

Hierarchical clustering of market risk models

Ludvig Pucek

Viktor Sonebäck

June 1, 2017

Abstract

This thesis aims to discern what factors and assumptions are the most important in market risk modeling through examining a broad range of models, for different risk measures ($\text{VaR}_{0.01}$, $\text{ES}_{0.01}$ and $\text{ES}_{0.025}$) and using hierarchical clustering to identify similarities and dissimilarities between the models. The data used is daily log returns for OMXS30 stock index and Bloomberg Barclays US aggregate bond index (AGG) from which daily risk estimates are simulated.

In total, 33 market risk models are included in the study. These models consist of unconditional variance models (Student's t distribution, Normal distribution, Historical simulation and Extreme Value Theory (EVT) with Generalized Pareto Tails (GPD)) and conditional variance models (ARCH, GARCH, GJR-GARCH and EGARCH). The conditional models are used in filtered and unfiltered market risk models.

The hierarchical clustering is done for all risk measures and for both time series, and a comparison is made between $\text{VaR}_{0.01}$ and $\text{ES}_{0.025}$.

The thesis shows that the most important assumption is whether the models have conditional or unconditional variance. The hierarchy for assumptions then differ depending on time series and risk measure. For OMXS30, the clusters for $\text{VaR}_{0.01}$ and $\text{ES}_{0.025}$ are the same and the largest dividing factors for the conditional models are (in descending order):

- Leverage component (EGARCH or GJR-GARCH models) or no leverage component (GARCH or ARCH)
- Filtered or unfiltered models
- Type of variance model (EGARCH/GJR-GARCH and GARCH/ARCH)

The $\text{ES}_{0.01}$ cluster shows that $\text{ES}_{0.01}$ puts a higher emphasis on normality or non-normality assumptions in the models.

The similarities in the different clusters are more prominent for OMXS30 than for AGG. The hierarchical clustering for AGG is also more sensitive to the choice of risk measure. For AGG the variance models are generally less important and more focus lies in the assumed distributions in the variance models (normal innovations or student's t innovations) and the assumed final log return distribution (Normal, Student's t, HS or EVT-tails).

In the lowest level clusters, the transition from $\text{VaR}_{0.01}$ to $\text{ES}_{0.025}$ result in a smaller model disagreement.

Sammanfattning - Hierarkisk klustring av marknadsriskmodeller

Denna uppsats syfte är att utröna vilka faktorer och antaganden som är de viktigaste i marknadsriskmodellering genom att undersöka en mängd modeller, för riskmått ($\text{VaR}_{0.01}$, $\text{ES}_{0.01}$ and $\text{ES}_{0.025}$) och genom hierarkisk klustring identifiera likheter och skillnader mellan modellerna

Datan som används är dagliga log-returns för OMXS30 och Bloomberg Barclays US aggregate bond index (AGG) från vilka dagliga riskestimat simuleras.

Totalt används 33 marknadsriskmodeller i denna studie. Dessa modeller består av modeller med obetingad varians (Student's t-fördelning, normalfördelning, historisk simulering och extremevärdeteori med Generalized Pareto svansar i fördelningen (GPD)) och modeller med betingad varians (ARCH, GARCH, GJR-GARCH och EGARCH). De betingade variansmodellerna används som filtrerade och ofiltrerade modeller.

Den hierarkiska klustringen görs för alla riskmått och för båda tidsserierna. En jämförelse görs mellan $\text{VaR}_{0.01}$ and $\text{ES}_{0.025}$.

Denna studie visar att det viktigaste antagande är om modellerna har betingad eller obetingad varians. Sedan skiljer hierarkin gällande vilka antaganden som är viktigast beroende på tidsserie och riskmått. För OMXS30 är klustrena för $\text{VaR}_{0.01}$ och $\text{ES}_{0.025}$ likadana och de viktigaste faktorerna i modelleringen är (i sjunkande ordning):

- Leverage-komponent (EGARCH or GJR-GARCH models) eller ingen leverage-komponent (GARCH or ARCH)
- Filtrerad eller ofiltrerad modell
- Typ av variansmodellering (EGARCH/GJR-GARCH and GARCH/ARCH)

Klustret för $\text{ES}_{0.01}$ visar att $\text{ES}_{0.01}$ sätter en större vikt vid antagandet om normalfördelning eller inte normalfördelning i modellerna.

Likheterna i de olika klustrena är mer framträdande för OMXS30 än för AGG. Klustren för AGG är även mer känsliga för valet av riskmått. För AGG är de olika valen av variansmodell generellt sett mindre viktiga och fokus ligger istället på den antagna fördelningen i variansmodellerna (normalfördelade eller Student t-fördelade) och den antagna slutgiltiga fördelningen (normal, Student's t, historisk simulering eller EVT).

I de lägsta nivåerna i klustrena resulterar bytet från $\text{VaR}_{0.01}$ till $\text{ES}_{0.025}$ i en mindre spridning mellan modellerna.

Acknowledgements

We would like to thank our supervisor Boualem Djehiche, at KTH Royal Institute of Technology, for the inspiring meetings and helpful discussions.

Furthermore we would like to extend our sincerest thanks and appreciation to Magnus Hansson, at Handelsbanken Group Risk Control, for his valuable help and encouragement throughout this thesis.

Contents

1	Introduction	3
2	Literature Review	5
2.1	Market risk models	5
2.1.1	Unconditional variance models	5
2.1.2	Conditional variance models	5
2.2	Methods for comparing market risk models	6
3	Theoretical Framework	7
3.1	Time series models	7
3.1.1	AR(p)	7
3.1.2	ARCH(q)	7
3.1.3	GARCH(p, q)	7
3.1.4	GJR-GARCH(p, q)	7
3.1.5	EGARCH(p, q)	8
3.2	Extreme Value Theory (EVT) - Peaks over threshold (POT)	8
3.3	Empirical Value at Risk	8
3.4	Empirical Expected Shortfall	8
3.5	Hierarchical clustering	9
4	Method	10
4.1	Research process	10
4.2	Market risk models	10
4.2.1	Selected models	10
4.2.2	Model descriptions	12
4.3	Estimation of the risk measures	13
4.4	Clustering analysis and the risk ratio	13
4.5	Comparing VaR _{0.01} and ES _{0.025}	14
4.6	Data analysis	14
4.6.1	Stock return data (OMXS30)	15
4.6.2	Interest rate return data (Bloomberg Barclays US Aggregate Bond Index)	17
5	Results	19
5.1	Stock return modeling (OMXS30)	20
5.1.1	VaR _{0.01}	20
5.1.2	ES _{0.01}	24
5.1.3	ES _{0.025}	27
5.1.4	Comparison between VaR _{0.01} and ES _{0.025}	31
5.2	Bond index return modeling (AGG)	33
5.2.1	VaR _{0.01}	33
5.2.2	ES _{0.01}	37
5.2.3	ES _{0.025}	40
5.2.4	Comparison between VaR _{0.01} and ES _{0.025}	43
6	Discussion	46
6.1	Difference between clusters for OMXS30 and AGG	46
6.2	Difference between clusters for the different risk measures	47
6.3	The study and choosing a market risk model	47
6.4	The methodology's impact on the result	47
6.4.1	Risk ratio	48

6.4.2	The dissimilarity measure of the hierarchical clustering and its effect on the clustering	48
6.4.3	Time horizon	48
6.4.4	Model choices	49
6.4.5	Monte Carlo estimation of the risk measures	49
7	Conclusion	50

1 Introduction

Over the years there has been an increasing amount of regulation regarding financial institutions. There are numerous governing documents such as the Basel accords, FRTB (Financial Review of the Trading Book), IFRS (International Financial Reporting Standards) to only name a few. Some of these documents have placed limitations on the banking sector regarding minimum capital requirements. The purpose of these is to make sure that banks are solvent enough to withstand certain risky events that might occur, with a specified level of probabilistic certainty. There are standardized models dictated by IFRS9 which provide banks with a capital requirement level that is most often high enough that the banks have incentives to construct their own internal models for risk measurement, since having too much extra capital unused raises capital costs. During the recent years, these internal models have decreased the amount of required capital for banks (Sveriges Riksbank, 2015).

The internal quantitative risk models rest upon different assumptions regarding, for example, volatility modeling or type of probability distribution for the returns. The models are therefore very varying from the aspect of complexity. These assumptions (or choices in model construction) cause different levels of disagreement between the models. The level of disagreement can depend on, for example, the current situation in the market (Danielsson et al., 2016), the characteristics of the time series in question and/or the choice of risk measure.

A market risk model can be very precise during specific time periods due to numerous factors, but might be misleading during others, for example during times of financial turmoil. This can be due to that the models fail to take into account the characteristics of the data. There are several model assumptions that can cause the model to fail during certain situations (Gibson, 2000). An important note on the subject is that there is no model that has the best performance for all periods in the time series or across different time series.

Since we generally do not know which model will perform the best, it can be important to evaluate the level of impact that the different model choices can have on the risk estimation. This can be done in order to, for example, highlight potential risky and crucial choices in the modeling which therefore should receive extra focus, or to explain the similarity/dissimilarity between the model and a benchmark or other models.

The purpose of this thesis is to investigate the similarity of different market risk models and finding key choices in model construction that to a large extent explains the disagreement between the models. Furthermore, we show how these choices and their impact can differ depending on the characteristics of the times series and what risk measure that is used. In the latter, the transition from Value at Risk (1%) to Expected Shortfall (2.5%) and its effects on the subject is also covered.

In order to fulfill the purpose of this study, the following research question must be answered:

- What assumptions in market risk models have the largest effect on the estimated VaR/ES?

The subject is further explored by answering the following related questions:

- How much of the difference in the estimates can be explained by these assumptions?
- How does the transition from $\text{VaR}_{0.01}$ to $\text{ES}_{0.025}$ affect the impact of these assumptions?

There are numerous different types of market risk models and the possible choices of these must therefore be restricted; 28 and 33 models are implemented for times series 1 and 2, respectively. The models are chosen using mainly the literature study and the data analysis in the methodology section. This data analysis also explains why there are 28 models for one of the time series and 33 models for the other.

The number of possible financial times series that could be used in the analysis is also incalculable. The analysis is therefore restricted to two times series that have different characteristics, are well known, and represents two different asset types; equity and bonds. The times series are OMXS30 and Bloomberg Barclays US Aggregate Bond Index.

Furthermore, the analysis is restricted to the risk measures Value at Risk and Expected Shortfall. There are several different possible probability levels for VaR and ES that could be used in the study. The chosen levels are 1% for both risk measures and 2.5% for ES in order for the analysis to include the effect of the transition from VaR to ES.

The contribution can roughly be divided in two parts. The first part is insights in the different levels of importance for the choices in market risk model construction and how much of the disagreement between the models that is explained by these choices and assumptions. This is, as previously mentioned, done for both VaR and ES, and also for two different times series. Secondly, a framework for analyzing the similarities/differences between a large number of models is provided.

This study will hopefully lead to a better understanding of the similarities/differences between market risk models, and highlight important aspects of the modeling. This will in turn help banks and other financial institutions to be better able to model their risk exposure. Ultimately this leads to a safer and more efficient risk management.

2 Literature Review

This section covers relevant literature that is of importance for this study. The main focus of the presented literature is the different market risk models and ways to compare these. This section aims to give an understanding of previous work that has been done in the field and acts as a foundation for the theoretical framework and the methodology sections.

2.1 Market risk models

In the following two subsections, some of the literature on market risk models is presented. In the first one, the models have an unconditional variance. In the section thereafter, the conditional models, such as GARCH are covered.

2.1.1 Unconditional variance models

There are a wide variety of methods to model the marginal distribution and forecast market risks. The far most commonly used non-parametric method is historical simulation (Danielsson et al 2016). This method has been questioned because of, for example, the i.i.d. assumption which causes it to fail many backtests (Alexander and Sheedy, 2008) (Pritsker, 2006).

Other common approaches include fitting the historical data to a probability distribution. The normal distribution has been widely used for this purpose. However, it has also been disputed by many due to, for example, its inability of considering the excess kurtosis and skewness that financial returns often exhibit (Aparicio and Estrada, 1997) (Verhoeven and McAleer, 2004). This causes a high bias for low level VaR/ES-estimates. Among other distributions that provide a better fit to the tails is the Student's t distribution (Lin and Shen, 2006).

Another set of approaches is the extreme value theory (EVT) where you model the tails of the distribution separately. These methods can be divided into two main categories: Block maxima and Peaks of threshold (POT). Critique against the block maxima approach is that it is wasteful of data. However, the POT method utilizes all the extreme value data and is therefore often preferred (McNeil et al., 2005).

The i.i.d. assumptions that the methods above utilize can be unproblematic for, for example, insurance and operational risk. For market risk the case is different due to heteroskedasticity in the form of volatility clustering and more. Marimoutou et al (2009) among others show the importance of filtering for HS, EVT etc (Marimoutou et al., 2009). Therefore, a number of methods with conditional variance are discussed in the following sections.

2.1.2 Conditional variance models

Due to the violation of the i.i.d. assumptions for financial returns, conditional variance approaches are often used as a remedy. The ARCH model was first introduced (Engle, 1982) and was later followed by GARCH and GARCH-t (Bollerslev, 1987). Several further extensions/variations of GARCH have been proposed. For example, the GJR-GARCH and others also take into account the possible asymmetry in the ARCH process (Glosten et al., 1993), and the AR-GARCH considers possible autocorrelation in the time series.

GARCH is one of the most common models to forecast volatility (Danielsson et al., 2016). It has, however, been criticised for not adequately modeling the tail events. Thus, the Student's t

GARCH (GARCH-t) is recommended to accommodate for the leptokurtosis of returns. Furthermore, to account for skewness the skewed generalized student-t is proposed (Harris et al., 2004) (Bauwens and Laurent, 2005). These GARCH models can be combined with the approaches covered in the earlier section. One can thus create, for example, a GARCH-filtered EVT model. Kuester et al (2006) show that some VaR forecasts can be greatly improved by using these combinations (Kuester et al., 2006). Some models that are highlighted:

- Normal GARCH filtered HS
- Student's t GARCH filtered HS
- Normal GARCH filtered EVT
- Student's t GARCH filtered EVT

All these conditional VaR models have substantially greater volatility for the estimates. Therefore, it could be problematic to adjust the amount of allocated capital reserve (Danielsson and Morimoto, 2000).

2.2 Methods for comparing market risk models

The focus on the relative output of different models has been a large part of the research on model risks. The comparison is mainly done in two ways: comparing models with each other, and comparing models with a benchmark, where the latter is less common. However, the benchmark approach is used by Alexander & Sarabia (2012) among others. They use a benchmark that is supposed to represent the knowledge of the authority that is responsible for model risk. They then compare the models with this benchmark and defines the discrepancy as a model risk. The drawback with the benchmark approach is that it is heavily dependent on the choice of benchmark and that it at the same time is difficult to choose this model. Therefore, the other approach is more common.

The risk ratio is a measure that assesses the model risk in the market by investigating the disagreement between models (highest estimate divided by the lowest estimate). It is thus not a measure for a specific model. It was proposed by Danielsson et al (2016), and they show that the risk ratio is much higher in times of financial distress compared to more normal market conditions (Danielsson et al., 2016). The measure is naturally sensitive to which and how many models that are used as inputs. The measure does not assess the individual models, and the usefulness of the measure in its original form can therefore be discussed. Other approaches include a method by Kerkhof et al. (2010) where they calculate the distance from a model's estimate to the worst-case estimate and defines this as a measure of model risk.

3 Theoretical Framework

The relevant theory of the study is presented in the following subsections. The times series models, extreme value theory (EVT) and the risk measures are presented. Lastly, some theory on unsupervised learning in the form of hierarchical clustering is covered.

3.1 Time series models

3.1.1 AR(p)

A process $\{X_t, t \in \mathbb{Z}\}$ is an AR(p) if it is stationary and for every t ,

$$X_t - \theta_1 X_{t-1} - \dots - \theta_p X_{t-p} = Z_t, \quad \{Z_t\} \sim WN(0, \sigma^2).$$

Where $WN(0, \sigma^2)$ denotes the white noise with mean zero and variance σ^2 .

3.1.2 ARCH(q)

The ARCH(q) is a time series model for modeling conditional variance. It has the following construction:

$$\sigma_t^2 = c + \sum_{i=1}^q \alpha_i \epsilon_{t-i}^2,$$

where $\epsilon_t = \sigma_t Z_t$ and Z_t has an assumed probability distribution, often assumed to be $N(0, 1)$ but can be, for example, Student's t distributed.

3.1.3 GARCH(p, q)

The GARCH(p, q) process is generalized version of the ARCH(p) process and includes a parameter for the previous value of σ , meaning that the predicted volatility is conditional on both the mean process and the variance. The volatility is modeled according to the following formula:

$$\sigma_t^2 = c + \sum_{i=1}^p \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2,$$

where $\epsilon_t = \sigma_t Z_t$ with Z_t as i.i.d. $c > 0$ and $\alpha_j, \beta_j \geq 0, j = 1, 2, \dots$. The Z_t are often assumed to be $N(0, 1)$ but can be, for example, Student's t distributed.

3.1.4 GJR-GARCH(p, q)

The GJR-GARCH(p, q) is an asymmetric version of the GARCH(p, q) model. The volatility is modeled according to the following formula:

$$\sigma_t^2 = c + \sum_{i=1}^p \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2 + \sum_{i=1}^p I_{\{\epsilon_{t-i} < 0\}} \gamma_i \epsilon_{t-i}^2,$$

where $I_{\{\epsilon_{t-i} < 0\}}$ is the indicator function. This extra term containing the indicator function $I_{\{\epsilon_{t-i} < 0\}}$ means that negative deviations from the mean are weighted higher than positive ones, called the leverage parameter.

3.1.5 EGARCH(p, q)

The exponential GARCH, (EGARCH) model is specified according to the following formula:

$$\log(\sigma_t^2) = c + \sum_{i=1}^p \alpha_i \log(\sigma_{t-i}^2) + \alpha_0 + \sum_{j=1}^q \beta_j \left(\frac{|\epsilon_{t-j}|}{\sigma_{t-j}} - E \left[\frac{|\epsilon_{t-j}|}{\sigma_{t-j}} \right] \right) + \sum_{j=1}^q \xi_j \left(\frac{\epsilon_{t-j}}{\sigma_{t-j}} \right)$$

where $\epsilon = \sigma_t Z_t$. The Z_t are often assumed to be $N(0, 1)$ but can be, for example, Student's t distributed. The parameters ξ_j are the leverage component coefficients. In the GARCH and GJR-GARCH models there are non-negativity constraints on the parameters α_i and β_j . In the EGARCH models however there is no such constraint on the parameters.

3.2 Extreme Value Theory (EVT) - Peaks over threshold (POT)

The peak over threshold (POT) method seeks to determine the distribution of X over a certain threshold u , in other words we seek to model:

$$F_u(x) = P(X - u \leq x | X > u) \text{ for } x \geq 0.$$

Given n observations of I.I.D random variables (X_1, \dots, X_n) with common unknown distribution function F where F has a regularly varying right tail $\bar{F} = P(X_k > x)$ we can approximate the excess, $X_k - u$, distribution over a threshold u with a generalized Pareto distribution (GPD). The GPD function is given by:

$$G_{\gamma, \beta}(x) = 1 - \left(1 + \frac{\gamma x}{\beta}\right)^{-1/\gamma} \text{ for } x \leq 0$$

where $\gamma > 0$ and $\beta > 0$.

3.3 Empirical Value at Risk

Let X be the value of a financial portfolio at time 1. Then we have $VaR_p(X) = F_L^{-1}(1-p)$ where $L = -X/R_0$, where R_0 is the return of the reference instrument. Given a sample of L_1, \dots, L_n of independent copies of L , we get the empirical estimate of VaR as

$$\widehat{VaR}_p(X) = L_{[np]+1, n},$$

where $L_{1, n} \geq \dots \geq L_{n, n}$ is the ordered sample and $[np]$ is the integer of np .

