Master's Thesis

Economics and Business Administration

MSc in Finance and Investments

Market Condition Based Modelling of Risk

How does the quantified output of VaR and CVaR perform under normal and non-normal market conditions?

Copenhagen Business School

May 15, 2017

Andreas Barfod

Sondre Valle Hestvik

Supervisor:

Michael Ahm

219,622 characters

109 pages

Abstract

The thesis finds evidence supporting that the assumptions surrounding the Basel framework are inadequate. Both risk measures applied in Basel, value-at-risk and expected shortfall, fails in their most standard form to provide estimates which are sufficiently reducing losses. In particular, the assumption of Gaussian distributed returns surrounding the Basel framework is found to be wrong. Both the homoscedastic volatility estimates provided by the model and the assumption of identically and independent returns are found to be violated.

The returns are found to be relatively stable from day-to-day over the whole period, but is at times quite volatile. Analyses of the observations show that, when divided into two subsamples, a normal and extreme market, the empirical distribution of the two differs significantly from each other and from a Gaussian distribution. Both are found to have high kurtosis and negative skewness, in addition to volatility clustering. The two market states differ the most in tail properties, as the extreme market is described by the large magnitude of the returns at each end of the distribution, whereas the normal market displays more moderate values. The findings suggest that it may not be suitable to use one single model to describe a market characterised by large differences in stability, but rather apply two models conditionally of the given market state.

To adjust the approach of the Basel framework, two models are proposed for the normal and extreme market respectively. A student's t-distributed GARCH (1,1) model is used during normal market conditions to incorporate the distribution properties of the sample, while at the same time incorporating heteroskedastic volatility estimates. Extreme value theory is proposed applied for extreme market conditions in the form of a conditional peak-over-threshold model to account for both the extremity of the tails and to incorporate heteroskedastic volatility. The two models are also combined into one applicable model to provide a realistic setup for a portfolio manager. The models are found to violate the risk limits of VaR and ES far less than the Basel framework, during both normal and non-normal market conditions. In addition, the two models prove to be more adaptable to the state of the market, quickly incorporating changing volatility and thus effectively notifying a portfolio manager of adjusted risk exposure.

Table of Contents

Chapter 1

1 Introduction

This chapter is meant to give the reader an introduction to thesis. Insight into the reflections done in the process of writing the thesis will be presented. First, the choice of topic and reasoning behind the choice will be presented. In section 1.2 the specific problem in which the thesis is trying to solve will be elaborated upon. Then, the topic will be delimited and the structure of the thesis explained.

To motivate and introduce the reader to each respective subject, every chapter throughout the thesis will begin with a short introduction.

1.1 Motivation

"A fraud." This is the characteristic Nassim Taleb, the author of "The Black Swan", describes the risk measure value-at-risk (VaR) in the NY Times article "Risk Mismanagement" (2009). Other quotes originated from the same article such as "*VaR is a very limited tool*." Or: "*Risk modelling didn't help as much as it should have,*" and: "[..] *relatively useless as a risk-management tool and potentially catastrophic when its use creates a false sense of security [..] This is like an air bag that works all the time, except when you have a car accident."* Quotes which stems from a risk consultant, a former risk manager at Morgan Stanley and a hedge fund founder respectively tells a story of a risk measure which, in the least, is disputed. The article has its origin as a backdrop to the Financial Crisis of '08 where VaR was one of the factors blamed for the magnitude of the crisis.

Value-at-risk was introduced in banking regulations in 1996. In the run-up to the financial crisis it failed completely to capture the risks, which for the largest banks and investment firms were excessive. As VaR was used to set capital requirements this had severe consequences, when companies failed to model their own actual risk levels. A result of the crisis and the failure of the model to capture the true market risk lead to regulatory reactions. The VaR requirements was changed to include a "Stressed VaR measure" and later the Basel committee has suggested the inclusion of the required use of another risk measure, conditional value-at-risk (CVaR), designed to deal with 'black swan' loss events, or so-called tail risk.[1]

As VaR measures portfolio risk along the Gaussian distribution curve it makes certain assumptions, one being, in short, that tomorrow will be more or less like today. The result being, VaR estimates the risk under normal market conditions. This raises the question of the ability of VaR to measure risk in a turbulent market such as the Financial Crisis of '08, dot-com bubble and Black Monday. If limited by certain market conditions one might think that another measure is needed in the case of non-normal markets.

The greatest risks are not the ones you can see or measure, but the ones you cannot see and as a result not measure. These are the risks found to be far outside the limits of Gaussian probability, and as such are not thought to happen. Yet they do. Only, historically, the input into risk-management models has been recent information rather than historic periods of stress, which has led to low capital requirements. Further, VaR as a risk measure in the Basel framework deals with 99% of the cases, as this is the standard confidence level. As a result, it tells you nothing about the cases that happens in the other 1%, the extreme observations or the 'black swans'. In other words, VaR tells you that you are unlikely to lose more than a certain amount 99% of the time, but says nothing about what could potentially happen the remaining 1% of the time.

In the 1% you could lose two million instead of one million, and this could happen once or twice a year without consequences. The problem is if you lose 10 or 100 times that, and end up going bankrupt. VaR has no way of measuring this. As a consequence, CVaR has been introduced. It tells you the expected loss, conditional on the losses exceeding the 1% threshold set by the VaR. The introduction of CVaR should as a result be, by definition, a more conservative risk measure. Which raises the question of its ability where VaR fails; does it handle turbulent markets and do we obtain information about the losses beyond the 99% limit as is intended? Further, does CVaR also suffer beneath the assumptions of normality as has been the critique of VaR?

The starting point of the thesis is the failure of VaR and the subsequent increased requirements of risk management, in particular the introduction of CVaR. The failure of the financial systems has broad consequences for the society, and better risk

[1] See [1]

management tools, at the disposal of large banks and investment firms, could help mitigate the risks of such failures. As such, it would help not only the companies itself, but also the stakeholders, which in the broadest sense also mean society.

1.2 Problem

Throughout the thesis, theoretical risk models will be described and elaborated upon using our dataset. [2] In particular, VaR and CVaR will, as motivated upon above, be analysed thoroughly. The performance of these models will be tested to be able to show potential shortcomings of each methodology. As a result, when knowing the potential problems, we will be able to adjust the models to mitigate any flaws, which in turn should give more accurate market risk exposure estimates.

We know we will observe periods with continued increased volatility, which will make it possible to distinguish between two market states; first, normal market conditions where volatility per definition should be lower, and second, non-normal or extreme market conditions, where we will observe higher and more fluctuating volatility levels. This distinction in market conditions makes it possible to statistically define two different market states, to see if there is a difference in modelling the two. If so, the best fit model, given the market setting, must be identified.

Summarised this has led to the following question:

How does the quantified output of VaR and CVaR perform under normal and non-normal market conditions?

Both value-at-risk (VaR) and conditional value-at-risk (CVaR) or expected shortfall (ES) as it will be denoted from this point on, will be tested in different forms. The corresponding assumptions, advantages and disadvantages will be described to be able to answer the question. By presenting a set of models we believe we will be able to present a general view of the two market states true risk profiles. To answer the main problem defined above, we will throughout the thesis seek to answer a set of questions:

- What is, or what can be assumed to be, a standard value-at-risk and expected shortfall model?

[2] See section 2.1

- What is the characteristics of normal and extreme market conditions, and how can they be statistically defined?
- Is there a difference in how the risk models perform during normal and extreme market periods? If so, which type of model do best estimate risk under each condition?
- How can we best decide which model to apply at different points in time?
- Is it possible to make qualified guesses on future market movements, to be able to choose which model to apply?

In addition, much emphasis will be laid upon the choice of distribution to accurately model the risk. We will seek to explore whether the often assumed Gaussian distribution fit our empirical data and is able to give precise risk estimates. If not, other choices of distributions will be investigated to see which best describe the empirical data.

1.3 Delimitation

The focus of the thesis is to effectively estimate VaR and ES through the appropriate distribution models during different market states. Therefore, the main priority is to explore how the underlying models' ability of estimating future risk exposure is affected by market movements, given the empirical distribution of returns. As a result, since the market movements from day-to-day is crucial, the overall volatility of the returns and its direct impact is important for the estimation of risk exposure. Because of this, the modelling of explicit changes in correlation, interest and exchange rates and their effect on the risk measures are not of primary focus. We know all three factors to continuously influence portfolio values, and as a result they are important to take into consideration. Yet, the explicitly given market movements and the challenge in how to model these are the primary focus.

Further, the scope is limited to model the risk of stock indices, represented by a diversified, long positioned portfolio[3] in the thesis. As a consequence, financial derivatives or other financial products are not considered. The selection of the portfolio is not subject to any optimisation, and we assume the portfolio as explicitly given. The portfolio weights will be held constant and the portfolio value will evolve freely in line with the market movements throughout time. As a result, all estimations

[3] Defined in chapter 2.1

related to the risk measures are done on the portfolio as it is, the result being that it is not possible to directly compare the size of the portfolio values at the end of the period with the beginning, as the portfolio evolves with time.

The findings of the thesis are applied to two sets of data, the original sample and an out-of-sample. Because of this, the results are mainly applicable to the contents of the two datasets, and only the tendencies found can be argued to be relevant to similar data. The validity is the empirical data used, and cannot as a result be generalised to other markets and financial products.

Lastly, the area of application for the thesis is to see how changes in risk measures affect institutional investors, in particular portfolio managers, which could have constraints regarding risk exposure. A specific risk constraint or cap will not be defined, but consequences of shifting risk exposure will be discussed. In other words, no specific calculations on the consequences of the risk exposure found throughout the thesis will be done, rather the findings will be thoroughly discussed at each point in time.

1.4 Structure

Chapter 2 & 3: Market characteristics, VaR and ES

The characteristics of the data are explored. Normal and extreme market conditions are defined. Then, value-at-risk and expected shortfall are described and a standard approach to modelling the two introduced.

Chapter 4 & 5: Normality, GARCH, extreme value theory

In this part of the thesis, the normality and statistical characteristics of the normal and extreme samples are investigated. Alternative ways to describe the two subsamples are explored.

Chapter 6 & 7: Thorough testing of discovered models

The properties of the different market conditions and the appropriate models are put together. Through chapter 6 and 7 the assumptions, advantages and disadvantages of the models are explained and their performance tested. The findings are further explored through an out-of-sample test to better be able to draw conclusions on the results.

Chapter 8: A deeper look into extreme periods

In this chapter we look further into the characteristics of the extreme periods, as we have defined them in chapter 2. The properties of the extreme periods are explained and the time leading up to extreme market conditions explained, in an attempt to make qualified guesses on future extreme returns. Furthermore, we seek to test the findings in the chapter and its abilities in guessing future extreme market movements.

Data processing

To gather the empirical data used in the thesis, Thomson Reuters Datastream will be used as primary source of information. Implementation and calculations of empirical data will be performed using the RStudio and Microsoft Excel. R-codes can be acquired by email at: andreasbarfod[at]gmail.com or sondre.valle[at]hestvik.no.

Chapter 2

2 Data characteristics

As we seek to find how value-at-risk and expected shortfall perform under normal and non-normal market conditions, we need to first define a dataset so that we are able to investigate both market states. The observations have to span over a large enough time period in order for there to be enough non-normal or extreme returns to model properly. Further, the definition of extreme have to be determined such that it is possible to, in a statistical manner, pick up the returns to best model their performance against the 'normal' market returns. When investigating the performance of each model later, we emphasise that we do not see risk management as precise science, but rather a tool intended to provide an overview of risk exposure, with the purpose to manage it.

2.1 The Dataset

The rationale behind our choice of data is explained in this section. As the estimates of VaR and ES need to span over a large time period to get accurate estimates, it is also important to know the market development over the period. The dataset forms the basis of the context in which estimates are conducted, which makes it important that the data best resembles the real world, as we wish to build models on the correct empirical basis.

The basis for our chosen dataset was to construct an investment universe as close to that of the average investor as possible, which also should concern the likely risk a portfolio manager could be exposed to. Some investors will invest in less risky products such as bonds and indices. Others, on the other hand, will be interested in an increased risk exposure with positions in stocks, by gearing investments or by taking long or short positions in derivatives. We have, as a consequence, chosen to base our data on stock indices alone, in an attempt to capture the "average" of the different risk exposure. In addition, we think it is important not to overcomplicate the estimation with complex products that is difficult to model and does not in itself contribute to the overall problem, namely the estimation of VaR and ES.

The use of indices gives us indirect diversification effects, by the pooling of large amounts of stocks these conduct. With the inclusion of several geographically diversified indices, this effect is further amplified, while capturing a larger investment universe at the same time, again, with the underlying thought being that the data should best reflect an average risk profile.

Our data spans over a period of 40 years from, 1977 to year-end 2016, and consists of three indices, the S&P 500, DAX 30 and NIKKEI 225. We have chosen indices with as long a history of returns as possible, while still being relevant to the average portfolio manager. The length of our time series affect the accuracy of our statistical analysis and, as such, the robustness of our thesis. This affects the availability of indices to our dataset. An Emerging Markets index, for instance, which would further expand the investment universe could not be included as we could not find one with long enough history of returns to justify its inclusion.

Therefore, the investment universe consists of America, with the S&P 500 consisting of the 500 largest companies listed on NYSE or NASDAQ; Europe, where DAX 30 consists of the 30 major German companies listed on the Frankfurt Stock Exchange; and Asia where the NIKKEI 225 index represents the 225 stocks in the 1st section of the Tokyo Stock Exchange, which is for the large companies listed on the exchange.

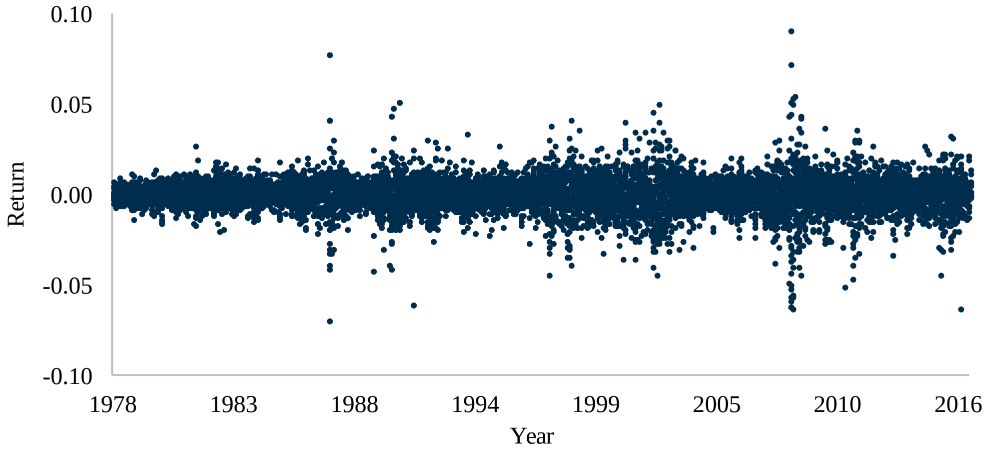
The dataset has daily returns from the beginning of 1977 while the cut-off date is December 31st 2016. The total sample is therefore 10,160 daily returns on each index, and a total of 40,640 daily returns. This gives an average of 254 trading days per year. The data uses daily close prices from Datastream Reuters. All returns have been calculated from prices quoted in the local currency of the index. We have chosen to use log-returns, as this is the often applied method in the financial literature today. From this point on log-returns will be referred to as returns.

We have chosen a simple combination of our indices to form the portfolio used to conduct calculations in the thesis, with equal weights attributed to each of the three. The portfolio is not recalculated at any point during the time period, meaning that each index has a constant weight of one-third during the whole period. The reasoning behind choosing such a simple portfolio is that portfolio optimisation is not a focus of the thesis, and we see the portfolio as explicitly given from this point on, meaning that our concern is to use it to find the best estimate for market risk.

2.2 Market Characteristics

Figure 2.1 displays the daily returns of the portfolio from 1977-2016 on the y-axis while the x-axis is time displayed as years. Most intraday returns are centred around zero, with a large majority within the interval of plus/minus 1%. A few observations are outside this interval, with most still being within a 5% intraday movement in both directions.

Figure 2.1: Return plot dataset



The dataset contains returns from a large period in time, and observations over 40 years should include a large variety of market states. The time period of the dataset is covering several crises in global financial markets, such as Black Monday (1987), the Dot-com bubble (2000) and the Late 2000s Financial Crisis (2008). All of which can be observed as relatively large vertical movements in the plot, expanding the range of returns observed significantly. This should provide us with an adequate number of extreme observations, as all three periods were characterised by large fluctuations, and combined they spanned over a large number of trading days. Further, the markets regularly experience periods of fluctuations with larger corrections or events such as Brexit, which, especially if compared with a short-term historical mean and standard deviation, might be seen as extreme. In the plot they are typically displayed as dots outside the large cluster around +/- 1%, but still well within the interval stretching from -5% to +5%.

A 40-year long period of stock returns will mostly contain observations which could not, in any way, be defined as extreme. A majority of the data will be subject to small intraday movements in the periods between the corrections and crises the markets experience. Such periods are typically characterised by trends, where the markets move in one direction when seen over a larger timespan, but where this trend is not observable in the day-to-day returns. These periods will, especially when compared to the observations that fall into the extreme category described above, be well within the normal state of the market.

As a major objective of the thesis is to explore how VaR and ES are affected, in particular by so called "extreme conditions", the dataset will be divided into two subsamples; one for normal conditions and one for extreme conditions. The reasoning behind this is twofold; first, as will be investigated further in the next subchapter, we believe financial markets have different statistical properties during “extreme” conditions and "normal" states. This could possibly affect the choice of distribution model, and bias results if not taken into consideration. Second, as we later will investigate VaR and ES' characteristics when conditions in the financial markets change, it could be advantageously to be able to distinguish each period's statistical properties from one another, and as such use a different approach to each market state.

Even though it is possible to observe when the occurrence of so called extreme periods discussed earlier happens in the market by simply looking at the returns, it would not be a quantitative definition of extremeness, but rather a qualitative assessment. In addition, we would not be able to pick up all market periods where the fluctuations are extreme at that exact time, as this is not observable in one return alone, but rather needs to be observed over a larger time frame. As such, we would not be able to define one return as extreme as a result of the state of the market, but rather as a result of the return’s extremeness when compared to the whole sample. In the next section, we will therefore further investigate the possibility of defining the extremeness of the returns, based on a quantitative measure that takes into account the current state of the market.

2.3 Dividing into Subsamples

Built on the statistical properties of the returns we want to define them as normal or extreme to be able to model VaR and ES accurately, in particular, when markets experience large fluctuations, as this has proved to be where the risk measures in their standard form experience problems.[4] Distinguishing between normal and extreme returns could provide us with an option to better measure the risk in the periods characterised by large fluctuations, as we get the ability to customise the modelling of these observations. This gives us the opportunity to use the specific characteristics of

[4] See [2] and [3]

each subsample to measure the risk through separate choice of distributions fitted individually on both selections.

When defining the returns, it is important to distinguish between events happening in the market every 50-100 years, which there is no meaning in modelling as they are so rare, and extreme events happening a handful of days every year on average. In other words, the meaning of the word extreme here is not events which are only likely to happen every few years, but rather can be observed a few times a year. As such, they are extreme compared to the state of the market at that time, for instance over the last year, and not extreme over a larger historic perspective. Therefore, when defining the extreme observations, we have to make sure that they are not defined too extreme, as we rather want too many, than too few extreme observations, since these are the interesting events to investigate further.

When dividing the complete sample in two, we use the properties of a Gaussian distribution when defining an extreme period. The Gaussian distribution is a common distribution used to represent real-value random variables with unknown distributions.[5] Since we have no information of our distribution yet this is a natural starting point. Therefore, when finding an appropriate probability to define as extreme, we will use the probability theory surrounding such a distribution. However, we cannot expect a theoretical probability of a certain number of returns to be accurate, as this would assume the distribution to be exactly equal to a Gaussian distribution. We assume the distribution to be Gaussian up until this point, as we have done no empirical estimations to confirm otherwise, but this is merely a tool to find probabilities within an approximate interval, and as such the exact fit of distribution is not important yet.

We want our definition of extreme to contain some information of the short-term state of the market, i.e. over the last year, as the extremeness should be relative to that of the market condition at any given point in time. Further, we want a definition that include a certain percentage of the values that lie outside the given measure to be flagged as extreme, such that the extreme returns are defined relative to the properties of the given dataset. An appropriate measure, then, would involve an assessment of a given standard deviation around the mean, as we would be able to both take into account the current market conditions through a rolling mean and standard deviation,

[5] See [4]

and the extreme sample would be relative to the properties of the dataset as a whole, and as such applicable on other data.

In other words, we want a confidence interval giving us a probability of a return being within a certain interval around a rolling mean. The probability, given that the distribution is normal, i.e. that an observation is within two standard deviations of the mean, is given as:

$$\mathbb{P}(\mu - 2\sigma \leq X \leq \mu - 2\sigma) = \phi(2) - \phi(-2) \approx 0.9772 - (1 - 0.9772) \approx 0.9545 \qquad (2.1)$$

So that approximately 5% of the returns will be more than +/- two standard deviations from the mean. Choosing the returns to be either +/- one or +/- three standard deviations from the mean would lead to probabilities of approximately 68.27% and 99.73% respectively and would as such be too loosely and tightly defined to find the appropriate number of extreme returns. We further assume that the returns are iid, i.e. independent and identically distributed, so that each return has the same probability distribution as others and all returns are mutually independent.[6] Then the probability of seeing one such return over, say one week, or five trading days is given as:

$$\binom{n}{x} p^x (1-p)^{n-x} = \binom{5}{1} 0.05^1 \cdot 0.95^4 = 20.36\% \qquad (2.2)$$

Assuming that returns are iid, the probability of observing one return approximately +/- two standard deviations from the mean is 20.36%. A sensitivity analysis, where the columns represent number of trading days, i.e. a week, two weeks and a month, and the rows represent number of extreme observations during the same periods, yields the following probabilities:

Table 2.1: Gaussian probability of extreme returns

		Trading days		
		5	10	20
# Observations	1	20.36%	31.51%	37.74%
	2	2.14%	7.46%	18.87%
	3	0.11%	1.05%	5.96%
	4	0.00%	0.10%	1.33%
	5	0.00%	0.01%	0.22%

As these are theoretical figures only, we apply the same input parameters to the empirical data, i.e. we count the number of occurrences where we see 1-5 returns that are +/- two standard deviations from the mean over 5-20 trading days. The mean and

[6] See [5]

standard deviation are rolling averages over the last year (254 trading days), so that the extremeness of the returns is a constant factor of the current state of the market. This results in the following figures:

Table 2.2: Empirical probability of extreme returns

		Trading days		
		5	10	20
# Observations	1	21.76%	34.55%	52.53%
	2	5.25%	13.64%	28.69%
	3	1.67%	5.96%	15.15%
	4	0.35%	2.42%	9.90%
	5	0.10%	1.21%	5.86%

The actual percentage of the returns that are classified as extreme, i.e. the number of observations being +/- two standard deviations from the mean, are significantly higher than the theoretical probabilities suggested by a Gaussian distribution. However, the probability of observing one extreme return is more similar to a Gaussian distribution compared to the probability of observing several extreme returns over a given period, which is increasing empirically. It seems as if extreme returns are more likely to occur after one another and thus increasing the actual probability of observing several extreme observations over a short time period.

We have several possibilities when defining the extreme sample which yields a percentage of extreme returns that are within an interval where we have a handful of observations each year, on average. A figure around 5% would result in the number of extreme returns each year being around 13 out of the 254 trading days per year we have in our sample. Further, we need to take into account the sample size of the extreme returns. A too low share of returns being defined as extreme would lead to a small sample, which in turn could lead to difficulties in both fitting a distribution to the observations and drawing conclusions about it in general. Another factor to consider is the length of the period in which we want to count the number of observations being +/- two standard deviations from the mean. A month long perspective (~20 trading days) would require a large number of observations, five or more, and as such the number of "normal" returns included in the twenty day period then being defined as extreme would be up to 75%. With a week long period (five trading days), up to 60% of the returns would be normal, with two returns being defined as extreme. A two-week long perspective (10 trading days) with three extreme returns would result in up to 70% of the returns being normal. At the same time, the period in itself is long enough to draw conclusion on each period on its own. With 10-

day long periods we also have twice as many individual intervals to analyse as the 20-day long periods.

