserved random variable [7].

The Taylor 82 model can be fitted to the data by using a Kalman filter. By squaring and taking the logarithm of (79), one see similarities to a state space system (81).

$$u_t = \log y_t^2 = h_t + \log z_t^2 \quad z_t \sim N(0, 1) \tag{81}$$

where $E(\log z_t^2) = -1,27$ and $V(\log z_t^2) = \frac{\pi^2}{2}$. This yields (82) which have clear similarities to the state space system defined in (35) in the section about the Kalman filter.

$$\begin{aligned} u_t &= h_t + \log z_t^2 \quad z_t \sim N(0,1) \\ h_{t+1} &= a_0 + a_1 h_t + \sigma \eta_t \quad \eta_t \sim N(0,1) \end{aligned} \tag{82}$$

The initial estimates of the mean, (39), and variance, (40), are seen in (83) and (84) respectively.

$$m_0 = E(x_1) = \frac{a_0}{1 - a_1} \tag{83}$$

$$V_0 = V(x_1) = \frac{\sigma^2}{1 - a_1^2} \tag{84}$$

This knowledge yields an opportunity to use a Kalman filter to estimate the unknown parameters a_0, a_1 and σ.

3 Simulations

The main part of the thesis is based on simulations performed in Matlab. This section will consist of a summary of the simulation steps.

3.1 Creation of the NIG-CIR data

In section 2.6.1 it is described how to create NIG-CIR distributed data. The data set that was generated consisted of two NIG-CIR processes, with slightly different parameters, that were concatenated to one. The reason for the slight change of parameters in the data is to see how well the method can follow when the conditions change.

The parameters for the two processes were chosen as

First set of parameters:	**Second set of parameters:**
$\alpha = 21.1975$	$\alpha = 20.1975$
$\beta = -1.1804$	$\beta = -1.3804$
$\delta = 7.0867$	$\delta = 7.0867$
$\kappa = 5.7101$	$\kappa = 5.2101$
$\eta = 5.5507$	$\eta = 5.5507$
$\lambda = 2.7864$	$\lambda = 2.7864$
$y_0 = 1$	$y_0 = 1$

The parameter choices was influenced by parameters found in [3]. Generating the NIG-CIR data with these parameters yields the plots seen in Figure 1 and Figure 2 shows a closer look at the first 500 samples. This NIG-CIR data process is the one that is considered to be the true loss function and it is this one that the VaR and ES measurements are supposed to be calculated for using model averaging.

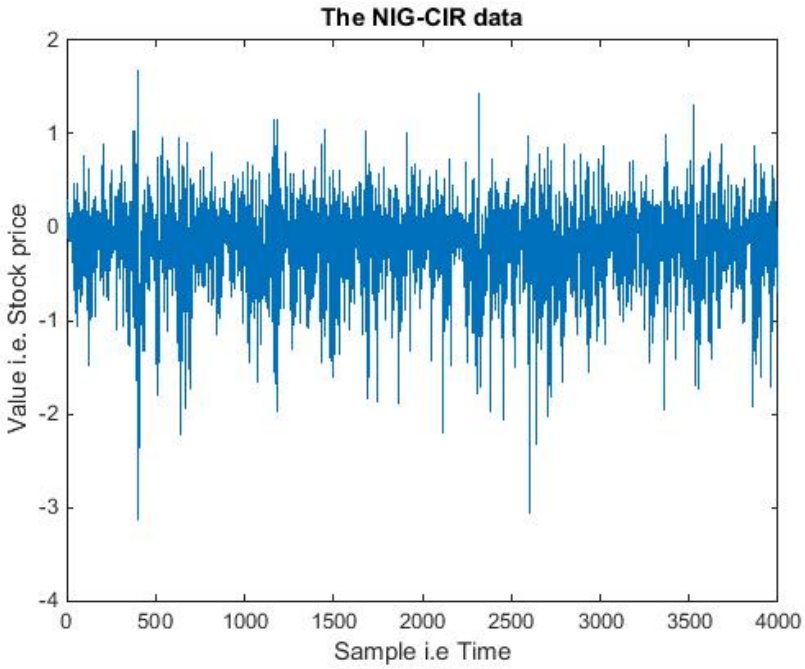

Figure 1: A plot over the data set used to estimate the parameters of the different processes. The first parameter set is used to generate the first 2000 samples and the remaining samples are generated using the second parameter set, seen on the previous page.

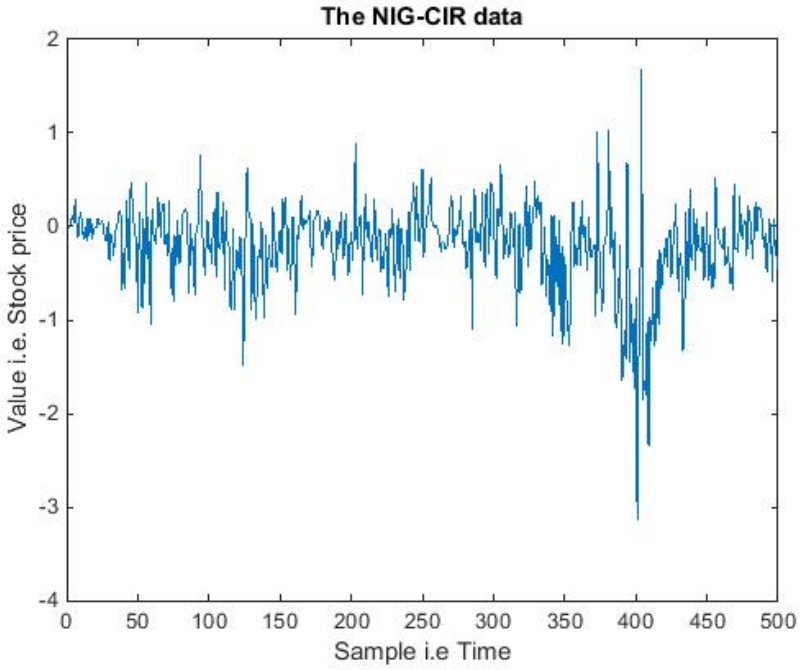

Figure 2: A closer look at the first 500 samples of the NIG-CIR data set. In this plot it is possible to see that the process have stochastic volatility.

3.2 Fitting distributions to the data

Once the data that was to be used as foundation for the simulations was generated, the next step was to fit the different distributions to the data. The generated data set was 4000 samples long and for each fitting of parameter estimates 1000 samples were used, iterating forward with 50 samples at a time finally yielding 60 parameter estimates for each distribution. The distributions that were fitted are:

- Normal Distribution
- Student's t Distribution
- GARCH process with Normal Distribution
- GARCH process with Student's t Distribution
- EGARCH with Normal Distribution
- Stochastic Volatility process - Taylor 82

Although six different distributions were fitted, only five were used in the proceeding calculations. This is due to that the parameter estimates generated for the GARCH process with student's t distribution sometimes yielded an unstable processes. It is known from [14] that fitting stable parameters to a GARCH process with student's t distribution could be tricky and they have come up with a solution to this problem based on Generalized Auto regressive Scoring algorithms. But this article was found rather late in the process of this thesis and the method is hence not used in this thesis, instead the GARCH with Student's t distribution is excluded when calculating the average. The interested reader can read more about the Generalized Autoregressive Scoring algorithms in [14].