3.4 Empirical Expected Shortfall

Expected shortfall at level p of a portfolio with value X at time 1 is given by

$$ES_p(X) = \frac{1}{p} \int_0^p VaR_u(X) du$$

so to get the empirical ES we replace $VaR_u(X)$ with its empirical estimator and get

$$\widehat{ES}_p(X) = \frac{1}{p} \int_0^p L_{[np]+1, n} du = \frac{1}{p} \left(\sum_{k=1}^{[np]} \frac{L_{k, n}}{n} + \left(p - \frac{[np]}{n} \right) L_{[np]+1, n} \right)$$

3.5 Hierarchical clustering

Hierarchical clustering is a method within the area of unsupervised machine learning. It is used to find subgroups (clusters) of observations in a given data set. The result of the hierarchical clustering is often presented in a dendrogram which is a tree-like visual representation of the observations and their similarities. Hierarchical clustering requires no pre-specified number of clusters as, for example, K-means clustering does.

Three decisions that affect the result of the clustering are:

- Choice of dissimilarity measure between observations (for example, Euclidean or Manhattan).
- Choice of type of algorithm; agglomerative (bottom up) or divisive (top down).
- Choice of dissimilarity measure between clusters (for example, complete or single linkage).

The choice of dissimilarity measure between the observation is simply how to measure the distance between two observations.

In an agglomerative hierarchical clustering all the observations start in their own clusters and the clusters are thereafter recursively merged together at different levels depending on the similarities between the clusters. If there are N observations then there are $N - 1$ steps of the merging. In a divisive hierarchical clustering it is the other way around, all the observations start in the same cluster and are thereafter split into smaller and smaller clusters until they are all in their own cluster, thus $N - 1$ splits.

In order to decide in which order the clusters should be merged/divided a dissimilarity measure between the clusters is needed. There are several different options but three examples are; complete, single and average linkage. Complete linkage is used in this study and it defines the dissimilarity between cluster A and B as:

$$d(A, B) = \max_{i \in A, j \in B} d_{i,j}.$$

Where $d_{i,j}$ is the distance between observation i in cluster A and j in the cluster B . Thus, the distance between the clusters is defined as the maximum intergroup dissimilarity.

4 Method

This section covers the scientific method used in the study. First, the research process is described. In the sections thereafter, some important steps of the research process are explained in more detail. The data analysis for the two chosen time series is presented at the end of the chapter.

4.1 Research process

The research process can be summarized in four steps:

1. Choosing and implementing the market risk models.
2. For each day of the historical data, estimating the parameters of the models and forecasting the daily VaR/ES. This is done with a parameter estimation window of 1000 days.
3. Evaluating the difference/similarities between the models and finding key assumptions/model choices and estimating how much of the total dissimilarity between the models that might be explained by these choices.
4. Compare the results from step 3 for the different time series and risk measures.

In the first step, the models are chosen using the literature study but also by finding models that have some similarities with the other models but are different in some aspect of the modeling. This is done in order to isolate some assumptions/choices, because if all the models are completely different then it will be problematic to draw any conclusions regarding which of the assumptions that actually causes the largest part of the dissimilarity between the models' estimated risk.

In the second step, the models are estimated with historical data and used to forecast the daily VaR/ES throughout the entire time series. This yields new time series of VaR/ES forecasts for each of the models.

In the third step, the difference/similarities between the models are evaluated for the two time series separately. The first part of this step involves unsupervised learning by using hierarchical clustering in order to provide insights in which models that tend to be different/similar and also which model assumptions/choices that influence the estimation of the risk measure the most, thus splitting the models into smaller and smaller clusters. Several choices of the model construction are then further investigated by analyzing how much of the maximum dissimilarity that is decreased when excluding a cluster of models.

The fourth step consists of comparing how the results differ depending on which risk measure and time series that is analyzed.

In the following sections, the steps of the research process are explained further. The chapter ends with the data analysis for the chosen time series.

4.2 Market risk models

4.2.1 Selected models

The following section describes the models that are implemented in the study. In figure 1 and 2, all models are presented that are used for time series 1 and 2, with 28 and 33 models respectively. The reason why there are more models for time series 2 (AGG) is that the data analysis shows

signs of autocorrelation in returns, meaning that models containing an AR process are added. The choices in model construction for the models in this thesis are summarized below:

1. Choose a mean process (AR(1) or mean offset).
2. Choose the variance modeling assumption (unconditional or conditional variance)
 - If unconditional variance: Choose probability distribution, Normal, Student's t, Peak Over Threshold (POT) with Pareto tails, or an empirical distribution using historical simulation with bootstrapped data.
 - If conditional variance: Choose the conditional variance model; ARCH, GARCH, GJR-GARCH or EGARCH. Then choose if the innovations should have a Student's t-distribution or normal distribution.
 - If filtration: choose EVT or HS.

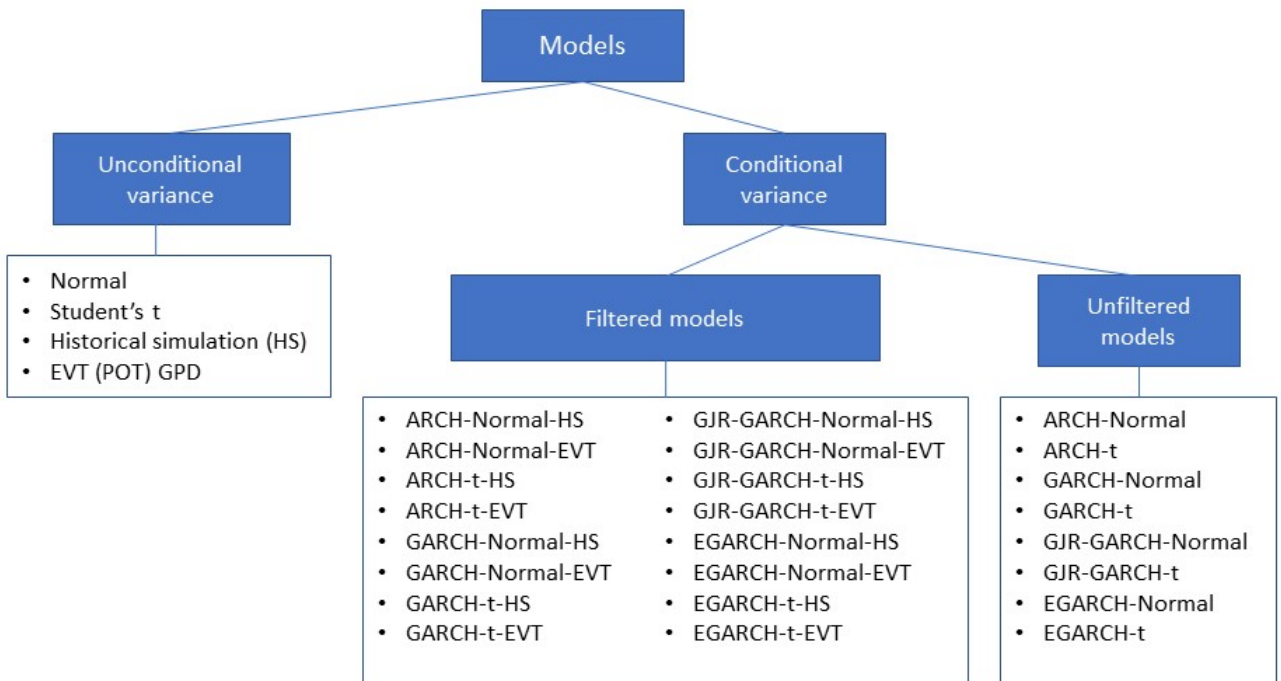


Figure 1: Map of the models for time series 1.

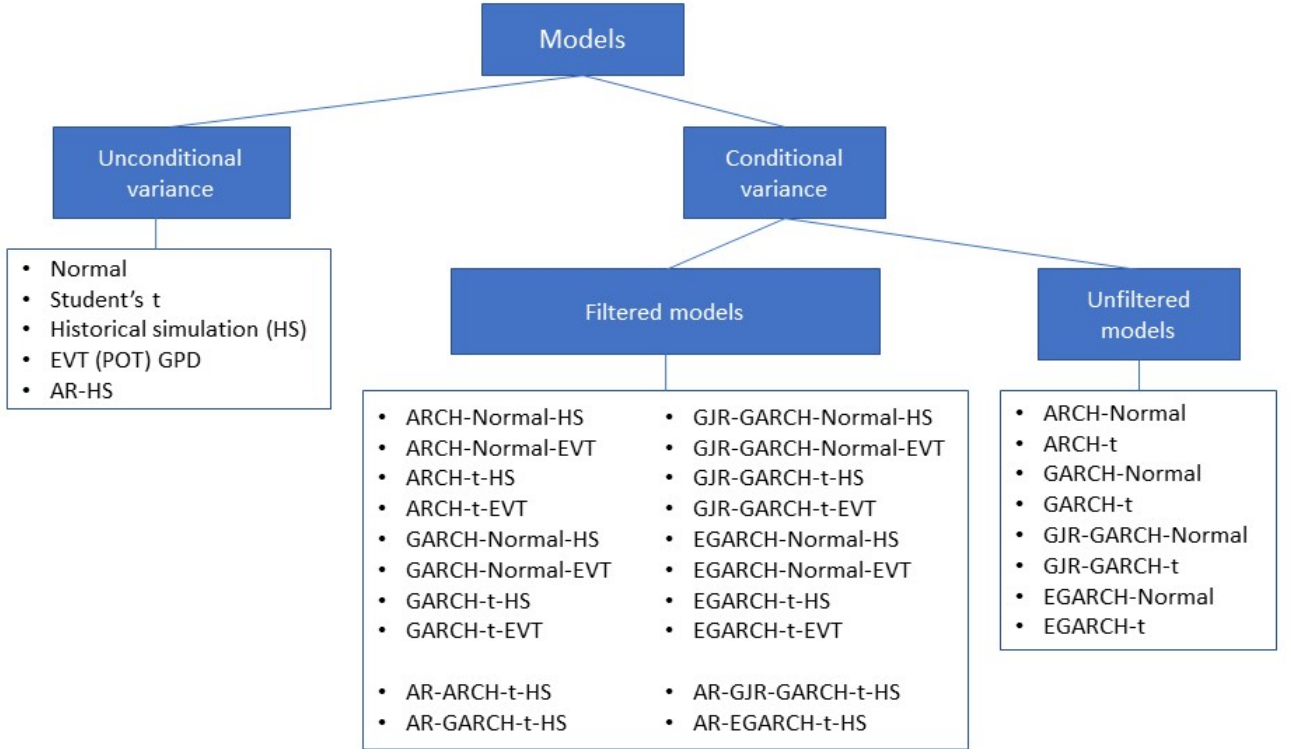


Figure 2: Map of the models for time series 2.

4.2.2 Model descriptions

In figure 1 and 2, the models are divided into three groups and the structure of the models is described below:

- **Unconditional variance:** Let y_t be the response, which in this thesis is the asset returns. If the mean is assumed to be unconditional, then:

$$y_t = \mu + \epsilon_t,$$

where ϵ_t is assumed to follow a Normal, Student's t or Empirical distribution. It can also be assumed to follow a Generalized Pareto distribution in the tails and an interpolated empirical distribution between the tails. If the mean is assumed to be conditional instead (an AR(1)-process in this case), then:

$$y_t = c + \theta y_{t-1} + \epsilon_t,$$

where ϵ_t can have the same types of distributions as above. The parameters of the models are estimated from the historical data in the parameter estimation window of 1000 days using maximum likelihood.

- **Conditional variance - Unfiltered models:** Depending on the choice of mean process, the models follow:

$$y_t = \mu + \epsilon_t, \text{ OR}$$

$$y_t = c + \theta y_{t-1} + \epsilon_t,$$

where $\epsilon_t = \sigma_t Z_t$, and Z_t is either $N(0, 1)$ or $t(v)$ distributed. The σ_t :s can in this study follow four different types of conditional variance models: ARCH(20), GARCH(1,1), GJR-GARCH(1,1), EGARCH(1,1) which corresponds to the formulas below,

- ARCH(20): $\sigma_t^2 = c + \sum_{i=1}^{20} \alpha_i \epsilon_{t-i}^2$,
- GARCH(1,1): $\sigma_t^2 = c + \alpha \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2$,
- GJR-GARCH(1,1): $\sigma_t^2 = c + \alpha \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2 + I_{\{\epsilon_{t-1} < 0\}} \gamma \epsilon_{t-1}^2$,
- EGARCH(1,1): $\log(\sigma_t^2) = c + \alpha \log(\sigma_{t-1}^2) + \beta \left(\frac{|\epsilon_{t-1}|}{\sigma_{t-1}} - E \left[\frac{|\epsilon_{t-1}|}{\sigma_{t-1}} \right] \right) + \xi \left(\frac{\epsilon_{t-1}}{\sigma_{t-1}} \right)$.

The parameters of the models are estimated from the historical data in the parameter estimation window of 1000 days using maximum likelihood.

- **Conditional variance - Filtered models:** The filtered models have the same structure as for the Unfiltered models but with one difference. The difference is that when the conditional variance models have been estimated, they are used to infer the historical conditional volatilities in the parameter estimation window, yielding $\{\sigma_{-t} \mid 1 \leq t \leq 1000\}$. These historical volatilities are then used to create the standardized residuals $\{\frac{\epsilon_{-t}}{\sigma_{-t}} \mid 1 \leq t \leq 1000\}$ which are then used to estimate the EVT or the empirical distribution. These distributions are then scaled with the volatility forecast. As previously mentioned, the parameters of the models are estimated from the historical data in the parameter estimation window of 1000 days using maximum likelihood.

4.3 Estimation of the risk measures

The risk measures can now be estimated using the different fitted distributions for the asset return. This is done using the Monte Carlo method and calculating the empirical VaR/ES using the generated samples from the different distributions. Let F_i be the cdf of model i . Draw n samples, $\{y_i \mid 1 \leq i \leq n\}$, from F_i . Calculate the estimates of VaR and ES from their empirical formulas:

$$\widehat{VaR}_p(X) = y_{[np]+1,n}^*$$

$$\widehat{ES}_p(X) = \frac{1}{p} \left(\sum_{k=1}^{[np]} \frac{y_{k,n}^*}{n} + \left(p - \frac{[np]}{n} \right) y_{[np]+1,n}^* \right),$$

where $\{y_i^* \mid 1 \leq i \leq n\}$ is the ordered sample and p is the chosen probability level.

4.4 Clustering analysis and the risk ratio

The market risk models are clustered using agglomerative hierarchical clustering with complete linkage as the dissimilarity measure with Euclidean distance. The data that is used in the clustering is all the daily estimates for $VaR_{0.01}$, $ES_{0.01}$ and $ES_{0.025}$. The models are clustered separately for each risk measure and time series.

The clustering analysis is visualized in a dendrogram that consist of $m - 1$ splits, where m is the number of models. Starting from the top of the dendrogram, all models are in the same cluster. Moving down the tree the clusters reduce in size for each split/level. Several of these cluster levels are analyzed further by calculating the maximum disagreement between the models' risk estimates for each day in the time series. This is done using the risk ratio, RR, which is defined below. Let M_i be the set of models in the cluster at level i and let $m_i^j \in M_i$ be model j at the

same cluster level. Also, let $\rho(m_i^j, k)$ be the estimated risk (VaR or ES) for model m_i^j at day k . The risk ratio at level i on day k is defined as:

$$RR_i^k = \frac{\max_{m_i^j \in M_i} \rho(m_i^j, k)}{\min_{m_i^j \in M_i} \rho(m_i^j, k)},$$

where the nominator is the worst case estimate and the denominator is the best case estimate. The complete linkage as dissimilarity measure for the hierarchical clustering is chosen since it tends to produce clusters that are compact and where the maximum dissimilarity in the cluster is low (Hastie et al., 2016). Thus, it should be a suitable choice in order to produce clusters with low risk ratio.

By following this method, more and more of the disagreement between the models is explained by the splits in the dendrogram. These splits can be viewed as a choice or assumption in the model construction. Each of the levels in the dendrogram has an empirical distribution consisting of all the calculated risk ratios for the time series/risk measure in focus. The mean and standard deviation of these distributions are further analyzed and the decrease in these moments is summarized, thus showing how much of the model disagreement that is removed by excluding a cluster of models (e.g. making a choice/assumption in the model construction).

4.5 Comparing VaR_{0.01} and ES_{0.025}

For both time series used in this thesis, the effects of changing from VaR_{0.01} to ES_{0.025} as a result of new regulations are analyzed. This is done through using the hierarchical clustering dendrograms stemming from using VaR_{0.01} as a risk measure, and comparing the risk ratios that occur when one estimates VaR_{0.01} and ES_{0.025}. This change in risk estimation is expressed both in absolute terms (how both VaR_{0.01} and ES_{0.025} are estimated for the specific cluster) and also in relative terms, i.e. how much the risk ratio has changed (in percentage) from VaR_{0.01} to ES_{0.025}.

4.6 Data analysis

The data used in this study came from the Nasdaq Nordic and Yahoo Finance. This data consist of historical prices for the index OMXS30 from 1993-04-02 to 2017-03-20. The bond index used is Bloomberg Barclays US Aggregate Bond (AGG) with time series from 2003-09-30 to 2017-04-24.

The following part consists of a detailed description and analysis of the stock, bond and currency data. For each data series the following is presented and analyzed:

- Plot of log-returns
- Histogram of log-returns
- QQ-plot where empirical data is compared to the quantiles of a fitted normal distribution and a fitted Student's t-distribution
- Plots of sample autocorrelation function (ACF) and partial autocorrelation function (PACF)
- Plots of ACF and PACF for squared returns

Given the breakdown of the data according to the above plots, a brief description of the implications is presented.

4.6.1 Stock return data (OMXS30)

The following is an analysis of the OMXS30 data used in this thesis. We show the returns, a histogram of returns, the autocorrelation function (ACF) and the partial autocorrelation function (PACF) for returns as well as squared returns. Furthermore we present a table with data characteristics such as mean, standard deviation, skewness and kurtosis.

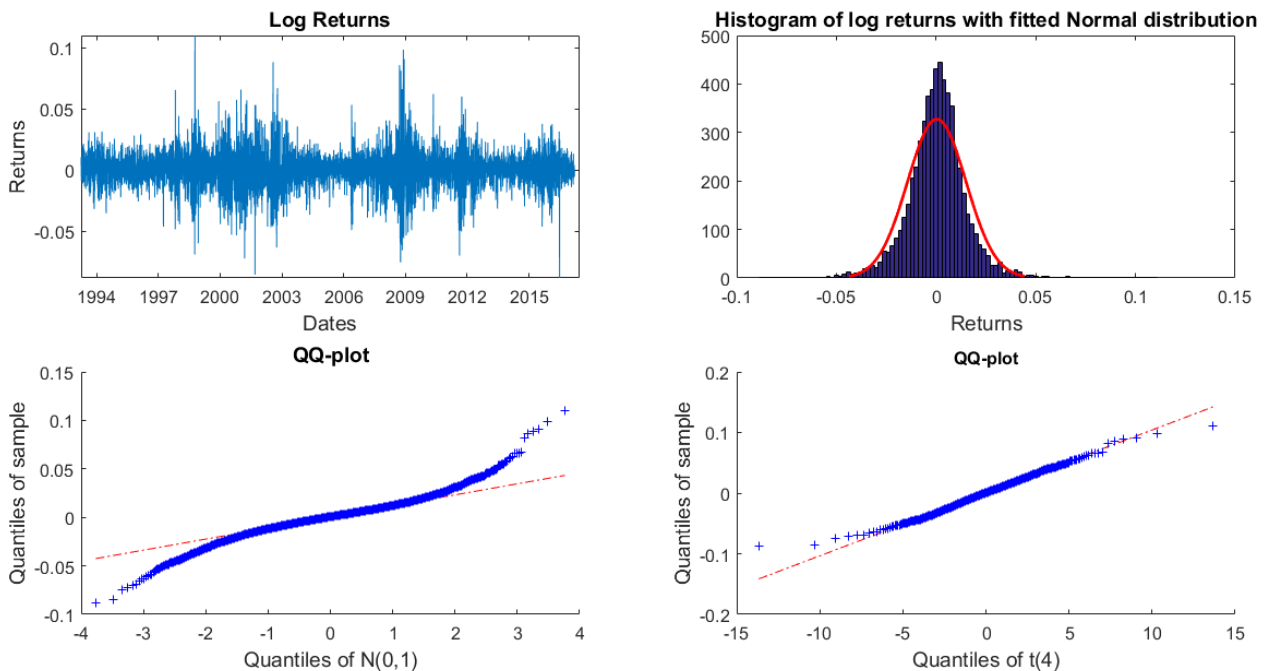


Figure 3: OMXS30. Graph of returns, histogram with a fitted Normal distribution and QQ-plots against a standard normal distribution and a standard student's t -distribution with estimated degrees of freedom.