Therefore, the conclusion is to use 10-day periods as the defined length in which we count extreme returns. To have around 5% extreme events to observe each year on average, the constraint must be defined such that we observe three returns or more being +/- two standard deviations from the mean over each 10-day period. The result is an extreme sample containing 590 returns, or 5.96% of the complete sample. The normal sample then, which include observations that does not fulfil the extreme criterion consists of 9310 returns. In total the normal and extreme subsamples corresponds to 39 years of data or 9900 observations, as we use a rolling mean and standard deviation of one year when computing two return sets.

Figure 2.2: Return distribution subsamples

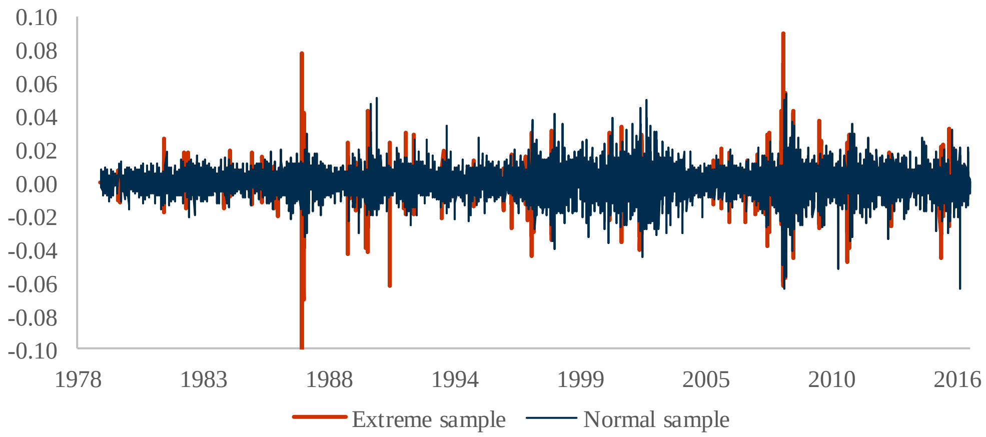


Figure 2.2 illustrates the two subsamples together, where the blue lines represent the normal returns, while the red lines are the extreme observations. The definition of the extreme sample seems to fit well with the largest fluctuating returns in the dataset. Some apparently extreme returns, based on their deviation from the mean of the sample, are marked in blue and as such defined in the normal sample, but is then extreme individual events not occurring closely enough in time to be defined within an extreme period.

As found earlier, the share of extreme returns in the sample, compared to the theoretical probabilities of the Gaussian distribution, are significantly higher throughout the varying number of trading days and observations displayed in the sensitivity analysis. This could be an indication of the sample not resembling the properties found in a Gaussian distribution, and as such it could not be described as one. A further investigation into the distribution properties of the normal and extreme subsamples will be conducted in the following sections in an attempt to further confirm the non-normality of the returns.

2.4 Concluding Remarks

With a large dataset spanning over 40 years combined into a diversified portfolio of three large indices, we find the sample to contain enough daily observations to divide it into two subsamples. The ‘extreme sample’, in particular, is found large enough to model. In addition, the period the dataset is spanning over covers market corrections, crashes and crises so that the returns in the extreme sample itself should be defined such that they are exactly that – extreme.

With the observations collected into a dataset, and the normal and non-normal market conditions defined, the risk measures, VaR and ES, which quantified output will be investigated, have to be explained thoroughly. Both will be defined and a standard approach to estimate market risk will be proposed in the next chapter.

Chapter 3

3 Value-at-Risk and Expected Shortfall

The two risk measures value-at-risk (VaR) and expected shortfall (ES) provides an opportunity to produce statistical data on the market risk during the defined normal and extreme periods. Both are widely used and standardised through the Basel Committee on Banking Supervision, and as such they serve as a good benchmark throughout the thesis. Both measures will be introduced and the results of a modified version of the Basel VaR and ES run on the dataset presented.

Both sides of the distribution will be presented, i.e. the loss and gain side of each risk measure, denoted VaR-/ES- and VaR+/ES+ respectively, which makes it possible to further investigate the differences in the return properties for negative and positive observations during normal and extreme market states.

3.1 Value-at-Risk

Value-at-risk (VaR) measures the potential loss in the value of a portfolio over the next, defined period of time, for a given confidence interval. It is a threshold value indicating that by a probability *p*, one will not lose more than the VaR loss over a given time period *t*. Or, defined the other way around, VaR is the minimum potential loss in *1-p* cases. What happens beyond this threshold and probability *p*, is not predicted by the framework of VaR. The Basel Committee on Banking Supervision applies VaR as a method of quantifying risk exposure. As VaR is a both a required and reputable method of quantifying risk, we find it natural to apply VaR when determining the risk and measuring the performance of modelling normal and extreme market conditions.

3.1.1 Confidence Level

A common measurement unit, a time horizon and a probability is needed to calculate VaR. The probability of loss chosen usually ranges between one and five percent, while the time horizon can be of any length, but with the assumption that the portfolio composition does not change during the holding period. The confidence level will be calculated under the same assumptions as the Basel committee guides, namely at a level of 99%.[7]

3.1.2 Investment Horizon

The standard of the Basel model is to calculate a daily VaR over a 10-day investment horizon.[7] The intention of the length of the horizon is such that there is a possibility of liquidating the portfolio, in case it is necessary to notably adjust risk exposure. Therefore, the investment horizon represents the liquidity level of the portfolio, and the more illiquid assets are, the more time is needed to sell of the assets. In other words, when new market information is incorporated, risk levels changes and it could take time to adjust the portfolio to an appropriate risk level. However, the assets combined in the portfolio of the thesis are highly liquid, all of them being stock indices in developed economies. For the scope of the thesis, and assuming that buying or selling assets does not affect markets dynamics, a one-day horizon for liquidating the portfolio should be sufficient.

Calculating 10-day VaR and ES such as proposed by Basel, implies the use of overlapping observations, meaning that observations are counted and included several times. This could potentially bias the results, indicating more violations than actually present. Hypothetically, one extreme loss (gain) could affect the nine following days of observations and thus make the 10-day loss (gain) exceed VaR and ES 10 times in total, even though it was caused by only one extreme loss (gain). Due to this potential bias, we find a one-day horizon more appropriate and accurate for the intention of the thesis, namely modelling distributions conditionally for normal and non-normal market states, and measure the performance through VaR and ES. Calculating VaR and ES for a one-day horizon gives more precise estimates, making it possible to directly compare daily loss (gain) with the corresponding estimate for VaR and ES. Therefore, a one-day horizon, rather than a 10-day horizon, will be applied as the investment perspective throughout the thesis.

3.1.3 Formula

$$VaR_{\alpha\,\%} = \pi(\mu_i + Z_\alpha \sigma_i) \tag{3.1}$$

Where α is the confidence level, π is the portfolio value, Z is the number of standard deviations corresponding to a confidence level of α, and σ is the volatility. The Basel

[7] See [6]

committee advices the use a confidence level at 99%. This implies that a portfolio manager, if regulated by Basel directives, should be able to withstand a loss equal to 99% VaR over a one-day horizon. If the model at 99% confidence level is accurate, 99% of all observed losses should be less than the VaR threshold, implying 1% loss violations. When modelling Basel VaR, the model is indifferent of whether gains or losses are calculated, due to the symmetrical properties of the Gaussian distribution.

3.1.4 Results

We will backtest the dataset to examine how the Basel VaR performs throughout the period. The VaR is measured daily with a risk perspective of one day into the future. The portfolio value is calculated by hypothetically investing one million at the start of 1978 and letting this value evolve by the daily returns of the portfolio. The portfolio value at day *i-1* is used to calculate the VaR for the following day, *i*. Actual loss (gain) is measured by taking the portfolio value at time *i*, subtracted by the portfolio value at day *i-1*. Doing this, VaR estimates at day *i* will be compared with actual (loss) at the same day, *i*. Mean and volatility are calculated using a rolling window of one year.[8] At each time point the mean and volatility will be calculated based on the preceding 254 days, i.e. the average number of trading days per year present in the sample. This means that our first observation is on the 255th day, at the beginning of 1978.

3.1.4.1 Violation Analysis

To quantitatively measure the performance of this VaR model, we will investigate how many times actual losses (gains) are exceeding their corresponding VaR thresholds. For each time it exceeds, this will be counted as a violation of VaR. A violation rate can then be derived, simply by dividing by the sample size. By definition, when calculating 99% VaR, 1% of the losses (gains) should intentionally exceed VaR, resulting in a target violation rate of 1%. To see how VaR performed during different market states, the definition of normal and extreme markets will be applied. This makes it possible to identify whether the violation occurred during a normal or extreme market state. Combining the normal and extreme sample, the complete sample is found. As such, the sum of violations and sample sizes for respectively normal and extreme market states should be identical to the number of violation and sample size for the complete sample.

[8] See [7]

Table 3.1: VaR violation rates using Basel model

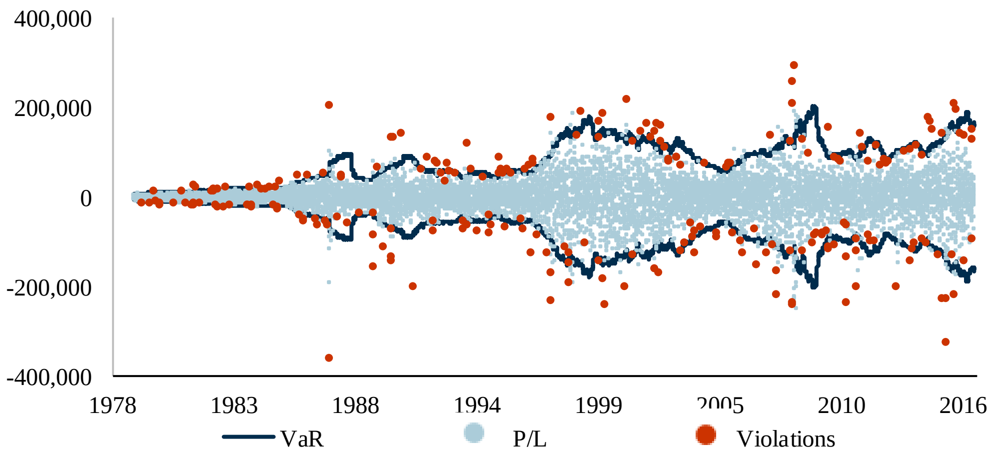
	Complete sample		Normal sample		Extreme sample	
	VaR-	VaR+	VaR-	VaR+	VaR-	VaR+
Violations	206	149	105	97	101	52
Violation rate	2.08%	1.51%	1.13%	1.04%	17.12%	8.81%
Sample size	9900	9900	9310	9310	590	590

Examining the numbers more closely, as displayed in table 3.1, results are not as accurate as the model projected them to be, as we have far more violations than 1% of the total observations for each quantile VaR+ and VaR-. VaR-, the loss side VaR, is over the complete sample violated twice as many time as the target rate of 1%, thus underestimating the risk. VaR+, the gain side VaR, is more accurate, but still underestimating the risk. Investigating further what causes the violations, it is evident during which market states most violations occur. For the extreme market states, the violations are approximately 17 and 9 times as many, for VaR- and VaR+ respectively, as the target violation rate. This is highly inaccurate and makes the model as a whole, over the complete sample, violate more often. Looking at violations during normal market conditions, these are found to be on a more reasonable level than during extreme market states, approximately at the target violation rate. However, the slightly higher violation rate suggests an overestimation of the risk, which could be a result of a constant high VaR rather than the accuracy of the model. This is easier investigated with a plot of the estimates.

3.1.4.2 Graphical Analysis

By analysing violation rates it is possible to get an impression of the performance of VaR throughout the time period. However, it gives no information about the level of VaR in those scenarios where it is not violated. For a portfolio manager the size of the VaR might have implications for the exposure to the market, in order not to exceed potential constraints regarding VaR levels. As such, it is of great importance that VaR is not estimated unnecessary high. Ideally, it should be as low as possible, but still derive VaR estimates with a 99% confidence. To see how it performs, it is possible to graphically plot VaR estimates and actual losses (gains) for the time period. Figure 3.1 below, illustrates VaR, losses (gains) and violations over the period 1978-2016. The vertical y-axis represents one-day portfolio profit/loss (P/L), while the horizontal x-axis denotes days *i* in years. The dark blue lines are the upper and lower quantile of the value-at-risk. Between the dark blue lines, the sky blue points represent losses (gains) not violating 99% one-day VaR+ or VaR-. Moreover, the red dots represent losses (gains) violating 99% one-day VaR+ or VaR-.

Figure 3.1: Basel value-at-risk



The visualisation of the Basel VaR show indications of a slow model with long memory. The fluctuations in the VaR estimates are small and seem to react slowly back to normal levels after spikes or sustained periods of high estimates occur. This signalises that the use of one-year historical mean and volatility, and an underlying Gaussian distribution is potentially not the best fit for estimating VaR. Using one year historical calculations, means that the model is incorporating information that may no longer be relevant. Large losses cause an upward bias in VaR estimates, while longer periods of returns around zero causes a downward bias. The implications are a model reacting too late and underestimating VaR when markets turn volatile, but overestimates VaR when markets are less volatile. In addition, the likelihood of extreme observations found in the empirical data, is significantly higher than the theoretical probability derived by the underlying Gaussian distribution of the model. Which confirms the findings in chapter 2.3, namely that the frequency of abnormal returns is severely underestimated by a Gaussian distribution.

3.1.5 Conclusion

The performance of VaR when applying the Basel framework does indicate a bad fit of the model to the data sample, in particular during extreme market states. Violations are found to be significantly higher than the target violation rate, which intentionally should be 1% for a 99% VaR model. The main reason for the underestimation of risk is due to the underlying Gaussian distribution not being able to explain the likelihood of extreme events, as the theoretical probability of such abnormal returns is so low that it should in practice almost never happen. Still, empirically it is quite frequently observed, potentially causing great losses for a portfolio manager. Further, the graphical analysis displayed that the use of one year of equally weighted historical data, made the model slow, not incorporating information fast enough.

3.2 Expected Shortfall

The second risk measure, expected shortfall (ES) will be presented in this chapter. Expected shortfall (ES), or conditional value-at-risk (CVaR), is a risk measure extending the properties of value-at-risk (VaR). While VaR is concerned with the value occurring at the threshold of a quantile q, ES focuses on what happens beyond this threshold. In other words, ES is the expected loss once VaR is reached. Therefore, at the same confidence level, ES is by definition always larger than the value-at-risk. The Basel Committee has proposed ES to be accompanying VaR as risk measure for regulatory standards, due to VaR's lack of ability to capture tail risk.[9] As such, the analysis of ES will also be an evaluation of its performance compared to VaR, in order to see whether it provides complementary estimates. If a given distribution has fat tails, meaning higher probability of extreme events, ES should capture the potential riskiness of this distribution and thus the magnitude of the tail loss such that it mitigates the underestimation of risk exposure during extreme periods VaR is found to do.

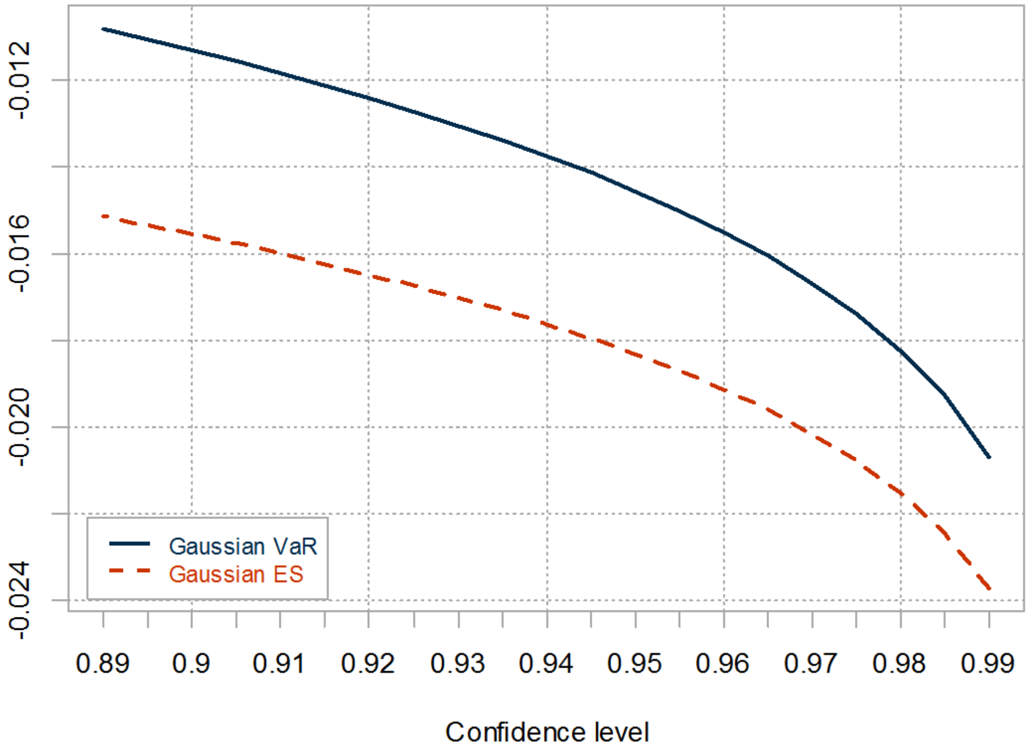
The tail loss is the extreme loss events of the distribution. These extreme events may be crucial for the owner of the portfolio, since the impact of tail events could be substantial, where a worst case scenario could wipe out a large part of the investment. If a good estimate of the portfolio distribution is found, it could extend information availability of tail risk and thus capture fat tails, making the portfolio manager aware of hidden risk not exposed by VaR. Even though uncovering losses behind heavy tails is, in theory, the main objective of ES, it is difficult to do it in practice for an unknown distribution. This is an issue of lack of observations as tail events are infrequent occurrences, and the outcome of the model is only as good as the input. To explain heavy tails, a distribution which is capturing the characteristics and actual movements of the returns, would have to be modelled. However, anticipating these returns is difficult, as they are rare, only occurring few times per year. Hence, the frequency is low, yet sensitivity and impact significant.

[9] See [8]

3.2.1 Confidence Level

The risk sensitivity does not increase linearly with a higher confidence level. Moving towards the tail, risk sensitivity increases exponentially and when approaching a theoretical value of a 100% confidence level, VaR and ES move towards each other.

Figure 3.2: Risk sensitivity VaR and ES



While the Basel Committee guides the use of a 99% confidence level for VaR, they advise the use of 97.5% confidence level for ES.[10] The reason why Basel uses 97.5% ES, is that it provides approximately the same risk sensitivity as 99% VaR, as the risk sensitivity is lower for ES than for VaR. Eyeballing the graph above, this detail can be confirmed. However, while using 99% VaR and 97.5% ES the risk sensitivity might be the same, but ES is no longer a direct extension of the 99% VaR. The size of the tail is not the same, having 99% VaR estimating a 1% tail quantile, while 97.5% ES estimating a 2.5% tail quantile. As the objective of VaR and ES in the thesis is to measure the performance of modelling riskiness during normal and non-normal market states, we believe consistency between VaR and ES is more important, than having equal risk sensitivity. Using 99% VaR and 99% ES would provide alternative information about the riskiness at the same confidence level, thus they would complement each other. As a result, a 99% ES will be applied, which makes ES an

[10] See [9]

extension of VaR, providing additional information about the riskiness during normal and non-normal market states.

3.2.2 Investment Horizon

As for VaR, the Basel Committee advices the use of a 10-day investment horizon for ES. This is to be able to liquidate the portfolio over the given horizon, and thus adjust the risk exposure in an appropriate manner. However, due to the intentional use of ES as a measure of determining risk in normal and non-normal market states, there are downsides of using a 10-day investment horizon. Observations are counted several times, possibly causing a bias in the risk estimates. Thus, this would weaken the accuracy of modelling risk during normal and non-normal market states. A one-day investment horizon would be better for the precision of measuring the performance of the distribution modelling, and as a result this will be applied. It should be an appropriate investment horizon, as the indices used in the portfolio of the thesis are highly liquid and could be sold in a day.

3.2.3 Formula

$$ES_\alpha = \mathbb{E}[\, L \mid L \geq VaR_\alpha] \tag{3.2}$$

$$ES_\alpha = \pi \left(\mu_i + \sigma_i \frac{\phi\left(\Phi^{-1}(\alpha)\right)}{1-\alpha} \right) \tag{3.3}$$

As displayed by the first equation, ES is the expected loss given that the loss is higher than the VaR. The second equation displays the mathematical formula of ES, where π is the portfolio value, μ is the daily mean, σ is the daily standard deviation, α is the confidence level, ϕ is the density function for a standard distribution and Φ^{-1} is the inverse of the cumulative Gaussian distribution. Inserting 99% for α provides a 99% confidence level. The investment horizon is set to one day. This implies, given a constraint on the market risk exposure based on ES, that a portfolio manager should be able to withstand a loss equal to the ES-estimate in case of a 1% tail event.

3.2.4 Results

In order to determine the performance of modelling ES, results need to be interpreted correctly. For VaR, it is possible to simply compare the violation rates against a given confidence level. The intention of VaR is producing a confidence interval which contains α% of all profits and losses, implying an ideal violation rate of 1-α. For expected shortfall this is not the case, as there is no optimal target level of violation rates. ES is a measure of the average loss (gain) beyond a given threshold (VaR), hence it is the expected tail loss (gain). Backtesting expected shortfall causes difficulties, as the performance is measured daily, while the model is an average expectation measure. Therefore, on a day-to-day basis, it is not possible to measure average loss (gain) against ES estimates directly, and as such the violation analysis of ES cannot be interpreted the same way as for VaR. Hence, it is not possible to compare the violations with an ideal violation rate. Yet, analysing violation rates will give a clue as to how the risk measure is performing. If violation rates are high, the model is not successful because the actual loss is exceeding ES too many times. If violation rates are low, the model is conditionally successful, as it is not possible to explain the low violation rates without investigating the numbers behind more closely. The number of violations only provides information about the frequency of violations, but not the seriousness of each violation. Few violations could indicate the confidence level of ES being unnecessary large, making it too unlikely to violate ES, or it could indicate that ES is adapting very well and provides appropriate risk estimates for each time period.

3.2.4.1 Violation Analysis

Preferably, a portfolio manager would have a low number of violations, in addition to low confidence levels of ES, as this indicates lower risk levels and more room for market exposure. To investigate this, an analysis of violation rates should be supplemented with a graphical analysis, as well as an exploration of the loss in the cases where ES is violated. The analysis will start by looking into violation rates, which will be specifically analysed and addressed to their respective subsample, normal or extreme.

Table 3.2: ES violation rates using Basel

	Complete sample		Normal sample		Extreme sample	
	ES-	ES+	ES-	ES+	ES-	ES+
Violations	137	88	64	51	73	37
Violation rate	1.38%	0.89%	0.69%	0.55%	12.37%	6.27%
Sample size	9900	9900	9310	9310	590	590

ES provide lower violation rates than the corresponding statistics for VaR. As mentioned earlier, given the mathematical relationship between VaR and ES, this tendency is logical. By definition ES estimates are always higher in absolute terms than the corresponding VaR estimates, given that they have the same confidence level. In other words, when comparing the estimates to the same daily losses, violations for

ES should per definition be lower than for VaR. Violations for ES over the complete sample are approximately half the size of the corresponding VaR figures. For normal market periods, violations are relatively rare. However, due to a high frequency of violations during extreme market periods, the total number of violations are quite high for the complete sample. An interesting finding is that there are significantly more violations for *ES-* than *ES+*, especially during extreme market states, which means that the loss side is severely underestimated compared to the gain side. This is concerning, as the loss side is the vital aspect for a portfolio manager. A systematical underestimation of risk could cause critical consequences. The higher number of violations for ES- could indicate tendencies of a left skewed data. This will be investigated further when analysing statistical properties in chapter 4.

3.2.4.2 Loss Analysis

As stated earlier, the accuracy of ES cannot be determined by looking at violation rates only, since the consequences of the violations are not known. To explore the outcomes of the violations, a loss analysis will be performed, investigating the seriousness of violating ES. The total loss is found to be the sum of the difference between the ES estimate at each point of violation and the actual observed loss. It does only indicate the magnitude of the potential of loss in the portfolio and is not netted against gains or adjusted for increases in portfolio value over the period. The average loss is total loss over total number of violations.