The Normal and Student's t distribution were generated using the built in functions `makedist` and `fitdist` in Matlab. The fitting of the GARCH and EGARCH processes was done using MLE as explained in 2.6.2. To establish numerical stability and make sure that the estimator converged to the correct values, some tricks were performed. I.e. when taking the inverse of the Fisher matrix an eigenvalue matrix with small values were added to avoid zeros along the diagonal and the pseudo inverse, `pinv`, was used to calculate the inverse. And in order for the estimator not to take too big steps at each iteration, each step size was reduced and the number of iterations were increased to complement the reduced step size. In figure 3 and 4 the iteration process of the parameter estimates of the MLE procedure is showed.

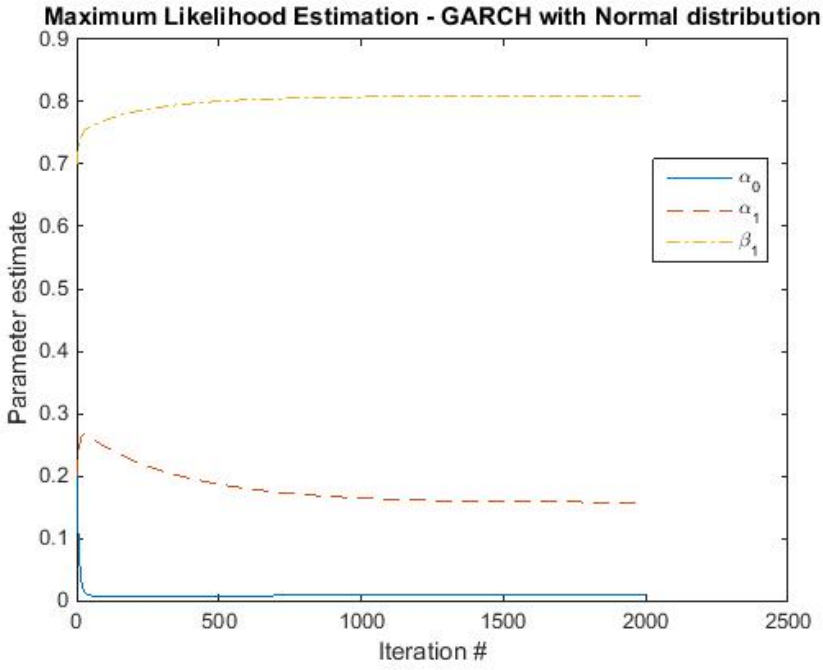

Figure 3: The estimation of the parameters of the GARCH(1,1) with Normal distribution as the iteration proceeds

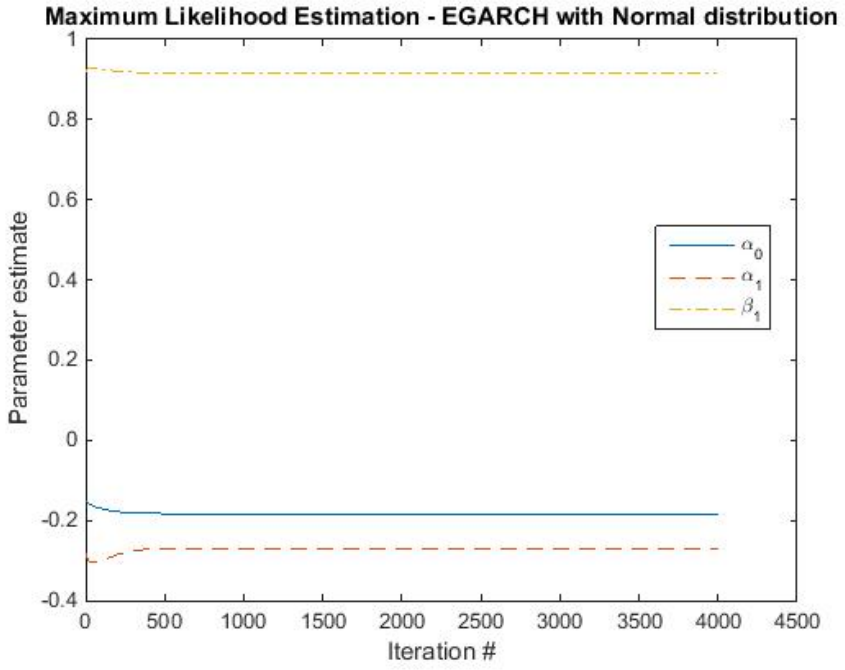

Figure 4: The estimation of the parameters of the EGARCH(1,1) as the iteration proceeds.

To calculate the parameters for the Taylor 82 process a Kalman filter was used.

Figure 5 to 9 shows the parameter estimates of the five different processes for the 60 different data sets. Comments on the different plots can be found in the captions.

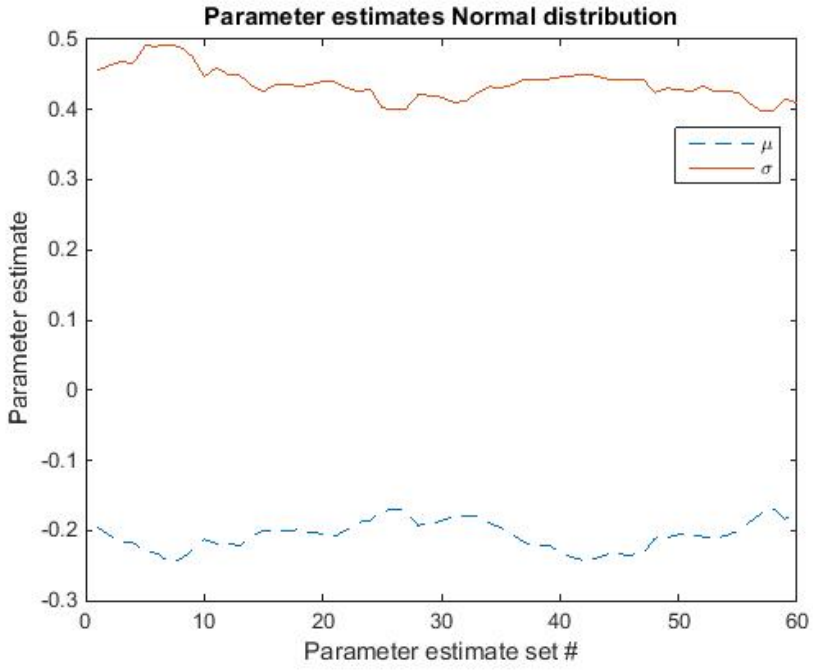

Figure 5: The parameter estimates of the Normal distribution for each of the 60 data sets. As seen the estimated parameters seem to be rather stable over the entire data set, even when the parameters have changed slightly it remains stable.