Data characteristics for OMXS30	
Number of days	6019
Mean	0.0003500
Std dev (daily)	0.01467
Skewness	0.02991
Kurtosis	6.876
Excess tails (1%)	1.877 %

Table 1: OMXS30 data table. Excess tails is the percent of actual daily returns falling in the lowest (1%) percentile of a fitted normal distribution.

From the top left graph of daily log returns we can quite clearly see that the time series does not just consist of noise. There are clear signs of volatility clustering which occur during crisis periods. Prime examples are during the IT bubble of late 1990s and early 2000s as well as during the financial crisis in the period of 2008-2010.

Judging by the look of the histogram of log returns and the data characteristics table we see clear signs of excess kurtosis in the empirical distribution. Furthermore the empirical tails are considerably heavier than those of a fitted normal distribution. In the lowest 1-percentile we

find approximately 1.9% of actual historical returns, indicating that a normal distribution gives a poor representation of the tails of the actual distribution, even if it is fitted using MLE.

The QQ-plots show that the standard normal distribution fails to take the heavier tails of the empirical distribution into account, whereas the fitted student's t -distribution accomplishes this better. For the t -distribution however, the QQ-plot shows that the tails are over-estimated compared to the empirical distribution.

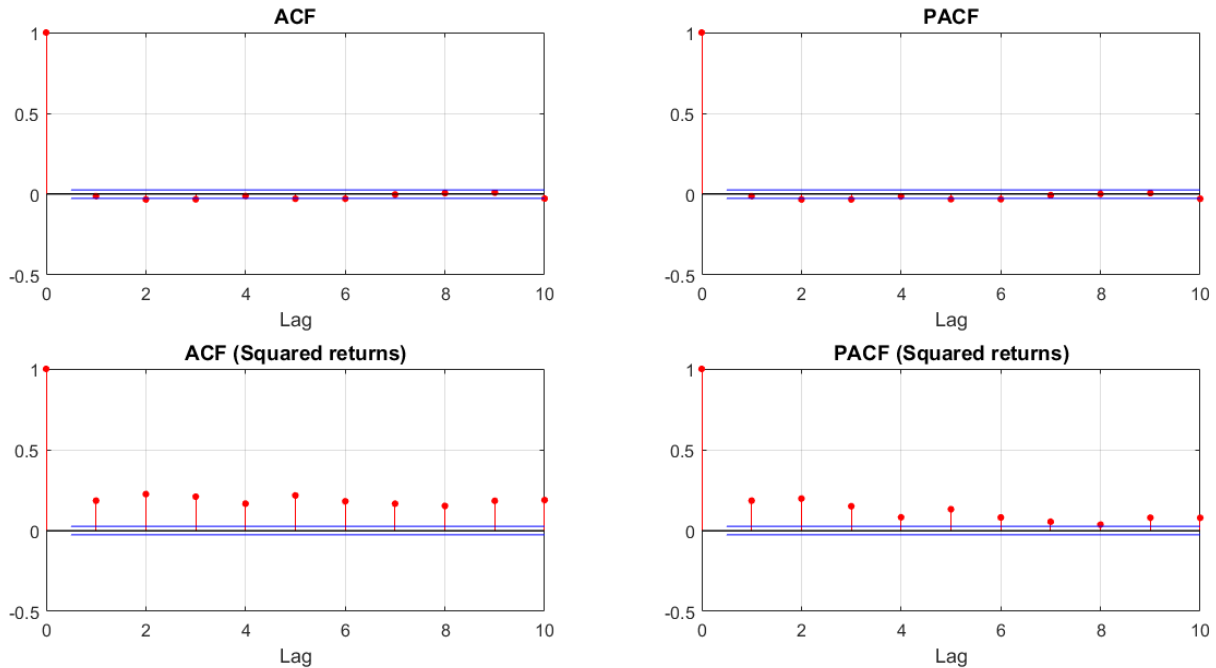


Figure 4: OMXS30. Autocorrelation function (ACF) and partial autocorrelation function (PACF) for the sample of daily log returns, both normal and squared.

From the top graphs, the ACF and PACF of daily log returns, there is no clear signs for any significant lags, suggesting that there is neither an $AR(p)$ or $MA(q)$ process present in the data. However, this does not mean that there is no autocorrelation present in the sample, because if we inspect the bottom two graphs, the ACF and PACF for the squared returns, there is evidently strong signs of autocorrelation. What can be said from these graphs is that, clearly, previous daily returns have an impact on the size of the following daily returns. This gives further merit to the occurrence of volatility clusters that are evident in the top left graph of figure 3, suggesting that the distribution of the data may benefit from being modeled with a conditional variance model.

4.6.2 Interest rate return data (Bloomberg Barclays US Aggregate Bond Index)

The following is an analysis of the Bloomberg Barclays US Aggregate Bond Index (AGG) data used in this thesis. We show the returns, a histogram of returns, the autocorrelation function (ACF) and the partial autocorrelation function (PACF) for returns as well as squared returns. Furthermore we present a table with data characteristics such as mean, standard deviation, skewness and kurtosis.

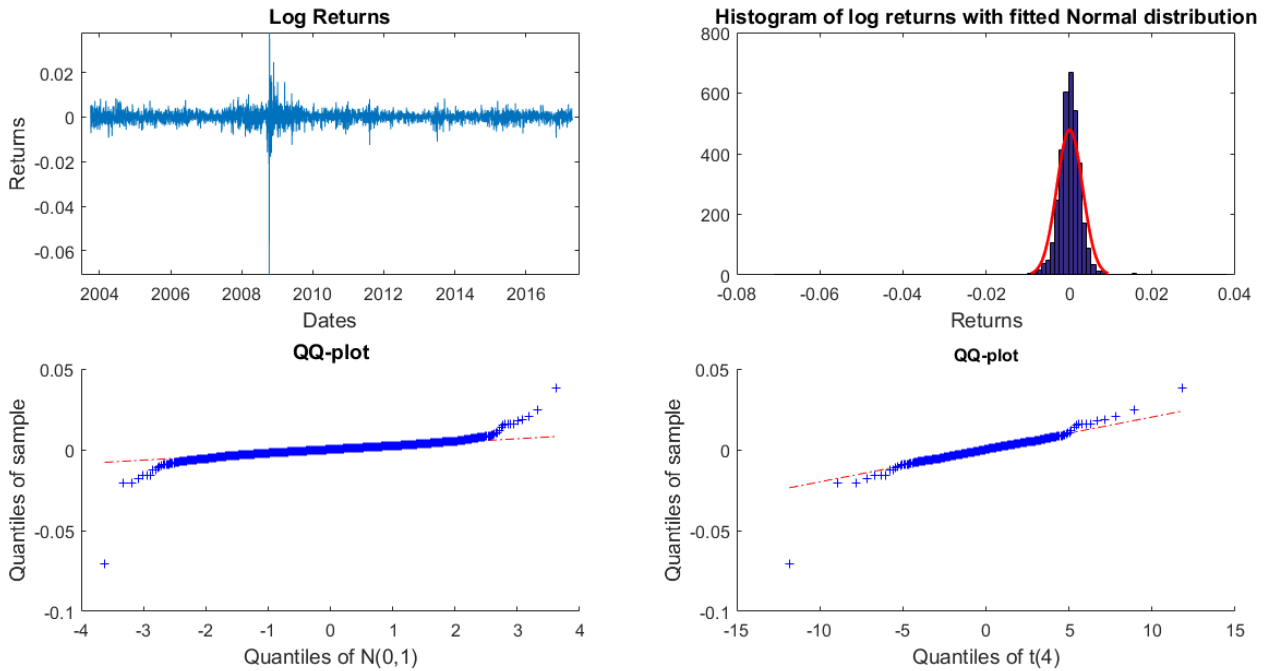


Figure 5: AGG. Graph of returns, histogram with a fitted Normal distribution and QQ-plots against a standard normal distribution and a standard student's t -distribution with estimated degrees of freedom.

Data characteristics for AGG	
Number of days	3415
Mean	0.0001557
Std dev (daily)	0.003010
Skewness	-2.882
Kurtosis	94.76
Excess tails (1%)	0.9078 %

Table 2: AGG data table. Excess tails is the percent of actual daily returns falling in the lowest (1%) percentile of a fitted normal distribution.

From the top left graph of daily log returns it is evident that the data series has rather constant, low, volatility and does not experience any particularly larger volatility clusters. The one exception would be during late 2008.

Approximately 0.9% of actual historical returns can be found in the lowest 1-percentile, indicating that a fitted normal distribution gives a rather good representation of the tail's of the actual distribution. However, the skewness apparent in the distribution and the very high kurtosis leads to the argument that a normal distribution is not a very good fit.

The QQ-plots show that the empirical distribution fits quite well to a standard normal distribution but that the standard normal distribution fails to take the slightly heavier tails of the empirical distribution into account. The fitted student's t-distribution accomplishes this better and shows an overall good fit.

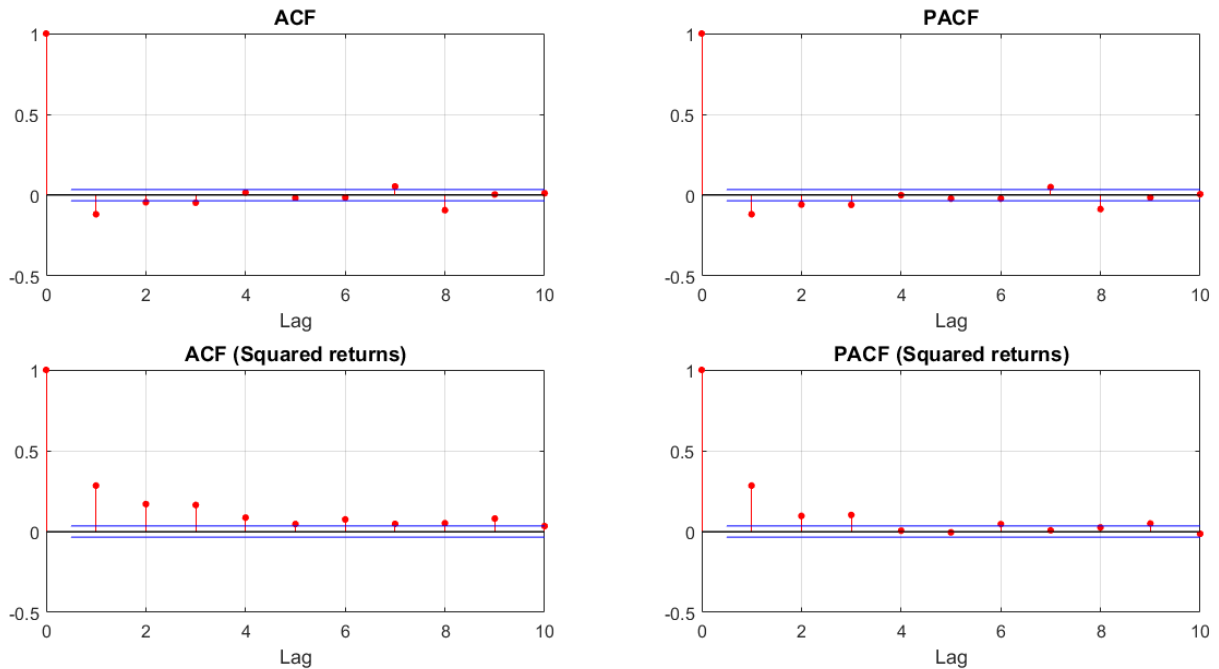


Figure 6: AGG. Autocorrelation function (ACF) and partial autocorrelation function (PACF) for the sample of daily log returns, both normal and squared.

From the top graphs, the ACF and PACF of daily log returns, there are slight signs of $AR(p)$ and $MA(q)$ processes being present in the data. However, there is no striking signals that the process should be modeled as in a certain way. If we inspect the bottom two graphs however, the ACF and PACF for the squared returns, there are signs of autocorrelation in the in the squared sample. What can be said from these graphs is that the size of previous daily returns have an impact on the size of the following daily returns. This slight volatility cluster is not as evident in the top left graph of Figure 5 as it is in the stock return series. However, this still suggests that the distribution of the data may benefit from being modeled using a conditional variance model.

5 Results

This section contains numerous dendrograms/graphs/tables presenting the breakdown and relations of the market risk models and their estimates of both VaR and ES at α -level 0.01, and 0.025 for ES. The purpose of the graphs is to present similarities and dissimilarities between models to determine what choices/assumptions causes the largest disagreement between models.

The result section is divided into two subsections, the first for stock index return data and the second for bond index return data. These subsections are structured in the following way:

1. Results for $\text{VaR}_{0.01}$.
2. Results for $\text{ES}_{0.01}$.
3. Results for $\text{ES}_{0.025}$.
4. Complementary results for the comparison between $\text{VaR}_{0.01}$ and $\text{ES}_{0.025}$.

Furthermore, the results for nr. 1-3 are each structured in the following way:

- a. A dendrogram from the hierarchical clustering containing all the market risk models.
- b. Graphs with max/min VaR_α or ES_α -estimates and the risk ratio for the entire set of models.
- c. Table with all the chosen cluster levels and their key figures for the risk ratio.

The result for nr. 4 consists of graphs and a table showing how the risk ratio for $\text{ES}_{0.025}$ would look compared to $\text{VaR}_{0.01}$ and using the clusters/levels from $\text{VaR}_{0.01}$.

Please note that VaR and ES are calculated from the return distribution instead of the loss distribution, meaning that VaR and ES are negative in the graphs.

5.1 Stock return modeling (OMXS30)

In the following section, the graphs and tables are explained in more detail than in the following sections where the structure and types of graphs are repeated.

5.1.1 $\text{VaR}_{0.01}$

In figure 7, the dendrogram for the hierarchical clustering of the market risk models is presented for $\text{VaR}_{0.01}$. Thirteen clusters at different levels are further analyzed. The first cluster (at level 1) contains all models. The models are thereafter divided into smaller and smaller clusters. By inspecting which models fall into the same cluster, one can identify similarities in the construction of the models. These similarities are, if possible, commented in the dendrogram. The dendrogram thus provides a map from which the similarities and the dissimilarities between market risk models can be visualised. The arrows in the dendrogram are added in order to more clearly illustrate that the cluster is further divided into sub-clusters.

At level 1, the models are split into clusters 2 and 3. All models in cluster 3 have unconditional variance and all models in cluster 2 have a conditional variance. Cluster 3 is not further divided/explored, but one can see that the normal-model has a relatively large dissimilarity with the other models of that cluster.

Cluster 2 is further divided into clusters 4 and 5. All models in cluster 4 have a leverage component in their conditional variance model since it only contains models with either a GJR-GARCH or EGARCH model. In cluster 4, there are only models with ARCH and GARCH as the conditional variance model, and the lack of a leverage component is thus one of the similarities in this cluster.

Clusters 4 and 5 are split further into clusters 6 & 7 and 8 & 9, respectively. One of the differences between cluster 6 and 7 is that cluster 6 only contains models with variance filtering while cluster 7 contains only unfiltered models. The same branching occurs for cluster 5, with cluster 8 only containing unfiltered models and cluster 9 only containing filtered models.

Clusters 7 and 8 are not further divided. However, one can see that they could be further branched first by the volatility specification (EGARCH or GJR-GARCH, ARCH or GARCH). Lastly, they are split depending on the distribution (Student's t or Normal).

In the clusters with the filtered models, i.e. clusters 6 and 9, the models are further divided into the clusters 10 & 11 and 12 & 13, respectively. The similarities in these clusters are that they all have the same type of variance model (EGARCH or GJR-GARCH, ARCH or GARCH) but with different distributions or filtration distributions¹.

Clusters 10, 11, 12 and 13 differ in which order the final splits are made. Two of them, 11 and 13, first split depending on the choice of HS or EVT. This is compared with clusters 10 and 12 which are further divided by the choice of filtration distribution, Normal or Student's t. These are, however, not further explored.

¹The "filtration distribution" is the distribution assumed in the variance model of the filtered models, thus Normal or Student's t.

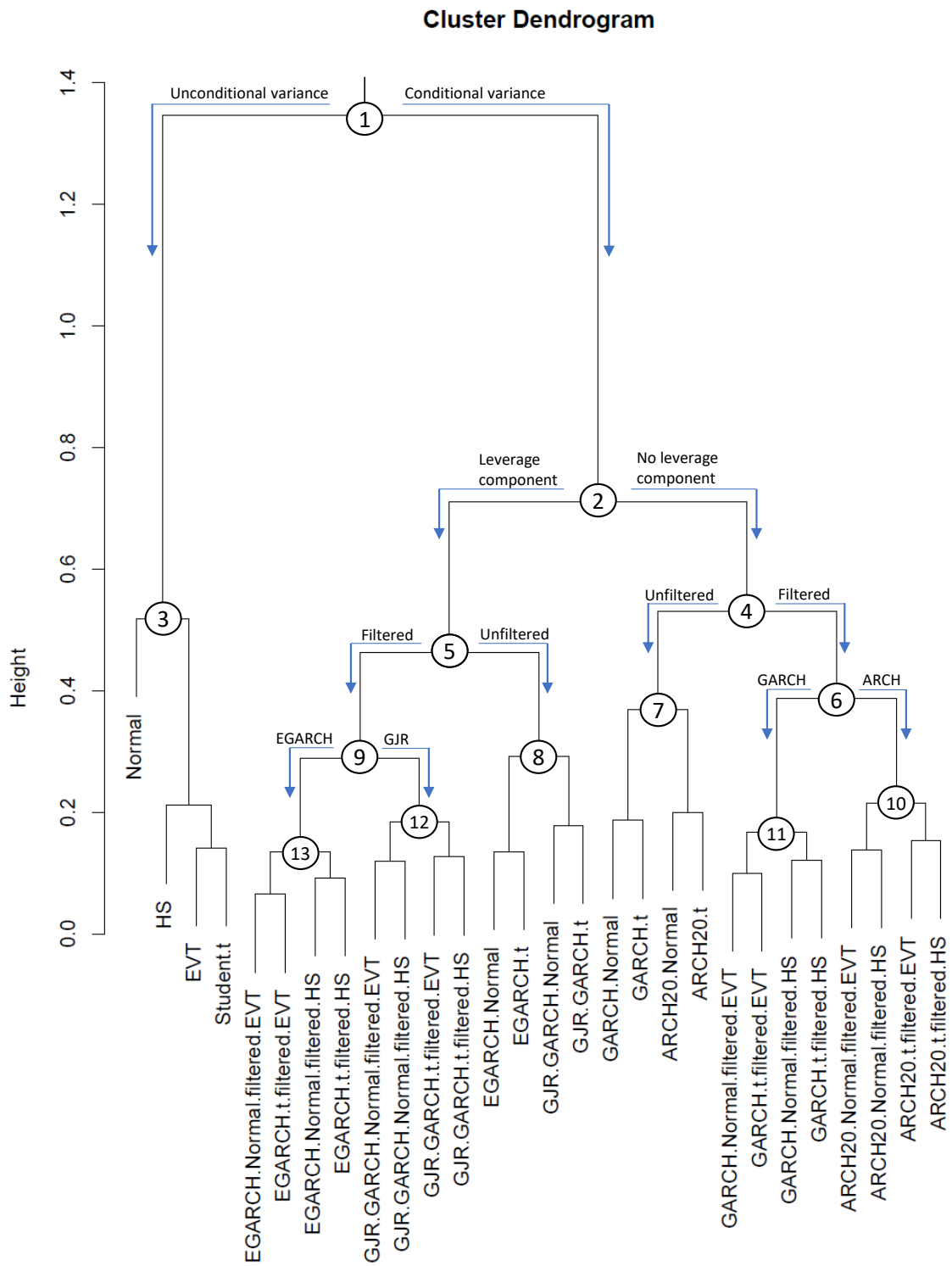


Figure 7: Clustering of models for $\text{VaR}_{0.01}$.