Table 3.3: Loss analysis using Basel

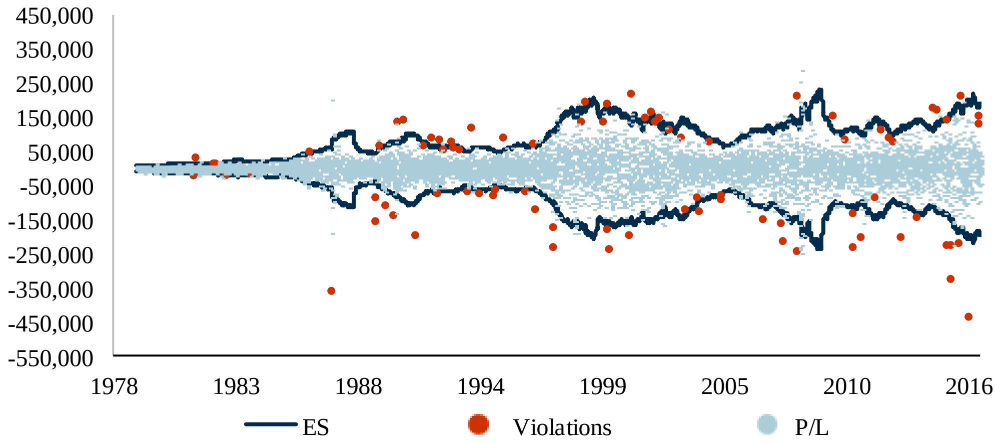
ES-	Normal	Extreme
Total loss	-1,485,165	-2,725,256
Average loss	-23,206	-37,332

As the portfolio value is not constant and changes throughout the time series, the size of the losses itself is of little interest to interpret. However, the trend is interesting and by matching normal and extreme periods it is possible to compare whether the consequences are differing for different market states. Previously, violations during extreme periods was found to be more frequent than during normal periods. The loss statistics suggests that the seriousness of the extreme periods is also greater, having a larger average loss, in addition to a higher number of violations. The larger average loss during the extreme periods indicate that during such market conditions the consequence of violating ES is more severe. Later, when modelling distributions for normal and extreme market states, the loss statistics will be compared using these figures as a benchmark.

3.2.4.3 Graphical Analysis

Up to this point, the accuracy of ES has been determined by the use of violation and loss statistics. However, little information has been provided to determine whether estimates of ES are unnecessary high, overestimating risk exposure and as such causing too many constraints for a portfolio manager. The final step in analysing the performance of the Basel model will be a graphical analysis of ES, in order to see how ES evolves and fits the data sample throughout the time period.

Figure 3.3: Expected shortfall Basel



The use of one year of equally weighted historical data, implicates a slow model reacting too late to market movements. The ES model seems to be overestimating risk in periods where returns are less volatile, with too much distance between ES thresholds and a substantial portion of the returns. For a portfolio manager this would indicate a high risk exposure, through large ES estimates, while the risk is in fact lower. This forces a portfolio manager to be more cautious than necessary and thus potentially lowering the profitability. On the other hand, for periods where returns are more volatile and the risk is higher, ES applies too low estimates, causing an underestimation of risk. This underestimation could make a portfolio manager take on too much risk and thus suffer larger losses. Again, the reasoning is that the use of one year of equally weighted historical data incorporates information too slowly. In addition to the use of a Gaussian distribution, which in general underestimates the likelihood of extreme market events.

3.2.5 Conclusion

Expected shortfall has to a great extent confirmed the discoveries found when analysing value-at-risk, namely that the Gaussian distribution severely underestimates the probability of extreme tail events. In addition, the consequence of these losses are showed to be more severe during extreme market states, compared to normal market conditions. Further, the use of one year equally weighted historical data results in too long a memory in the model, creating an upward bias in the ES estimates after extreme market periods, and a downward bias after less volatile periods.

3.3 Concluding Remarks

The key takeaway from the chapter is that the Basel model provides inaccurate estimates of both VaR and ES over the dataset, severely underestimating risk when the volatility in the market suddenly increases, while overestimating risk during less volatile market states. This is partly due to a one-year long, equally weighted memory, causing a slowly adaptable model, which uses old information no longer relevant for the current state of the market. In addition, the use of a Gaussian distribution does not seem to facilitate the high frequency of abnormal returns observed during extreme market states in particular, potentially resulting in losses for a portfolio manager.

Expected shortfall extends information about the tail distribution compared to the sole use of VaR estimates. Nevertheless, ES displays the same weaknesses as VaR, in that the risk estimates during extreme periods are inaccurate. This might be suggestive of the underlying assumption of distribution model, equal for both models in the Basel framework, not explaining the tail events appropriately.

In the next section we will investigate the statistical properties of the normal and extreme subsamples, as we seek to identify why the Basel model does not perform as well as is intended.

Chapter 4

4 Statistical Properties

In this subchapter, statistical properties of the normal and extreme samples will be analysed. The Basel framework has shown tendencies of not explaining risk accurately. Overall, the models displayed signs of responding late and an incapability of incorporating recent market information. This caused an overestimation of risk during less volatile market periods, while they did not adapt fast enough when markets suddenly changed, resulting in an underestimation of risk in periods with higher volatility. As a result, the risk during extreme market periods were severely underestimated, indicating that the underlying Gaussian distribution did not explain such periods adequately.

The motivation of this chapter is to investigate how appropriate a Gaussian distribution is in explaining the dynamics of the dataset. If not sufficient, we want to identify the best possible models to fit the distributions and market characteristics of the normal and extreme samples. Exploring the statistical properties of each subsample separately, could increase our knowledge of the dynamics during different market states for the data sample. As such, we investigate how to choose the appropriate models to best model VaR and ES for both market states. Each subsample will be analysed separately.

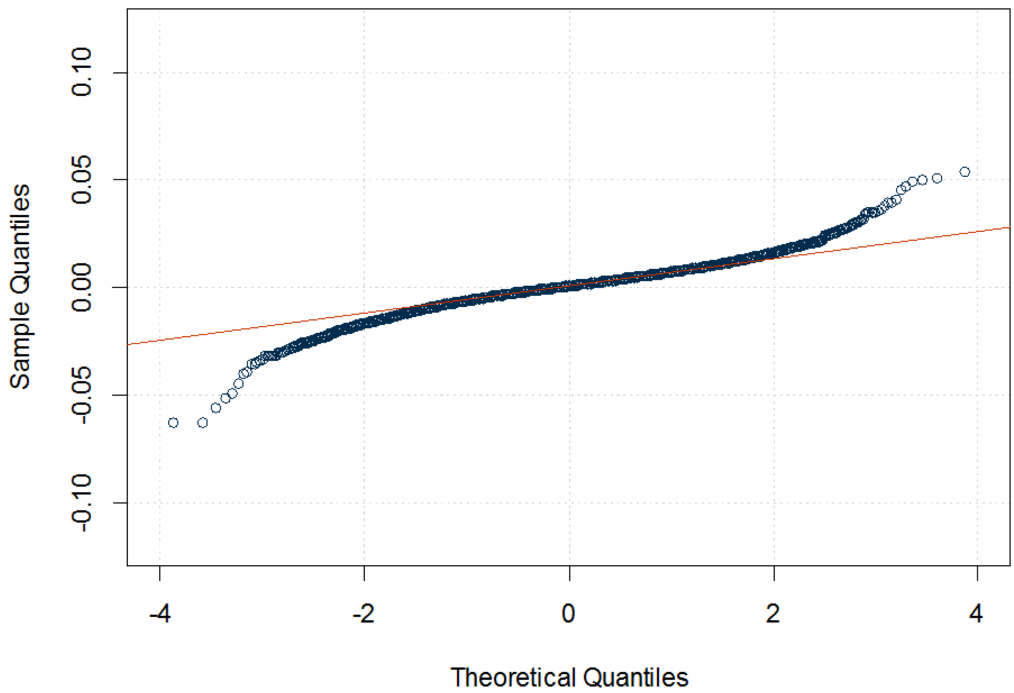
4.1 The Normal Sample

The normal sample is defined as 10-day periods which does not have at least three observations +/- two standard deviations from the mean. Intentionally it should mostly be a sample consisting of normal, regular portfolio returns, i.e. returns which are not extreme. Observations in the sample are typically from the middle 50% of the complete sample and thus the size of the absolute returns are small. However, the normal sample could contain single occurrences of abnormal returns, which are not qualified to be part of the extreme sample, since an extreme period is defined such that at least three non-normal returns must be observed. As the normal sample is much larger than the extreme sample and thus make up a large percentage of the total returns, it is of great importance to model normal market returns as accurate as possible. A small mismatch in predicting one single normal market return, would not cause much concern, since the absolute returns are not large. However, due to the high frequency of normal market returns, approximately 94% of the total portfolio returns, a small prediction error during all normal market states could have critical consequences when using the model to estimate risk at an accurate level. For example, a small prediction error may not be much for one single day, but with 9,310 occurrences of normal market returns, the aggregated consequences could potentially be severe. In order to gain more information about the data and thus explore which distribution that would be the best fit, a statistical analysis of the normal sample's distribution will be performed.

4.1.1 Normality Testing

The Gaussian distribution is by far the most applied distribution in statistics. Many other distribution models are built upon, or modified versions of the Gaussian distribution. Because of this, it is natural to start the statistical analysis by comparing the normal sample, with a Gaussian distribution, to see whether there are any noticeable differences in frequency returns. If the normal sample distribution is Gaussian, it should follow the red line in the graph below very precisely, as the red line is a generated Gaussian distribution with an identical mean and standard deviation as the normal sample.

Figure 4.1: Q-Q plot normal sample



The normal sample distribution is following the line very accurately between quantile -2 to quantile 2, i.e. the middle 50% of the total observations around the sample mean. However, when exceeding these quantiles, returns start to wander off from the Gaussian distribution prediction, becoming more extreme both negative and positive. To explore why this mismatch occurs, several statistical test will be performed to investigate further whether the Gaussian distribution is a bad fit for the normal sample.

4.1.1.1 Mean

In order to get a brief overview of the statistical properties of the normal periods, the moment statistics will be presented.

Table 4.1: Mean normal sample

	Mean
Normal sample	0.04%

The table above displays the daily mean of the normal period between 1978 and 2016. The interpretation of the daily mean is that it is the daily return you would yield on average over the normal period, i.e. you would get a positive average return on your investment of 0.04% every day. In other words, under normal market conditions you would have positive return on the portfolio.

4.1.1.2 Standard Deviation

The mean is the average value of a normal sample return, but returns are not a constant value. Standard deviation is the square root of the second moment statistics, the sample variance, displaying on average how much each day's return deviates from the mean of the normal sample, a return which will vary from day to day.

Table 4.2: Standard deviation normal sample

	Standard deviation
Normal sample	0.79%

On average, returns will over the normal sample deviate by 0.79% from the mean of 0.04%. However, it is difficult to interpret the size of the standard deviation, without comparing it to another sample. Later, the standard deviations of the normal and extreme sample will be compared, in order to see whether there is any noticeable observation to make.

4.1.1.3 Skewness

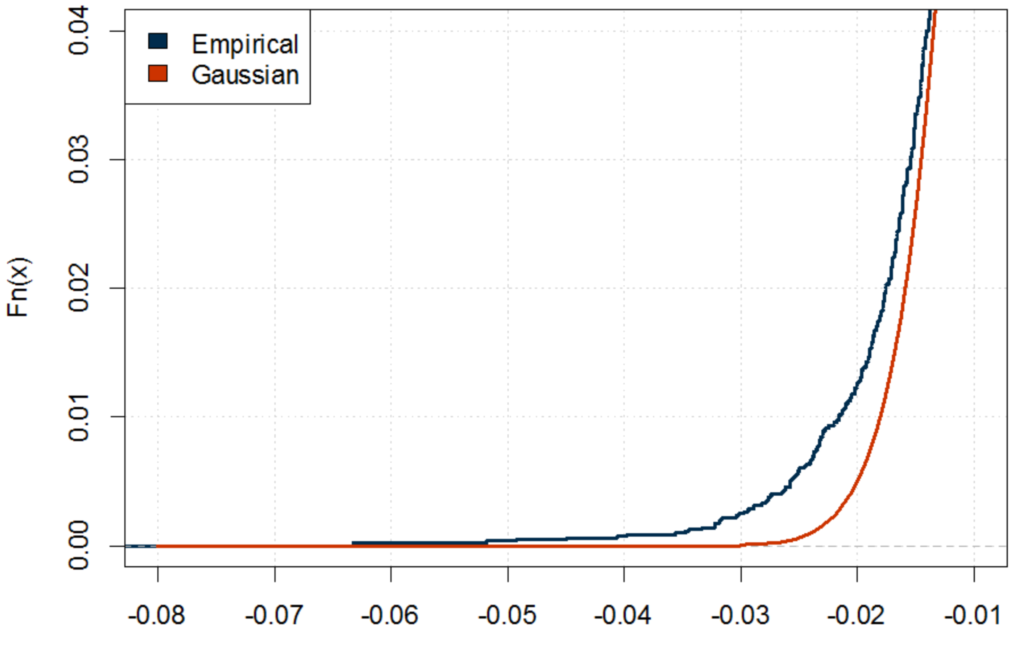
As we seek to find a distribution which is appropriate for the normal sample, an important factor is gaining information about the shape of the distribution. The third moment statistics, the skewness of the sample, measures how asymmetric a distribution is around its mean. A Gaussian distribution has a skewness of zero, implying that the distribution is perfectly symmetrical with equal amounts of return observations on each side of the sample mean. By exploring the skewness, it is possible to determine if the sample has similar properties as a Gaussian distribution or if other distributions could be more fitting to describe this characteristic subsample.

Table 4.3: Skewness normal sample

	Skewness
Normal sample	-0.2206

The normal sample is weakly negatively skewed, meaning that the left tail of the distribution is longer, comparatively to the right tail. This could be due to a presence of negative outliers in the distribution, pulling down the average of the normal sample, resulting in a longer left tail and a more asymmetric shape than a Gaussian distribution. Figure 4.2 below confirms this, as the sample distribution exceed several predicted values of a Gaussian distribution in the lower quantile. When returns get low, the probability predicted by a Gaussian distribution is almost zero, while the empirical distribution of the normal sample shows that the probability is in fact considerably higher than zero. The negative skew, though weak, could indicate that the distribution does not follow a Gaussian distribution and that an appropriate model for the normal sample should take the asymmetry of this distribution into consideration, allowing for a longer left tail. This supports the findings in chapter 3, where the loss side was found to have more violations than the gain side, indicating an underestimation of the risk.

Figure 4.2: Left-tail cumulative return distribution plot



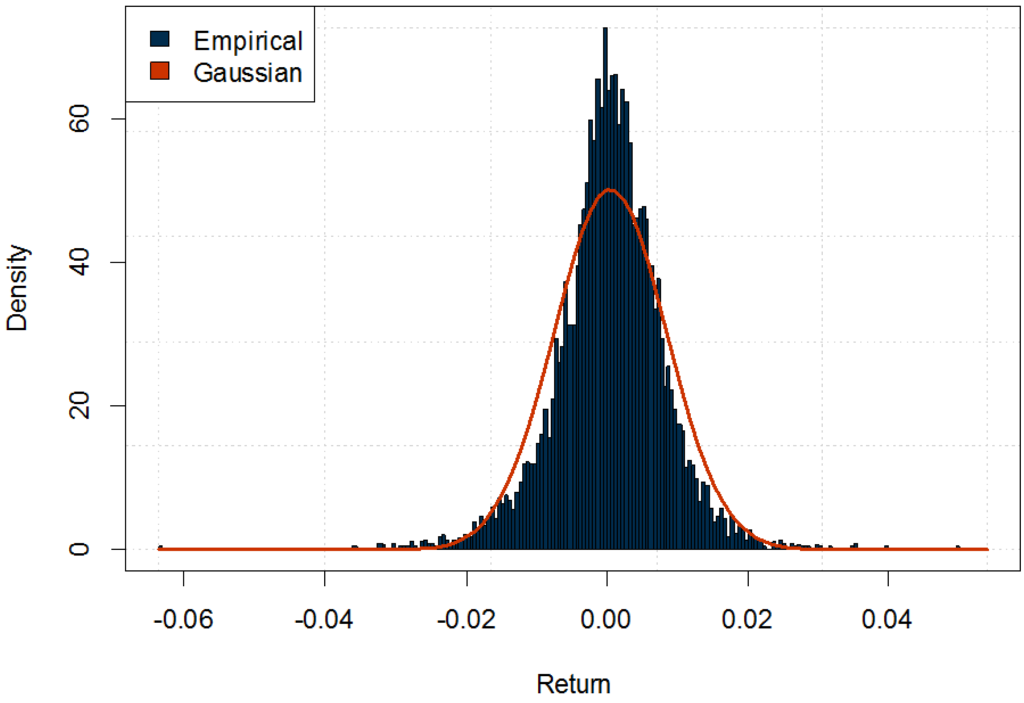
4.1.1.4 Kurtosis

The fourth moment statistics, the kurtosis, quantifies how heavy-tailed a distribution is, relatively to a Gaussian distribution. The heaviness of the tails is essential information when finding a distribution to explain financial return series. Using a distribution with too light tails, implicates an underestimation of the probability of outlier returns. Outlier returns is of importance for a portfolio manager, since it results in high risk and potentially large losses. If the kurtosis is large, the distribution has heavy tails, meaning that the frequency of outlier observations is higher than for a Gaussian distribution. As the Gaussian distribution has a kurtosis of three, the excess kurtosis will be in focus, defined as kurtosis subtracted the value of three.

Table 4.4: Excess kurtosis normal sample

	Excess kurtosis
Normal sample	4.2080

Figure 4.3: Distribution plot normal sample



The distribution of the normal sample is leptokurtic, meaning that excess kurtosis is positive. The distribution is featuring heavier tails than a Gaussian distribution, due to presence of a substantial amount of outliers, increasing the probability of abnormal returns. Moreover, the shape of the distribution is more pointed and less flat than a Gaussian distribution. Figure 4.3 confirms this analysis, where the normal sample displays properties of both heavier tails and higher peakedness, than a Gaussian distribution. As such, a distribution which adjusts for this could be more appropriate for the normal sample.

4.1.1.5 Jarque-Bera Test

As a validation of previous results, a joint test of both skewness and kurtosis could be done by performing a Jarque-Bera test. This is a goodness of fit test, investigating whether the dataset has a shape correspondingly to a Gaussian distribution.[11] Up to this point skewness and kurtosis has been calculated individually, but not whether they are significantly different from a Gaussian distribution. This will be tested by a Jarque-Bera test, where the null hypothesis is whether the normal sample has distribution shape similar to a Gaussian distribution.

$$H_0 : X \sim N(\mu, \sigma^2) \tag{4.1}$$

Table 4.5: Jarque-Bera test normal sample

Jarque-Bera	Test statistic	P-value
Normal sample	6,944.5784	0.0000

By performing a Jarque-Bera test for the normal sample, the null hypothesis of Gaussian normality is rejected. This supports previous findings of negative skewness and excess kurtosis, but now also proved to be statistically significant for the sample. The output from the test suggests that the underlying distribution of the normal sample is not Gaussian distributed and that there might be other distribution models more suitable, in order to find a better goodness-of-fit.

The leptokurtic shape of the distribution, suggests a sizeable frequency of outlier returns, found in the tail of the distribution. Since the Gaussian distribution is not detecting this characteristic, and as we have observed a negative skewness, we should seek an alternative distribution allowing for heavier tails. One such distribution could be a student’s t-distribution.

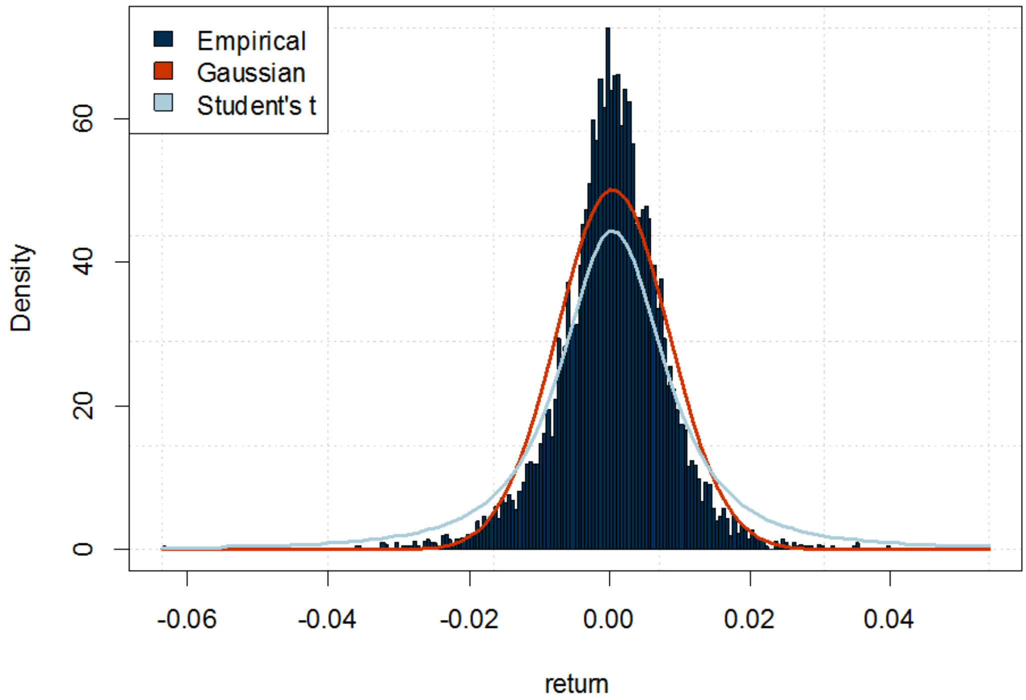
4.1.1.6 Student’s t-distribution

Student’s t-distribution is a modified version of a Gaussian distribution, allowing for heavier tails and as a result more abnormal observations. The original objective of the student’s t-distribution was to be applied to small samples and give more conservative distribution estimates, in form of heavier tails. A variable called degrees of freedom was implemented, making it possible to adjust the heaviness of tails, depending on the

[11] See [10]

sample size. When degrees of freedom approaches infinity, student's t-distribution converges into a Gaussian distribution. By effective use of degrees of freedom, it is possible to adjust the length of each tail, and thus capture outliers to a greater extent. To display the difference in properties between the Gaussian distribution and student's t-distribution, both will be compared to the empirical data of the normal sample in figure 4.4.

Figure 4.4: Distribution plot normal sample



The chart above displays the density function of the normal sample, compared to a Gaussian distribution and a student's t-distribution with two degrees of freedom. This is only an indicative illustration and the plots are not adjusted to best-fit the empirical data of the normal sample. The tails of the student's t-distribution (sky blue line) is clearly heavier than the corresponding Gaussian distribution (red line). In other words, the probability of abnormal returns, found in the tails, are much higher for the student's t-distribution, than for the Gaussian distribution. The t-distribution captures the outlier returns of the normal sample to a greater extent than the Gaussian distribution, and for this reason we believe a student's t-distribution to be a better distribution model for the normal sample.

Despite modelling heavy tails better, student's t-distribution is not improving prediction of the high frequency of returns around the sample mean. The Gaussian distribution predicts this pointed shape slightly better, but neither model is a good fit for this characteristic. If student's t-distribution is to be an appropriate model to use, it needs to be adjusted to the normal sample, so that it takes this peakedness into account, as well as the heavy tails.

4.1.1.7 Conclusion

The shape of the normal sample distribution is weakly skewed, suggesting an asymmetric figure with a long left tail, relative to the right tail. Further, the distribution is found to be leptokurtic, containing heavier tails and more peakedness than a Gaussian distribution. By performing a joint test of skewness of kurtosis, these are found to be significantly different from the properties of a Gaussian distribution, resulting in rejecting normality for the empirical data of the normal sample. As an alternative, student's t-distribution is proposed as a better fit for the subsample.