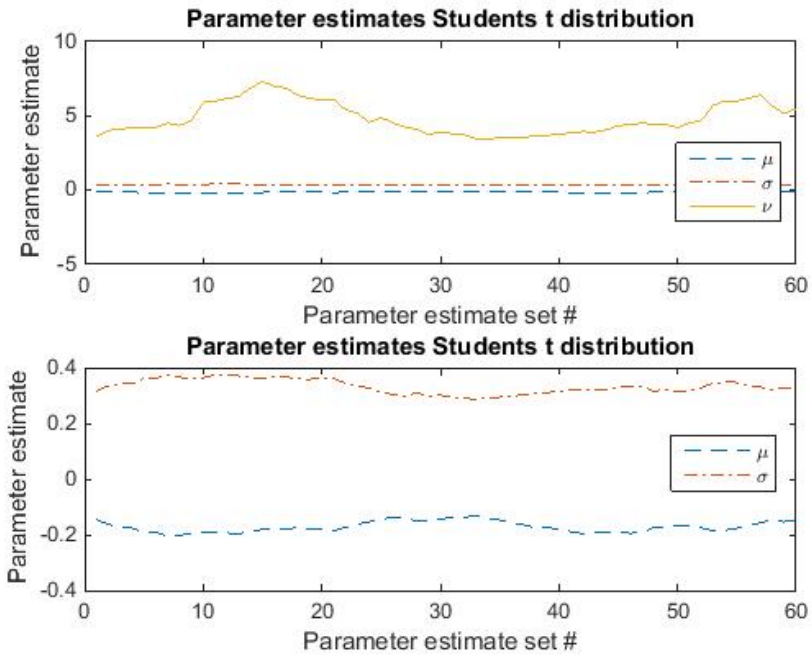

Figure 6: The parameter estimates of the Student's t distribution for each of the 60 data sets. The most notable in this plot is how much the degrees of freedom, ν changes. The reason for this is not something discussed closer in this thesis, but it is a behavior that is present rather often.

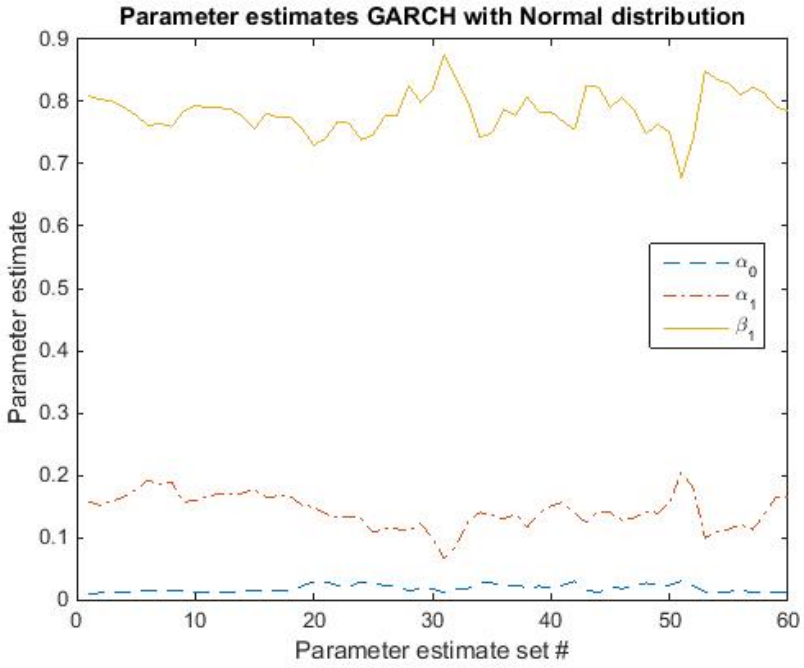

Figure 7: The parameter estimates of the GARCH(1,1) with Normal distribution for each of the 60 data sets. The parameter estimates are rather consistent through out the plot and it is possible to see a small peak and dip in the estimates around sample no. 30. This is when there have been a slight change in the parameters generating the NIG-CIR process.

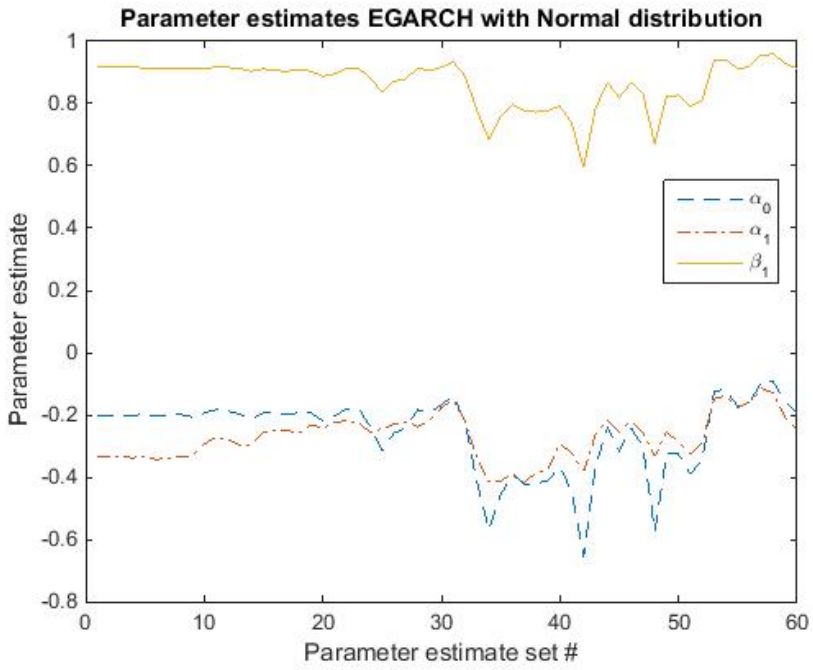

Figure 8: The parameter estimates of the EGARCH(1,1) distribution for each of the 60 data sets. The parameter estimates for the EGARCH process is not as stable for the second half of the data. Around 30 a clear dip is seen which is followed with more varying behavior. But it is possible to see that if there is a dip/peak in one of the estimated values, the other parameter estimates will also have a dip/peak at that time.

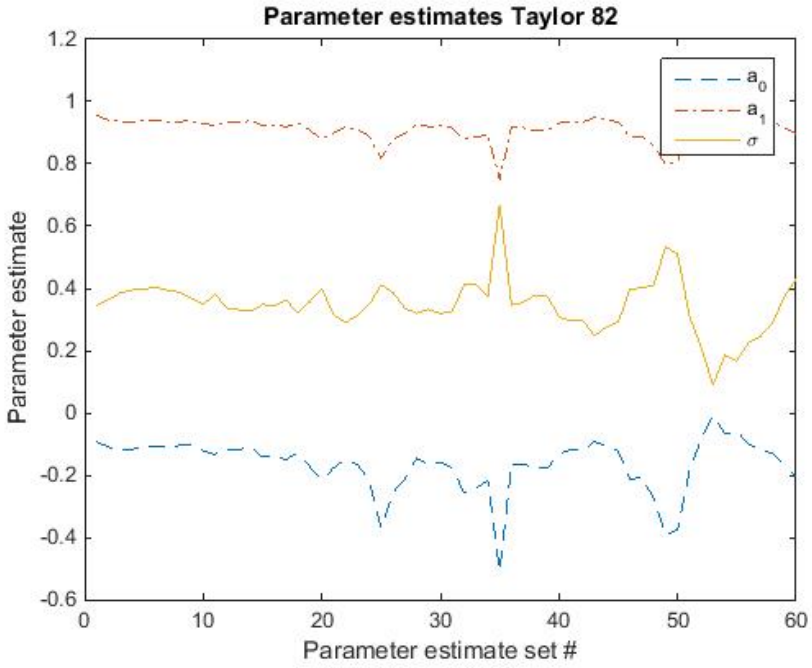

Figure 9: The parameter estimates of the Taylor 82 process for each of the 60 data sets. The parameter estimates for the Taylor 82 model show similarities to the EGARCH estimates in the sense that they are more stable in the first half, have a clear peak/dip around 30 and a more varying behavior in the second part.