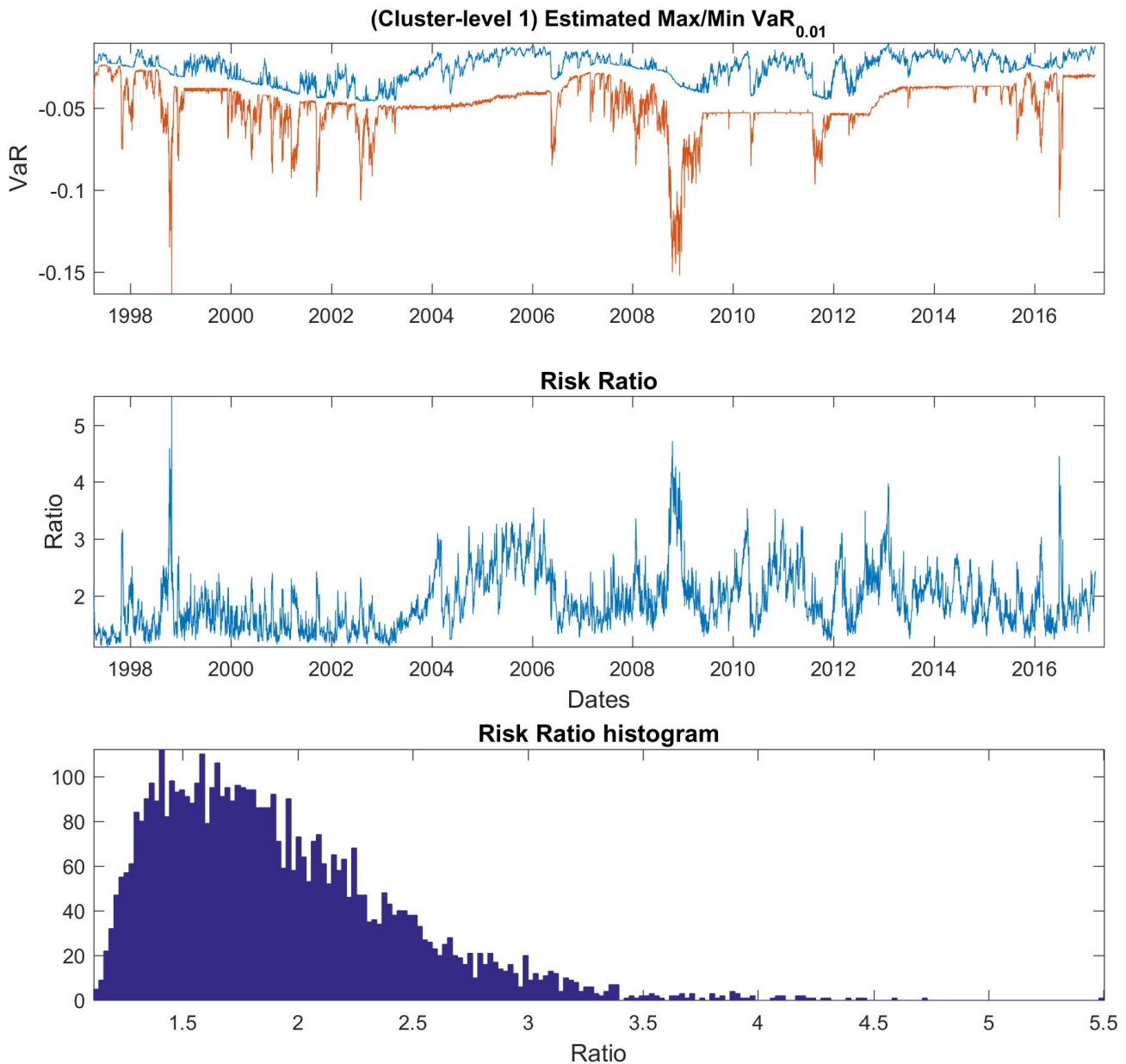


Figure 8: Max and Min VaR-estimates for cluster level 1.

In figure 8, the maximum dissimilarity between all models (cluster level 1) is presented in three different ways. In the first graph, the maximum and minimum $VaR_{0.01}$ estimates for each day in the time series are plotted. Note that the model which has the highest/lowest estimate changes numerous times.

In the second graph, the risk ratio time series for cluster 1 is shown, RR_1 . It is the quotient between the maximum and minimum estimates displayed in the first graph. In the third graph, RR_1 is summarized in a histogram which gives an idea of how the risk ratio is distributed. The mean of RR_1 is 1.92 and the standard deviation is 0.52, which is shown in the table 3 below together with all the sub-clusters. The lowest possible risk ratio is 1 and means that all models have estimated the risk the same.

In table 3, four key figures and their decrease in percentage is summarized for each cluster level. Note that the minimum value of the mean, 10-percentile and 90-percentile is 1 and that the decrease in percentage is adjusted for this fact.

The same graphs as 8 but for the other cluster levels all have the same characteristics but with a reduction of the key figures of RR_i . One exception is, however, for cluster level 3 which contains

only unconditional variance models. In this cluster, the risk ratio exhibits a much stronger autocorrelation compared with the other clusters.

The characteristics of the other clusters' risk ratios are, as mentioned, very similar. Also, there seems to be little difference in the way the different key figures decrease (e.g. a decrease in the mean with 85% roughly correspond to a decrease for the other key figures with 85%), but with the exception of the 10-percentile which has smaller percentage decrease in the early splits of the dendrogram. The decrease in percentage is declining, where the maximal decrease in the risk ratio achieved in the first split. Furthermore, all the clusters seem to be affected by financial turmoil which can be seen in the risk ratio graphs (the middle graphs of the figures) around, for example, 2008.

The mean at the different chosen cluster levels ranges from 1.92 to 1.06 where the lowest value corresponds to cluster 11 which contains four different GARCH-filtered models. The standard deviation ranges from 0.524 to 0.0346, where the lowest value was achieved in cluster 13 which consists of four EGARCH-filtered models.

Key figures of RR for VaR _{0.01}					
Level	Prev. level	Mean	Std. dev.	10-percentile	90-percentile
1		1.9169	0.5236	1.3458	2.6188
2	1	1.3851 (−58.0%)	0.1967 (−62.4%)	1.1845 (−46.7%)	1.6386 (−60.6%)
3	1	1.2142 (−76.6%)	0.0874 (−83.3%)	1.0892 (−74.2%)	1.3168 (−80.4%)
4	2	1.2686 (−70.7%)	0.1423 (−72.8%)	1.1201 (−65.3%)	1.4509 (−72.1%)
5	2	1.2167 (−76.4%)	0.1184 (−77.4%)	1.0899 (−74.0%)	1.3582 (−77.9%)
6	4	1.1722 (−81.2%)	0.1044 (−80.1%)	1.0706 (−79.6%)	1.3007 (−81.4%)
7	4	1.1655 (−82.0%)	0.1042 (−80.1%)	1.0631 (−81.7%)	1.2960 (−81.7%)
8	5	1.1159 (−87.4%)	0.0819 (−84.4%)	1.0378 (−89.1%)	1.2177 (−86.5%)
9	5	1.1396 (−84.8%)	0.0870 (−83.4%)	1.0594 (−82.8%)	1.2378 (−85.3%)
10	6	1.0697 (−92.4%)	0.0521 (−90.0%)	1.0231 (−93.3%)	1.1284 (−92.1%)
11	6	1.0600 (−93.5%)	0.0418 (−92.0%)	1.0178 (−94.9%)	1.1099 (−93.2%)
12	9	1.0718 (−92.2%)	0.0627 (−88.0%)	1.0224 (−93.5%)	1.1293 (−92.0%)
13	9	1.0615 (−93.3%)	0.0346 (−93.4%)	1.0237 (−93.2%)	1.1008 (−93.8%)

Table 3: Figures for the risk ratio distribution for VaR_{0.01}. Decrease from level 1 in percentage in the parentheses. Note that the minimum value of the mean, 10-percentile and 90-percentile is 1 and that the decrease in percentage is adjusted for this fact.

In the dendrogram, figure 7, the vertical distance (named height in the figure) roughly corresponds to how much the risk ratio decreases between clusters. For example, when comparing the dendrogram with table 3 it can be seen that the risk ratio has decreased more in cluster 5 than in cluster 4, and the same holds true when comparing cluster 10 with cluster 11.

All in all, the dendrogram in figure 7 illustrates a hierarchy of importance amongst different assumptions, and table 3 shows just how much of an effect these assumptions can have on the spread in risk ratio.

5.1.2 $ES_{0.01}$

In figure 9, the dendrogram for the hierarchical clustering of the market risk models is presented for $ES_{0.01}$. Eleven clusters at different levels are further analyzed. The first cluster (at level 1) contains all models.

At level 1, the models are split into the clusters 2 and 3. All models in cluster 2 have unconditional variance and all models in cluster 3 have a conditional variance model. Cluster 2 is not further divided/explored, but one can see that the Normal-model has a relatively large dissimilarity with the other models of that cluster. These splits are very similar to the ones of $VaR_{0.01}$.

Cluster 3 is further divided into clusters 4 and 5. The similarities in the clusters are not as clear as for $VaR_{0.01}$. Anyhow, one can see that all EGARCH models are placed in cluster 4 together with the unfiltered conditional variance models with normal distributions.

Cluster 4 is split further into clusters 6 & 7. One of the differences between cluster 6 and 7 is that cluster 6 only contains unfiltered conditional variance models with normal distributions. Cluster 6 and 7 are not further explored, however, one can see that cluster 6 is further divided where GARCH and ARCH are placed in the same cluster. In the next split of cluster 7, the unfiltered models are separated from the filtered.

Cluster 5 is split into clusters 8 & 9, where cluster 9 only contains ARCH models. Cluster 8 contains either GJR-GARCH or GARCH models. This differs from $VaR_{0.01}$ where GARCH and ARCH were separated from EGARCH and GJR-GARCH. Cluster 9 is not further explored, but one can however see that it is divided by filtered/unfiltered and then which filtration distribution that is used.

Cluster 8 is split into clusters 10 & 11, where cluster 10 only contains GARCH models, and cluster 11 only contains GJR-GARCH models. Cluster 10 is then divided by filtered/unfiltered followed by which filtration distribution that is used. Cluster 11 only contains filtered models and is further divided by the choice of filtration distribution.

Cluster Dendrogram

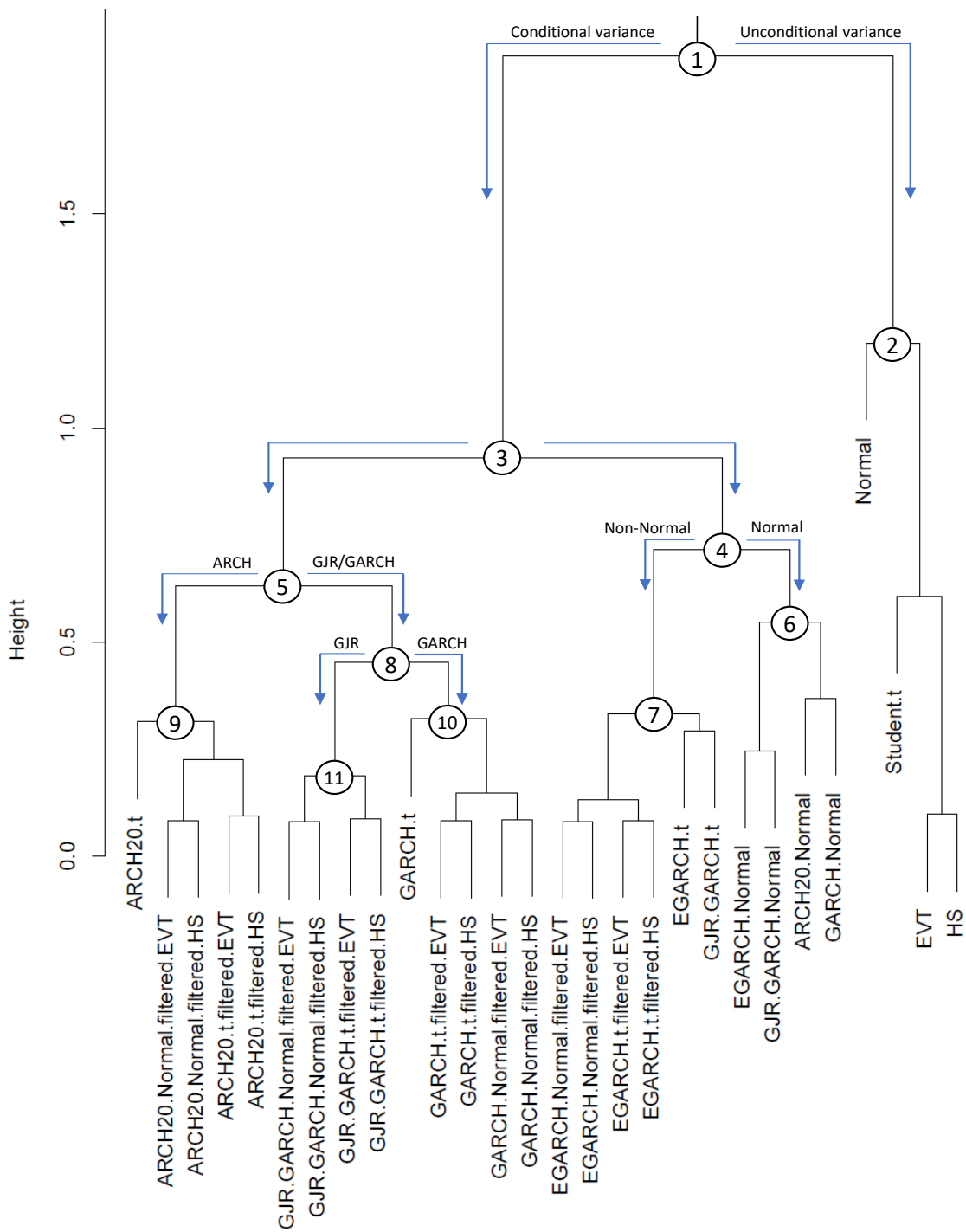


Figure 9: Clustering of models for $ES_{0.01}$.

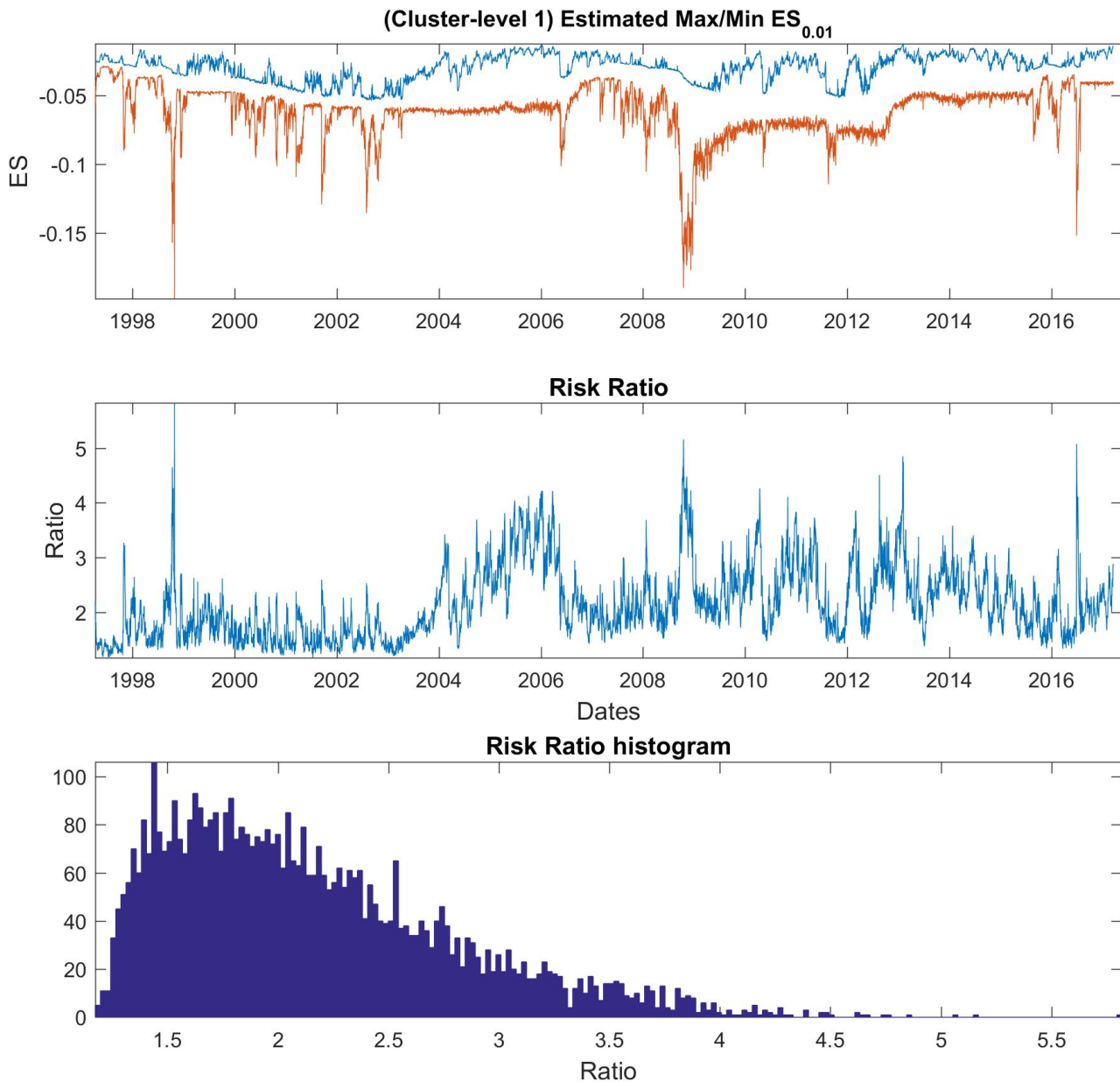


Figure 10: Max and Min ES-estimates for cluster level 1.

In figure 10, the maximum dissimilarity between all models (cluster level 1) is, as in the previous section, presented in three different ways. The key figures of the risk ratio at different cluster levels are summarized in table 4.

All the key ratios at cluster level 1 are larger than the ones for $VaR_{0.01}$. The mean at the different chosen cluster levels ranges from 2.15 to 1.05. The standard deviation ranges from 0.65 to 0.055.

Key figures of RR for $ES_{0.01}$					
Level	Prev. level	Mean	Std. dev.	10-percentile	90-percentile
1		2.1514	0.6503	1.4304	3.0766
2	1	1.3953 (−65.7%)	0.1713 (−73.7%)	1.1705 (−60.4%)	1.6035 (−70.9%)
3	1	1.4475 (−61.1%)	0.2151 (−66.9%)	1.2261 (−47.5%)	1.7306 (−64.8%)
4	3	1.3343 (−71.0%)	0.1792 (−72.4%)	1.1458 (−66.1%)	1.5739 (−72.4%)
5	3	1.2627 (−77.2%)	0.1371 (−78.9%)	1.1273 (−70.4%)	1.4369 (−79.0%)
6	4	1.2027 (−82.4%)	0.1325 (−79.6%)	1.0715 (−83.4%)	1.3699 (−82.2%)
7	4	1.1417 (−87.7%)	0.0832 (−87.2%)	1.0558 (−87.0%)	1.2478 (−88.1%)
8	5	1.1754 (−84.8%)	0.0958 (−85.3%)	1.0765 (−82.2%)	1.2989 (−85.6%)
9	5	1.0924 (−92.0%)	0.0592 (−90.9%)	1.0368 (−91.4%)	1.1621 (−92.2%)
10	8	1.0975 (−91.5%)	0.0568 (−91.3%)	1.0347 (−91.9%)	1.1696 (−91.8%)
11	8	1.0499 (−95.7%)	0.0547 (−91.6%)	1.0152 (−96.5%)	1.0870 (−95.8%)

Table 4: Figures for the risk ratio distribution for $ES_{0.01}$. Decrease from level 1 in percentage in the parentheses. Note that the minimum value of the mean, 10-percentile and 90-percentile is 1 and that the decrease in percentage is adjusted for this fact.