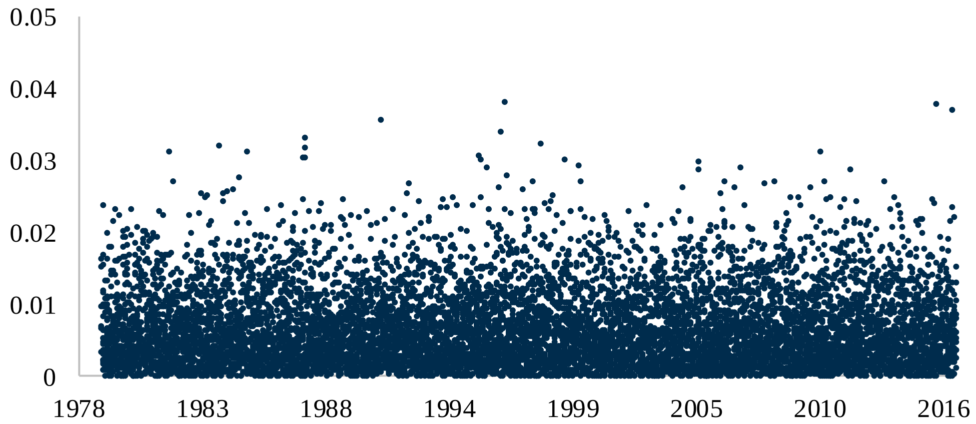
Up to this point the normal sample has been analysed in a static manner, focusing on probabilities and the frequency of observations. To further improve the statistical understanding of normal market states, we will explore how the normal market returns are distributed as a function of time throughout the sample.

4.1.2 Volatility Clustering

In this subchapter, returns will be presented as functions of time, rather than frequencies. The normal sample has a much higher frequency of observations around the sample mean, than both the Gaussian and student's t-distribution predicts. This is a result of the normal sample containing returns from normal market states, which by our own definition are small sized returns mostly clustering around the middle 50% of observations. However, for the distribution to be as accurate as possible in such normal market states, this peakedness has to be incorporated in a distribution model. An important factor when finding the appropriate distribution, is to exploit whether error terms of returns are independent and identically distributed variables (iid), as this expresses how the variance changes with time. Error terms are defined as the actual return, subtracted the mean of the distribution. An important assumption of the student's t-distribution, is iid error terms, and as a result constant variance. A violation of this assumption is therefore a violation of student's t-distribution.

To explore how returns with iid error terms might look like, a corresponding return-set has been generated based on a student's t-distribution. The sample mean, standard deviation and number of observations are identical to those of the normal sample, but they are based on randomly generated variables, rather than empirical returns. In this way it is possible to illustrate how a student's t-distribution distribution could look like.

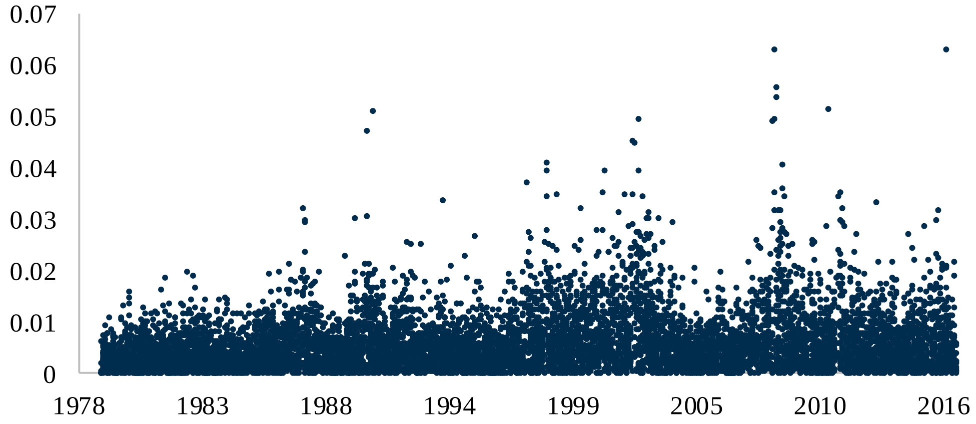
Figure 4.5: Random generated student's t-distribution absolute returns



This is only an indicative illustration of a student's t-distribution, intended for comparison, and as such each respective return cannot be compared directly, but rather over the sample as a whole. As shown by the chart above, absolute returns are distributed with a high degree of symmetry throughout time. Graphically, it looks like returns are iid, since there is no distinct pattern where returns occur together in groups. There are some outliers in the randomly generated sample, but they seem to occur alone and are spread out over the sample, and are thus not causing periods of abnormal returns. For instance, absolute returns from year 1980-1990, appears to be very similarly distributed, as the returns distributed between year 1990-2000 and year 2000-2010. If returns are clustered together, the error terms would not be iid and random, but rather systematically dependent on each other for some periods of time. More specifically, absolute returns for some time periods would be considerably higher than the overall sample mean. Other time periods would yield absolute returns significantly lower than the sample mean. A further implication of error terms not being iid, would be a non-constant variance throughout the sample and thus presence of volatility clustering. If heteroscedasticity is the case, assumptions of both constant variance and iid error terms in the student's t-distribution would be violated.

In order to explore whether the normal sample have presence of volatility clustering, the absolute returns from the normal sample will be plotted as a function of time, by the same procedure as the standard t-distribution. In this way, any noticeable differences between the two datasets, concerning the distribution of returns as a function of time, should be disclosed.

Figure 4.6: Absolute returns normal sample



This distribution of the empirical data of the normal sample has to greater extent periods where absolute returns are higher than other periods, as if abnormal returns are followed and correlated with other abnormal returns. The plot displays significantly lower absolute returns than the sample mean for longer periods of time, while other periods show much higher absolute returns than the overall mean of the sample. This indicates that the error terms, the actual return subtracted by the sample mean, are not iid and truly random, but rather dependent on time. If error terms are dependent on time, it implies that variance is not constant for all periods of the sample, i.e. the normal sample would be heteroscedastic. Graphically there are significant differences between the plot of the normal sample and a standard t-distribution. It looks as if error terms are not equally distributed throughout time, but rather show signs of volatility clustering.

Volatility clustering is a phenomenon where volatility is conditional of time, i.e. heteroscedastic, rather than unconditional and constant. Empirical research of stock market mechanisms show that volatility appears to rise in recessions and fall during peaks, hence it varies over time and depends on the state of the economy.[12] Large moves in returns are often followed by large moves in returns the next days, i.e. the absolute values of returns are correlated. This feature of conditional volatility, is not taken into consideration in the Gaussian distribution, where the volatility is a constant value and not a variable changing throughout the dataset. As constant variance is a vital assumption in the student’s t-distribution as well, this is an important factor to investigate further. The graph suggest that volatility is not constant throughout the data sample. However, eyeballing a graph is not enough to statistically test volatility

[12] See [11]

clustering. To analyse this properly, it is possible to explore whether there is autocorrelation present in the normal sample.

4.1.2.1 Autocorrelation

By analysing autocorrelation, presence of volatility clustering in the normal sample should be discovered. Autocorrelation measures whether the dataset is linearly dependent on itself at different points in time, using a time lag k. The intention is to explore whether a return at time i, denoted R_i, is affected by the returns of the past values at time *i*-*k*, R_{i-k}. Put differently, the purpose is to see if there is any pattern of correlation between returns at time *i* and *i*-*k*, or if they are independently distributed of each other. As autocorrelation is a standardised measure, the coefficient value is a number between -1 and 1. In the calculations of autocorrelation, returns of the normal sample has been transformed to absolute sizes. The reasoning behind this is to exploit whether large returns are correlated with other large returns, positive or negative, and thus indicating volatility clustering.

Figure 4.7: Autocorrelation normal sample

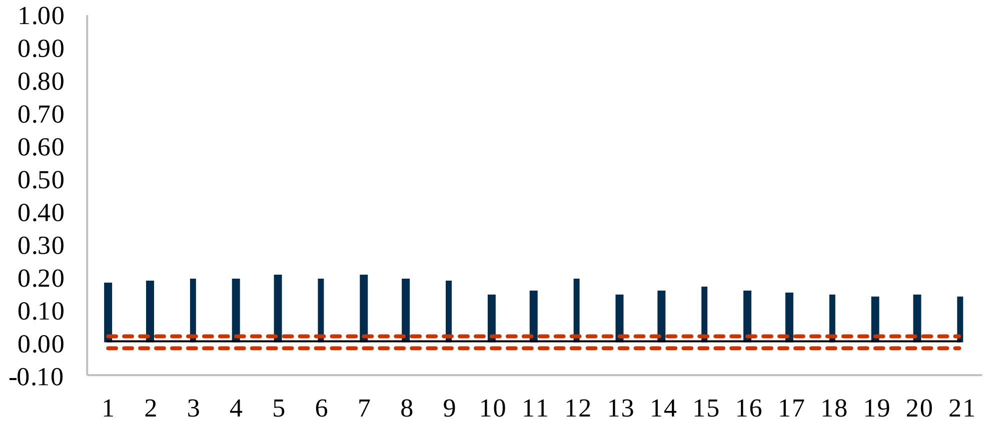


Figure 4.7 shows a clear tendency of autocorrelation in the normal sample. The dark blue spikes in the chart are the size of the autocorrelation coefficients, while the dotted red line is the confidence interval, which a value must exceed in order to be significant at a 95% confidence level. For lags of 1 to 21 trading days, representing one month of data, all autocorrelation coefficients are significant, indicating autocorrelation. This means that returns are conditionally dependent of time and are thus violating an assumption of the student’s t-distribution.

The autocorrelation plot is a test of each correlation coefficient respectively, but to confirm whether this is applicable for the normal sample as one, a joint test will be performed, testing all coefficients together.

4.1.2.2 Ljung-Box Test

Testing autocorrelation for each time lag, gives an indication of whether there is presence of autocorrelation in the dataset. However, independent on whether the autocorrelation coefficients were significant or not, they were only tested for each respective time lag variable. In order to get a more unambiguous conclusion a simultaneous analysis of all the autocorrelation coefficients will be performed, in a so-called joint test. The joint test performed is the Ljung-Box test [13], with a significance level of 95% and up to 21 lags:

$$H_0: (\rho_1, \dots, \rho_{21}) = 0 \tag{4.2}$$

$$H_A: (\rho_1, \dots, \rho_{21}) \neq 0 \tag{4.3}$$

$$Q^* = T(T+2)\sum_{k=1}^{m} \frac{\hat{\rho}_k}{T-k} \sim \chi_m^2 when\, T \to \infty \tag{4.4}$$

Table 4.6: Ljung-Box test normal sample

Ljung-Box	Test statistic	P-value
Normal sample	285.5000	0.0000

The test statistic is higher than the critical value, meaning that the null hypothesis should be rejected. The result is supported by the theory in Brooks (2008) p. 210, stating “*only one autocorrelation coefficient needs to be statistically significant for the test to result in a rejection*”.[14] Since all 21 coefficients was rejected in the last section, results are aligned with theory. The joint test rejects the hypothesis of zero autocorrelation in the time series, strengthening the conclusion that the normal sample does in fact have presence of autocorrelation. We should, as a result seek a distribution model that incorporates autocorrelation when explaining the distribution of the normal sample returns.

4.1.2.3 Conclusion

In contrast to a standard student’s t-distribution, where it is assumed iid error terms and homoscedasticity, there is found to be significant presence of heteroscedasticity in the normal sample. This indicates volatility clustering, meaning that large returns in

[13] See [12]
[14] See [13]

the normal sample has a tendency to be followed by other large returns of either sign. In order to successfully predict normal market returns, a volatility model should be implemented in the student's t-distribution. Without using such a volatility model, which adjusts for autocorrelation in the returns, the likelihood of misestimating market risk increases.

4.2 The Extreme Sample

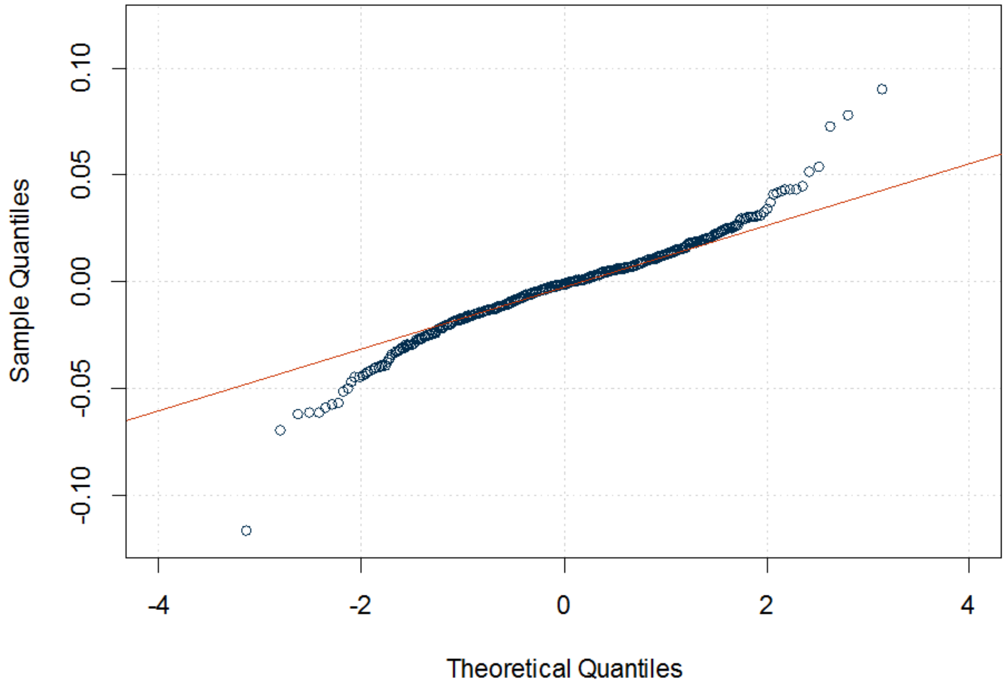
The extreme sample consists of the returns which are two standard deviations away from the mean, and occurs in clusters of three or more over the last two weeks. In addition, up to seven normal returns are included for each three extreme observations, as each 10-day interval is incorporated as its own period. As such, the sample will also consist of some less extreme returns, which could potentially affect the dynamics of the sample as a whole. In general, however, the subsample consists of the absolute most extreme returns, such that this should be reflected in the statistics of the sample and the distribution of the returns, both which will be investigated further in this chapter.

4.2.1 Normality Testing

Up until now, no investigation of the distribution properties of the extreme sample has been performed, but rather the Gaussian distribution has been used as a base in deciding the characteristics of the returns when assigning to subsamples. However, as seen in the share of extreme returns found in the empirical data, compared to the theoretical probabilities of the Gaussian distribution, it seems as if the data does not fit a Gaussian distribution. As such, we will in this chapter further investigate the potential non-normality of the extreme sample and attempt to find an alternative distribution to describe it.

Figure 4.8 is a quantile-quantile(QQ) plot which compares the empirical data of the extreme sample with the theoretical Gaussian distribution of the same observations, through the plotting of sample and theoretical quantiles. Had the returns fitted a Gaussian distribution perfectly, the blue dots (returns) would have been aligned with the red line (theoretical Gaussian distribution). However, the returns seem to only fit a Gaussian distribution for the middle values, approximately the middle 50% of observations, found in the +/- two quantiles of the distribution. When moving outside of the middle 50%, i.e. the tails of the distribution, the data does not resemble a Gaussian distribution, but rather indicates a significant deviation.

Figure 4.8: Q-Q plot extreme sample



As the tails deviates severely from the red line, the plot does seem to indicate the data to fit a Gaussian distribution badly. It also seems as if the tails of the empirical data are asymmetrical as the returns on the left and right side are differently distributed. The plot supports our earlier findings when calculating the share of extreme returns in the complete sample, where the empirical data was found to contain a much higher number of extreme observations than the theoretical probabilities of a Gaussian distribution suggested. In the next sections, further investigations of the statistical characteristics of the extreme sample will be done, in an attempt to find a proper description of the data.

4.2.1.1 Mean

Table 4.6: Mean extreme sample

	Mean
Extreme sample	-0.22%

The daily mean of the extreme sample is -0.22% and thus considerably lower than the corresponding mean of the normal sample. Compared to the normal sample it is approximately fivefold lower. This significant difference is a result of the definition of the extreme sample where three or more returns over the last 10 trading days are required to be +/- two standard deviations from the mean to be included in the sample. However, the fact that the mean is negative indicates that the extreme sample is either heavily negatively skewed or that a few large observations pulls down the mean, or both. On average, sitting with a position in the portfolio of the three indices on the days included in the extreme sample, will lose you money.

4.2.1.2 Standard Deviation

Table 4.7: Standard deviation extreme sample

	Standard deviation
Extreme sample	1.90%

The standard deviation of the returns in the extreme sample is 1.90%, over twice as high as the corresponding figure for the normal sample. The observed standard deviation is a natural result of the definition of the sample, since it is largely based on extreme returns being +/- two standard deviations from the mean. As such, since the extreme sample is known to comprise of larger fluctuations than the normal sample, the standard deviation figures are merely a confirmation of the definition. A high standard deviation also fits well with the intention of the properties of the extreme sample, where the returns are expected to have large fluctuations on a day-to-day basis, on average, which definitely is the case and thus conclusions can be drawn on the properties of the returns through the extreme sample.

Mean and standard deviation are convenient tools to get a brief impression of the data sample and its mean values. However, skewness and kurtosis of the distribution will be tested in order to determine the shape of the distribution and the spread of returns.

4.2.1.3 Skewness

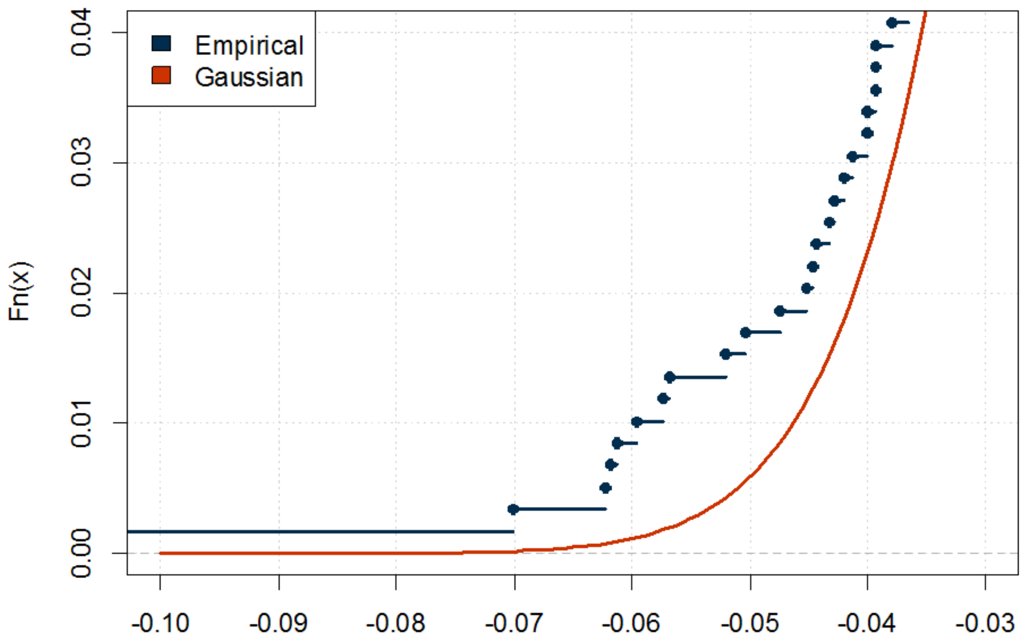
Skewness is measuring the asymmetry of the probability distribution. A skewness of zero shows that the distribution function has the same shape on both sides of the mean.

Table 4.8: Skewness extreme sample

	Skewness
Extreme sample	-0.2808

The extreme sample has a negative skewness, indicating a longer left tail compared to the right tail of the distribution. As such, the shape is more asymmetrical than a Gaussian distribution where the shape is equal on both sides. It is clear that the losses occur with greater frequency than a Gaussian distribution allows. This could be a result of large negative observations in the sample, reducing the average and increasing the size of the left tail. The negative skewness can be observed in figure 4.9.

Figure 4.9: Left-tail cumulative return distribution plot extreme sample



The plot shows the distribution of the empirical dataset side by side with a theoretical Gaussian distribution of the same data, i.e. the Gaussian line is based on the mean and standard deviation of the extreme sample. The negative skewness of the sample is clearly displayed as the returns lie consistently on the left-hand side of the normal line, again, indicating that the extreme observations have a larger probability of being low and negative, with the sample distribution exceeding the predicted values of the Gaussian distribution in the lower quantile. The negative skew, does as a result indicate that the distribution does not follow a Gaussian distribution and an appropriate model for the extreme sample should take the asymmetry of the distribution into consideration.

4.2.1.4 Kurtosis

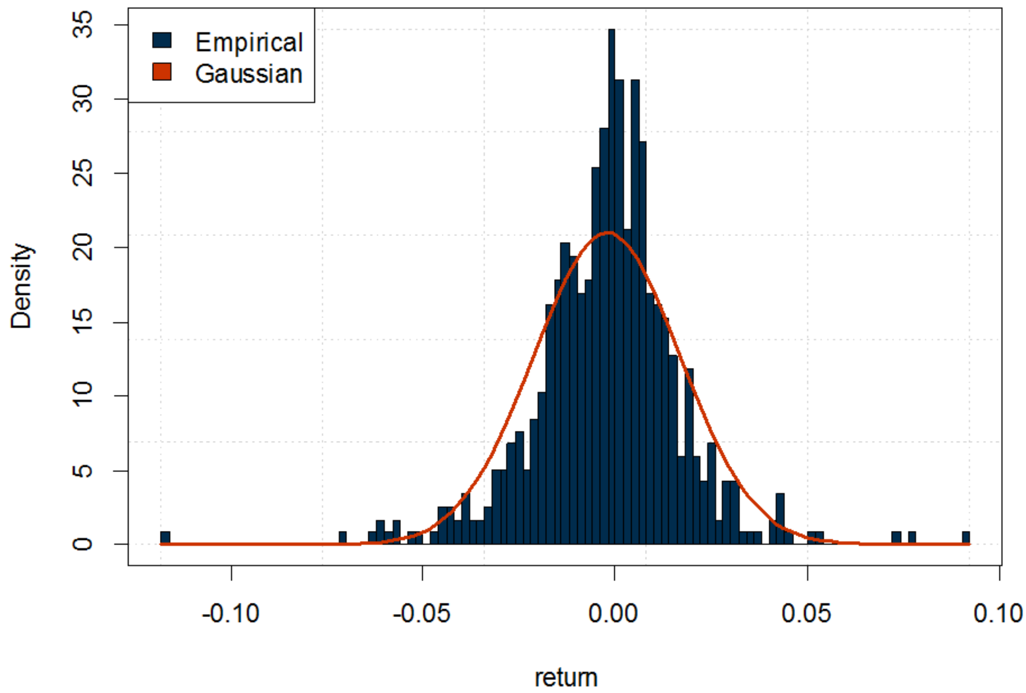
Kurtosis measures the tail size in the distribution, i.e. the likelihood of extreme values.

Table 4.9: Excess kurtosis extreme sample

	Excess kurtosis
Extreme sample	4.3669

The extreme sample is slightly more leptokurtic than the normal sample, implying that the extreme distribution is even more fat-tailed. This suggests a wide spread in the observed extreme returns. In other words, the sample contains a large amount of outliers, which increases the probability of abnormal returns. As such, the distribution is more pointed and less flat than a Gaussian distribution. This can easily be observed in the plot, which displays the distribution of the returns in the extreme sample, compared to a fitted Gaussian distribution on the same data.

Figure 4.10: Distribution plot extreme sample



The frequency of outlier returns in the tail of the distribution is not picked up by a Gaussian distribution, suggesting that a Gaussian shape does not take into account the heavier tails. A consequence of the lack of fit is, in combination with the negative skewness, that other distributions may be more fitting in order to model the data properly.

4.2.1.5 Jarque-Bera Test

A Jarque-Bera test is a goodness of fit test, investigating whether the dataset have kurtosis and skewness corresponding to a Gaussian distribution. So far, no tests have been done to see whether the skewness and kurtosis of the extreme sample are significantly different from a Gaussian distribution. The null hypothesis of the test is as such, whether the sample is distributed as a Gaussian distribution.

$$H_0 : X \sim N(\mu, \sigma^2) \tag{4.5}$$

Table 4.10: Jarque-Bera test extreme sample

Jarque-Bera	Test statistic	P-value
Extreme sample	476.5444	0.0000

The Jarque-Bera test for the extreme sample reject the null hypothesis of normality, which supports our earlier findings of negative skewness and excess kurtosis with statistical significance. As such there might be other distribution models which are more suitable to find better goodness-of-fit.