3.3 BMA

The next step is to calculate weights used for the BMA. When calculating the weights BIC is used, the calculated BIC measures is of magnitude BIC $\sim 1000 - 4000$. If calculating the weights using (10) the magnitude of the BIC measure will become a problem, since the weights then become very large and approach infinity. Adjusting this as in (85) is the same thing as when using the unadjusted weights. This can be seen when normalizing the weights as in (86).

$$w = e^{(-\frac{\text{BIC}}{2})} = e^{(-\frac{\text{BIC}\pm A}{2})} = e^{\frac{A}{2}} e^{(-\frac{\text{BIC}-A}{2})} \tag{85}$$

Which yields

$$\begin{aligned} w_j &= \frac{e^{(-\frac{\text{BIC}_j}{2})}}{\sum_{l=1}^{n} e^{(-\frac{\text{BIC}_l}{2})}} \\ &= \frac{e^{\frac{A}{2}} e^{(-\frac{\text{BIC}_j - A}{2})}}{\sum_{l=1}^{n} e^{\frac{A}{2}} e^{(-\frac{\text{BIC}_l - A}{2})}} \\ &= \frac{e^{(-\frac{\text{BIC}_j - A}{2})}}{\sum_{l=1}^{n} e^{(-\frac{\text{BIC}_l - A}{2})}} \end{aligned} \tag{86}$$

A problem with this way of calculating the weights is that due to the considerably different sizes of BIC only one of the models is considered, that is the Taylor 82 model which can be seen in 10. The way of calculating the log-Likelihood function have been validated by testing it on GARCH data, in which case the GARCH with Normal distribution and Taylor 82 model preforms equally well. The conclusion to draw from this is that considered the data in this thesis the Taylor 82 model is a significantly better fit than the other models, and hence it is the only one used.

But if instead adjusting the BIC according to (87)

$$w = e^{-\frac{(\text{BIC}/A)}{2}} \tag{87}$$

all models get considered, even though the BIC values differs rather much. In both cases the constant A was set to be BIC_{max} which is the maximum BIC from this set of values. Different choices of the value A was tested, i.e. the mean of the BIC values, but it was decided to use BIC_{max} since it yielded good results. The weights that were used was calculated according to (87) and can be seen in Figure 11. As seen in the figure this will yield a result that is rather close to equal weights. But with a slight consideration of the BIC measurements.

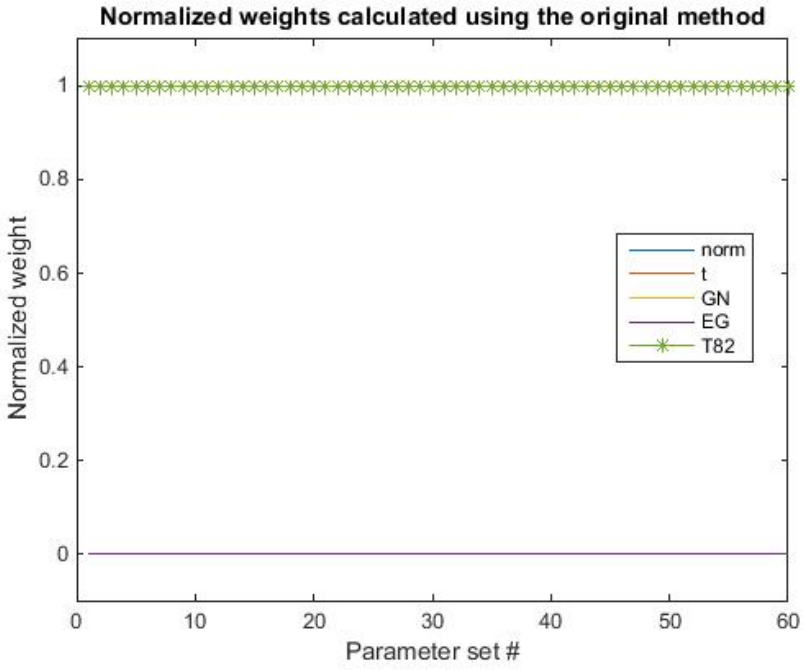

Figure 10: The calculated weights using the weight in (10). As seen in the figure only Taylor 82 is chosen.

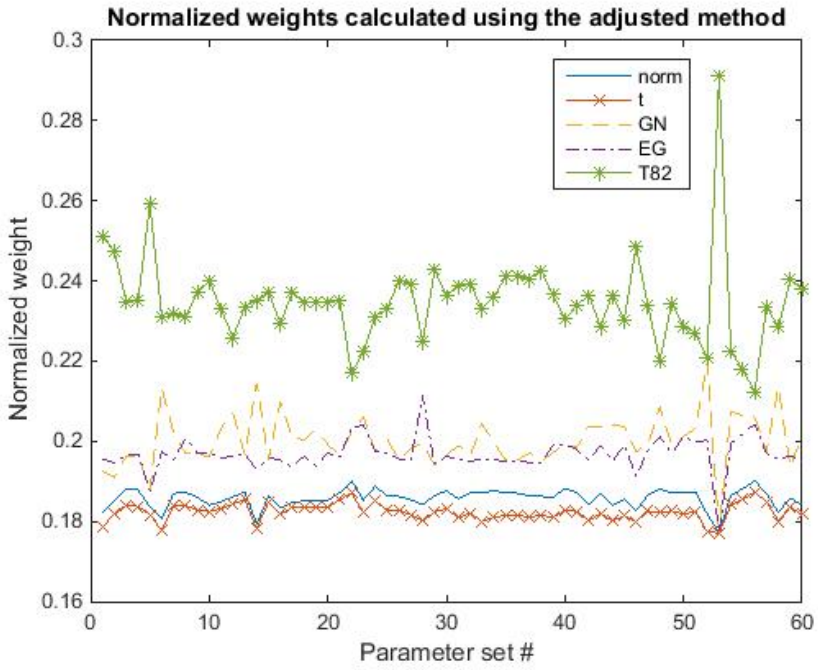

Figure 11: The calculated weights using the weight in (87). All models will be considered almost equally, but the will be a preference of the Taylor 82, and a slight preference of the GARCH and EGARCH process.