5.1.3 $ES_{0.025}$

In figure 11, the dendrogram for the hierarchical clustering of the market risk models is presented for $ES_{0.025}$. Thirteen clusters at different levels are further analyzed. The first cluster (at level 1) contains all models.

The dendrogram for $ES_{0.025}$ is very similar to the one of $VaR_{0.01}$. So, the reader is referred to section 5.1.1 for descriptions of the similarities in the clusters at the different levels. However, there is one difference in the lower level splits for the filtered models. For $ES_{0.025}$, the filtered models are always separated depending on the filtration distribution, see for example at level 13 where $VaR_{0.01}$ is divided by HS or EVT and in $ES_{0.025}$ it is instead divided by the choice of Normal or Student's t as a filtration distribution.

Cluster Dendrogram

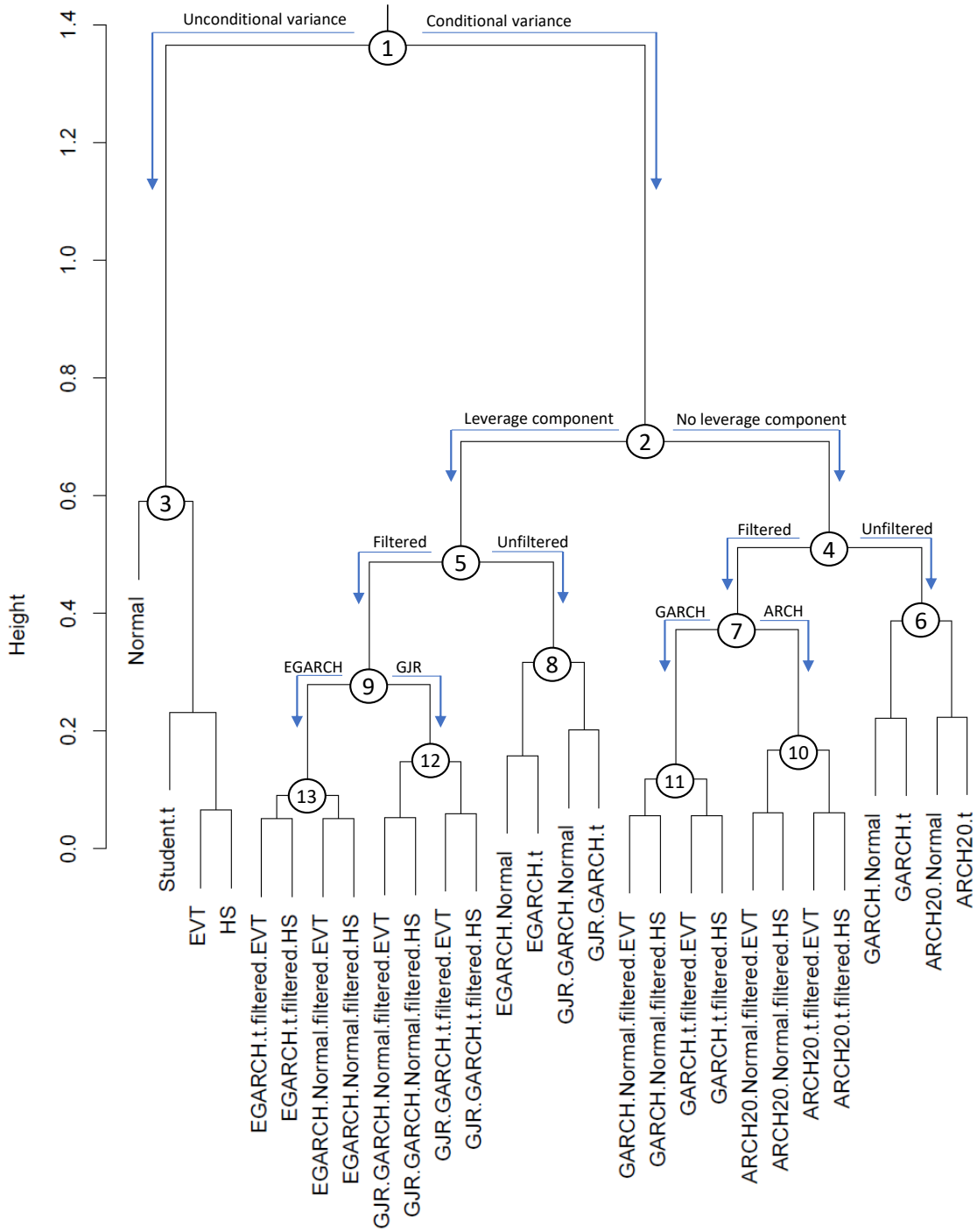


Figure 11: Clustering of models for $ES_{0.025}$.

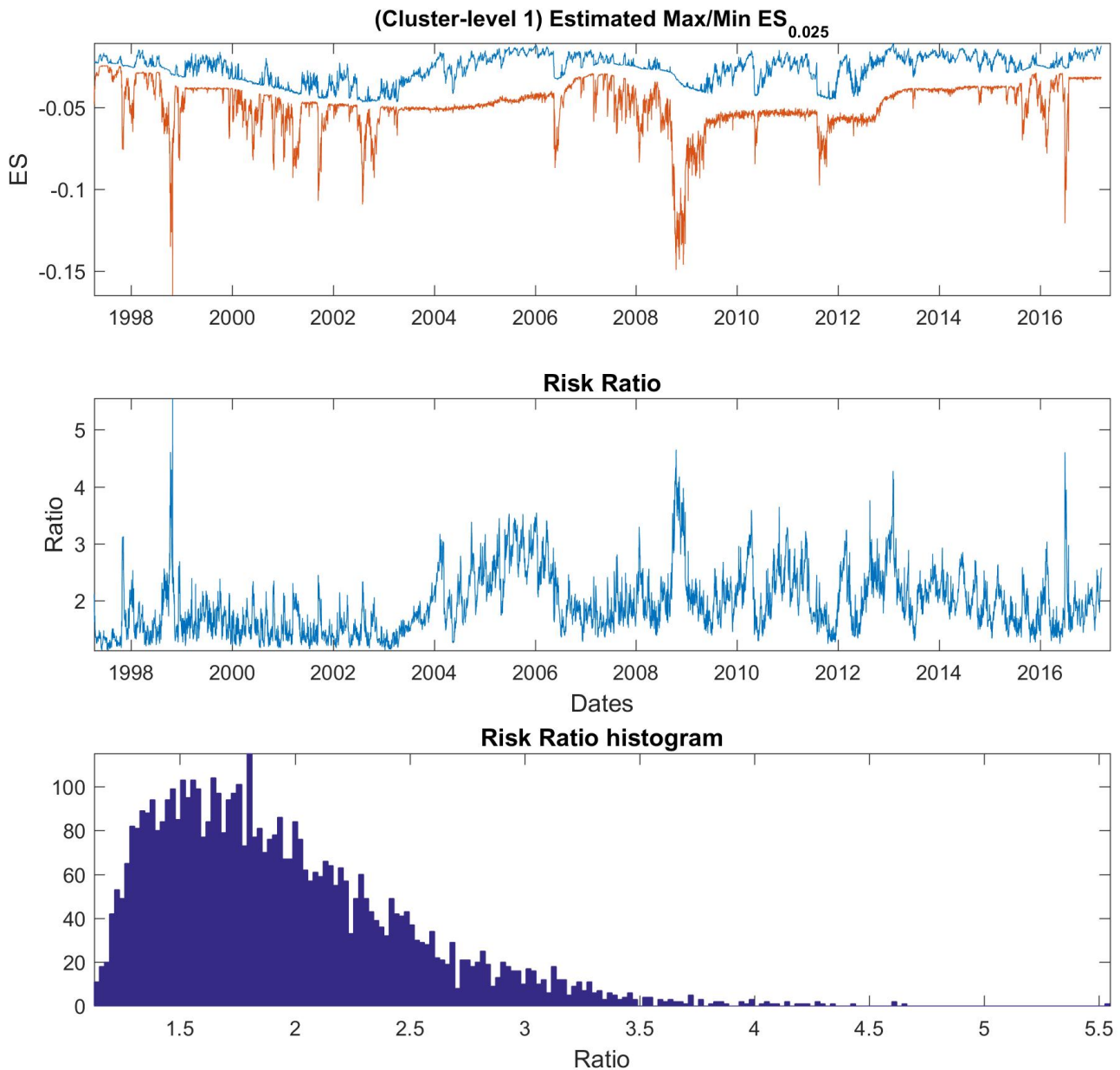


Figure 12: Max and Min ES-estimates for cluster level 1.

In figure 12, the maximum dissimilarity between all models (cluster level 1) is, as in the previous sections, presented in three different ways. The key figures of the risk ratio at different cluster levels are summarized in table 5.

The key figures of RR seem to decrease at the same rate, but with the exception of the 10-percentile which has smaller percentage decrease in the early splits of the dendrogram.

All the key ratios at cluster level 1 are larger than the ones for $VaR_{0.01}$ but smaller than the ones corresponding to $ES_{0.025}$. The mean at the different chosen cluster levels ranges from 1.94 to 1.03. The standard deviation ranges from 0.54 to 0.02.

Key figures of RR for $ES_{0.025}$					
Level	Prev. level	Mean	Std. dev.	10-percentile	90-percentile
1		1.9401	0.5423	1.3455	2.6848
2	1	1.3791 (−59.7%)	0.1945 (−64.1%)	1.1764 (−48.9%)	1.6361 (−62.2%)
3	1	1.2364 (−74.9%)	0.0940 (−82.7%)	1.1041 (−69.9%)	1.3457 (−79.5%)
4	2	1.2650 (−71.8%)	0.1374 (−74.7%)	1.1189 (−65.6%)	1.4409 (−73.8%)
5	2	1.2121 (−77.4%)	0.1265 (−76.7%)	1.0809 (−76.6%)	1.3678 (−78.2%)
6	4	1.1784 (−81.0%)	0.1076 (−80.2%)	1.0720 (−79.2%)	1.3140 (−81.4%)
7	4	1.1460 (−84.5%)	0.1006 (−81.4%)	1.0509 (−85.3%)	1.2741 (−83.7%)
8	5	1.1253 (−86.7%)	0.0841 (−84.5%)	1.0439 (−87.3%)	1.2308 (−86.3%)
9	5	1.1088 (−88.4%)	0.0821 (−84.9%)	1.0364 (−89.5%)	1.2034 (−87.9%)
10	7	1.0452 (−95.2%)	0.0422 (−92.2%)	1.0150 (−95.7%)	1.0838 (−95.0%)
11	7	1.0395 (−95.8%)	0.0334 (−93.8%)	1.0114 (−96.7%)	1.0731 (−95.7%)
12	9	1.0459 (−95.1%)	0.0568 (−89.5%)	1.0124 (−96.4%)	1.0844 (−95.0%)
13	9	1.0335 (−96.4%)	0.0230 (−95.8%)	1.0111 (−96.8%)	1.0611 (−96.4%)

Table 5: Figures for the risk ratio distribution for $ES_{0.025}$. Decrease from level 1 in percentage in the parentheses. Note that the minimum value of the mean, 10-percentile and 90-percentile is 1 and that the decrease in percentage is adjusted for this fact.

5.1.4 Comparison between $VaR_{0.01}$ and $ES_{0.025}$

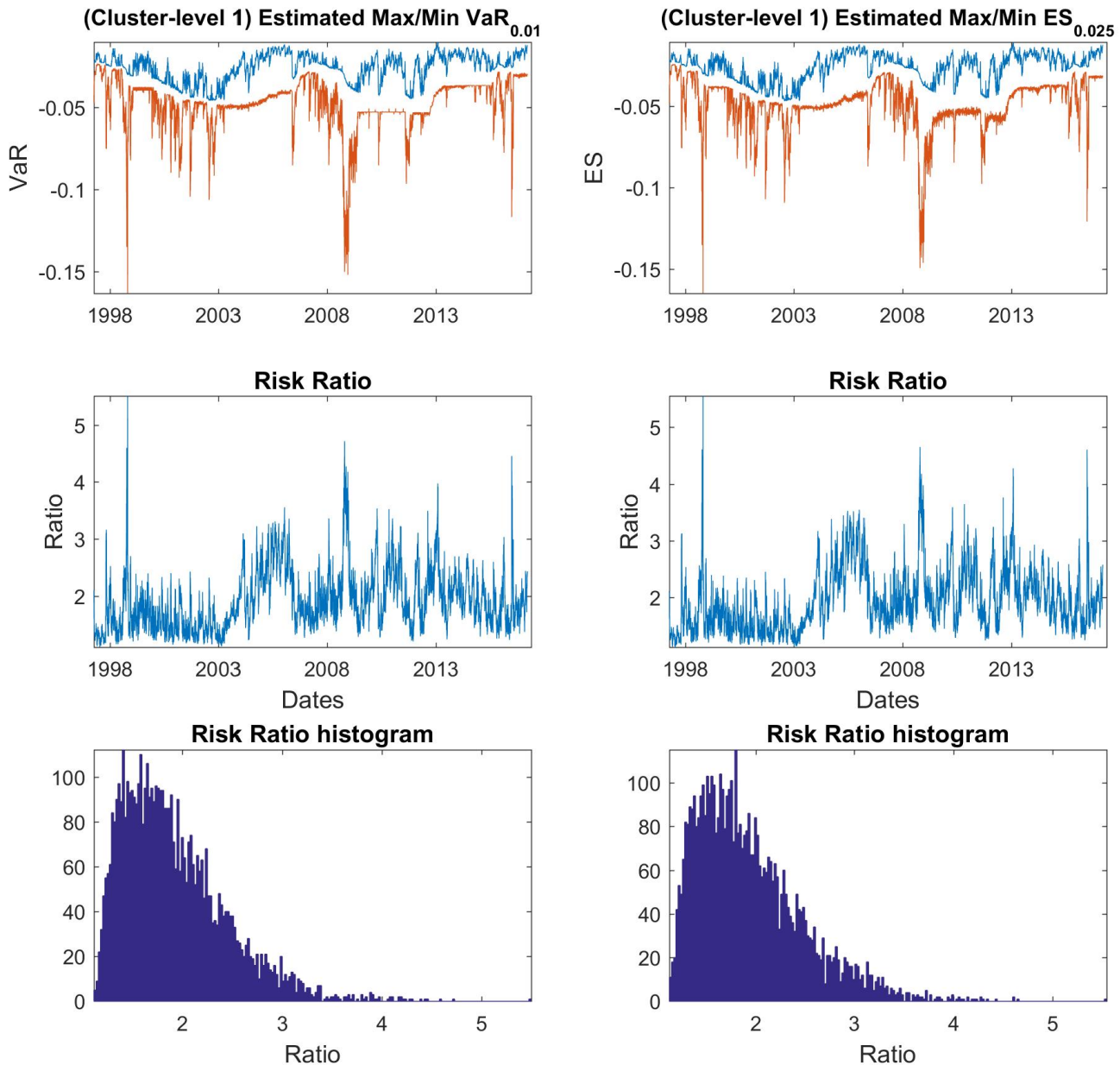


Figure 13: Max and Min $VaR_{0.01}$ and $ES_{0.025}$ -estimates for cluster level 1.

Figure 13 above shows the risk ratio for $VaR_{0.01}$ estimation for cluster 1 to the left and for $ES_{0.025}$ to the right.

Key takeaways from are that the risk ratios for $VaR_{0.01}$ and $ES_{0.025}$ generally look very similar, but with slight deviations. At the higher cluster levels, dissimilarities are harder to spot, but at the lower levels, for example clusters 10 and 13 there is a larger dissimilarity. Looking at the dendrograms corresponding to $VaR_{0.01}$ and $ES_{0.025}$, figures 7 and 11 one can see that the clusters look very much the same. This indicates that the larger spread in RR when $ES_{0.025}$ is used as a risk measure for the clusters for $VaR_{0.01}$ does not stem from the models being more dissimilar than before but rather the way $ES_{0.025}$ estimates risk. The difference caused by the choice of risk measure is presented in the table below, table 6 where the mean, standard deviation, 10-percentile and 90-percentile for the estimation of RR are presented.

Key figures of RR for $ES_{0.025}$ compared to $VaR_{0.01}$					
Level	Prev. level	Mean	Std. dev.	10-percentile	90-percentile
1		1.9401 (2.53%)	0.5423 (-3.57%)	1.3455 (-0.01%)	2.6848 (4.08%)
2	1	1.3791 (-1.56%)	0.1945 (-1.15%)	1.1764 (-4.36%)	1.6361 (-0.38%)
3	1	1.2364 (10.40%)	0.0940 (7.56%)	1.1041 (-16.70%)	1.3457 (9.15%)
4	2	1.2650 (-1.33%)	0.1374 (-3.44%)	1.1189 (-1.01%)	1.4409 (-2.2%)
5	2	1.2121 (-2.13%)	0.1265 (6.87%)	1.0809 (-10.00%)	1.3678 (2.68%)
6	4	1.1460 (-15.2%)	0.1006 (-3.57%)	1.0509 (-27.80%)	1.2741 (-8.84%)
7	4	1.1784 (7.81%)	0.1076 (3.28%)	1.0720 (13.97%)	1.3140 (6.11%)
8	5	1.1253 (8.09%)	0.0841 (2.63%)	1.0439 (16.05%)	1.2308 (5.97%)
9	5	1.1088 (-22.03%)	0.08214 (-5.62%)	1.0364 (-38.78%)	1.2034 (-14.44%)
10	6	1.0452 (-35.13%)	0.0422 (-19.00%)	1.0150 (-34.98%)	1.0838 (-34.71%)
11	6	1.0395 (-34.09%)	0.0335 (-19.93%)	1.0114 (-35.74%)	1.0731 (-33.43%)
12	9	1.0459 (-36.06%)	0.0568 (-9.39%)	1.0124 (-44.69%)	1.0844 (-34.74%)
13	9	1.0335 (-45.50%)	0.0230 (-33.56%)	1.0111 (-53.04%)	1.0611 (-39.34%)

Table 6: Figures for the risk ratio distribution for $ES_{0.025}$ using the clusters from $VaR_{0.01}$. The change, in percentage between $VaR_{0.01}$ and $ES_{0.025}$ are presented in the parentheses.

For some clusters there is not a large difference in RR , for example clusters 1, 2, 4 and 5. However, for the other clusters the change from $VaR_{0.01}$ to $ES_{0.025}$ seems to have a larger effect. The more extreme differences occur the further down one goes in the cluster levels, which is evident in the table above where the change in risk measure has a large effect on the mean RR for clusters 9, 10, 11, 12 and 13. Generally speaking, the different key figures seem follow the same pattern where they all have the same sign and are of similar size. The exception to being of similar size in change is the standard deviation, but the sign (positive or negative) is generally the same as for the other key figures.

5.2 Bond index return modeling (AGG)

5.2.1 $\text{VaR}_{0.01}$

The dendrogram for the hierarchical clustering of the market risk models for $\text{VaR}_{0.01}$ is presented in figure 14. The dendrogram is divided into eleven cluster levels where each cluster level contains all the models connected to it below. In the case of cluster level 1, it contains all possible models.

At cluster level 1, the models get split into clusters 2 and 3. All models in cluster 3 have unconditional variance and all models in cluster 3 have a conditional variance model. All the models in this cluster are relatively similar but the Normal and Student's t are the closest neighbors. However, this cluster is not further explored.

Cluster 2 divides into clusters 4 and 5. Cluster 5 contains only ARCH models while cluster 4 contains the rest of the conditional variance models (GARCH, GJR-GARCH and EGARCH). Cluster 5 is not further divided into lower levels. However, one can see that the following split would have been to exclude an outlier from the cluster (AR-ARCH20-t-filtered-HS) and after that the ARCH models are divided again where all the models that include a normal distribution are placed in the same cluster.