4.2.1.6 Extreme Value Theory

So far the analyses have shown a bad fit between the extreme sample and a Gaussian distribution. The extreme sample has a negative, large mean, experience large daily fluctuations on average and is negatively skewed. Combined this makes a case for the need of a model capable of modelling the extremeness of the returns clearly present in the sample. Mainly, we are most interested in the left tail events of the extreme sample, which represent the challenge in handling unusually steep losses, as they are by far the most extreme observations in the whole sample. These returns represent the presence of so called tail risk in the sample, namely the risk caused by relatively rare events, but which can have substantial impact on the portfolio. Since an important part of the thesis is the estimation of VaR and ES for non-normal market conditions, finding a distribution that is able to fit and model the tails of the extreme sample is vital to capture the risk found in the extreme observations.

Extreme value theory (EVT) is the study of the tails of distributions.[15] As the tails of the extreme sample contains far more extreme observations than the normal sample, finding accurate estimates of the risk concerned with these is crucial. A portfolio loss during a tail event has the potential to be catastrophic, due to the large size of the returns in the left tail of the extreme sample. EVT is considered a more reliable and robust methodology when estimating the outliers of a return distribution for assets compared to the properties of a Gaussian distribution. It supplies a framework for modelling financial market risks, and extreme losses in particular as it models each tail individually.[16] It does therefore fit especially well with asymmetrical data, as was found to be the case for the extreme sample earlier.

One of the challenges when analysing left-tail events is the insufficiency of data. The extreme sample consists of 590 observations, or approximately 6% of the complete dataset. Of the 590 extreme returns, the middle 50% are returns between -1.2% and 0.7%, meaning that the tails of the sample are found well outside this interval.

[15] See [14]
[16] See [15]

However, the size of the returns is relatively modest, and even when moving out towards the ends of the distribution to the 10% and 90% quantiles respectively, the size of the returns is only -2.4% and 1.8%. This indicates that the real steep losses found in the tails are extremely rare events. For instance, there is only approximately 1.5% losses greater than 5% in the extreme sample, while the corresponding figure on the right side of the distribution, namely gains greater than 5%, are observed 0.8% times in the extreme sample. As such, the availability of data to model is sparse, and modelling events that are as infrequent as the empirical data of the extreme sample shows, is statistically challenging. Therefore, the estimates found need to be significant with little data to analyse. Regardless of the underlying distribution of the returns EVT identifies one for fat-tailed returns, as such it does not require any information or assumption about the actual return distribution. [17]

Another argument to be made for EVT is its ability to offer information regarding the magnitude and probability of potential values being more extreme than those seen previously throughout history. When considering the modelling of extreme events, the distribution should not be constrained by history, but rather use the previous extreme values to offer information about potential worse scenarios. Standard models used to measure risk may, by design, fail to warn of the possibility of a return being significantly worse than the outliers present in the historical data. The steepest daily loss in our portfolio, present in the extreme sample, is approximately 12%. A naive viewpoint would be that this is the worst-case scenario. Therefore, there is need for a methodology which quantifies the potential ‘black swans’ seen in the historical extremes. EVT does, in general, characterise the distribution of values above a given threshold, where the returns found above this threshold is the tail events or extreme observations. As a result, it should be especially well fitted to model the extreme sample.

By using a distribution which takes into account the properties of each tail on its own, it could better describe the challenging properties of the extreme sample. For instance, by using a standard historical VaR measure, as the Basel framework applies, the ’87 crash would cause a large absolute estimate when taking the data of the crash into account. But since a crash like Black Monday occurs rarely, say, every 50-100 years, it will create an upward bias in the subsequent VaR estimates and as a result the consequence will be too conservative risk management. However, when using data

---

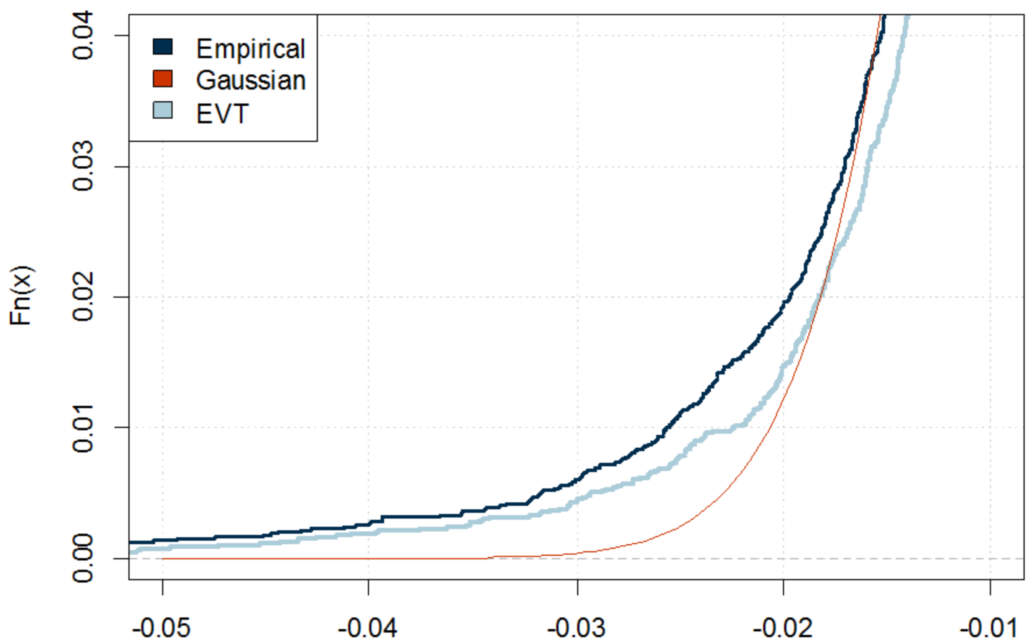
[17] See [16]

from the tail distribution, the probability of such an event will be much smaller and should lead to better estimates.[18]

The Gaussian distribution is shown to underestimate the probability of outlier returns. A conditional heteroskedastic model such as a standard GARCH approach which uses a normal likelihood function to estimate its parameters have bad tail properties. This is a result of the likelihood function weighting values close to zero higher than large values, so that the contribution of the larger values to the likelihood function is relatively small.[19] As most of the observations are in the centre they dominate the estimation. There is need for a model that does the exact opposite of the Gaussian and GARCH setup, since the focus of modelling the extreme sample need to be the tails, and the extremeness of the returns found here.

The plot below displays the empirical cumulative distribution function, for the empirical data, the Gaussian distribution and an EVT model fitted to the dataset. The whole portfolio is used when modelling, to have a large enough sample to display properly.

Figure 4.11: Cumulative left-tail return distribution plot extreme sample



As the whole dataset is used, rather than only the extreme sample, the plot is only indicative of the fit of the model. But, by looking into the left tail of the distribution of

[18] See [14]
[19] See [14]

the portfolio this is where the extreme sample is most present, and as such it is a good indication. The EVT-model is a considerable better fit than the Gaussian distribution, especially when moving left in the tail, and as the returns become larger, the fit seem to increase.

4.2.1.7 Conclusion

Both the mean and standard deviation of the extreme sample was found to be in line with the definition of the sample itself, showcasing that the returns flagged as extreme earlier has a larger, negative mean and high standard deviation. The negative skewness and high kurtosis was confirmed by the Jarque-Bera test, which, significantly, showed that the shape does not fit as a Gaussian distribution. Further, the extreme sample is best described by a distribution which takes into account the asymmetrical tails and the extremeness of the returns, especially considering the tail risk which the returns in the extreme sample contains. As a result, the need for a distribution that does not potentially contribute to an increased upward bias in the estimates of the risk measures is clearly present in the sample. In an attempt to mitigate the difficulties in describing the data an EVT model was proposed as an alternative to the Gaussian distribution.

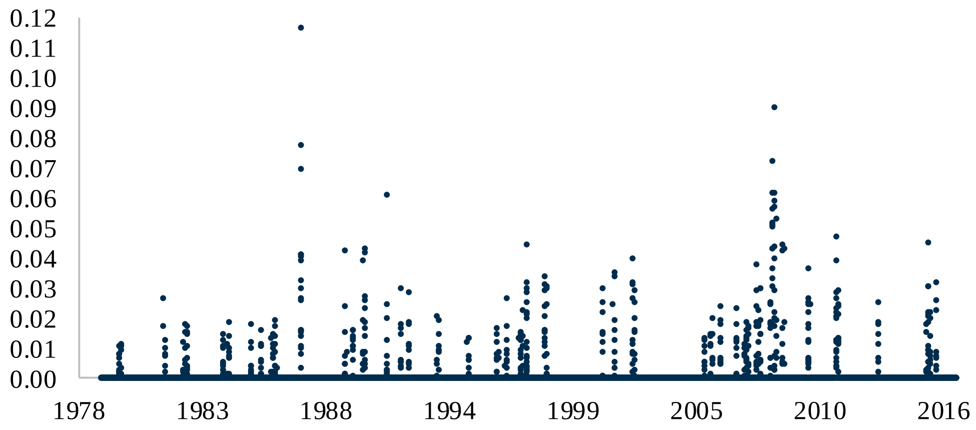
The statistical properties of the extreme sample found so far gives a good indication of its characteristics and distribution over the dataset as a whole, but does not take into account its evolving properties through time. As a feature of the extreme sample is large fluctuations in the returns, often found in clusters, it would be natural to assume it to contain characteristics where some of the observations are dependent on each other. This could potentially affect the measuring of risk in the portfolio, as it would create a bias in an up- or downward direction, and would have to be taken into account. We will therefore investigate the possibility of volatility clustering in the sample.

4.2.2 Volatility Clustering

Volatility clustering is an observation where the volatility is conditional of time, i.e. heteroscedastic, rather than unconditional and constant. Put differently, large moves in returns are often followed by large moves in returns the next days, meaning that the absolute values of the observations are correlated.

Figure 4.12 shows the distribution of the of the returns in the extreme sample, plotted as a function of time. The dots occur as they do in the dataset, i.e. the distance between the dots in the plot is equivalent to their distance in time. Even more so, than for the normal sample, there seems to be a clear presence of the returns being correlated with each other. The vertical movement of the returns over the same period in time indicates that, large fluctuations in the returns of the portfolio, is followed by other large movements in the returns.

Figure 4.12: Absolute returns extreme sample



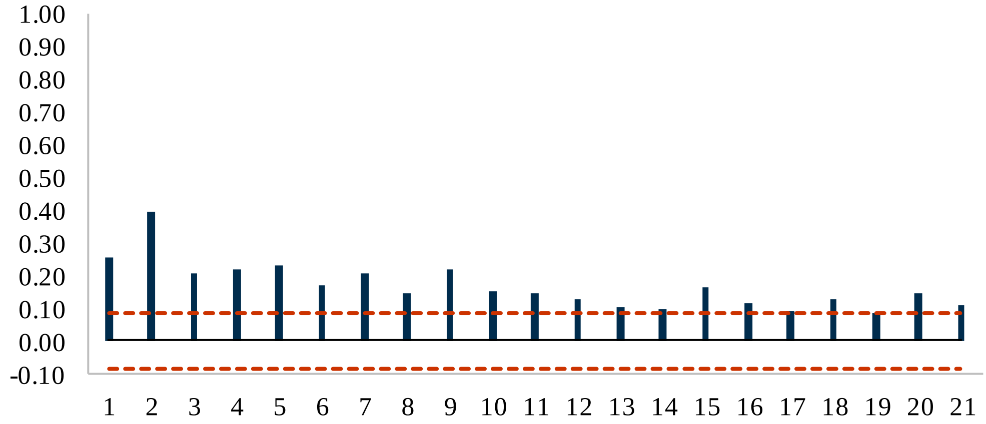
It may seem as if the variance is not constant, but rather conditionally dependent of time, which could mean that there is presence of volatility clustering in the extreme sample. Conditional volatility is not a property of a Gaussian normal distribution, where the volatility is a constant value and not a variable changing throughout the dataset. As such, this is a further verification of the findings in the previous subchapter confirming the non-normality of the extreme sample. In addition, this is an indication that the returns are not iid, as each return does not seem to be mutually independent, but rather dependent on each other through time.

For the estimation of VaR and ES possible volatility clustering present in the subsample should be taken into account, and the graph suggests that volatility is not constant throughout the sample. However, the plot needs to be supplemented with further statistical evidence, which is where autocorrelation can be used.

4.2.2.1 Autocorrelation

Autocorrelation explores further whether the sample has presence of volatility clustering. It measures whether the dataset is linearly dependent of itself at different points in time, using a time lag k. The test is performed using absolute returns, in order to find correlation between large returns with either sign, which in turn indicates volatility clustering.

Figure 4.12: Autocorrelation extreme sample



There seems to be a clear indication of autocorrelation in the extreme sample. The absolute returns are significant until the 21st lag, indicating a clear correlation between the returns, which means that the returns are conditionally dependent of time. This is a violation of the assumption of the Gaussian distribution where each observation must be mutually independent. The plot also shows that the correlation between the returns are larger during the extreme periods than for the normal observations, with autocorrelation coefficients of absolute returns being between 0.1 and 0.4, compared to coefficients of a maximum of approximately 0.2 for the corresponding figures in the normal sample. This suggests a higher dependence in the returns of the extreme sample. The dependence could be a result of the larger fluctuations in the subsample, which in turn means that the large changes observed tend to be followed by large changes of either sign.

Testing for autocorrelation is a test of each correlation coefficient respectively, but to confirm its applicability for the whole extreme sample, a joint test will have to be performed. As the correlation coefficients above seem to be volatile and only slightly above the confidence interval of the sample, testing all coefficients together is needed, to be able to draw a definite conclusion.

4.2.2.2 Ljung-Box Test

Table 4.11: Ljung-Box test extreme sample

Ljung-Box	Test statistic	P-value
Extreme sample	60.2120	0.0000

The test statistic is higher than the critical value for the portfolio, meaning that the null hypothesis should be rejected. The conclusion is that hypothesis of the dataset containing zero autocorrelation in the time series, is rejected. As a result, presence of autocorrelation in the extreme sample is confirmed, and thus also the previous indications of volatility clustering.

4.2.2.3 Conclusion

There is found presence of volatility clustering in the extreme sample, confirmed by a significant presence of autocorrelation. This has a clear influence on how to approach the estimation of VaR and ES later in chapter 5, as the lack of mutually independent returns needs to be taken into consideration. As a result, observations which are dependent on each other could create a potential upward bias, for instance, with too high VaR estimates as a result of the clustering of tail returns.

4.3 Concluding Remarks

Based on the statistical analysis in this chapter it is clear that normality is rejected for both the normal and extreme sample. The lack of normality helps explain the results from chapter 3, where the Gaussian Basel model's VaR and ES estimates were worse than expected. Instead, student's t-distribution and EVT is found to better describe the data found in the normal and extreme samples respectively. Further, heteroscedasticity is detected in both subsamples, implying that volatility is conditional of time, rather than constant. To model the risk more accurately, a time dependent volatility model should be considered.

In the next chapter, student's t-distribution will be fitted and applied for normal market periods, while the extreme value theory will be adjusted to fit extreme market periods. The use of these distribution models should intentionally increase the likelihood of predicting tail events. In addition, a GARCH volatility model will be applied on both samples, in order to address the issue of the slow reacting Basel model. The conditional volatility model will be more adaptable and incorporate new market information faster by weighting recent returns to a greater extent.

.

Chapter 5

5 Models & Distributions

The Basel model has proved to estimate risk insufficiently for both the normal and extreme sample, providing inaccurate estimates for VaR and ES in chapter 3. The analysis of the fit of a Gaussian distribution in chapter 4 concluded to reject the hypothesis of normality for both the normal and extreme sample. As an alternative, student's t-distribution was proposed for the normal sample, while extreme value theory (EVT) was suggested as appropriate for the extreme sample. In addition, presence of volatility clustering has been detected in both subsamples, indicating that a volatility model should be combined with the distribution models in order to explain returns adequately.

The normal sample will be explained using a student's t-distribution in order to find a good fit with the data. Second, EVT will be adjusted and applied to the extreme sample. The distribution models for both subsamples will be supplemented with a GARCH volatility model, intended to address the issue of volatility clustering found in the returns.

5.1 The Normal Sample

In this subchapter an appropriate distribution and volatility model, namely student's t-distribution and GARCH, will be customised to fit the normal sample. Student's t-distribution is proposed as a suitable distribution due to the property of allowing for heavier tails. Moreover, the normal sample was found to contain volatility clustering, indicating that variance is not constant. This heteroscedasticity is a violation of iid error terms and constant variance assumed by student's t-distribution. A GARCH volatility model will be implemented as an attempt to better explain the variance of the normal sample. Combining the student's t-distribution with GARCH a student's t-GARCH is derived. This model allows the distribution to have both heavier tails and conditional variance throughout the time period. The t-GARCH will be applied on the normal sample, in order to see how well it fits the dataset. To test the performance, VaR and ES will be used as indicators, making us able to compare the t-GARCH to the benchmark Basel framework from chapter 3.

5.1.1 Student's t-distribution

Students t-distribution will allow for longer tails in order to capture the excess kurtosis. More specifically, it will allow for an adjustment to the skewness of the sample by using different degrees of freedom on each side of the distribution. The fewer degrees of freedom the more the distribution allows for heavier tails and more extreme observations when compared to a Gaussian distribution. As analysed in section 4.1.1.3, the distribution of the normal sample is left-skewed, indicating an asymmetric distribution shape with a longer left tail. Because of this, less degrees of freedom will be applied to the left tail or the loss side, than for the right tail or gain side. The best fit for the normal sample is found by giving the left tail 20 degrees of freedom, while the right tail is found to be best described by 100 degrees of freedom.

5.1.2 GARCH

When finding an appropriate distribution for the normal sample, it is, as found earlier, crucial to take volatility clustering into account. Without using a volatility model which adjusts for autocorrelation in the returns, the likelihood of underestimating the risk increases. The volatility needs to be conditionally distributed, rather than unconditional and constant throughout time. The most common model when dealing these kind of problems in time series, is GARCH – generalised autoregressive conditional heteroscedasticity model.

A GARCH model makes assumptions on the conditional distribution, meaning it assumes a dependence through time via a feedback in the volatility. Both conditional and unconditional variance are measures of volatility, but the time dependency is what distinguish GARCH-modelling from models assuming homoscedasticity. The unconditional variance does not account for the current market trends, but is rather an equally weighted average measure of variance over a large sample. The conditional variance on the other hand, incorporates market movements up until the point where the estimation is done, meaning such a variance measure will be superior when dealing with heteroskedastic returns. More recent market movements are given much greater impact than less recent ones. Here the variances of the error terms will not be equally distributed, but rather reasonably expected to be larger for some points of data than for others. As seen earlier, in subchapter 4.1.2, the assumption of normal market returns being iid was violated due to the autocorrelation and heteroskedastic properties of the returns, therefore the use of conditional variance should provide better estimates.

5.1.2.1 Formula

A standard GARCH(*p*, *q*)-model uses lag-parameters *p* and *q* in order to determine how long memory the volatility model should have. The memory is a measure of how much history, i.e. how many days, of error terms are included into predictions of volatility. When *p* and *q* are given higher values, more days of historical error terms are included in the model.

$$GARCH(p,q): \quad \sigma_t^2 = \omega + \beta(L)\sigma_t^2 + \alpha(L)\eta_{t+1}^2 \tag{5.1}$$

where $\alpha(L)$ is the lag-operator of the *p*th order, while $\beta(L)$ is the lag-polynomial of the qth order. The first value *p* refers to how many autoregressive lags (ARCH) terms appears in the equation, while the second value *q* refers to how many moving average (MA) lags are specified. Lag parameters *p=1* and *q=1* is found to be most suitable for the normal sample[20] and by implementing this a GARCH(*p*, *q*) could be made into a GARCH(1, 1).

$$GARCH(1,1): \quad \sigma_t^2 = \omega + \beta\sigma_{t-1}^2 + \alpha\eta_t^2 \tag{5.2}$$

Where ω is a constant determining the level of the long-time daily variance of the time series. β is a constant quantifying the persistence of the volatility at previous time points. σ_{t-1}^2 is the variance at the previous point in time, *t-1*. $\propto$ is a constant representing the sensitivity of the lag return error term. η_t^2 is the squared difference between the expected return and the actual return of the portfolio, also known as residuals.

5.1.2.2 Parameters

The GARCH-parameters are estimated with maximum likelihood estimation to give the best fit for the distribution of the normal return sample.

Table 5.1: GARCH (1, 1) parameters normal sample

GARCH(1, 1)	ω	α	β
Parameters	4.67E-07	6.32E-02	9.30E-01

If everything else is held equal, in this case the error terms of expected returns and volatility both being zero, the variance will converge to the value of ω throughout time. The value of ω is positive, such that the variance is always positive. β has a value close to one, indicating that high volatility in the foregoing time point will result

[20] Appendix table 1: Goodness-of-fit test run combining different *p* and *q*. GARCH(1, 1) provided highest test results.

in high volatility at the current point in time. In other words, the variance will be highly dependent of its preceding values and thus market shocks integrated in future estimations. The size of α determines the reaction of misestimating returns, thus a large α makes the variance highly sensitive to unexpected returns in the market. The sum of α and β is a vital statistic in GARCH-modelling. Together they display how long fluctuations in returns are incorporated in future estimations of volatility, i.e. how fast volatility decays after-market shocks.

To see graphically how the GARCH(1, 1) performs when applying the parameters above, it is possible to compare the historical absolute returns of the normal sample with the estimated volatility by GARCH(1, 1). Ideally, the volatility should increase over the same periods as when the absolute values of market returns increase.

Figure 5.1: Absolute returns normal sample

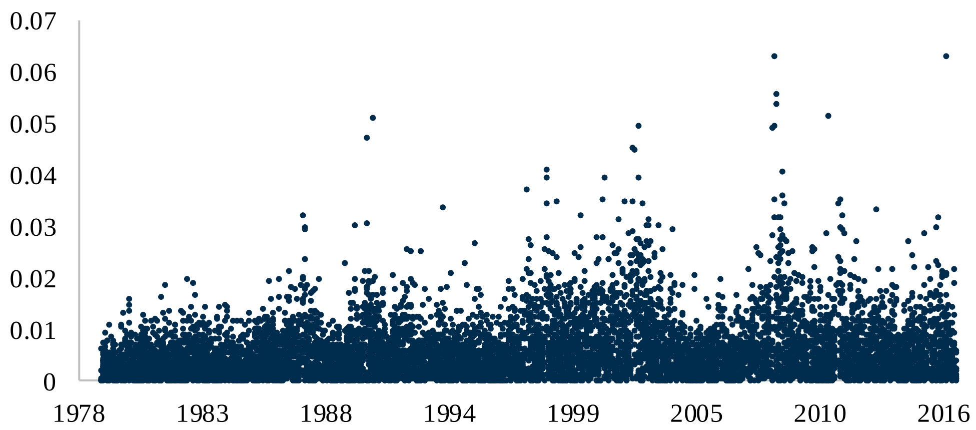
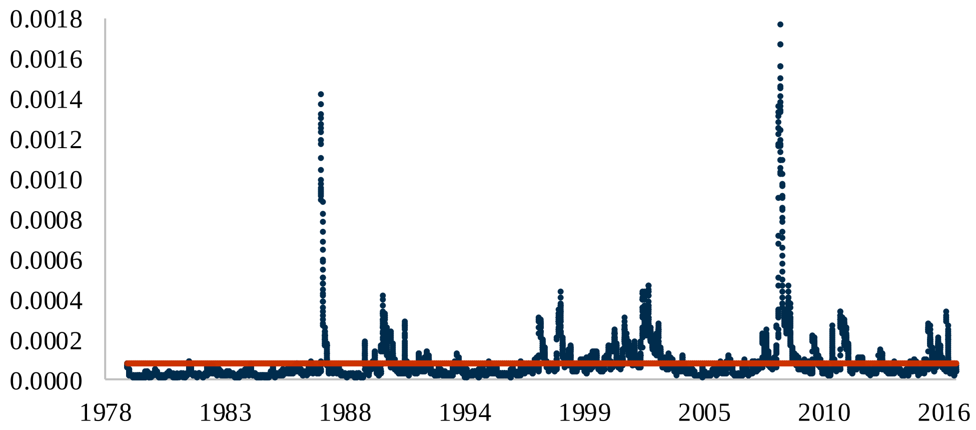


Figure 5.2: GARCH estimated volatility normal sample

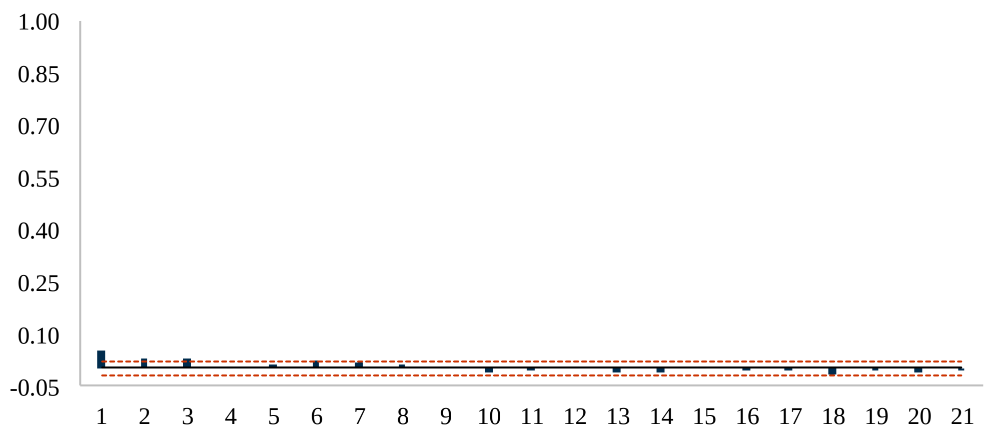


The two graphs above have very much the same trends. When absolute returns increases in figure 5.1, the estimated conditional variance (figure 5.2) for the same point in time displays a similar trend. In comparison, finding volatility based on maximum likelihood estimation would give a constant variance (red line), not adjusting itself to the current state of the market. As seen in the charts, it looks as if sudden changes in portfolio returns are successfully incorporated into the GARCH(1, 1) conditional volatility modelling, meaning that heteroscedasticity is effectively dealt with.