Before finally calculating the BMA estimates for VaR and ES, the sample size used for the BMA need to be established. This was done according to (31), with the constant $C = 10^{-2}$ and $\alpha = 0.1$ the sample size N_α was calculated to be

$$N_\alpha = \frac{1}{0.1 \cdot (10^{-}2)^2} = 100\,000 \tag{88}$$

This means that using an α-level of 0.1 there will be $\sim$ 10 000 samples that are used to calculate the ES estimate, which was deemed a sufficient amount of samples for the result to be relevant. Since VaR essentially is the α-quantile of the data, the matlab function `prctile` was used to calculate the VaR estimates. When knowing the VaR estimates it is straight forward to calculate the ES estimates using (26). To be sure that the averaging of the VaR and ES estimates was not a problem the averaging was performed by creating a new mixed data sample, where the mixture was based on the weights. It was from this new weighted and mixed data sample the VaR and ES estimates finally were calculated.

4 Results

The calculated estimates for VaR and ES can be seen in Figure 12 to 15. It is the unconditional 1-step prediction of VaR and ES that is calculated. Due to that Monte Carlo methods is used to calculate VaR and ES, it is trivial to calculate the n-step prediction. As previously mentioned, when using the unadjusted weights only the Taylor 82 model is chosen is chosen, since it has the best fit for the data. That means that when it is written Taylor 82 in the figures, that correspond to the original weights and the parts where it is written weighted corresponds to the adjusted weights.

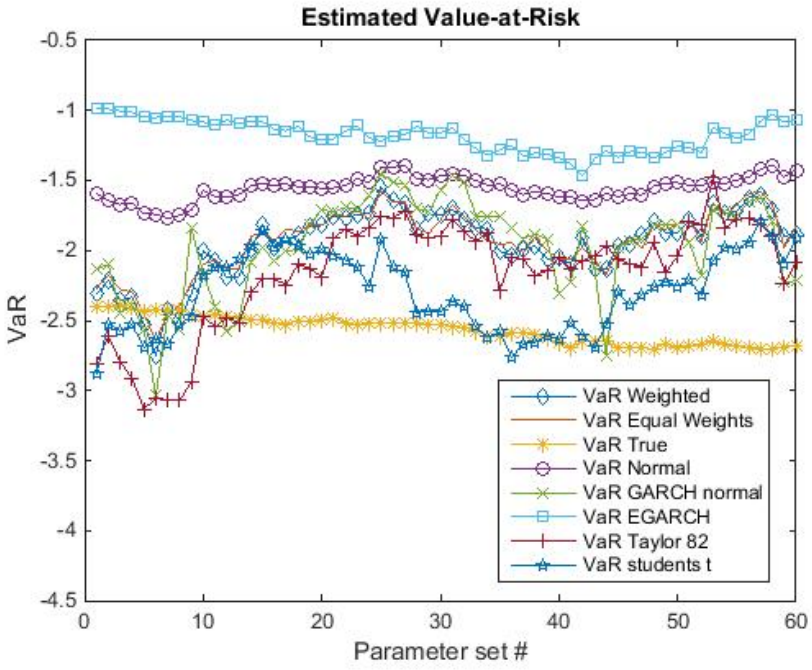

Figure 12: The different calculated VaR estimates. In the picture it is possible to see both the estimated values and the true value. One thing that is possible to see in this figure is that most of the functions under estimate the VaR, the idea was that GARCH with student's t were to compensate for this, since the t distribution usually yields higher estimates, which also is possible to see in the picture.

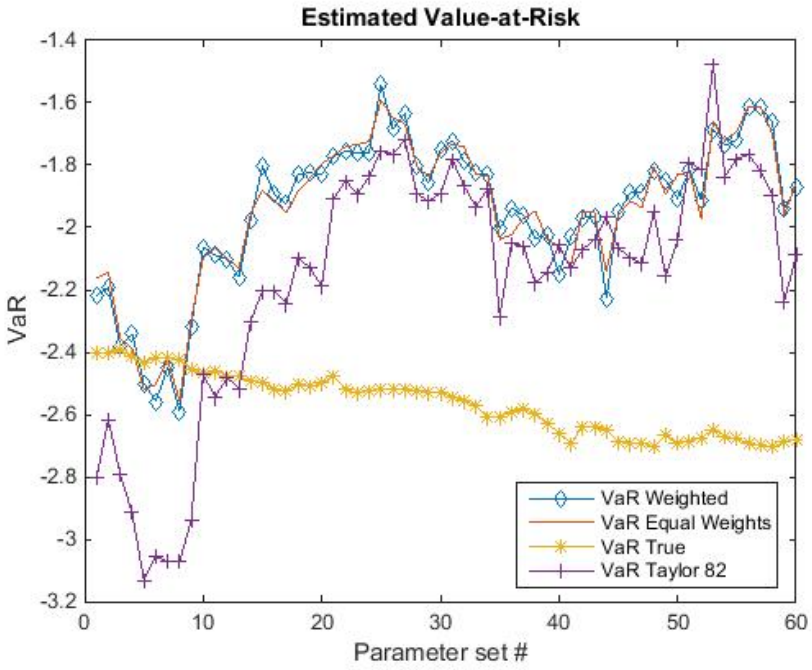

Figure 13: A closer look at the more interesting VaR estimates, the different weighted averages, where the line corresponding to Taylor 82 also corresponds to the original weights.

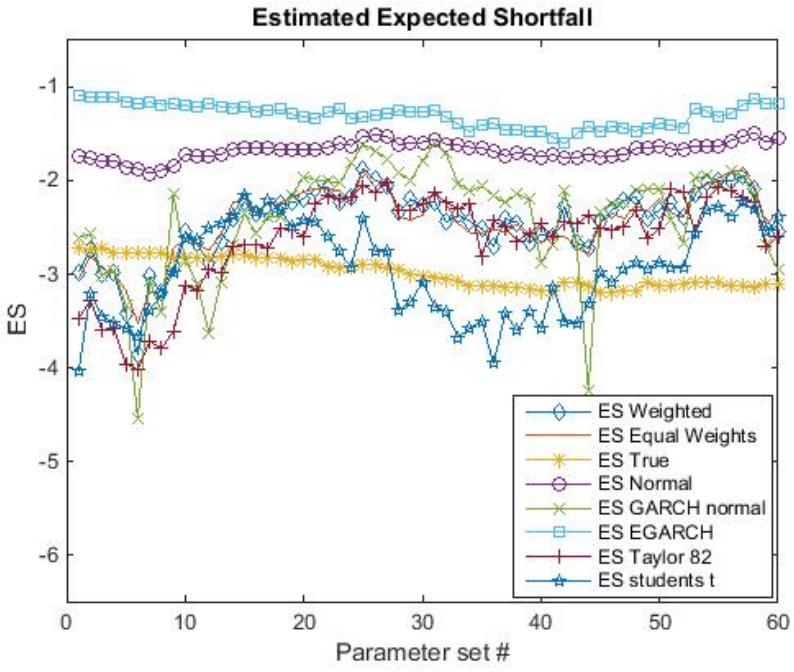

Figure 14: The calculated ES estimates for the different models.