Cluster 4 divides into clusters 6 and 7, where the dividing factor is that cluster 7 contains only models based on normal distributions or normal filtration with the exception of GJR-GARCH-t-filtered-HS. The models in cluster 6 are those that have a non-normal filtration or final distribution, with the one exception of EGARCH-Normal. One can also note that cluster 7 only contains GJR-GARCH and GARCH models, and not a single EGARCH model.

Cluster 6 splits up into cluster 8 and cluster 9, where cluster 9 contains only EGARCH models.

Cluster 8 contains all types of GARCH models, but is further divided into clusters 10 and 11, where the EGARCH models are separated from the GARCH and GJR-GARCH models.

Cluster 10 contains only EGARCH models and gets further split up into two sub-clusters, each where both models have the same innovations distribution, either Student's t or normal distributed. Also, the closest neighbor to AR-EGARCH-t-filtered-HS is the EGARCH-t-filtered-HS.

Cluster 11 further gets split up, where the two models at the top both have an autoregressive mean modeling while the bottom four do not. All distributions are also non-normal.

Cluster Dendrogram

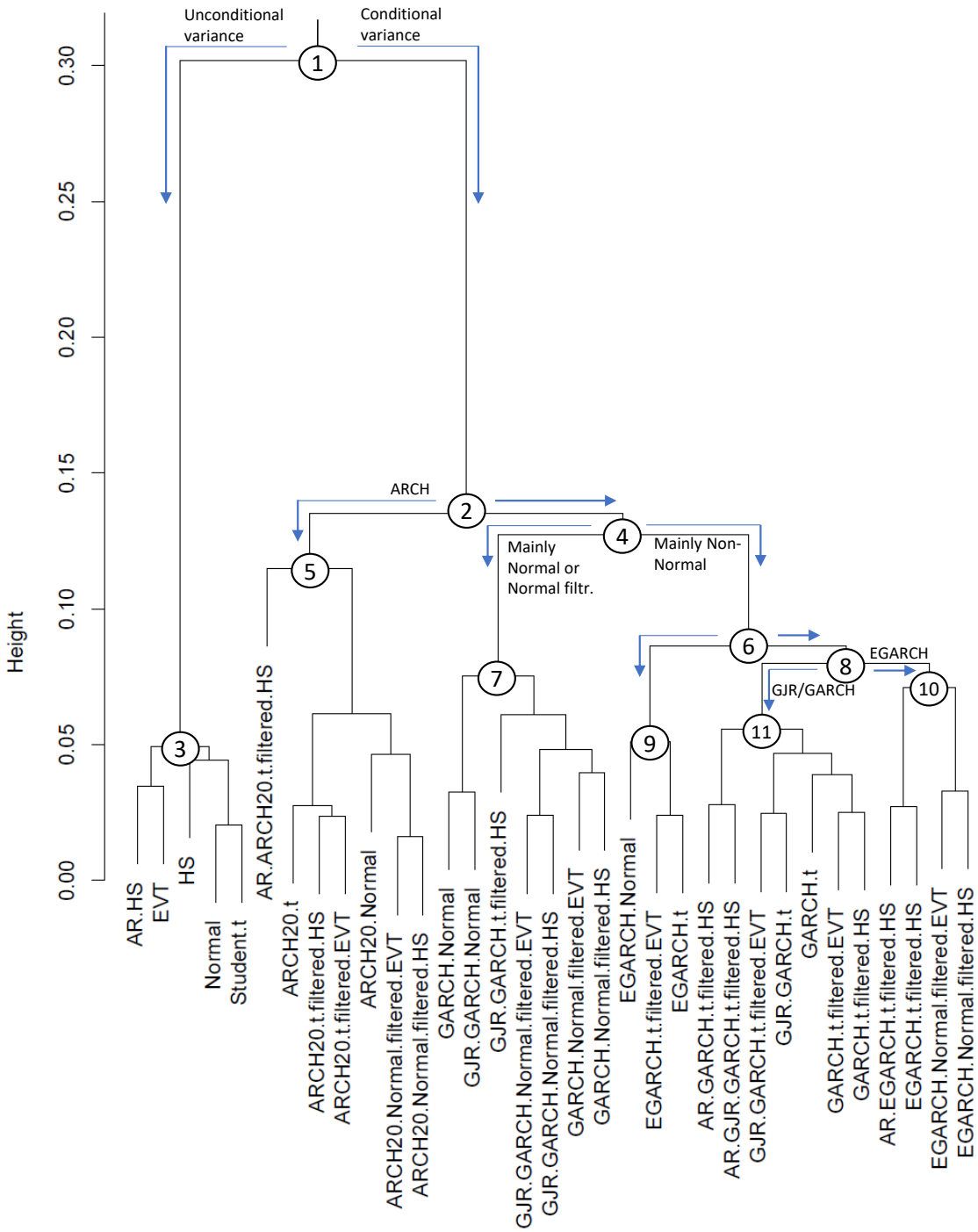


Figure 14: Clustering of models for VaR_{0.01}.

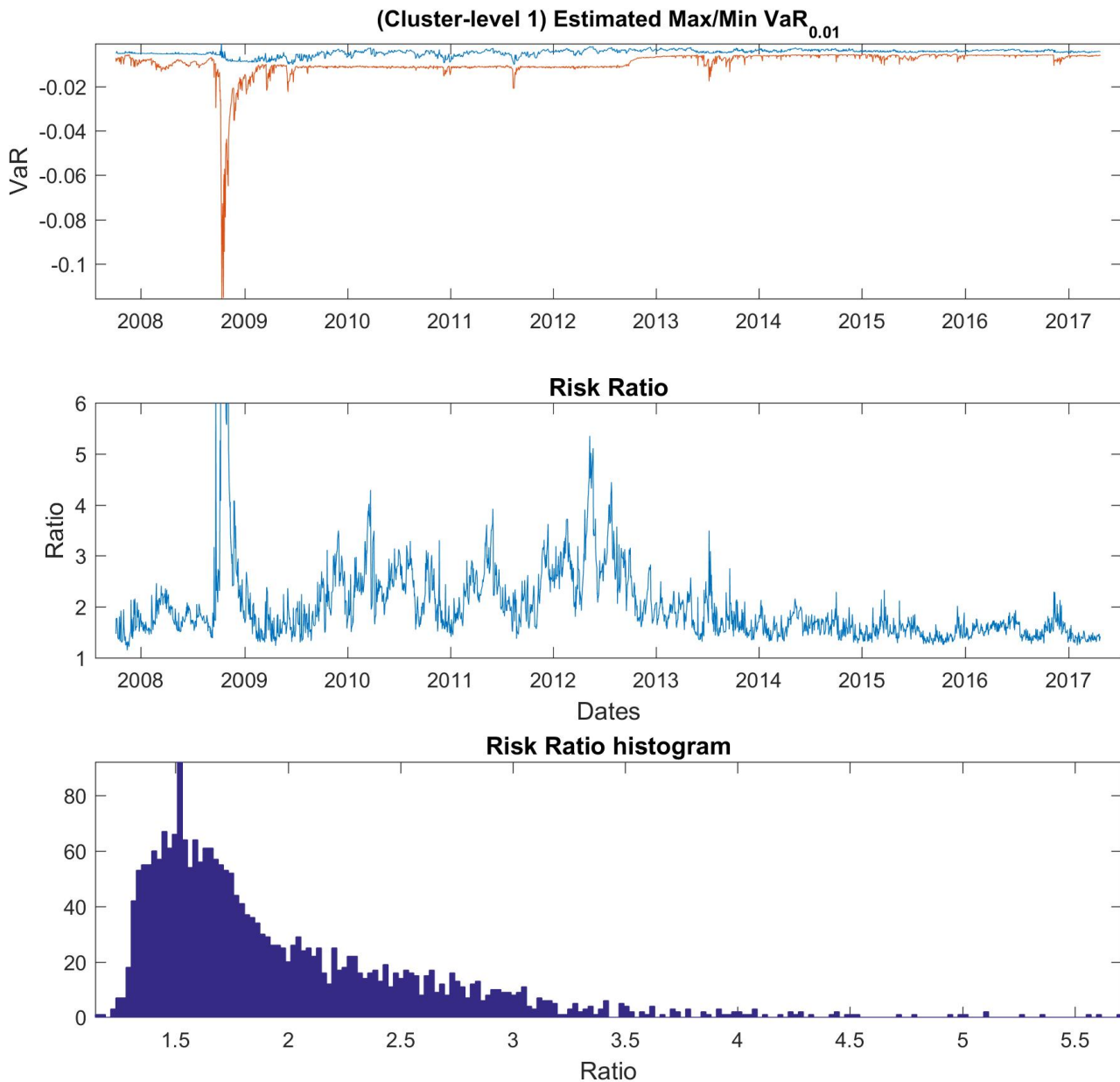


Figure 15: Max and Min VaR-estimates for cluster level 1. The y-axis in the Risk Ratio graph and the x-axis in the risk ratio histogram have both been set to a maximum of 6 to better illustrate the behavior of RR over time. The actual spike in RR in late 2008 is around 117.

In figure 15, the maximum dissimilarity between all models (cluster level 1) is, as in the previous sections, presented in three different ways. The key figures of the risk ratio at different cluster levels are summarized in table 7.

The key figures of RR seem to decrease at roughly the same rate, but with a larger variation than for OMXS30. Also, the cluster with the unconditional variance models have an outlier for the risk ratio early in the time series. This outlier was caused by the AR-HS model in the cluster that naturally reacted much stronger than its neighbors to the large negative return that can be seen in the data analysis of AGG, figure 5.

Key figures of RR for VaR _{0.01}					
Level	Prev. level	Mean	Std. dev.	10-percentile	90-percentile
1		1.9572	0.6213	1.3937	2.8296
2	1	1.4993 (−47.8%)	0.2101 (−66.2%)	1.2794 (−29.0%)	1.7461 (−59.2%)
3	1	1.2079 (−78.3%)	0.2298 (−63.0%)	1.1058 (−73.1%)	1.2923 (−84.0%)
4	2	1.4108 (−57.1%)	0.1722 (−72.3%)	1.2203 (−44.0%)	1.5982 (−67.3%)
5	2	1.1983 (−79.3%)	0.1304 (−79.0%)	1.0947 (−75.9%)	1.3267 (−82.1%)
6	4	1.3183 (−66.7%)	0.1588 (−74.4%)	1.1545 (−60.8%)	1.5102 (−72.1%)
7	4	1.2664 (−72.2%)	0.1270 (−79.6%)	1.1148 (−70.8%)	1.4177 (−77.2%)
8	6	1.2680 (−72.0%)	0.1435 (−76.9%)	1.1295 (−67.1%)	1.4370 (−76.1%)
9	6	1.1344 (−86.0%)	0.0620 (−90.0%)	1.0629 (−84.0%)	1.2158 (−88.2%)
10	8	1.1085 (−88.7%)	0.0848 (−86.4%)	1.0288 (−92.7%)	1.2218 (−87.9%)
11	8	1.1835 (−80.8%)	0.1069 (−82.8%)	1.0860 (−78.1%)	1.3028 (−83.5%)

Table 7: Figures for the risk ratio distribution for VaR_{0.01}. Total decrease from cluster level 1, in percentage, are in the parentheses. The numbers exclude all risk ratios above 6 since the large spike that can be seen in figure 15 is deemed an unlikely outlier and distorts the comparison presented in the table.

In the dendrogram, figure 14, the vertical distance (named height in the figure) roughly corresponds to how much the risk ratio decreases between clusters. For example, when comparing the dendrogram with table 7 it can be seen that the risk ratio has decreased more in cluster 7 than in cluster 6, and the same holds true when comparing cluster 7 with cluster 3.

All in all, the dendrogram in figure 14 illustrates a hierarchy of importance amongst different assumptions, and table 7 shows just how much of an effect these assumptions have on the spread in risk ratio.

5.2.2 $ES_{0.01}$

In figure 16 the dendrogram is presented for the hierarchical clustering of the market risk models for $ES_{0.01}$.

At cluster level 1 the models are divided due to the same factor as in previous dendrograms, conditional or unconditional variance. Cluster 1 splits up into clusters 2 and 3, and in cluster 2 there is evidently a large difference between the normal distributed model and the non-normal models.

Cluster 3 has a clear split into clusters 4 and 5, where cluster 5 only contains the unfiltered conditional variance models. This cluster then separates the ARCH-models, which form their own cluster, from the GARCH models. The EGARCH models end up in the same cluster while there is a large similarity between the GJR-GARCH and GARCH models which are separated according to the distribution assumed in the variance modeling (Normal or Student's t).

Between cluster levels 3 and 4 there is an outlier model, AR-ARCH20-t-filtered-HS, which is excluded from further analysis since it is a single-model-cluster which naturally causes the risk ratio to constantly be equal to 1.

Cluster 4 divides into clusters 6 and 7, where cluster contains all filtered ARCH models. This cluster is further divided according to the assumed distribution in the variance process (Normal or Student's t).

Cluster 6 divides into clusters 8 and 9, where cluster 8 only contains EGARCH models and cluster 9 contains GJR-GARCH and GARCH models. In the same way as in cluster 7, cluster 9 is further divided into smaller clusters according to the distribution in the variance process.

Cluster 8 contains GJR-GARCH and GARCH models and splits into clusters 10 and 11, where cluster 11 contains only t-filtered models (both GARCH and GJR-GARCH). Cluster 10 contains only normal filtered models with the exception of GJR-GARCH-t-filtered-HS, which is separated from the normal filtered models but still in cluster 10. In cluster 10, the normal distributed models are further divided according to variance model, GJR-GARCH and GARCH models respectively.

Cluster Dendrogram

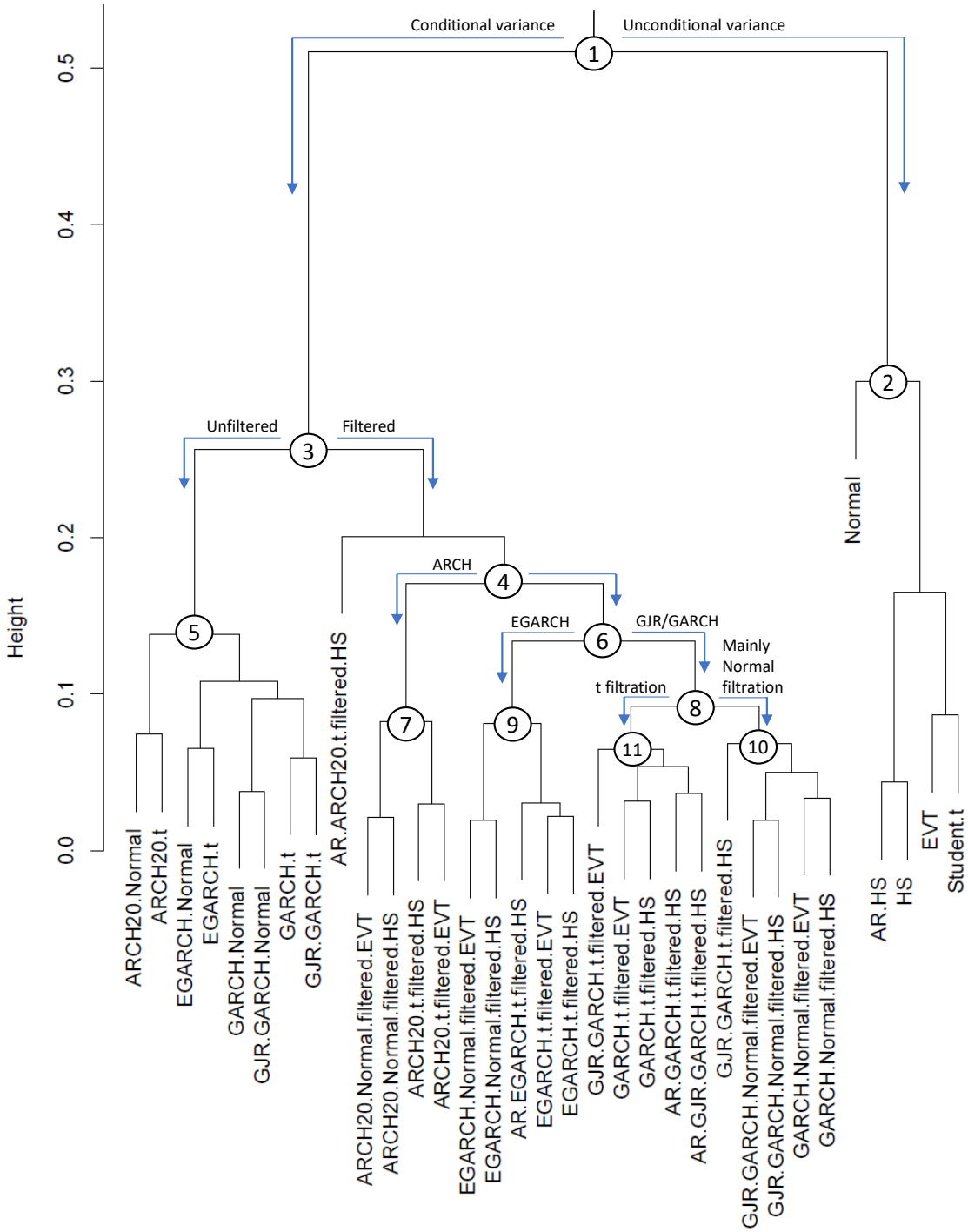


Figure 16: Clustering of models for $ES_{0.01}$.

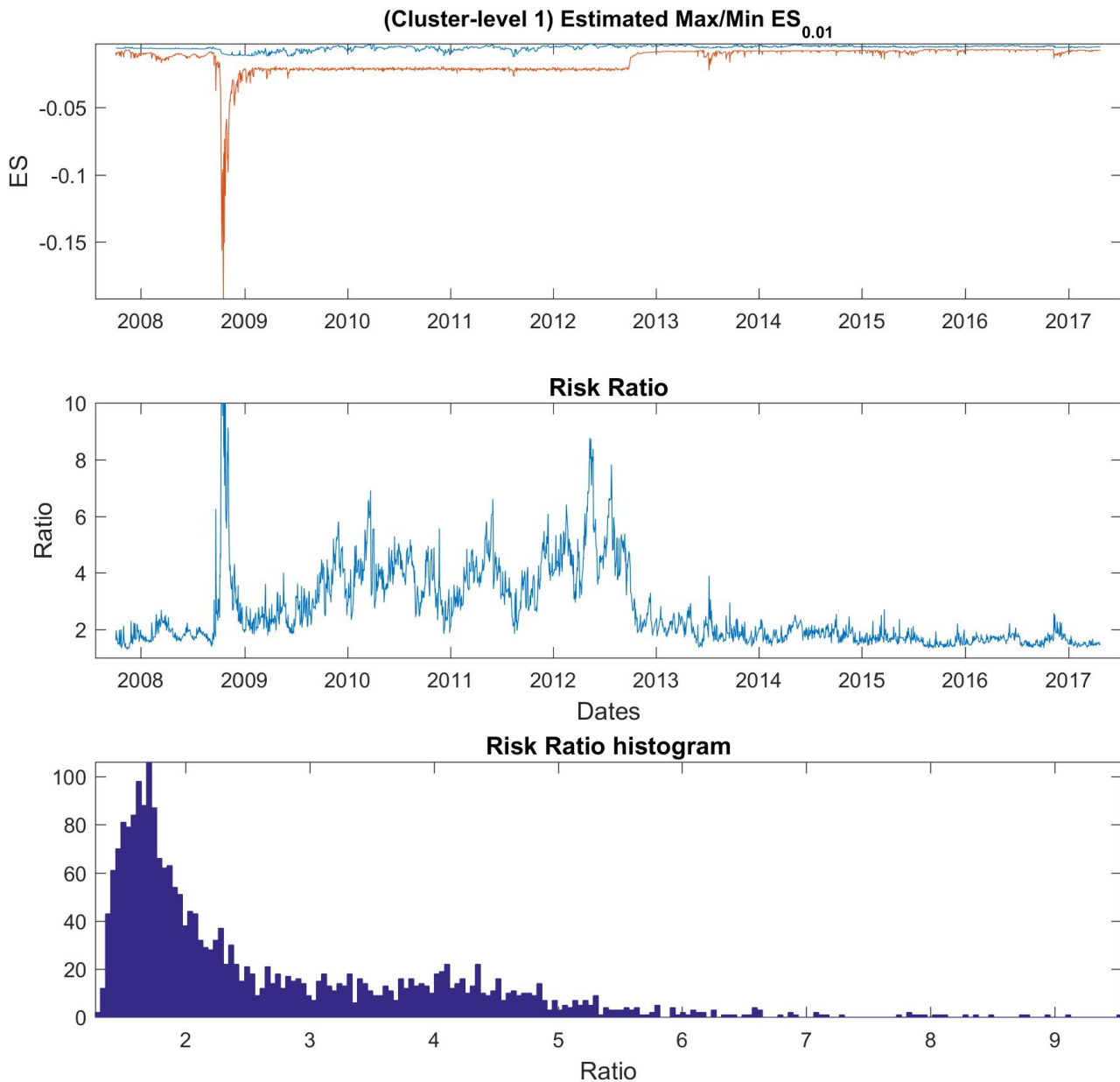


Figure 17: Max and Min ES-estimates for cluster level 1. The y-axis in the Risk Ratio graph and the x-axis in the risk ratio histogram have both been set to a maximum of 10 to better illustrate the behavior of RR over time. The actual spike in RR in late 2008 is around 19.