5.1.2.3 Error Terms

In subchapter 4.1.2, an analysis of autocorrelation was performed, proving that the error terms of the returns were not iid. Hence, they do not have constant variance, and does therefore violate important assumptions for a Gaussian distribution. However, GARCH has a similar assumption of iid. More specifically, the error term η should be strict white noise, meaning that it has a mean of zero, but vary around the mean of the error terms with a conditional standard deviation σ_t. Since the error terms, also known as residuals, are assumed to be white noise there should no autocorrelation present. The variance itself should not be constant, but the error terms should not be serially correlated when distributed with a conditional volatility. To test whether this assumption holds, another plot of the autocorrelation will be performed. In order to compare the error term residuals to one another, they are standardised by dividing with a standard deviation. More technically; η_t, the residual at time t, is divided by σ_{t-1}, the conditional standard deviation at time *t-1*.

Figure 5.3: Autocorrelation error terms normal sample



As displayed above there is little or no presence of autocorrelation in the conditionally standardised error terms. For lags between one and three there is a weak autocorrelation. However, for lags above this there is no presence of autocorrelation. Compared to the original autocorrelation of the normal sample, all coefficients are considerably reduced. This indicates that the implementation of GARCH has to a great extent removed serial correlation and thus improved the explanation power of the volatility dynamics in the normal sample.

5.1.2.4 Drawbacks

As for any statistical model, GARCH is based on historical numbers and will not be able to anticipate the future without flaws. A large movement in returns will not be incorporated in the model before the next estimation of volatility and thus it will react one day too late in order to successfully protect a portfolio manager from taking on too much risk

The benefit of GARCH could also be one of its main disadvantages. The GARCH-framework should intentionally react quickly to market changes and incorporate these changes when estimating future values of volatility. As such, estimates might change quickly and the risk measures as a result could end up being substantially higher. On the upside this quickly prepares a portfolio manager for higher risk, meaning that reduction of risk exposure is needed. However, the downside is that such frequent actions of composing the portfolio comes with lower potential upside and gain, in addition to higher transaction costs. Also, portfolio managers might argue that the rapid changes in risk estimates is only due to a temporary volatility spike and that they should hold their portfolio in order to stay profitable. These conflicting opinions and interests challenges the use of GARCH-models in the real world. In contrast, the Gaussian Basel model estimates of VaR and ES are more stable and has a less responsive evolvement. Information is incorporated slowly, making it ascend more carefully. A portfolio manager would rarely need to make significant changes in the composition of the portfolio, implying less trading costs. However, the long memory of the method could potentially force a portfolio manager to comply with information no longer relevant for the current state of the market. This property of Basel could lead to risk limits being unnecessary high, and therefore reducing the efficiency and profitability of the portfolio.

Even though the practical implementation of GARCH could cause some challenges, these are necessary consequences. A portfolio manager cannot have it both ways, with a quickly adapting risk model, and infrequent changes in risk exposure. As such, we believe that a GARCH model is a more suitable model to estimate tail risk during normal market states. In addition, the adaptable properties provide a portfolio manager with accurate risk estimates, capping the potential upside as little as possible.

5.1.2.5 Conclusion

The implementation of a GARCH(1, 1) volatility model has successfully removed the presence of autocorrelation in the normal sample. This indicates that the model is explaining dynamics of volatility and incorporating information at a sufficient level, quickly adapting to new market states.

The next step in the process of explaining the dynamics of normal market conditions, is testing the performance of the modelling done so far. In order to see whether the modelling is improving risk estimates compared to the Gaussian distribution procedure used in the Basel framework, VaR and ES will be applied with the t-GARCH model.

5.1.3 Value-at-Risk

A student's t-GARCH(1, 1) model is run on the normal sample, calculating one-day 99% VaR.

Table 5.2: VaR violation rates using student's t-GARCH(1, 1) [21]

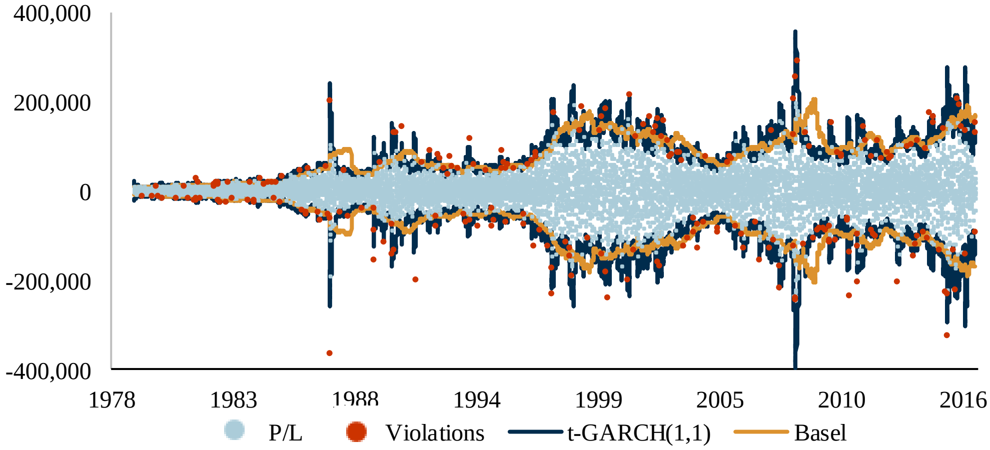
	Normal sample	
	VaR-	VaR+
Violations	93	89
Violation rate	1.00%	0.96%
Sample size	9310	9310

As displayed in table 5.2, the violation rates for VaR during normal market conditions are very accurate, at a level close to the target rate of 1%. VaR- is spot on 1%, while VaR+ is very slightly overestimating the risk. It seems as if the model succeeds in estimating an acceptable number of violations for the dataset under normal market conditions. Despite the degrees of freedom being fitted for this subsample in particular, it is possible to see a weak trend of the distribution skewing left, as we have more violations of VaR-, than VaR+. Considering the risk management perspective of the thesis, it is more important that VaR- is modelled correctly than VaR+, due to the more severe consequences of exceeding VaR-. If VaR- is exceeded, a portfolio could potentially experience severe losses. As seen in chapter 3, Basel underestimated the risk for both VaR- and VaR+ during normal market states. By comparing the violation rates of the two approaches, student's t-GARCH(1, 1) is found to be more accurate, which could be due to the heavier tails of the student's t-distribution.

[21] Appendix table 2: Student's t-GARCH VaR results for complete and extreme sample.

The second drawback of using the Basel model was the incorporation of new market information, which resulted in overestimated risk during less volatile market periods, and oppositely underestimated during highly volatile periods. To see how student's t-GARCH(1, 1) incorporates new market information, it will be graphically compared to the Basel VaR model in figure 5.4.

Figure 5.4: Value-at-risk t-GARCH and Basel



The GARCH model effectively adjusts and incorporates new market information. When market shocks occur, it seems to be rapidly taken into consideration by the volatility modelling performed by GARCH. Intentionally, the VaR should be as low as possible, but still estimate risk levels at a 1% violation rate. A low VaR, means lower risk exposure and thus higher potential upside for a portfolio manager. By having an effective, adaptable VaR, it is possible to react quickly to the market and increase exposure when markets cool down, and oppositely reduce risk exposure rapidly when markets become more volatile and extreme. This seems to be the case for the GARCH model, in particular when compared to the Basel VaR, thus we believe it is more flexible and suitable for a portfolio manager to apply. Next, we will analyse ES, in order to further discover the performance of the student's t-GARCH-model.

5.1.4 Expected Shortfall

ES will be used as a performance indicator for how tail events are predicted by student's t-GARCH. ES is an extension of VaR and thus provides information, not only about the threshold value of the 1% tail event, but also regarding what happens beyond this threshold. This will provide more information about the consequences of violating VaR, and as such be a supplement to the previously performed analysis.

5.1.4.1 Violation Analysis

Table 5.3: ES violation rates using student's t-GARCH(1, 1) [22]

	Normal sample	
	ES-	ES+
Violations	27	35
Violation rate	0.29%	0.38%
Sample size	9310	9310

The number of losses violating ES during normal market states is approximately halved, compared to Basel ES. The number of violations is low, indicating that few gains and losses are exceeding the estimated expected shortfalls. However, as discussed previously, there is no target violation rate for ES, as is the case for VaR. By looking at violation rates only, it is not possible to directly determine whether the ES estimates in general are too high, or if the conditional variance makes ES successfully adaptable as intended.

5.1.4.2 Loss Analysis

To further investigate the few violations found in the ES model, total and average loss in the case of violation are displayed below. The figures are only indicative of the downside of the model, as they do not account for any potential upside. As the perspective of the thesis is from the viewpoint of a portfolio manager, consequences of violating the left tail of losses will be analysed.

Table 5.4: Loss analysis using student's t-GARCH (1,1)

ES-	Normal sample
Total loss	-795,512
Average loss	-29,463

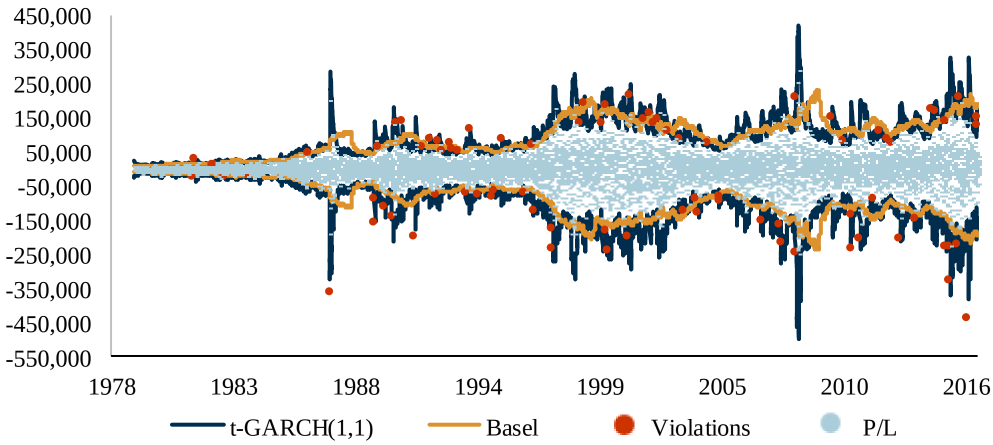
The average loss for ES- is slightly larger than the corresponding figure for the Basel ES model presented earlier in section 3.2.4.2. The total loss is significantly reduced, and is almost halved compared to the Basel model. The average loss is higher because the number of violations are far lower than for the Basel model, implying a lower numerator and thus higher average loss. Hence, the losses are rare, but when occurring they are more severe. As such, the conclusion of a significant improvement compared to the Basel model is supported by the loss analysis.

[22] Appendix table 3: Student's t-GARCH ES results for complete and extreme sample.

5.1.4.3 Graphical Analysis

Ideally, the model should be adaptable in such a way that it yields low ES estimates under stable market conditions, but high estimates under risky market conditions. This would make ES flexible throughout the time period, forcing the portfolio manager to reduce exposure only when the actual risk increases. To investigate further, the ES student's t-GARCH model will be compared to the Basel ES model, displayed with the P/L's of the portfolio.

Figure 5.5: Expected shortfall t-GARCH and Basel



As the plot shows, there are significant differences in the dynamics of the two measures. The contrasting dynamics have implications for a portfolio manager, which could be required to comply with risk limits given by ES. ES derived by a student's t-GARCH (1,1) is far more sensitive with respect to the return movements. In periods where returns are more volatile, ES reacts quickly and incorporates the risk by increasing the confidence level swiftly. Moreover, when the volatility decreases, it quickly lowers the confidence level. In this way a portfolio manager has a more dynamic risk measure, highly adaptable and adjusting for the state of the market.

5.1.5 Conclusion

For the normal sample, the student's t-GARCH (1,1) is significantly improving risk estimates, displaying lower violation rates for both VaR and ES than the corresponding figures found for the Gaussian Basel model. The GARCH model handles the volatility clustering, while the student's t-distribution allows for returns to be more leptokurtic, i.e. have fatter tails and experience more extreme outcomes. The conditional volatility implemented by GARCH makes it a flexible measure which adapts quickly and thus prepares a portfolio manager for both higher and lower risk in the market. This property is confirmed by the loss analysis, which shows that total losses during normal market periods is nearly cut in half. In the next chapter we will seek to find the best approach to risk management during extreme market conditions.

5.2 The Extreme Sample

The extreme sample contains the returns that differs the most, measured in standard deviation, from the mean. While differing the most from the mean, they are at the same time occurring so often, three times or more, over a two-week period, that the theoretical probability is very low, as found in chapter 2, around 1%. Yet, as seen before, the extreme returns comprise of 6% of the total return sample. While this is still a low percentage of the complete dataset, accurate modelling of the outer 50% of the extreme sample on both sides of the distribution is inconvenient, as these are even rarer tail observations. The returns found here are infrequent, fluctuations large and the losses, as a result, steep, indicating a high risk of large losses on the portfolio. As a consequence, Extreme value theory (EVT) was introduced in chapter 4, as a model that could provide good estimates with little data available to analyse and a high degree of extremeness found in the returns. One central part of the extreme value theory is the peak-over-threshold (POT) model.

5.2.1 POT-model

There are, in general, two main models under EVT; the block maxima and the peak-over-threshold (POT) model. Both deal with the extreme deviations from the median of the probability distribution.[23] As the block maxima approach requires a too large sample for accurate estimates compared to the size of the extreme sample, the POT-model will be applied. The peak-over-threshold approach is also generally preferred because of its practical functionality.[24] There are two approaches to POT; fully parametric models (e.g. the generalized Pareto distribution); and semi-parametric models (e.g. the Hill estimator). A fully-parametric model which uses the generalised Pareto distribution (GPD) is, as the name suggests, the most general of the two approaches. It can be applied to a wide variety of data and does not require prior knowledge of the tails, in contrast to the Hill estimator, which is only valid for a certain type of distributed data. As a result, we will use a POT-model with GPD fitted to the empirical tail distribution.

[23] See [17]
[24] See [16]

An underlying assumption of the POT model is that returns over a given threshold are iid. As investigated in the previous analyses of the extreme sample this is problematic, as it was showed that this was not necessarily the case. A result of this assumption is that the model does not take into account current market volatility, but rather only takes a given α, i.e. a quantile value, and the current portfolio value into account when estimating VaR and ES. As the return sample is relatively dispersing and unevenly distributed, combined with the large, rapid fluctuations that tend to occur during the extreme market conditions, this could quickly become a limitation of the POT-model. We therefore choose to tweak the approach of the POT-model, to be able to take the heteroscedasticity into account, with a conditional POT (c-POT) model.

5.2.2 Conditional POT-model

The conditional POT (c-POT) model integrates extreme value theory and GARCH-modelling. The hybrid combines dynamic volatility, i.e. volatility clustering, and non-normality with the presence of fat tails in the return distribution. In the conditional POT model, GARCH is used to fit the return data with maximum likelihood in order to estimate the current conditional volatility. The GPD is approximated to model the tail of the innovations, i.e. the standardised residuals, of the GARCH model. Since the daily returns, as shown earlier, exhibit both autocorrelation and heteroscedasticity, the implementation of time-varying volatility into the calculations is needed. Therefore, using a dynamic volatility model (GARCH), which incorporates volatility clustering, should make the error terms iid. Then, the assumption of normal standard distribution in the GARCH model is replaced by applying EVT, in the form of GPD, to the noise variable, which is suited for modeling the tails without assuming any specific shape of the distribution. As such, conditional VaR forecasts could be derived.

5.2.2.1 Defining the Model

The GARCH(1, 1) setup will be implemented, where a return at time *i*, X_i is described as:

$$X_i = \mu + \epsilon_i = \mu + \sigma_i Z_i \tag{5.3}$$

And where we have:

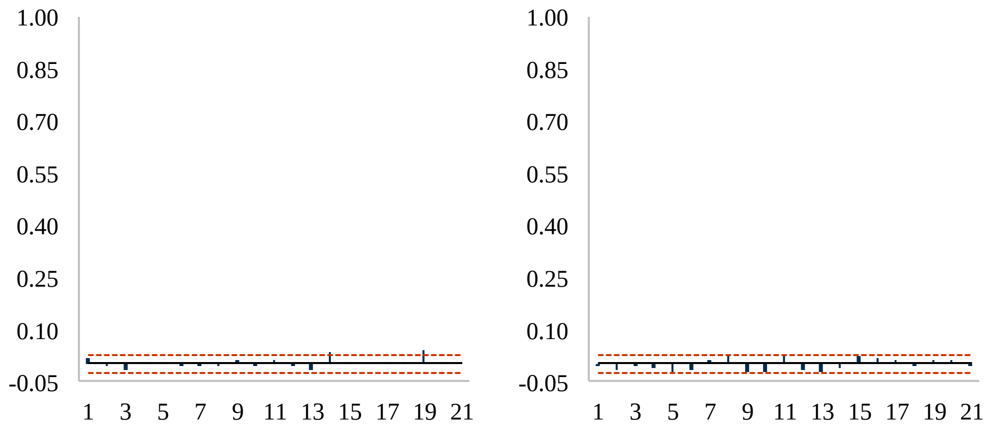
$$\sigma_i^2 = \omega + \beta\sigma_{i-1}^2 + \alpha\epsilon_{i-1}^2 \tag{5.4}$$

The expected return is μ and the volatility of the returns at time *i* is σ_i. So that, the conditional volatility today is dependent on yesterday's innovations, i.e. $\epsilon_i = X_i - \mu_i$, yesterday's conditional volatility and the unconditional volatility ω. The innovations ϵ_t are defined as the error term between the estimated return and the realised return. The standardised residuals or innovations, are then:

$$Z_i = \epsilon_i / \sigma_i \tag{5.5}$$

The innovations Z_i are, in the conventional GARCH model, assumed to be iid, independently and identically distributed, and to follow a Gaussian distribution. As the innovations must be independent of each other, the validity of the model will be tested by investigating whether there is autocorrelation in the innovations for various lags. Each side of the distribution is show below, with the negative innovations on the left side.

Figure 5.6: Autocorrelation innovations extreme sample



There is no significant tendency of autocorrelation on either side, as the coefficients of almost all lags are below the dotted lines, i.e. within the confidence interval of significance. As such the assumption must be considered to be fulfilled. Then, by minimising the ratio of error in the model, i.e. the innovations, the GARCH model with normal innovations is fitted with a maximum likelihood approach.

In the next sections, the standardised residuals are extracted, so that EVT can be applied to the residuals Z_i to model the tail behaviour.

5.2.2.2 Threshold Determination

When implementing EVT, i.e. the conditional POT model, a threshold value for the innovations has to be determined, as the c-POT is only applied to the tails of the distribution. Moving into the centre of the distribution the c-POT becomes increasingly inaccurate. There is, however, no universal rule telling when it becomes inaccurate, because this depends on the underlying distribution of the data. The most common approach when determining the optimal threshold, is the eyeball method.[25] Here, the focus is to find a region where the tail index seems to be stable. Each tail must be estimated individually since it was found that the empirical distribution is asymmetrical. From this point on, the tails are investigated separately, by first estimating the left side, with only negative returns, then the right side, using only the positive returns. Both sides are displayed below as absolute values.

Figure 5.7: QQ plot innovations extreme sample

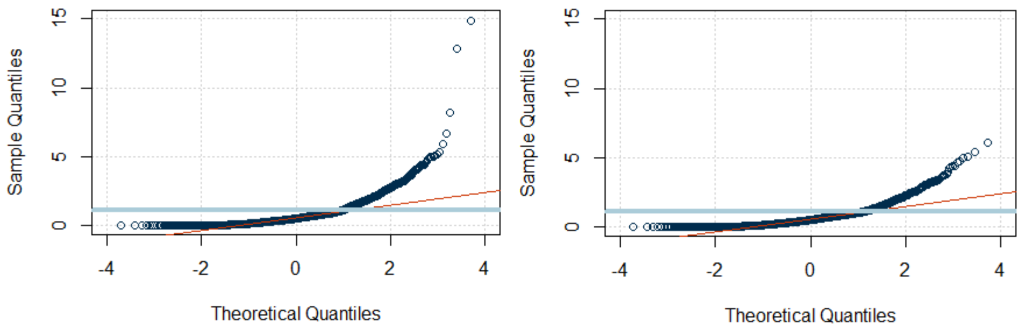


Figure 5.7 shows the Q-Q plots for the negative and positive innovations respectively. The plots compare the sample quantiles of the innovations with the theoretical quantiles of the data normally distributed. If the innovations (blue circles) were distributed normally they would align with the red line, meaning that the greater the difference between the two, the larger is the evidence for different distributions. While close to true for the centre of the distribution, the tails of the innovations show significant deviation. For c-POT modelling, the point where the deviation begins is that of interest, as we seek to find the point that appears to separate the tails from the rest of the distribution. The sky blue lines indicate the values where the deviation begins. The threshold for the negative innovations side is therefore determined to be, u_{-} = 1.10, while the threshold for the positive side is determined to be, u_{+} = 1.15. As a results, all innovation values of the respective sides of the extreme sample larger than the two thresholds will be used when modelling c-POT with a GPD.

5.2.2.3 Generalised Pareto Distribution

In this section, all values larger than the thresholds found previously, will be considered. The difference between these values and the thresholds are called

[25] See [16]

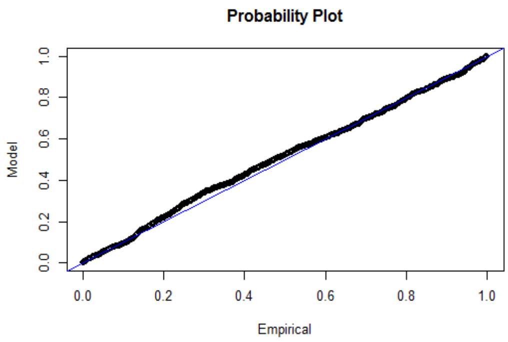
exceedances over the threshold. These exceedances are assumed to have a generalised Pareto distribution. The GPD has a distribution function defined by [26]:

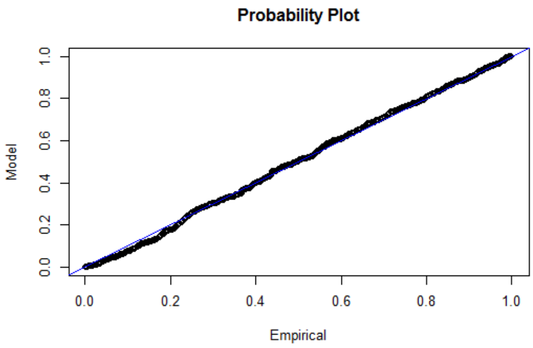
$$G_{\xi,\beta(x)} = \begin{cases} 1-\left(1+\frac{\xi x}{\beta}\right)^{-\frac{1}{\xi}} & if\ \xi \neq 0 \\ 1-\exp\left(-\frac{x}{\beta}\right) & if\ \xi = 0 \end{cases} \tag{5.6}$$

Where the parameters of the distribution are ξ and β, the shape and scale parameters respectively. A maximum likelihood estimate is applied on both sides of the distribution to find the optimal shape and scale parameters to describe the positive and negative innovations. [27]

To investigate the match of the GPD with the extreme sample, a probability plot will be applied on both sides of the distribution, where the negative (left hand plot) and positive (right hand plot) innovations are fitted to the GPD, using maximum likelihood estimation of the distribution function as described above.