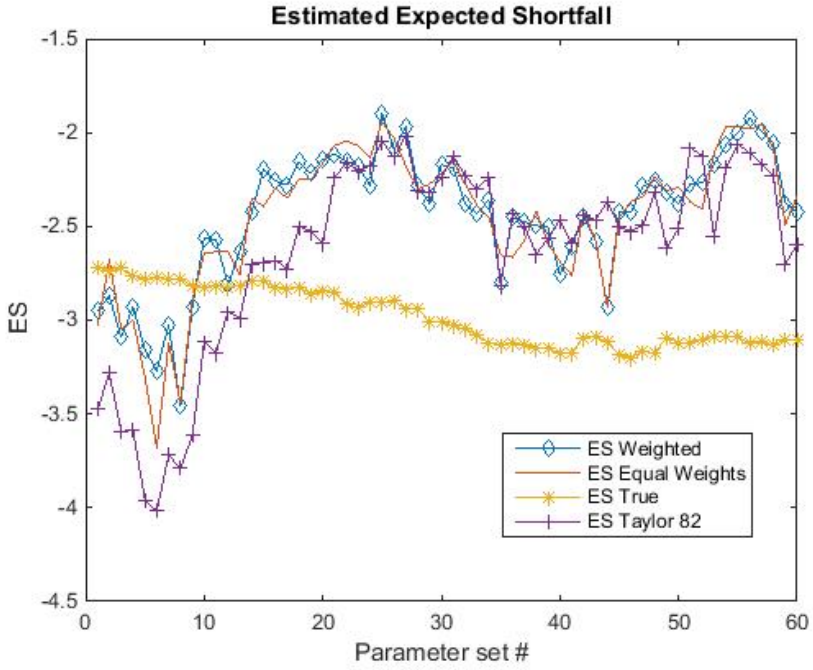

Figure 15: A closer look at the more interesting ES estimates calculated using the different averaging processes. The line corresponding to Taylor 82 also corresponds to the original weights.

The first thing that is noticed when studying the different pictures is that there is not any average or model that outperforms the others, but they are all rather similar. One thing that is easy to see is that using only the Normal distribution will make for a underestimation of the risk. This is due to that the data has heavier tails than the Normal distribution is able to predict. Another thing that is seen in the figures is that the different estimates follow each other, if the one of the estimates increases in value, most of the others does the same at approximately the same time. A further interesting thing is that the EGARCH estimates yields an even worse estimate than the Normal distribution, the reason for this is unclear. In order to see if there seems to be any model that is performing better than the others the absolute value of the errors is analysed and the results are presented in the tables below.

Error of the Value-at-Risk estimates		
Distribution	*Mean*	*SD*
Weighted average adjusted weights	0.6331	0.2741
Weighted average original weights	0.5641	0.2368
Equally weighted average	0.6320	0.2728
Normal distribution	1.0091	0.1569

Errors of the Expected Shortfall estimates		
Distribution	*Mean*	*SD*
Weighted average adjusted weights	0.6324	0.2752
Weighted average original weights	0.6594	0.2802
Equally weighted average	0.6363	0.2803
Normal distribution	1.3041	0.2066

When studying the errors the previous guess that the different averaging estimates were similarly good is confirmed. Their mean errors are very similar to each other and they deviate approximately equally much from the true curve. It is known that an equally weighted portfolio sometimes outperforms a minimum variance portfolio when it is out of sample [16]. Considering the results in this thesis it is reasonable to assume that the same principle might be valid to this area of application.

There might be several reasons for these slightly vague results. One of these could be that there were some problems when fitting the different models to the unknown data, as it was with the GARCH with Student's t distribution, so that the parameters did not converge to the optimal values. These problems could be due to the chosen parameters of the NIG-CIR process. Another problem is that the BIC values varied quite a lot, which lead to difficulties calculating the weights which in turn resulted in that solely the Taylor 82 process was chosen or that the averaging was performed with either equal or close to equal weights.

Figure 16 shows a histogram of the NIG-CIR data used as validation of the estimates. The red line in the figure is a fitted Normal distribution. A closer

study of this figure shows that it has several of the properties that are common to financial data i.e. leptokurtosis, skewness and heavy tails. This is best seen by comparing to the Normal distribution in the picture. The peak is more narrow and higher than for the Normal distribution yielding that is it leptokurtic and the highest peak of the histogram is slightly shifted from that of the Normal distribution. By a closer study of the tails, it is possible to see that the NIG-CIR data has heavier tails than the Normal distribution. This explains why the VaR and ES estimates using only the Normal distribution is underestimated, as mentioned in 2.1.

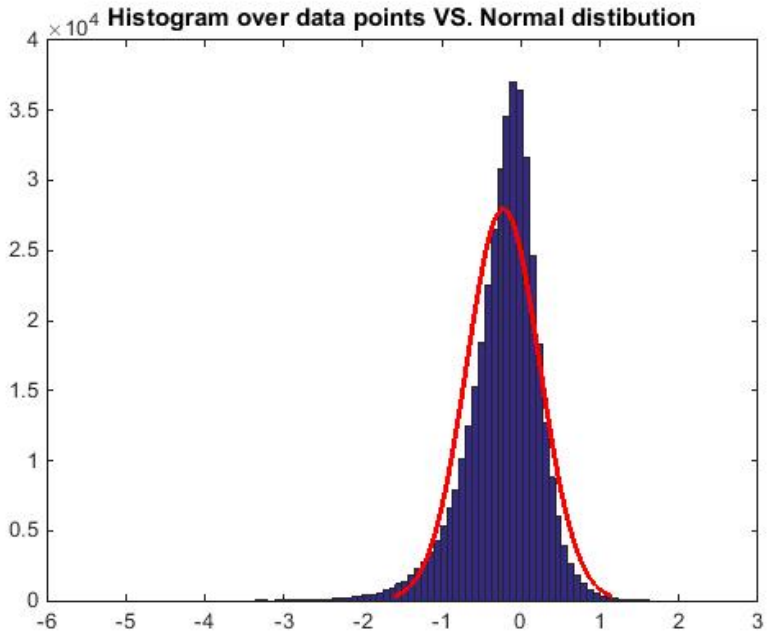

Figure 16: A histogram of the NIG-CIR data, the red line corresponds to a Normal distribution fitted to the histogram.

5 Conclusions

As stated in the results, in this thesis, there is not one averaging process calculating an average over several models or using one model, that fits remarkably better than the other. But calculating an average, regardless of the weights used, is extensively better than assuming that the data is Normal distributed. If the amount of time and work that is needed to calculate the different weights are taken into account when deciding on how good a fit the model is the conclusion, based on the results in this thesis, is that the equally weighted average would be preferable over the weighted ones. This is due to that the accuracy of the estimates differ very little between the weighted and the equally weighted averages.

The main problem with the result in this thesis is that when using the intended weights, the fit of the Taylor 82 model was significantly better than for the other models to the extent where the Taylor 82 model was the only model chosen. If more models would have been considered it is likely that several of these would have given an equally good or even better fit than the Taylor 82 model. More models would also change the estimated weights which in turn might yield even better estimates.

And even though this thesis did not shine any light on which weights are the best to use to get the most accurate estimation fit, it is possible to draw the conclusion that using model averaging for calculating Value-at-Risk and Expected Shortfall is a good method. More considered models might also yield a more stable estimate.

A way to continue the research within this topic is to consider how to calculate the weights. There are many different information criterion's that could be considered and that is just a small part all the possibilities for developing new weights. The interested reader can read more about different information criterions in [9].