In figure 17, the maximum dissimilarity between all models (cluster level 1) is, as in the previous sections, presented in three different ways. The key figures of the risk ratio at different cluster levels are summarized in table 8.

The key figures of RR seem to decrease at roughly the same rate, but with a larger variation than for the OMXS30. Also, the cluster with the unconditional variance models have an outlier for the risk ratio early in the time series. This outlier was caused by the AR-HS model in the cluster that naturally reacted much stronger than its neighbors to the large negative return that can be seen in the data analysis of AGG, figure 5.

All the key ratios at cluster level 1 are much larger than the ones for $VaR_{0.01}$. The mean at the different chosen cluster levels ranges from 2.64 to 1.11. The standard deviation ranges from 1.30 to 0.09.

Key figures of RR for $ES_{0.01}$					
Level	Prev. level	Mean	Std. dev.	10-percentile	90-percentile
1		2.6380	1.3045	1.5070	4.5394
2	1	1.5413 (−67.0%)	0.2506 (−80.8%)	1.2740 (−46.0%)	1.8660 (−75.5%)
3	1	1.6497 (−60.3%)	0.2576 (−80.3%)	1.3786 (−25.3%)	1.9241 (−73.9%)
4	3	1.3616 (−77.9%)	0.1851 (−85.8%)	1.1757 (−65.4%)	1.6045 (−82.9%)
5	3	1.3992 (−75.6%)	0.2044 (−84.3%)	1.1956 (−61.4%)	1.6599 (−81.3%)
6	4	1.2733 (−83.3%)	0.1484 (−88.6%)	1.1254 (−75.3%)	1.4542 (−87.2%)
7	4	1.0641 (−96.1%)	0.0611 (−95.3%)	1.0202 (−96.0%)	1.1287 (−96.4%)
8	6	1.2014 (−87.7%)	0.1315 (−89.9%)	1.0769 (−84.8%)	1.3598 (−89.8%)
9	6	1.0841 (−94.9%)	0.0666 (−94.9%)	1.0243 (−95.2%)	1.1838 (−94.8%)
10	8	1.1452 (−91.1%)	0.1150 (−91.2%)	1.0423 (−91.7%)	1.2961 (−91.6%)
11	8	1.1123 (−93.1%)	0.0943 (−92.8%)	1.0369 (−92.7%)	1.206 (−94.2%)

Table 8: Figures for the risk ratio distribution for $ES_{0.01}$. Decrease from the level above in percentage in the parenthesis. The numbers exclude all risk ratios above 10 since the large spike that can be seen in figure 17 is deemed an unlikely outlier and distorts the comparison presented in the table.

There is a rapid decline in each risk ratio key figure from cluster 1 and downwards, most notably in standard deviation. This decline continues in a somewhat large extent going from cluster 3 to cluster 4 and 5. From this cluster and onwards the changes are not as large.

5.2.3 $ES_{0.025}$

In figure 18 the dendrogram is presented for the hierarchical clustering for $ES_{0.025}$.

At cluster level 1 the models are divided into cluster 2 and 3 according to unconditional or conditional variance modeling, the same way as in previous clusters.

At cluster 2 the ARCH models are separated into their own cluster, while the other conditional variance models are grouped in cluster 4.

At cluster 4 the models are divided into clusters 6 and 7 according to variance model, EGARCH and GJR-GARCH/GARCH respectively.

Cluster 6 divides into clusters 8 and 9 where the normal distributed models end up in cluster 9. Cluster 8 consists of the models that have non-normal distributions in them. Some of the models have normal distributed innovations in the GARCH processes, but the similarity between them is that they all have either a Student's t, HS or EVT distribution assumption.

Cluster 8 further divides into clusters 10 and 11, where the t-filtered models are in cluster 10. Cluster 11 consists of mainly normal-filtered models with the notable exception of GJR-GARCH-t-filtered HS. In cluster 11 the models are divided one last time according to the variance model (GARCH or GJR-GARCH) and in cluster 10 we see the same effect, except that the AR models are grouped together regardless of one being a GJR-GARCH model and the other a GARCH model.

Cluster Dendrogram

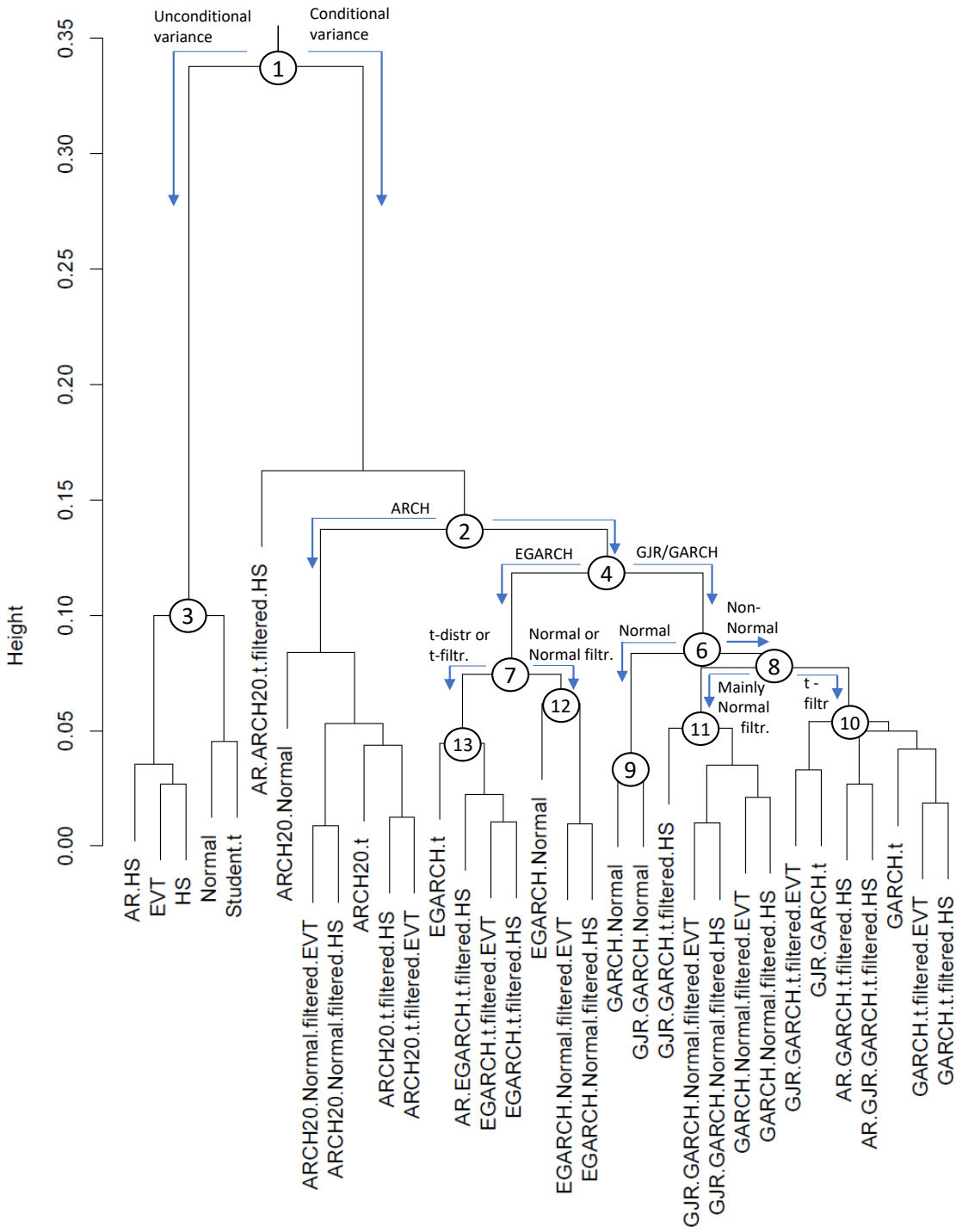


Figure 18: Clustering of models for $ES_{0.025}$.

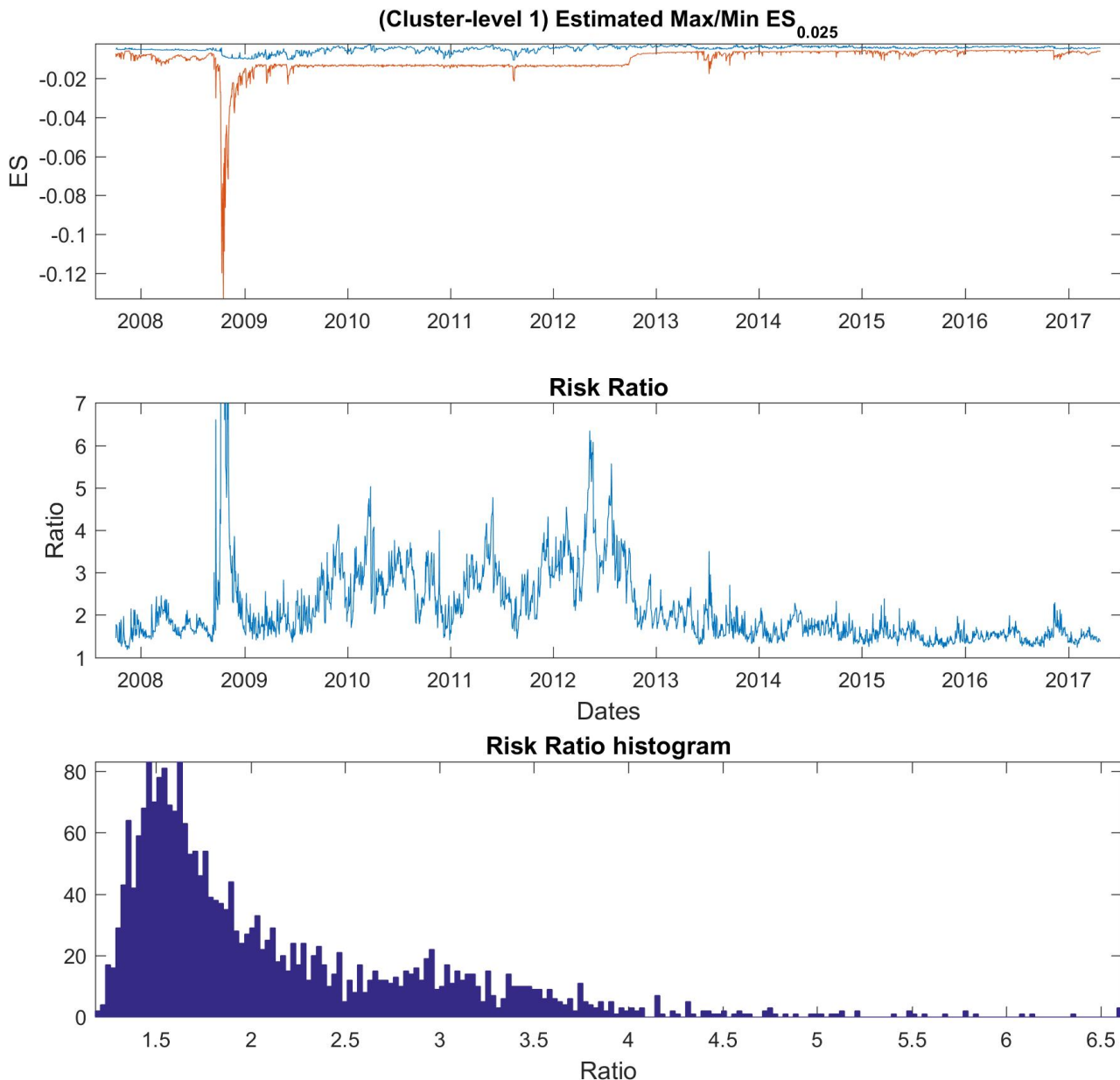


Figure 19: Max and Min ES-estimates for cluster level 1. The y-axis in the Risk Ratio graph and the x-axis in the risk ratio histogram have both been set to a maximum of 7 to better illustrate the behavior of RR over time. The actual spike in RR in late 2008 is around 21.

In figure 19, the maximum dissimilarity between all models (cluster level 1) is, as in the previous sections, presented in three different ways. The key figures of the risk ratio at different cluster levels are summarized in table 9.

The key figures of RR seem to decrease at roughly the same rate, but with a larger variation than for OMXS30. Also, the cluster with the unconditional variance models have an outlier for the risk ratio early in the time series. This outlier was caused by the AR-HS model in the cluster that naturally reacted much stronger than its neighbors to the large negative return that can be seen in the data analysis of AGG, figure 5.

All the key ratios at cluster level 1 are larger than the ones for $VaR_{0.01}$ but smaller than the ones corresponding to $ES_{0.025}$. The mean at the different chosen cluster levels ranges from 2.13 to 1.11. The standard deviation ranges from 0.84 to 0.05.

Key figures of RR for $ES_{0.025}$					
Level	Prev. level	Mean	Std. dev.	10-percentile	90-percentile
1		2.1317	0.84627	1.4072	3.3249
2	1	1.4913 (-56.6%)	0.2015 (-76.2%)	1.2846 (-30.1%)	1.7282 (-68.7%)
3	1	1.2597 (-77.1%)	0.0744 (-91.2%)	1.1870 (-54.1%)	1.3274 (-85.9%)
4	2	1.4063 (-64.1%)	0.1677 (-80.2%)	1.2345 (-42.4%)	1.6027 (-74.1%)
5	2	1.1733 (-84.7%)	0.0815 (-90.4%)	1.0831 (-79.6%)	1.2639 (-88.7%)
6	4	1.3190 (-71.8%)	0.1399 (-83.5%)	1.1783 (-56.2%)	1.4827 (-79.2%)
7	4	1.2135 (-81.1%)	0.0901 (-89.3%)	1.1210 (-70.3%)	1.3534 (-84.8%)
8	6	1.2233 (-80.3%)	0.1265 (-85.1%)	1.1034 (-74.6%)	1.3769 (-83.8%)
9	6	1.0369 (-96.7%)	0.0567 (-93.3%)	1.0031 (-99.2%)	1.0894 (-96.2%)
10	8	1.1658 (-85.2%)	0.0982 (-88.4%)	1.0779 (-80.9%)	1.2612 (-88.8%)
11	8	1.1194 (-89.5%)	0.0993 (-88.3%)	1.0332 (-91.8%)	1.2429 (-89.6%)
12	7	1.1681 (-85.1%)	0.0574 (-93.2%)	1.0999 (-75.5%)	1.2289 (-90.2%)
13	7	1.1128 (-90.0%)	0.0470 (-94.4%)	1.0579 (-85.8%)	1.1766 (-92.4%)

Table 9: Figures for the risk ratio distribution for $ES_{0.025}$. Decrease from cluster level 1 in the parentheses. The numbers exclude all risk ratios above 8 since the large spike that can be seen in figure 19 is deemed an unlikely outlier and distorts the comparison presented in the table.

5.2.4 Comparison between $VaR_{0.01}$ and $ES_{0.025}$

The following graphs show the difference in max and min estimates for $VaR_{0.01}$ and $ES_{0.025}$ for specific cluster levels where the clusters are determined according to the $VaR_{0.01}$ level.

New regulations have come out that dictate the change from $VaR_{0.01}$ to $ES_{0.025}$ as a risk measure. The following section shows the effect this has on the similarities and dissimilarities for the market risk models used in this thesis through comparing the risk estimate for $VaR_{0.01}$ and $ES_{0.025}$ using the clusters from $VaR_{0.01}$ as shown in the dendrogram in figure 14.

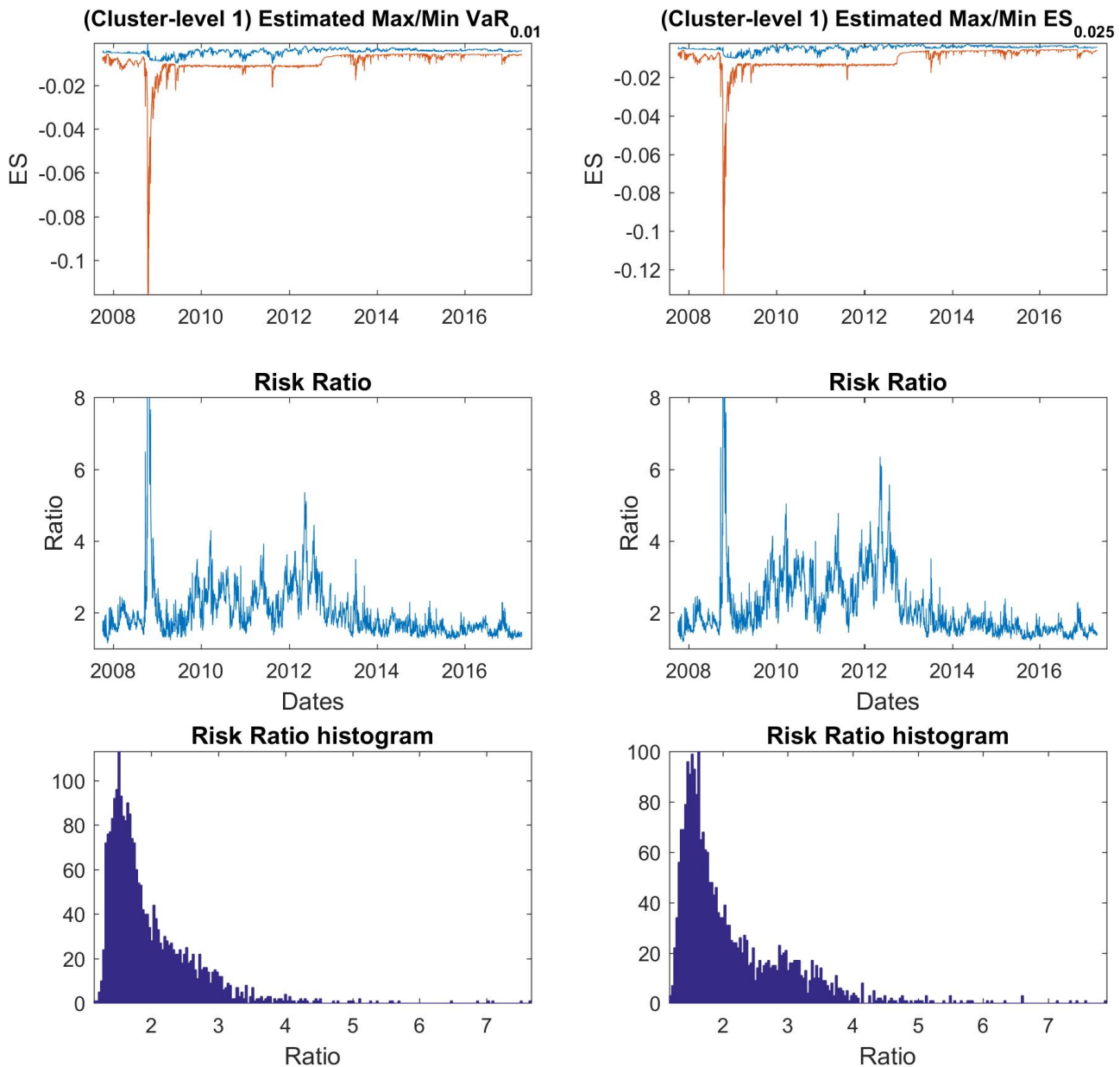


Figure 20: Max and Min VaR_{0.01} and ES_{0.025}-estimates for cluster level 1.