Figure 5.8: Probability plot extreme sample





The plots show how the data in the extreme sample follows the generalised Pareto distribution. The straight line displays the theoretical distribution of the GPD, and is estimated on basis of the given scale and shape parameters. Departures from the straight line indicate departures from the GPD. As such, the correlation coefficient between the theoretical distribution and the empirical data, i.e. the coefficient linked to the linear fit of the data, is a measure of the goodness of fit between the two. Both sides of the distribution seem to closely follow the GPD with small deviations from the straight line, which suggests that the GPD is able to explain the distribution of the innovations accurately.

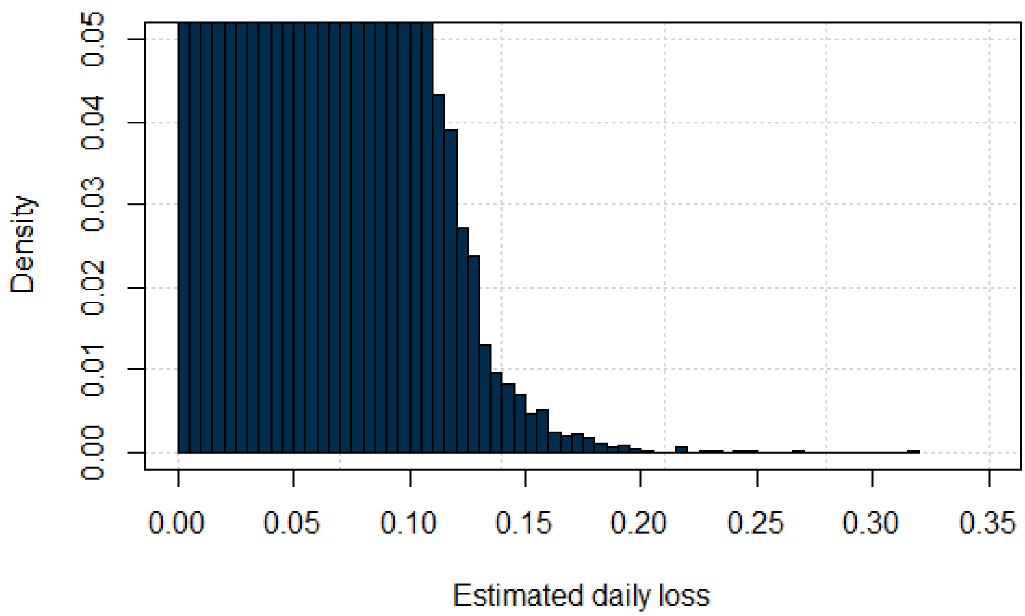
[26] See [18]
[27] Appendix table 4: The GPD parameters

The c-POT model is a bit of a black box, in terms of the results of its distribution modelling. It is difficult to, based on the above, know the accuracy and correctness of the model in describing the extreme sample without first estimating VaR and ES. We will therefore, as a complementary test to the risk measures, first look further into the tail possibilities of the model described above.

5.2.2.4 Investigating the Tail Possibilities

The conditional POT-model does in theory fit well to the extremeness of the returns in the subsample based on the above findings. However, this could be supplemented before moving into risk measure estimation. As the extreme sample consists of only 590 observations, a Monte Carlo simulation is performed in order to investigate the tail possibilities of the subsample further. The simulation is performed by drawing one million daily returns based on the EVT modelling above, i.e. by using the parameters found to describe the distribution. As the left side of the extreme sample is of particular interest for any risk aware investor, we will focus on the loss side in the simulation. The simulated left-tail density histogram is displayed, in absolute terms, below.

Figure 5.9: Left-tail generated extreme sample



The plot, displaying the Monte-Carlo simulations of the left tail of the extreme sample, suggests that the EVT model manages to capture the negative skewness, and extremeness of the data. The potential of large losses is higher compared to what the dataset from 1978-2016 implies. The steepest loss in our portfolio was approximately 12%, while, by contrast, the EVT-based simulation suggests that the worst daily loss on the portfolio could be over 20%. However, most of the simulated losses max out at around 15% and the probability of seeing anything larger is extremely low. And, as discussed earlier, in subchapter 4.2, the extreme modelling should not necessary be limited by the most extreme events observed in our dataset, but rather offer information regarding the magnitude and probability of potential values being more extreme than those seen previously throughout history. As such, the previous extreme values should offer information about potential worse scenarios. Therefore, it seems as if the EVT successfully takes into account and quantifies the potential 'black swans' seen in the historical extremes.

To further investigate the accuracy of the c-POT model, VaR and ES will be run through the dataset in the next two sections, and compared to the Basel figures found earlier in the thesis.

5.2.3 Value-at-Risk

The value-at-risk is estimated using [28]:

$$VaR_{\alpha}^{i} = \mu_{i+1} + \sigma_{i+1} q_{\alpha}(Z) \tag{5.7}$$

Where Z is the GARCH (1, 1) fitted innovations. The α quantile of the innovations are scaled with the estimated σ, such that heteroscedasticity is taken into account. Then the β and ξ are found from fitting the GPD over the innovations, so that the quantile can be estimated by using:

$$q_{\alpha} = u + \frac{\beta}{\xi}\left(\left(\frac{1-\alpha}{F(U)}^{-\xi}\right) - 1\right) \tag{5.8}$$

This results in the following VaR estimates:

Table 5.5: VaR violation rates using c-POT [29]

	Extreme sample	
	VaR-	VaR+
Violations	9	5
Violation rate	1.53%	0.85%
Sample size	590	590

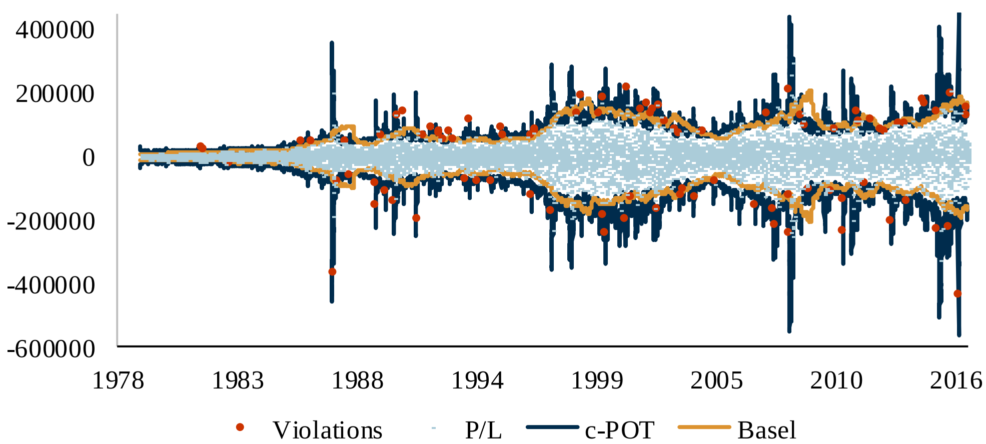
On the loss side, the model now tends to underestimate the risk, meaning that the violation rate is higher than the 1% rate the 99% confidence interval suggests. A

[28] See [19]

[29] Appendix table 5: Conditional-POT VaR results for complete and normal sample.

further investigation into the nine violations shows that eight out of the nine on the loss side are significantly worse than the corresponding VaR, meaning that the loss at each point in time is far greater than the VaR level. Which suggests that these are extreme observations, even for the extreme sample, which the model does not manage to pick up. This follows from the fact that, when accounting for the current volatility of the market as the c-POT model does, sudden large fluctuations in returns will not always be within the confidence interval of the model. On the other hand, for VaR+, the risk seems to be slightly overestimated, with a lower than 1% violation rate. Still, the VaR estimates for the conditional-POT model showcase much better estimates than the Basel estimates, suggesting the adjustment done to incorporate the extremeness of the tails and heteroscedasticity have had effect. This is also clearly evident in figure 5.10 below, where the current market volatility is taken into account, displayed by the large fluctuations of VaR estimates.

Figure 5.10: Value-at-risk c-POT and Basel



5.2.4 Expected Shortfall

Expected shortfall for the conditional-POT model, is found to be [30]:

$$ES_q^t = \mu_{t+1} + (\frac{VaR_q}{1-\xi} + \frac{\beta-\xi u}{1-\xi})\sigma_{t+1} \quad (5.9)$$

Where μ_{t+1} and σ_{t+1} are conditional GARCH estimates of mean and volatility.

5.2.4.1 Violation Analysis

The figures for the expected shortfall conditional-POT is:

Table 5.6: ES violation rates using c-POT [31]

	Extreme sample	
	ES-	ES+
Violations	6	0
Violation rate	1.02%	0.00%
Sample size	590	590

As found for the VaR results when applying the conditional POT model, the left side of the distribution causes problems for the ES model as well, and results in a relatively high violation rate of 1%. On the right side of the ES model, zero violations are found. The high number of violations for the left tail stems from the fact that the conditional variance modelling does not account for or pick up the changes in volatility to a high enough degree, meaning the ES- is significantly underestimating the risk at the points of the violations. The low reduction in violations from VaR to ES does indicate that the losses at the time of the violations are in fact very large, so that they are not even covered by ES.

5.2.4.2 Loss Analysis

A violation rate of approximately 1% is a relatively high rate for ES, resulting in a large total loss over the six violations of approximately 400,000, as seen in the table below. In addition, the average loss is displayed as a result of the total loss over the period. The total loss is found to be the sum of the difference between the ES estimate at each point of violation and the actual observed loss. As such, it does only indicate the magnitude of the potential of loss of the portfolio and is not netted against gains or adjusted for increases in portfolio value over the period.

[30] See [19]

[31] Appendix table 6: Conditional-POT ES results for complete and normal sample.

Table 5.7: VaR violation rates using c-POT

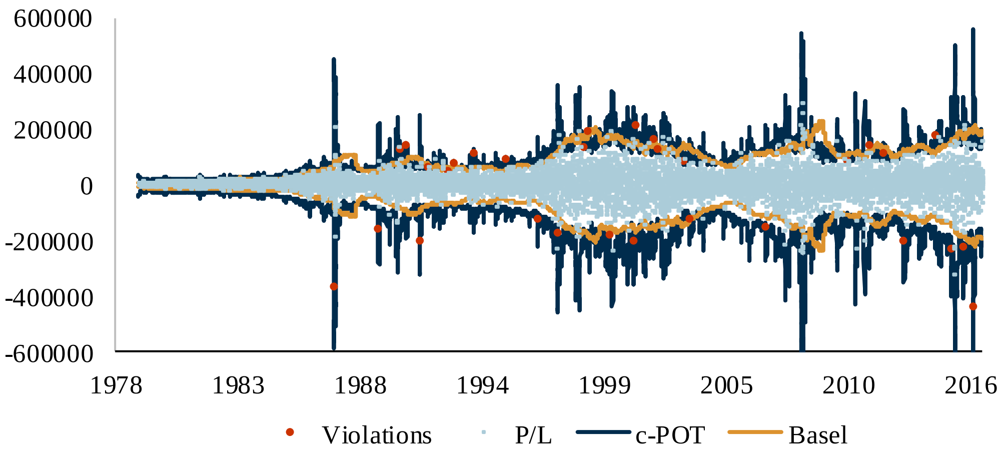
ES-	Extreme sample
Total loss	-395,468
Average loss	-65,911

The large average loss, compared to the previously observed values for student's t-GARCH and the Basel model, suggests that each violation on its own is of large magnitude, while the total loss over the period is small in comparison. As a whole, the ES modelling for the extreme sample better pick up the risk in the portfolio, and reduce the potential losses over the whole period, compared to Basel during the extreme market periods. The high average loss, however, suggests that the violations themselves are of large magnitude. Which could mean that even if ES adjusts to a large degree, the fluctuations in the returns are even larger and the risk measure does not manage to accurately estimate them.

5.2.4.3 Graphical Analysis

To further explore how ES evolve throughout the dataset as compared to the P/L, it is plotted together with the Basel ES for comparison in figure 5.11.

Figure 5.11: Expected shortfall c-POT and Basel



The conditional-POT ES estimates fluctuate to a large degree throughout the dataset, in particular when compared to the corresponding Basel ES estimates. It seems as if large fluctuations in the returns results in matching movements of the c-POT ES, displayed as sudden spikes in the graph. This means that the model is quick in adapting to changes in market volatility, and as a result reduces the number of violations significantly compared to the more constant Basel model. In addition to the rapid reaction rate to sudden increases in market volatility, the model seems to quickly reduce its ES estimates as the fluctuations in the market decreases. The quick reduction in ES estimates is important to reduce the potential risk overestimation in the model, as the model should not “cap” the potential upside by consistently providing too high ES estimates compared to the actual P/L in the market. However, the low violation rate in combination with the observable, continual larger ES estimates than Basel, as seen in the graph, suggests that risk overestimation might in fact be the case for the model. Yet, given that the model is designed and intended to deal with extreme observations only such an overestimation might in the end be preferable to reduce the probability of tail risk over the sample as a whole.

5.2.5 Conclusion

The conditional-POT model is a clear improvement of the standard Basel framework. This is first and foremost evident in the highly dynamic modelling of both VaR and ES, taking into account the current market volatility. Second, the risk measures do, with the c-POT model, not consistently overestimate the risk of the portfolio, so that a portfolio manager is not too risk averse. The model provides strong improvements in the results of the risk measuring of the extreme observations, but do still display problems with some extreme outliers leading to relatively high violation rates on the loss side of both VaR and ES.

5.3 Concluding Remarks

Through the modelling of the normal sample, where the student’s t-GARCH model was applied to estimate VaR and ES separately, and later when using c-POT to model the extreme sample, each sample has been processed separately. Modelling the two subsamples separately, reduces the applicability of using the models for risk management, as the market is not divided into two categories consequently. As such we have to apply the two models together to be able to draw conclusions on their utilisation for any given market state.

To further improve our setup, we will in the next chapter look to combine our approach by using both the normal and extreme sample. This should provide a model which is more applicable to real life use.

Chapter 6

6 Combining Models

Up to this point in the thesis, the two distribution models have been treated and analysed separately. In this chapter we will strive to prepare the applicability of the model, in order to make it more practical in terms of real life use. In particular, the focus will be on the timing of the switching between the normal and extreme distribution, as well as the procedure of collecting the data used to model the distributions for normal and extreme market states. The two respective distribution models will be implemented into one combined model, using student’s t-GARCH (1,1) for normal market conditions and conditional-POT for extreme market conditions. The intention of this chapter is, as a result, to determine whether the models are applicable for a portfolio manager, not only for each subsample, but for the complete sample as one.

The combination of the modelling procedures will be approached in two phases in order to arrive at a model that best describes a real life approach to VaR and ES estimation. Both in regard to information availability, in particular, but also in reducing the assumptions used so far in the thesis, when estimating the two risk measures. The first step, is combining the two models, student’s t-GARCH and c-POT, into in a “Paired model”. This model will have access to all information in the dataset available at all times, and as such serve as a concise, combined version of the calculations presented in chapter 5. Later the “Autonomous model” will be presented, where assumptions will be tightened and the information available limited to a realistic level. The tightening of these assumptions should intentionally produce less bias in the results and as such be closer to a real life approach.

6.1 Paired Model

When constructing a combined model, the “Paired model” should intentionally shift underlying distribution model depending on the market state of the portfolio. Student’s t-GARCH (1,1) will be used during normal periods, since this is fitted based on the normal sample. If the market state qualifies as extreme, the c-POT will be applied, as this is constructed based on the extreme sample. The model is assumed to switch between student’s t-GARCH (1,1) and c-POT at a perfect timing. In other words, when an extreme period is occurring the model switches from student's t-GARCH (1,1) to c-POT exactly when the extreme period begins, even though there is no indication up to that point in time that an extreme period actually is approaching.

6.1.1 Value-at-Risk

Table 6.1: VaR violation rates using Paired model

	Complete sample		Normal sample		Extreme sample	
	VaR-	VaR+	VaR-	VaR+	VaR-	VaR+
Violations	102	94	93	89	9	5
Violation rate	1.03%	0.95%	1.00%	0.96%	1.53%	0.85%
Sample size	9900	9900	9310	9310	590	590

VaR violation rates for normal periods are identical to the violation rates for the normal periods in the student's t-GARCH(1, 1) model presented in chapter 5.1.3, while the extreme period violation rates above are identical to the output from chapter 5.2.3, when calculating extreme period violations for c-POT model. The equivalent results follow from the fact that the Paired model now uses the two methodologies of student's t-GARCH and the c-POT together, only switching when the market conditions changes.

Violation rates over both the normal and extreme period are previously found to be yielding around the target rate of 1% violations. VaR- violations for extreme market periods are a little high, but this is, as touched upon earlier, caused by a few outliers which are difficult to detect by any distribution model. Therefore, when combining the two distribution models into the Paired model, violation rates for the complete sample are found to be satisfying, violating close to 1% of the total sample size, with a 99% confidence interval. It is the number of violations over the complete sample which are of the most interest when looking at the results of the combined model, as these are taking into account the complexity of modelling the whole period as one. Integrating the two models found in chapter 5, it is now possible to draw conclusions over the whole dataset, where the results suggests that the model, when presented with both normal and extreme market conditions, manages to apply both in an appropriately manner leading to almost perfect results. The close to ideal results, naturally stems from the timing of the switch between normal and extreme modelling being perfect, and the models being run on the same sample in which they have been defined. Both 'issues' will be dealt with, first in subchapter 6.2 and later in chapter 7, with the Autonomous model and an out-of-sample test respectively.

6.1.2 Expected Shortfall

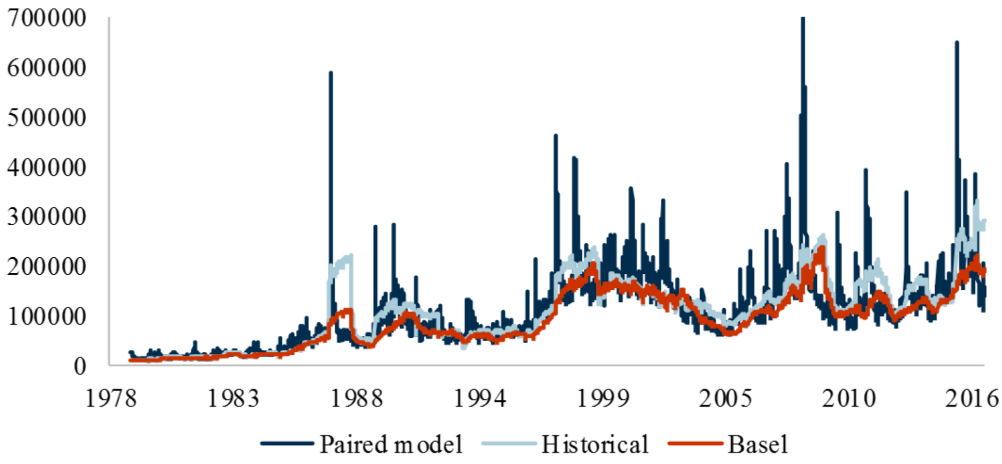
Table 6.2: ES violation rates using Paired model

	Complete sample		Normal sample		Extreme sample	
	ES-	ES+	ES-	ES+	ES-	ES+
Violations	33	35	27	35	6	0
Violation rate	0.33%	0.35%	0.29%	0.38%	1.02%	0.00%
Sample size	9900	9900	9310	9310	590	590

As for the value-at-risk figures presented above, the ES results of the normal and extreme market are identical to the findings derived in subchapters 5.1.4 and 5.2.4. Combining the two provides violation rates over the complete sample at approximately 0.3% for both sides of the distribution. The low number of violations are good news for any portfolio manager, greatly reducing the potential loss over the whole dataset when combining the two models. This further confirms the success of the two models found when running VaR above, namely that the two combined are able to explain the risk in the returns of complete the dataset accurately.

However, as seen through the analysis of ES done earlier in the thesis, it is difficult to know the nature of potential risk overestimation in the model presented, only by investigating the violations. Below, the viewpoint of the analysis is from a left-tail perspective, as such the focus is to draw conclusions on the potential losses for a portfolio manager. The violation rate is found to be low, but not the degree in which the model achieves the low rate, whether it is constantly too high or rather fluctuates with the movement in the market. VaR should be 1%, but how low the violation rate of ES should be to optimise risk management, is not set. Therefore, to get an indication of the ability of the Paired model to successfully adjust to market movements and as such not constantly overestimate risk, it will be compared with the chosen benchmark model, the Basel framework. Figure 6.1 is used to compare the actual value of the loss side of the expected shortfall calculations for the Paired model and the Basel model derived in chapter 3, in addition to the historical 99% ES estimated as a function of the portfolio. The figures given are displayed in absolute terms.

Figure 6.1: Expected shortfall comparison



The Paired model fluctuates extremely compared to the Basel and Historical ES values. It seems as if the model reacts rapidly to changes in the market with large, sudden spikes in the ES estimates when the market moves. The movements are in stark contrast to the relatively small movements in both the Basel model and the actual historical ES figures. This suggests that that the model is quick to adapt, both on the upside and downside of the risk measure, as the large spikes when volatility increases is followed by similar movements in the ES downwards. As such, the Paired model seem to fluctuate around the mean of the two other ES estimates, but over the time period as a whole, it provides higher ES estimates, on average. The quick downward move of the ES estimates when large upward movements occur, means that we do not cap our potential upside longer than necessary as both Basel and historical ES showcase to do.

6.1.3 Conclusion

As touched upon in the beginning of this chapter, this an all-knowing model assumed to foresee whether a normal or extreme market period is approaching, making it able to switch distribution method perfectly with respect to timing. By making use of information, not obtainable at each respective day of risk estimation, the Paired model relies on observations yet to come. This is not a realistic procedure of switching between different distribution methods. A genuinely applicable model cannot know the future and thus perfectly predict the state of the market tomorrow. In the next section this issue will be addressed, searching for alternative methods of switching between distributions, as well as finding a more realistic collection of data for fitting distributions; both in order to derive less biased results.

6.2 Autonomous Model

In this section we will present the “Autonomous model”, a model not knowing anything about the future when advancing through the dataset. The move from the Paired model towards a more realistic model will be two sided. First, we highlight the procedure of collecting data used to fit the distribution in normal and extreme periods. Second, we focus on the procedure of switching between normal and extreme modelling. Both are elements in the Paired model not being strictly forward-looking. This means that the decision making process when switching between normal and extreme modelling for applying VaR and ES will be made retrospective, not knowing when extreme market periods are approaching. Therefore, the Autonomous model will mitigate the weaknesses of the Paired model, from a real life perspective, and produce risk estimates only based on previously observed data.

6.2.1 Collection of Data

Up to this point the distribution has been approximated based on the entire dataset, and thus to some extent indirectly been founded on information based on the future. Now however, the distribution will be fitted only based on information available at the day of estimation. For each day passing, information from this day will be added to a data sample and used to improve future predictions. The model is autonomous because of its self-learning properties, meaning that it is able to automatically incorporate new information, in order to produce better prediction estimates and as a result improve itself when the data sample is increasing.