Appendices

A Derivation of the Hessian matrix - MLE

A.1 Normal distribution

Deriving the expressions for J originates from the log-Likelihood function. The log-likelihood function for the Normal distribution is seen in (89)

$$\ell_t^N(\theta) = -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma_t^2) - \frac{1}{2}\frac{y^2}{\sigma_t^2} \tag{89}$$

The first derivative is

$$\frac{\partial \ell_t^N}{\partial \theta} = \frac{1}{2\sigma_t^2}\frac{\partial \sigma_t^2}{\partial \theta}\left(\frac{y_t^2}{\sigma_t^2} - 1\right) \tag{90}$$

Yielding the second derivative to be

$$\begin{aligned}
\frac{\partial^2 \ell_t^N}{\partial\theta\partial\theta} &= \left(\frac{y_t^2}{\sigma_t^2} - 1\right)\frac{\partial}{\partial\theta}\left(\frac{1}{2\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta}\right) + \frac{1}{2\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta}\frac{\partial}{\partial\theta}\left(\frac{y_t^2}{\sigma_t^2} - 1\right) = \\
&= \left(\frac{y_t^2}{\sigma_t^2} - 1\right)\frac{\partial}{\partial\theta}\left(\frac{1}{2\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta}\right) + \frac{y_t^2}{2\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta}\frac{\partial(1/\sigma_t^2)}{\partial\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta} = \\
&= \left(\frac{y_t^2}{\sigma_t^2} - 1\right)\frac{\partial}{\partial\theta}\left(\frac{1}{2\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta}\right) - \frac{y_t^2}{2(\sigma_t^2)^3}\frac{\partial\sigma_t^2}{\partial\theta}\frac{\partial\sigma_t^2}{\partial\theta}
\end{aligned} \tag{91}$$

Defining the standardized residuals as $Z_t = y_t/\sigma_t$ where Z_t is SWN, so that we have $Z_t|\psi_{t-1} \sim SWN(0,1)$ [12]. The following results is known

$$E(Z_t|\psi_{t-1}) = 0, \quad E(Z_t^2|\psi_{t-1}) = 1, \quad E(Z_t^2 - 1|\psi_{t-1}) = 0 \tag{92}$$

Since σ_t^2 only depends on past values of the residuals Z_i and σ_t^2 are independent. Knowing this yields [12]

$$E\left(\frac{Z_t^2}{(\sigma_t^2)^2}\middle|\psi_t\right) = E(Z_t^2)E\left(\frac{1}{(\sigma_t^2)^2}\right) = \frac{1}{(\sigma_t^2)^2} \tag{93}$$

Combining the concept of standardized residuals with the expression in (91) yields the expression for J

$$\begin{aligned}
J &= -E\left(\frac{1}{2\sigma_t^2}(Z_t^2 - 1)\frac{\partial^2\sigma_t^2}{\partial\theta\partial\theta} - \frac{Z_t^2}{2(\sigma_t^2)^2}\frac{\partial\sigma_t^2}{\partial\theta}\frac{\partial\sigma_t^2}{\partial\theta}\right) \\
&= -\frac{1}{2\sigma_t^2}E(Z_t^2 - 1)\frac{\partial^2\sigma_t^2}{\partial\theta\partial\theta} + \frac{E(Z_t^2)}{2(\sigma_t^2)^2}\frac{\partial\sigma_t^2}{\partial\theta}\frac{\partial\sigma_t^2}{\partial\theta} \\
&= \frac{1}{2\sigma_t^4}\frac{\partial\sigma_t^2}{\partial\theta}\frac{\partial\sigma_t^2}{\partial\theta}
\end{aligned} \tag{94}$$

A.2 Student's t distribution

The same calculations can be performed on the Student's t distribution. Begin with the log-likelihhod function of the t distriution as seen in (95), the first derivative can then be seen in (96)

$$\begin{aligned} \ell_t^T(\theta) &= \log(\Gamma((\nu+1)/2)) - \log(\Gamma(\nu/2)) - \frac{1}{2}\log(\pi) \\ &-\frac{1}{2}\log(\nu-2) - \frac{1}{2}\log(\sigma_t^2) - \frac{\nu+1}{2}\log\left(1+\frac{y^2}{(\nu-2)\sigma_t^2}\right) \end{aligned} \tag{95}$$

$$\frac{\partial \ell_t^T}{\partial \theta} = \frac{1}{2\sigma_t^2}\frac{\partial \sigma_t^2}{\partial \theta}\left(\frac{y_t^2}{\sigma_t^2}\mathcal{G} - 1\right) \tag{96}$$

Where $\mathcal{G}$ is defined as in (97)

$$\mathcal{G} = \frac{\nu+1}{(\nu-2)+\frac{y^2}{\sigma_t^2}} \tag{97}$$

The second order derivative then becomes

$$\begin{aligned} \frac{\partial^2 \ell_t^T}{\partial\theta\partial\theta} &= \left(\frac{y_t^2}{\sigma_t^2}\mathcal{G}-1\right)\frac{\partial}{\partial\theta}\left(\frac{1}{2\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta}\right) + \frac{1}{2\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta}\frac{\partial}{\partial\theta}\left(\frac{y_t^2}{\sigma_t^2}\mathcal{G}-1\right) = \\ &= \left(\frac{y_t^2}{\sigma_t^2}\mathcal{G}-1\right)\frac{\partial}{\partial\theta}\left(\frac{1}{2\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta}\right) + \frac{y_t^2}{2\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta}\frac{\partial}{\partial\theta}\left(\frac{y_t^2}{\sigma_t^2}\frac{\nu+1}{(\nu-2)+\frac{y^2}{\sigma_t^2}}-1\right) \\ &= \left(\frac{y_t^2}{\sigma_t^2}\mathcal{G}-1\right)\frac{\partial}{\partial\theta}\left(\frac{1}{2\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta}\right) - \frac{y_t^2}{2\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta}\frac{\partial\sigma_t^2}{\partial\theta}\frac{(\nu+1)(\nu-2)}{(\sigma_t^2(\nu-2)+y_t^2)^2} \end{aligned} \tag{98}$$

Using standardized residuals yields an expression for J

$$\begin{aligned} J &= -E\left(\frac{1}{2\sigma_t^2}(Z_t^2\frac{\nu+1}{(\nu-2)+Z_t^2}-1)\frac{\partial^2\sigma_t^2}{\partial\theta\partial\theta} - \frac{Z_t^2}{2(\sigma_t^2)^2}\frac{\partial\sigma_t^2}{\partial\theta}\frac{\partial\sigma_t^2}{\partial\theta}\frac{(\nu+1)(\nu-2)}{(\nu-2)^2+2(\nu-2)Z_t^2+(Z_t^2)^2}\right) \\ &= -\frac{1}{2\sigma_t^2}(E(Z_t^2)\frac{\nu+1}{(\nu-2)+E(Z_t^2)}-1)\frac{\partial^2\sigma_t^2}{\partial\theta\partial\theta} + \frac{E(Z_t^2)}{2(\sigma_t^2)^2}\frac{\partial\sigma_t^2}{\partial\theta}\frac{\partial\sigma_t^2}{\partial\theta}\frac{(\nu+1)(\nu-2)}{(\nu-2)^2+2(\nu-2)E(Z_t^2)+E((Z_t^2)^2)} \\ &= \frac{1}{2(\sigma_t^2)^2}\frac{(\nu+1)(\nu-2)}{((\nu-2)+1)^2}\frac{\partial\sigma_t^2}{\partial\theta}\frac{\partial\sigma_t^2}{\partial\theta} \end{aligned} \tag{99}$$

References

[1] Alexander J. McNeil, Rüdiger Frey, Paul Embrechts *Qutantitative Risk Management* , Princeton University press, Princeton, 2005.