Figure 20 above shows the max min estimations of VaR_{0.01} and ES_{0.025} together with risk ratio and risk ratio histogram.

At cluster level 1, Both RR estimates look very similar throughout the time series, but the histogram tells a different story. The ES_{0.025} estimates seem to be more volatile given the larger tail of the histogram while the VaR_{0.01} RR estimates are centered more below an RR of 3. Looking at the middle graphs the Risk Ratio from ES_{0.025} seems to vary more in size. During the volatile period of late 2009 until 2013 the ES_{0.025} RR is generally estimated higher while also being more volatile. During the period after 2013 they both look the same however. Looking at the table below, table 10, we find supporting data for the graph. At cluster 1 ES_{0.025} estimates a mean of 2.1135 which is 16.33% higher than the VaR_{0.01}-estimate. The standard deviation for RR_1 for ES_{0.025} is 27.59% higher.

Key figures of RR for $ES_{0.025}$ compared to $VaR_{0.01}$					
Level	Prev. level	Mean	Std. dev.	10-percentile	90-percentile
1		2.1135 (16.33%)	0.7927 (27.59%)	1.4067 (3.84%)	3.2630 (26.15%)
2	1	1.4975 (-0.36%)	0.2118 (0.80%)	1.2886 (3.29%)	1.7363 (-1.31%)
3	1	1.2592 (27.37%)	0.0709 (-10.69%)	1.1870 (76.77%)	1.327 (11.93%)
4	2	1.4063 (-1.10%)	0.1677 (-2.62%)	1.2345 (6.41%)	1.6027 (0.75%)
5	2	1.2073 (4.56%)	0.1372 (5.18%)	1.0943 (-0.41%)	1.3538 (8.30%)
6	4	1.3072 (-3.48%)	0.1551 (-2.34%)	1.1568 (1.50%)	1.5021 (-1.60%)
7	4	1.2623 (-1.54%)	0.1145 (-9.80%)	1.1346 (17.30%)	1.3798 (-9.05%)
8	6	1.2347 (-12.42%)	0.1315 (-8.38%)	1.1160 (-10.41%)	1.3859 (-11.70%)
9	6	1.1755 (30.51%)	0.0811 (30.96%)	1.0890 (41.55%)	1.2957 (37.02%)
10	8	1.0725 (-33.11%)	0.0617 (-27.26%)	1.0199 (-30.77%)	1.1590 (-28.30%)
11	8	1.1658 (-9.65%)	0.0982 (-8.12%)	1.0779 (-9.47%)	1.2612 (-13.7%)

Table 10: Figures for the risk ratio distribution for $ES_{0.025}$ using the clusters from $VaR_{0.01}$. The change, in percentage between $VaR_{0.01}$ and $ES_{0.025}$ are presented in the parentheses. Due to outliers in the data where the risk ratio for cluster 1 and $VaR_{0.01}$ momentarily exceeded 120 in late 2008, all days with a $VaR_{0.01}$ risk ratio above 6 were excluded in the comparison.

For some cluster levels, for example levels 2, 4, 5, 6 and 7 there is no significant change in RR when using $ES_{0.025}$ as a risk measure. For other clusters there is a notable change in risk estimates and the spread within the clusters. Most notable is the changes in the risk ratio mean for RR_9 and RR_{10} where the mean changes with +30.6% and -33.1% respectively. Looking at the dendrogram in figure 14 cluster 9 contains three models, EGARCH-Normal, EGARCH-t-filtered-EVT and EGARCH-t.

6 Discussion

In the following sections, the results and the possible effects the methodology had on the result are discussed.

6.1 Difference between clusters for OMXS30 and AGG

The clusters for OMXS30 and AGG, with the same risk measures, differ significantly, meaning that the primary assumptions which have the largest effect on the models also differ. The first important modeling assumption is, in all clusters, whether the models have an unconditional variance or a conditional variance. Focusing on the conditional variance part of the dendrograms we notice that there are large dissimilarities in the clusters that arise for AGG and OMXS30. For example for the OMXS30 time series, with $\text{VaR}_{0.01}$, we see that the second most important assumption is whether the conditional variance models have a leverage component or not. A reason to this might be the large volatility and the many volatility clusters in this time series where large negative log returns are quite common, meaning that the leverage component becomes an important dividing factor. For the AGG time series, we see in the dendrogram in figure 14 that the second most important assumption is not the leverage component, but rather whether the models are GARCH-models or ARCH. Very soon after this point, at cluster 4, the dividing factor is still not about the leverage component, but instead if the models have a normal distribution or not. The fact that the clustering for AGG and $\text{VaR}_{0.01}$ puts little to no emphasis on the leverage component probably stems from the characteristics presented in the data analysis where it is obvious that the bond index time series does not have many large negative log returns and no significant volatility clustering. The fact that the OMXS30 time series has a considerably higher degree of volatility and volatility clustering seems to result in a stronger emphasis on the volatility models in the clusters. The dividing factors are more tilted towards the models themselves, i.e. if they are filtered or unfiltered, have a leverage component or not and so on. For AGG it seems that the underlying distribution assumptions are more important, in general. I.e. if the conditional variance distribution is Student's t or a Normal, or if the assumed marginal distribution is Student's t, Normal, HS or EVT.

Looking at the dendrogram for $\text{ES}_{0.01}$, figures 9 and 16 we find that the first split for OMXS30, at cluster 3, is not quite clear but that all EGARCH models end up in cluster 4. For AGG however, the first split occurs regarding filtered or unfiltered models. After this point, both dendrograms split up depending on variance model, ARCH is excluded while the remaining (GJR-GARCH, GARCH and EGARCH for AGG and GARCH and GJR-GARCH for OMXS30) fall in the same cluster. One can see in the dendrograms for $\text{ES}_{0.01}$ that AGG at the final steps divide first places EGARCH separate from GJR-GARCH and GARCH but that the distributions in the GJR-GARCH and GARCH models are more important than the variance model itself. For OMXS30, the variance model type is more important than the distribution, and we see that the dendrogram clusters the GJR-GARCH models separate from the GARCH models. What differs between the GJR-GARCH and GARCH models is the leverage term in the GJR-GARCH. What differs between the GJR-GARCH and the EGARCH is that the EGARCH model has no constraints on the parameters. It could therefore be argued that the leverage effect is quite limited for AGG while it has a larger effect for OMX. The effect of unconstrained parameters in the EGARCH seems to be important for both time series.

Looking at the dendrograms for OMXS30, figures 11 and 18, there are large dissimilarities. The first split occurs due to leverage component for OMXS30, but for AGG the dendrogram first separates the ARCH models from the rest. At cluster level 4 for AGG, the separation is between EGARCH and GJR-GARCH/GARCH, indicating that the unconstrainedness in the

EGARCH models has a large effect. The GARCH and GJR-GARCH models are then rather mixed together and one can see that the filtration distributions and final distributions have a larger impact than the model itself. For OMXS30 the leverage effect has a large impact meaning that the EGARCH and GJR-GARCH models are clustered together. Then, in both cluster 4 and 5, the dividing factor is if the models are filtered or unfiltered. Lastly, the models are divided based on variance model. This differs a lot from how the AGG dendrogram separates at different levels.

6.2 Difference between clusters for the different risk measures

Using $\text{VaR}_{0.01}$ as risk estimate might not take into account the differences in the distribution tails caused by the dissimilarities between the models. $\text{ES}_{0.025}$ however might capture the dissimilarity that occurs better, causing the RR to increase substantially as the models now differ more than before. Looking at the dendrogram in figure 18, EGARCH-Normal and EGARCH-t are still grouped together, as in cluster 9 in figure 14, but the EGARCH-t-filtered-EVT model has been separated far away from them. It seems that the filtration procedure has effects which $\text{VaR}_{0.01}$ does not capture well, but $\text{ES}_{0.025}$ manages to.

6.3 The study and choosing a market risk model

The study provides a hierarchy for the different assumptions which the models are based on. Through the clustering one can see that some seemingly simple models are often clustered together with more complicated models. This indicates a large similarity between the models and it could thus be questioned whether the extra complexity is necessary or not. The study therefore provides information to what assumptions are necessary in the models, and helps to avoid over complicated models where the added complexity does little to make the model any different. Two examples are given below.

Some of the most complicated models are the filtered EVT methods, such as the GJR-GARCH-t-filtered-EVT. In these models, the Generalized Pareto Distribution (GPD) is fitted to the tails of the standardized residuals of the historical data. In this method there are many assumptions and parameters which therefore increases the risk in the model estimation. For example, the choice of the threshold u can have a large effect on the estimated distribution. If u is chosen to be too small, then there will be too few data points for the estimation, and if the u is chosen to be too large, then "non-extreme" observations will affect the estimation of the tails of the distribution. So using the cluster (for OMXS30 $\text{VaR}_{0.01}$, for example) and the risk ratio of that level, one could therefore argue that the cluster neighbor GJR-GARCH-t-filtered-HS is a suitable replacement, with a possibility of an added margin. This would reduce the complexity and the number of parameters that are to be estimated.

A similar argument could be made for replacing the GJR-GARCH-filtered-EVT with the GJR-GARCH-t for AGG with the risk measures $\text{VaR}_{0.01}$ or $\text{ES}_{0.025}$. This replacement would reduce the complexity even more than before since the entire step with the standardization of the residuals is also removed.

6.4 The methodology's impact on the result

In the following sections some examples regarding how the result could have been affected by the chosen methodology are described.

6.4.1 Risk ratio

The risk ratio was chosen as a measure to describe the intra-cluster dissimilarity. This measure is, as mentioned, a measure of the maximum dissimilarity in the cluster and does not consider how the spread of the other models than the models with the maximum and minimum estimate behave. If another measure would have been chosen, then it might have altered how much of the model disagreement that is explained by each split/exclusion.

6.4.2 The dissimilarity measure of the hierarchical clustering and its effect on the clustering

The complete linkage measure was chosen in the hierarchical clustering algorithm since it produces clusters where the maximum dissimilarity is low. Therefore it can be argued that it is a suitable choice due to its connection to the risk ratio. However, complete linkage can produce clusters where a member of a cluster actually is more similar to members of other clusters than its own which questions this choice of dissimilarity measure if one was to choose something other than the risk ratio to measure the intra cluster similarity.

6.4.3 Time horizon

The time period used for parameter estimation in this thesis is 1000 days for all models. Using a different number of days might yield a different result than presented in this thesis, and may affect the models to a different extent. For example, the GARCH-based models should be much more insensitive to the parameter estimation windows than the unconditional variance models such as Normal distribution, Student's t , Historical Simulation (HS) or Extreme Value Theory (EVT). This stems from the fact that the GARCH-models put more emphasis on recent events while the static distributions do not. A clear sign of this is in cluster 3 where the max and min estimates fall sharply at the end of 2008, only to rise sharply in 2013, 1000 after a larger negative return. In the same figure, but cluster 2, we see that the GARCH-models are much more responsive to recent events and is less affected by a momentary large shock in the returns. Changing the parameter estimation time window to 500 would mean that the max and min estimates for cluster 3 would rise back up after only 500 days instead of 1000, significantly altering the risk ratio histogram.

As can be seen from the middle graph in figure, for example, 17 the overall risk ratio varies quite a bit from the beginning of the time series until the beginning of 2013. After this point the risk ratio is more steady and remain relatively low in comparison to the earlier time period. The earlier high volatility is reflected in the risk ratio histogram through the distribution having a long and heavy tail. The bulk of the distribution is located at a risk ratio below 3, where the majority lies even below 2. If this study were done with the same time series, but starting at 2013 the histogram and risk ratio would look very different. This high volatility is further reflected in table 8, where the standard deviation for RR at cluster 1 is as high as 1.3045. As soon as one goes further into the dendrogram, into cluster level 2 or 3, the difference between the models with conditional variance and those with unconditional variance becomes clearer. The high volatility in the time series at the end of 2008 creates a large difference between how the different model groups estimate ES, and eliminating this difference and focusing on the models that have the same variance modeling yields a much lower standard deviation in risk ratio, 0.2506 and 0.2576 for clusters 2 and 3 respectively.

6.4.4 Model choices

The result could have been partly changed if even more types of models were included in the study. Several more possible variations of the models could have been implemented, such as changing the number of parameters in the conditional variance model, increasing the number of different distributions, changing the threshold for the EVT-models and more. Two examples are given in the two paragraphs below.

By varying the number of ARCH, GARCH and leverage lags in the conditional variance models, one could have been able to see how this affected the clusters. Maybe a GARCH(2,2) is placed in the same cluster as a EGARCH(2,2) instead of GARCH(1,1), for example.

Furthermore, one could have included additional types of distributions, for example, skewed-t or skewed-normal. However, since the time series in our case did not exhibit any strong skewness, this would probably not have provided any additional interesting insights.

6.4.5 Monte Carlo estimation of the risk measures

As mentioned in the methodology section, the Monte Carlo method is used to estimate the chosen risk measures. Given a large enough sample size, this should have little impact on the results for the most part. However, it could still affect some parts. It might, for example, affect the comparison between $\text{VaR}_{0.01}$ and $\text{ES}_{0.025}$ when comparing the risk ratio at the lower level clusters since $\text{ES}_{0.025}$ is estimated with a mean comprised of multiple sample points while $\text{VaR}_{0.01}$ is estimated with a single sample.

7 Conclusion

In most clusters there is a very clear hierarchy for the assumptions in the models, where one can easily see which underlying assumptions have a significant effect on the risk estimation. This holds true for both time series in this thesis and when the models are clustered a clear structure emerges from which one can discern the most important aspects of the risk modeling. The clustering methodology provides a framework from which risk models can be chosen. Through showing which models are similar and dissimilar one can better understand the differences in the models, their effects on the risk estimation and avoid using over complicated models. For example, in cluster 9 in figure 14, the EGARCH-t and EGARCH-Normal models are clustered together with EGARCH-t-filtered-EVT, meaning that the models are very similar. Instead of using the much more complicated model EGARCH-t-filtered-EVT, which also an added model risk in the form of GPD assumptions, one can chose to use the much simpler GARCH-t model.

The two different time series provide different clusters, meaning that the dissimilarity and similarity between models differ depending on the time series on which the risk is estimated. Performing the same type of clustering on another data set which differs from a stock index and bond index (OMXS30 and AGG) would likely provide its own clusters. This means that different time series are affected differently by the assumptions in the models, which falls quite naturally from the data analysis. For example, in the OMXS30 time series, the leverage effect in the GJR-GARCH and EGARCH models have a much larger effect than it does for AGG. In the case of AGG, the GJR-GARCH models are clustered very close to the GARCH models, indicating that the leverage component has a limited effect. Looking at figure 5 and the graph with log returns there is no clear volatility clusters, giving further merit to the argument that the specific GARCH model is not as important in the AGG modeling as it is when estimating risk for OMXS30.

The clusters may differ quite a bit depending on what risk measure is used. In this thesis we have shown that there is a large difference between the clusters for $\text{VaR}_{0.01}$ and $\text{ES}_{0.01}$ for both time series. Generally, the final final distribution assumption (Normal, t, HS or EVT) is more important for the clustering when looking at Expected Shortfall. However, there does not seem to be the same large difference in clustering when comparing $\text{VaR}_{0.01}$ with $\text{ES}_{0.025}$. The models seem to keep their structures in the clusters. What does differ though, and sometimes with large numbers, is how $\text{VaR}_{0.01}$ and $\text{ES}_{0.025}$ estimate the risk levels for these clusters. Even though the clusters are somewhat the same, the lower one looks in the dendrogram ES tends to estimate the risk significantly lower than VaR does. This is positive for risk modelers, since changing from $\text{VaR}_{0.01}$ to $\text{ES}_{0.025}$ then leads to a lower spread between models at lower clusters and would mean that the model choice within the smaller clusters becomes less important since it would not affect the economic capital requirement as much as it does for $\text{VaR}_{0.01}$.

Further research could be done using a similar methodology but comparing different clustering methods. Altering the method might lead to other clusters than those presented in this thesis, from which further conclusions can be drawn. Furthermore, it would be an important part of analyzing and evaluating the methodology used in this thesis. If the clusters were to change drastically one might argue that the clustering method has a very large effect on the end result, meaning that the clustering method needs to be evaluated further in order to determine which one produces the most correct clusters.

The same methodology as used in this thesis can be used but for multiple models. There are numerous additional models than those analyzed in this thesis, and conducting the same study with a larger set of models would yield a more comprehensive view of the similarities and dissimilarities between market risk models.

The effect of time series length for parameter estimation can be analyzed further, to address whether a smaller or larger window for historical data would affect the clusters to any significant extent.

This thesis is done with two different time series, a bond index and a stock index. The same study could be done with other, fundamentally different time series in order to determine what assumptions are the most important depending on the characteristics of the time series being examined. Perhaps one can find a link stating that "if a time series has the characteristics A, B and C then assumptions X and Y will be the most important".

References

- Carol Alexander and Elizabeth Sheedy. Developing a stress testing framework based on market risk models. *Journal of Banking and Finance*, 32:2220–2236, 2008.
- Felipe Aparicio and Javier Estrada. Empirical distribution of stock returns: Scandinavian securities markets, 1990-95. Report, Department of Statistics and Econometrics and Department of Business, Carlos III University, Madrid, Spain, 1997.
- Luc Bauwens and Sébastien Laurent. A new class of multivariate skew densities, with application to generalized autoregressive conditional heteroscedasticity models. *Journal of Business and Economic Statistics*, 23:346–354, 2005.
- Tim Bollerslev. A conditionally heteroskedastic time series model for speculative price and rates of return. *The Review of Economics and Statistics*, 69:542–547, 1987.
- Jón Danielssón and Yuji Morimoto. Forecasting extreme financial risk: A critical analysis of practical methods for the japanese market. *Institute for Monetary and Economic Studies, Bank of Japan*, 18:25–48, 2000.
- Jon Danielsson, Kevin R. James, Marcela Valenzuela, and Ilknur Zer. Model risk of risk models. *Journal of Financial Stability*, 23:79–91, 2016.
- Robert F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50:987–1008, 1982.
- Rajna Gibson. *Model Risk: Concepts, Calibration and Pricing*. Risk Books, 2000.
- Lawrence R. Glosten, Ravi Jagannathan, and David E. Runkle. On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance*, 48:1779–1801, 1993.
- Richard D. F. Harris, C. Coskun Küçüközmen, and Fatih Yilmaz. Skewness in the conditional distribution of daily equity returns. *Applied Financial Economics*, 14:195–202, 2004.
- Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning -Second Edition*. Springer, 2016.
- Keith Kuester, Stefan Mittnik, and Marc S. Paoletta. Value-at-risk prediction: A comparison of alternative strategies. *Journal of Financial Econometrics*, 4:53–89, 2006.
- Chu-Hsiung Lin and Shan-Shan Shen. Can the student-t distribution provide accurate value at risk? *The Journal of Risk Finance*, 7:292–300, 2006.
- Velayoudoum Marimoutou, Bechir Raggad, and Abdelwahed Trabelsi. Extreme value theory and value at risk: Application to oil market. *The Journal of Risk and Insurance*, 31:519–530, 2009.
- Alexander McNeil, Rüdiger Frey, and Paul Embrechts. *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press, Princeton, 2005.
- Matthew Pritsker. The hidden dangers of historical simulation. *Journal of Banking and Finance*, 30:561–582, 2006.
- Sveriges Riksbank. Kapitalkrav för de svenska storbankerna - riksbankens ståndpunkt. Report, 2015.
- Peter Verhoeven and Michael McAleer. Fat tails and asymmetry in financial volatility models. *Mathematics and Computers in Simulation*, 64:351–361, 2004.