[2] Martin Sköld *Computer Intensive Statistical methods* , Lund University, Lund, 2005.

[3] Andreas Kyprianou, Wim Schoutens, Paul Wilmott *Exotic option pricing and advanced Lévy models* , John Wiley & Sons Ldt., Chichester, 2005.

[4] Andreas Jakobsson *Time Series Analysis and Signal modelling*, Lund University, Lund, Edition 1:1, 2012.

[5] Erik Lindström, Henrik Madsen, Jan Nygaard Nielsen *Statistics for Finance*, CRC Press, Boca Ranton, 2015.

[6] Christian Francq, Jean-Michel Zakoïan *GARCH Models - structure, statistical inference and financial applications* , John Wiley & Sons Ldt., Chichester, 2010.

[7] Luc Bauwens, Christian Hafner, Sebastien Laurent *Handbook of Volatility models and their applications*, John Wiley & Sons Ldt., Chichester, 2012.

[8] Nikolaus Hautsch, Yangguoyi Ou *Discrete-Time Stochastic Volatility Models and MCMC-Based Statistical Inference*, SFB 649 Discussion Paper 2008-063, http://sfb649.wiwi.hu-berlin.de/papers/pdf/SFB649DP2008-063.pdf , 2008.

[9] Gerda Claeskens, Nils Lid Hjorth *Model Selection and Model Averaging*, Cambridge University Press, Cambridge, 2008.

[10] Sadanori Konishi, Genshiro Kitagawa *Information Criteria and Statistical modeling*, Springer, New York, 2008.

[11] Jennifer A. Hoeting, David Madigan, Adrian E. Raftey, Chris T. Volinsky *Bayesian Model Averaging: A Tutorial*, Statistical Science, Vol 14, No. 4, 1999 ,p. 382-417.

[12] George Levy *Computational finance - numerical instruments for pricing financial instruments*, Butterworth-Heinemann, Elsevier, Oxford, 2004.

[13] R.I. Jennrich, P. F. Sampson *Newton-Raphson and related Algorithms for Maximum Likelihood Variance Component Estimation*, Technometrics, 18:1, 1976 ,p. 11-17.

[14] Drew Creal, Siem Jan Koopman, André Lucas *Generalized Autoregressive Score Models with applications*, Journal of applied econometrics, 28, 2013 ,p. 777-795.

[15] Carol Alexander *Market Risk Analysis Volume IV- Value-At-Risk Models*, John Wiley & Sons Ltd, Chichester, 2008.

[16] Victor DeMiguel, Lorenzo Garlappi, Raman Uppal *Optimal Versus Naive Diversification: How Inefficient is the 1/N Portfolio Stratergy*, Oxford University Press, Oxford, 2007.

Calculation of VaR and Expected Shortfall under model uncertainty

Alexandra Böttern *

*Lund University

A common problem for financial institutions is to decide which model should be used when modelling prices and risks, in this article we focus on the risk aspect. When you have unknown data, for example losses, it is always a problem to decide which model to use when making the simulations, this is a concept called model risk. In this thesis the two risk measures calculated are Value-at-Risk and Expected Shortfall and they are calculated using a model averaging approach.

Value-at-Risk | Expected Shortfall | Bayesian Model Averaging

Value-at-Risk, or VaR as it is more commonly called, is a very common risk measurement used by financial institutions. It is often used to calculate market risk. Market risk is the risk of losses that a financial institution obtains due to factors affecting the financial market. An example of a market risk is that there might be a natural disaster or terrorist attack, this will affect the market as a whole. In Basel II, the second Basel accord, it is written that VaR is the preferred method for calculating market risk. But there are some problems with using VaR, for example it does not capture tail risk. And Value-at-Risk is not a coherent risk measure. This means that it does not fulfill the four axioms of coherence that follow below, the axiom it does not fulfill is the one for subadditivity.

1. *Monotonicity* - which can be explained as, if portfolio A is consistently better than portfolio B under almost all scenarios then the risk of A should be less than the risk of B.
2. *Subadditivity* - meaning that the risk of portfolio A and B together cannot be worse than adding the risks for A and B separately.
3. *Positive homogeneity* - this means that if you double the size of your portfolio, you double the size of your risk.
4. *Translation invariance* - meaning that if you have one portfolio A with a guaranteed return and a portfolio B, the portfolio A is the same thing as adding cash to your portfolio B.

Since there is some problems with using VaR, the Basel committee have in Basel III decided that Expected Shortfall (ES) shall be used instead of VaR. Expected Shortfall is a risk measure that is closely related to Value-at-Risk, but it is a coherent risk measure and it captures a lot of the behaviours that VaR does not.

Financial data often behaves in a special way and contain many complex behaviours. An example of this is that financial data often is leptokurtic, meaning it have a more narrow and higher peak and heavier tails than the Normal distribution. It is also common that it is skewed and that it contains volatility clustering. That a data series have volatility clusters means that one extreme event is likely to be followed by several extreme events.

These complex behaviours of financial data can be hard to model and predict if the underlying model is unknown, which usually is the case. In this thesis we use model averaging to solve this problem. By using model averaging, or more specifically Bayesian Model Averaging, the model risk is reduced. This is due to that we now consider several models instead of just choosing one. By choosing several models each one of them get to contribute with their unique behavior and generate a more confident final prediction. But there are some problems one is faced with when performing the averaging and one of these is how to select the weights used to calculate the average. We tried two different approaches when choosing the weights. One more naive approach using equal weights, which have proven to be rather good in the past. And one where the weights are calculated using Bayes Information Criterion (BIC). Bayes Information Criterion is a method of deciding which model is the best fit to the data using a Bayesian framework. The measurement is based on the log-likelihood function and punishes based on model complexity.

The different simulations performed shows that it is advantageous to use BMA rather than using just one model. But due to some problems with the modelling the results are inconclusive regarding which weights are best to use. But if the time it took to calculate the weights is taken into consideration, the equally weighted average is to prefer since it is a lot faster to calculate and yielded an equally good result. Considering that Basel III states that Expected Shortfall is to be the recommended measurement to use when calculating market risk and that the BMA showed good results, I do believe that this method is something that financial institutions can benefit from implementing.

Master's Theses in Mathematical Sciences 2015:E37
ISSN 1404-6342

LUTFMS-3289-2015

Mathematical Statistics
Centre for Mathematical Sciences
Lund University
Box 118, SE-221 00 Lund, Sweden

http://www.maths.lth.se/