The starting point of estimation will be in 1978, with one year of data as basis. Then the model continuously adds observations to its available data day-by-day, when moving forward through to the end of the dataset, in year-end 2016. In practice this means that, if a normal day is observed, new information is incorporated into the normal data sample and used onwards to fit a student’s t-GARCH (1,1) applied in normal market periods. As time passes and new observations of normal periods are encountered, the normal data sample increases and the fit of distribution for normal periods is additionally improved. Equivalently, the same mechanism is applied for the extreme periods and the c-POT model. The reasoning behind using such a model is to apply a design that is more "real life like", and which could in fact be used to make actual calculations for a portfolio manager.

6.2.2 Switching Procedure

In the attempt of producing a more genuine test of the Paired model's performance, the procedure of switching between distribution models should be done in a way that is not all-knowing. In other words, when estimating VaR and ES the use of either student's t-GARCH or the conditional-POT must be determined by information known at the time of the decision. This means that the model must base the selection of normal or extreme modelling on the basis of history, in contrast to the previous subchapter, where information ahead of each decision point in time revealed that extreme observations were approaching. Instead now, recent observations in the market is used, suggesting whether the market qualifies as extreme, followed by a switch into extreme modelling.

The procedure of switching between normal and extreme modelling will be based on the same definition as applied in subchapter 2.3 when defining the subsamples. However, in contrast to earlier use, the switching cannot be triggered before the extreme observations actually have occurred, in order to maintain a historical perspective. As a result, the Autonomous model is now backward-looking, not applying any information about the future. More specifically, the model will switch into extreme modelling if there is, over the course of the last 10 trading days, detected three or more returns being +/- two standard deviations from the mean.[32] This procedure is repeated every day, potentially allowing the Autonomous model to, if necessary, switch between normal and extreme modelling daily. As such, the choice of model happens each day after the markets close, so that if three extreme returns are observed over the last 10 days the extreme model is applied immediately by the start of the next trading day. The result is that there is no lag after the extreme observations have in fact occurred, but the switch between the two models happens retrospectively.

The fact that switching to extreme modelling is performed after three extreme observations already have occurred, makes the model more realistic compared to the earlier applied methods in the thesis. However, optimally the model should be forward-looking, which implies an ability to predict the future or make qualified guesses and anticipate when the extreme observations will happen. We will investigate the possibility of this, when exploring the properties of the days before extreme periods in chapter 8.

[32] Estimates of mean and standard deviation are based on one year of data prior to the respective day.

For now, we assume that extreme market periods cannot be predicted. As such, switching to extreme modelling can only be done, after the extreme observations have occurred. This backward-looking method of switching distribution models is a drawback of the Autonomous model, as the switch happens retrospectively, thus the risk of not modelling the correct returns with the correct model to a high enough degree is present. For the model to work as desired, a high share of extreme periods should appear consecutively one after another, so that the retrospective switch of models succeeds. This method should be appropriate, as discoveries in chapter 4 suggests substantial presence of volatility clustering in both the normal and extreme sample. Put differently, it seems as if volatility today is highly correlated with yesterday's volatility, implying that normal periods are often followed by normal periods, and extreme periods are often followed by other extreme periods. Based on this, we believe the procedure of switching models retrospectively, despite the delayed reaction time, is a reasonable way of deciding whether normal or extreme modelling will be applied, without using information about the future.

By doing this, the procedure of both switching between normal and extreme modelling, and fitting the distribution is as realistic as it can be, in regard to the use of information available to a portfolio manager.

6.2.3 Value-at-Risk

Table 6.3: VaR violation rates using Autonomous model

	Complete sample		Normal sample		Extreme sample	
	VaR-	VaR+	VaR-	VaR+	VaR-	VaR+
Violations	140	126	108	109	32	17
Violation rate	1.41%	1.27%	1.16%	1.17%	5.42%	2.88%
Sample size	9900	9900	9310	9310	590	590

Value-at-risk violation rates over the complete sample are higher for the Autonomous model when compared to the Paired model. They are slightly above the desired 1% rate, but yet close to an acceptable level for the model as one. However, the violation rate being larger than one implies an underestimation of risk, which means the likelihood of larger losses is present. Using the results from Basel as a peer, the output from the Autonomous model is much closer to the ideal target rate of 1%, compared to the findings in chapter 3. However, returns in normal periods are violating the value-at-risk at a slightly higher level than the corresponding statistic from Basel, indicating that the total reduction comes from the extreme sample. The Autonomous model does experience significantly fewer violations during extreme periods, than the Basel model, indicating that such periods are modelled better by using the conditional-POT for extreme periods, also when the model is gradually increasing its data through time. Despite the improvement of the Autonomous model over the complete sample compared to the corresponding Basel figures, the underestimation of risk during the extreme periods, is too high, and displays the continued problems in correct estimation of risk for extreme observations.

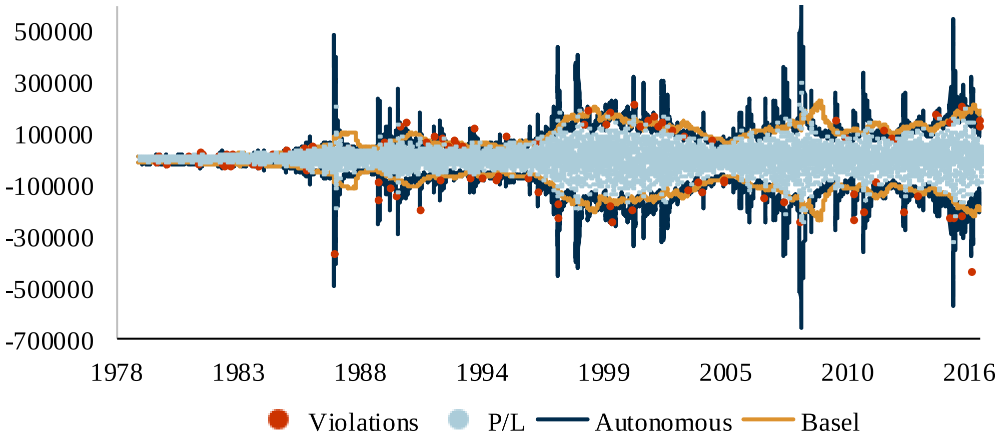
6.2.4 Expected Shortfall

Table 6.4: VaR violation rates using Autonomous model

	Complete sample		Normal sample		Extreme sample	
	ES-	ES+	ES-	ES+	ES-	ES+
Violations	47	54	31	45	16	9
Violation rate	0.47%	0.55%	0.33%	0.48%	2.71%	1.53%
Sample size	9900	9900	9310	9310	590	590

The expected shortfall violation rate is slightly up from the corresponding figures found for the Paired model, up to approximately 0.5% for the complete sample, from 0.3% found earlier. As the model is restricted to using only historical information observed at any point in time, the violation rate is expected to increase. The advantage of knowing the future is removed, and as a result the losses has to increase. However, when compared to the Basel model, the improvement is still clearly present, with the violations rate for the loss side, in particular reduced, down from the 1.4% found earlier in chapter 3. This is, again, mainly driven from reductions of violations in the extreme sample, which are reduced around six fold from the corresponding Basel violations. The plot of the Autonomous model's ES estimation below, is compared with the Basel ES and the P/L's of the portfolio.

Figure 6.2: Expected shortfall Autonomous and Basel



The Autonomous ES estimates seem to display approximately the same properties as the ES of both the student's t-GARCH and the c-POT, with a highly fluctuating, rapidly reacting model. Yet, compared to the latter in particular, the number of violations are up, suggesting that, even though the model seem to showcase the same trends of behaviour, it does not reach the same level of accuracy when the extreme observations occur.

6.2.5 Conclusion

Compared to the Paired model, violations are higher for both VaR and ES. The tightened assumptions have made the model more realistic, but the results are, not surprisingly, less accurate. Violations in normal markets are at a reasonable level, while the model still show difficulties when predicting risk during extreme market conditions. Despite violations being higher than desired, the Autonomous model produces far more precise results for both VaR and ES, than the corresponding output derived from the Basel approach. The model's opportunity to switch between normal and extreme modelling, makes it designed to effectively adapt into new market states and thus provide better risk estimates for a portfolio manager. The method of switching is far more applicable and realistic, than using the Paired model, since the procedure of switching now is retrospective and does not take information about the future into account when selecting between normal or extreme modelling. In addition, the procedure of fitting data only relies on information obtainable at the day of estimation.

6.3 Concluding Remarks

With the evolving setup of the models in this chapter, from separate units modelling each subsample, via an assembly of student's t-GARCH and the c-POT models combined, to the Autonomous model, we have moved from a purely theoretical approach to a potential useable model. The Autonomous model has mitigated the two parts of the theoretical approach that was highly unrealistic; the collection of data used to estimate the distributions of the two models and the switching between the two, now based on realised returns only. Moving away from the theoretical approach where all information in our dataset was known to the models, we have, at the same time, moved from highly accurate risk measure estimates to more underestimation of risk, in particular for the extreme observations. This trade-off displays the difference in a purely theoretical approach to testing VaR and ES, and a realistic take on finding the accuracy in using the two measures as a tool to reduce risk on the portfolio. Despite the less accurate risk estimates, the Autonomous model is still providing significantly improved risk estimates for all market states compared to the Basel model.

Even though the Autonomous model is by far the most genuine model, with the most likely real life applications, it is in this chapter run on the same data as it is built upon. Potentially, this could lead to a severe bias in the VaR and ES estimates as the model's results could reflect the data it is fitted on, rather than being independent of it. Despite that fact that the model is built such that it should only be dependent on the data it is run through, i.e. the previous observations in time, testing it on a new sample should be done, in order to be certain of its independence.

Chapter 7

7 Out-of-Sample

The need for an independent test of the VaR and ES estimates done so far is clearly present, as all parameter and model estimations done so far are applied on the same dataset. This could potentially lead to a bias in the results presented previously, as all estimations are based on parameters especially fitted to the particular returns in the dataset. Therefore, in this chapter, previous estimations will be applied on a new dataset with the goal to further strengthen, or discard the conclusions already drawn.

When deciding on the original data sample in chapter 2, we chose to include as large a time span as possible to have enough data to split it into two subsamples, in particular to have enough extreme observations to base the modelling on. This affects the possibility of setting up a dataset which in no way is influenced by the original sample data. As we are dealing with 40 years of data in the calculations made earlier in the thesis, the availability of indices not in the same time frame is sparse. Further, we wanted a diversified sample when putting together the original sample, this led to a portfolio consisting of three wide and geographically spread indices. The possibility of finding a sample which does not correlate with any of S&P, DAX or NIKKEI is therefore impossible. These two issues, of not having a separate time frame or an un-, or low correlated second sample to test the VaR and ES calculations on, makes the independence of the tests performed in this chapter weakened. However, a major reason to perform the out-of-sample tests regardless, is the setup of the Autonomous model. As discussed earlier it does not base its calculations on data ahead of time and its parameters is only based on previous observations, as such the model should not be affected by the dependence the two samples have, other than the probable similarity in the movement of returns over parts of the time period. To clarify further, the Autonomous model is not dependent on any data from the original sample and when run on the out-of-sample, all parameters will automatically be based on out-of-sample data, as the model makes use of the data when it adds it on a day-to-day basis over the period.

In an effort to reduce the independence on the original dataset as much as possible, while still keeping the sample size high, and taking into account the availability of data, the out-of-sample is put together as portfolio of three Northern-Europe based indices. Data from the Denmark based OMXC20, Finnish OMXH25 and the Swedish OMXS30 are used over a period from 1991 to year-end 2016. The previously explained Basel and Autonomous model will be used to calculate VaR and ES estimates for the out-of-sample. The Basel model comparison is necessary to be able to strengthen the analysis and conclusions drawn from performing a test of the Autonomous model, as it has previously served as benchmark for the Autonomous model on the original sample. Since the Autonomous model is previously identified as the most genuine model of the ones presented in the thesis, this is the natural choice of comparison to the original sample. In addition, as discussed above, the Autonomous model serves as a way to avoid some of the dependence in the out-of-sample dataset on the original data.

7.1 Basel Model

Through the thesis the Basel framework has been used as a peer to evaluate the performance of the models developed. This will be extended in this chapter.

Table 7.1: Out-of-sample VaR violation rates using Basel model

	Complete sample		Normal sample		Extreme sample	
	VaR-	VaR+	VaR-	VaR+	VaR-	VaR+
Violations	152	114	75	69	77	45
Violation rate	2.34%	1.75%	1.24%	1.14%	17.91%	10.47%
Sample size	6500	6500	6070	6070	430	430

Running the Basel model on the Nordic out-of-sample results in violation rates approximately twice the size of the target rate of 1%. The high violation rate is mostly caused by a bad fit in extreme market periods, yielding between 10 and 18 times the ideal rate of 1%. The tendency is the same as for the original dataset, i.e. the model has trouble estimating value-at-risk accurately, seen through a severe underestimation of the risk. Normal periods are violating more frequent than desirable, but is still on a more moderate level slightly over 1%.

Table 7.2: Out-of-sample ES violation rates using Basel model

	Complete sample		Normal sample		Extreme sample	
	ES-	ES+	ES-	ES+	ES-	ES+
Violations	108	67	51	35	57	32
Violation rate	1.66%	1.03%	0.84%	0.58%	13.26%	7.44%
Sample size	6500	6500	6070	6070	430	430

Naturally, the trend is similar for expected shortfall. Violation rates in normal periods seem to be on a reasonable level, although slightly overestimating the risk. However, a consequential mismatch during extreme periods is drastically increasing the overall violation rate of the model.

Summarising the results of using the Basel framework on the out-of-sample, the trends seem to be very much in line with the original dataset. Partly this is a result of the two samples having similar properties, as both includes international indices and have observations overlapping the same time period. Regardless of a potential correlation with the original sample, the performance of Basel in the out-of-sample test is not sufficient for a portfolio manager, with potential losses during extreme periods large. The out-of-sample test confirms the previous findings in chapter 3, namely that the Basel framework is too static and lack a method of incorporating new information quickly and thus adapt when volatility in the market unexpectedly changes.

Findings in chapter 6, suggested that the performance of the Autonomous model, was significantly better than the Basel framework in explaining extreme periods. However, as explained earlier, the Autonomous model was constructed based on the original sample. Testing the performance on the very same sample could create a bias. Next, we will test the Autonomous model on the Nordic out-of-sample, in order to find the potential presence of bias, or confirm the properties of the Autonomous model regardless of data.

7.2 Autonomous Model

As the Autonomous model is only dependent on the data it is run through, it should, in theory, not be dependent on the original dataset, even though it was created with it as the basis. As such, the results on the out-of-sample should be pretty similar to the original results from chapter 6, as the violations are, if the model works as designed, purely a result of the properties of the returns in the given dataset it is modelling.

Table 7.3: Out-of-sample VaR violation rates using Autonomous model

	Complete sample		Normal sample		Extreme sample	
	VaR-	VaR+	VaR-	VaR+	VaR-	VaR+
Violations	96	63	73	56	23	7
Violation rate	1.48%	0.97%	1.20%	0.92%	5.35%	1.63%
Sample size	6500	6500	6070	6070	430	430

The Autonomous model provides good estimates for the normal sample with 1.2% and 0.9% violations for the loss and gains side respectively. For the extreme sample, however the results are higher than the preferred violation rate. There is a gross underestimation of the risk on the left side of the tail, while violations on the right side are lower. Despite this, when comparing the Autonomous model and Basel modelling for the extreme periods in the out-of-sample, trends are consistent with the findings for the original sample. As the violations during extreme periods are heavily reduced, down from approximately 18% and 11% for the loss and gain side respectively. This suggests, again, that the extreme modelling in the Autonomous model, and the timing in the model of extreme versus normal modelling, produces far better results than the standard Basel model, also for the out-of-sample.

Compared to the extreme sample in the Autonomous model run on the original sample, the results are pretty similar. The loss side of the VaR calculations in the out-of-sample are in line with the original sample, while the right side of the distribution is slightly better estimated in the out-of-sample. In total, this indicates a consistency in the model when estimating the extreme observations, also when presented with a new sample. Since the two samples have probable similarities it is, with consistent results between the two datasets, difficult to distinguish where the consistency stems from. It could be that the model's structure when analysing data is working as intended or the extreme observations it analyses could be too closely matched. But, since the model is built as it is, there should be no doubt that the consistency is a result of its properties rather than the dependency between the two samples.

A key takeaway is, however, independent of the potential similarities of the two samples, the high violation rate of the Autonomous model on the extreme sample. The results above underline the difficulties in finding a more generalised model that is able to model extreme observations.

Table 7.4: Out-of-sample ES violation rates using Autonomous model

	Complete sample		Normal sample		Extreme sample	
	ES-	ES+	ES-	ES+	ES-	ES+
Violations	35	21	25	20	10	1
Violation rate	0.54%	0.32%	0.41%	0.33%	2.33%	0.23%
Sample size	6500	6500	6070	6070	430	430

When analysing ES the results over the complete return set of the original and out-of-sample are similar. For ES, as was the case for VaR, the out-of-sample modelling seems to provide less violations on the gain side of the distribution, this is a trend across both the normal and extreme sample, which could indicate a tendency of less outliers on the right side of the data for the out-of-sample.

Again, the extreme sample loss side of the distribution proves difficult to model with ES, as there is a large underestimation of the risk displayed through a violation rate of 2.33%. The right side of the extreme sample distribution have only one violation. Compared to the Basel ES, the risk estimation has once again improved when applying the Autonomous model.

Looking at the original sample modelling of the extreme sample, the ES figures were 2.7% and 1.5% for the loss and gain side respectively. The results for ES could therefore, as was the case for VaR, suggest that the Autonomous model is in fact only dependent on the properties of the particular returns it models, and as such, the small fluctuations in the results could stem from differences in the properties of the returns between the two datasets.

7.3 Concluding Remarks

In general, the results of the out-of-sample testing was very much in line with the results of the original sample. Both for each model and risk measure individually, but also in the trend between the Basel and Autonomous model. As suggested in the beginning of the chapter, this should be the case because of the way the Autonomous model works. As such, the working theory was for the Autonomous model to provide the needed results to confirm its function, while the Basel model was only a tool to compare the two. The deduction is that the consistency in the results across the two samples stems from the properties of the Autonomous model, in which it is not dependent on the original sample when estimating VaR and ES on other samples. The conclusion from chapter 6 is also strengthened, namely that the Autonomous model provides more accurate risk estimates than the standard Basel model.

Even though the results of both the VaR and ES for the Nordic indices are significantly better than the corresponding results for the Basel model, the results of the extreme sample prove the difficulties in finding a generalised model to successfully describe and measure extreme observations. As we believe the Autonomous model to be independent of the dataset it is applied upon, this suggests the difficulties in modelling the extreme returns to be independent of the return sample, but rather is model specific. As such, since the Autonomous model seems to have a particular trouble in modelling the extreme parts of the returns it is provided with, we want to further explore the properties of such observations in an attempt to enhance the results of the model.

Chapter 8

8 A Deeper Look into Extreme Periods

As we have seen in the analysis of the normal and extreme samples, modelling the extreme observations has proved difficult, in particular when applying realistic assumptions. Both the original and out-of-sample tests of the Autonomous model underlined the difficulties in providing a generalised model to describe the extreme returns accurately, independent of the dataset. Therefore, the motivation for the last part of the thesis is clear, we want to further investigate the extreme sample and aim to understand the non-normal observations in an attempt to uncover the possibilities of modelling such returns better.

The first part of this chapter will be closer to a qualitative analysis, rather than the quantitative focus of earlier, as we seek to understand the dynamics of extreme periods. A consequence of this is that the analysis is mainly applicable on our dataset, and results cannot be generalised to other datasets, without further investigation. The focus is mainly on improving the understanding of extreme periods, and help explain why they are so difficult to predict and model. In addition, we seek to thoroughly investigate whether the extreme sample does in fact display the extreme properties wanted, to be able to draw conclusions on the extreme returns of the whole dataset in general. In particular, to be able to draw conclusions on the time leading up to each extreme period, as the second part of the chapter will investigate the properties of the returns leading up to the defined extreme periods. Finally, we will seek to use the information obtained to test a new method of defining when to use normal and extreme modelling, both on the original sample and out-of-sample. Over the chapter as a whole, we will use the properties of each of the three indices in the portfolio used in the original sample, to further enhance the analysis with a broader comparison basis.

As explored earlier, there does not exist a perfect model to define the tails. It is not our aim to find one either, but rather explore how to better the models already applied on the dataset. Also, it is previously found how value-at-risk and expected shortfall, especially in its most general forms, experience problems during extreme periods, which in the end is a problem for a portfolio manager. Since the modelling of the normal sample mainly has appeared successful, a further exploration of the extreme periods is the remaining critical point in the understanding of the two risk measures.

When dividing the returns into the extreme sample, as done earlier, the extreme periods will only consist of returns defined as extreme, with the definition being that during a 10-day period, there is a minimum of three observations being +/- two standard deviations from the mean at that time. However, the nature of the mean or standard deviation used to estimate this extremeness, is not known, which could mean that some of the returns defined as extreme, might be coloured by a lower than average mean or standard deviation over the last year, used to define the returns. As such, certain periods could be defined as extreme while they, compared to the rest of the sample is not, but rather at that exact time is as a result of the characteristics of the market over the last year. Further, the properties of the returns "flagged" as extreme is only known because of their distance, measured in standard deviations, from the mean. However, there is no information of the extremeness of the remaining amount of returns defined as extreme in the same period. Of the 10 returns used to characterize the 10-day period as extreme, up to seven returns could be less than two standard deviations away, but still flagged as extreme due to their surrounding returns. The results over each 10-day extreme period could therefore be severely biased by the other returns, not flagged as extreme. Although, this bias likely is present in each 10-day extreme period, it is in many ways a natural bias as a period of extreme returns in the market rarely is characterised by exclusively large, negative or positive returns far from the mean, but rather some large movements followed by days of consolidation. This is observed by the short lasting fluctuations found in the return plots earlier. We will further investigate the properties of the extreme 10-day periods in subchapter 8.1 to see if they show the characteristics of being actual extreme returns.

8.1 Dynamics During Extreme Periods

Earlier, in chapter 2 and 4, when analysing the extreme sample, the focus has been the sample as a whole. Its statistical properties have been looked into via the distribution, its modelling and its characteristics when measuring VaR and ES. Through this, the extreme sample has been found to contain the characteristics of extreme observations, in that it, compared to the normal sample, has less of a Gaussian distribution, with fatter tails, more extreme movements in its return properties, and it has proved harder to achieve correct risk measuring. As such, it shown to be accurate, in regard to the subsample overall being more extreme than the normal sample. However, we have not investigated the specific characteristics of neither the 10-day periods it is divided into, as its own units, nor the particular return data on a thorough level as compared to the normal sample. Therefore, we will seek to better understand the extreme returns below, to be able to draw conclusions both on if the definition of the extreme sample is as accurate as we believe it to be, and to further enhance our grasp on the 10-day periods.

8.1.1 Frequency Analysis

To better understand the extreme periods, their extent should be explored, how often they occur and the timing of each period. Therefore, each 10-day period will be analysed in a frequency analysis.

Table 8.1: Frequency of 10-day extreme periods

10-day periods	S&P 500	NIKKEI 225	DAX 30
Total	990	990	990
Extreme	53	73	65
Frequency	5.35%	7.37%	6.57%

The frequency of the extreme periods is around 5-7% over the sample, meaning that 10-day extreme periods happen approximately 1.5 times every year on average. As seen in subchapter 2.3 the extreme sample is defined to be the periods of 10 trading days where three or more observations of returns more than +/- two standard deviations from the mean are observed. Both the standard deviation and mean are calculated as rolling averages over the last year. The theoretical probability of a 10-day period being defined as extreme is 1.05% as found in subchapter 2.3. However, as seen here the number is found to be much higher. This is, as touched upon earlier, because the distribution of the sample does not match the properties of the Gaussian distribution which is used when calculating the probabilities.

Table 8.2: Continuity of extreme periods

10-day periods	S&P 500	NIKKEI 225	DAX 30
Continous extreme	19	18	20
Frequency	35.85%	24.66%	30.77%

Of the extreme periods, a relatively high number, between one-fourth and one-third, of these are continuous, meaning that one extreme period is followed by another. This supports our earlier findings of volatility clustering in the returns. In this regard, meaning that large fluctuations in returns follows each other, which leads to continuous extreme periods per our definition. As such, the same trend is observed in the extreme sample, as in the raw return data. This makes it easier to draw conclusions from the extreme sample as the subsamples, according to our definition, are in fact representative for the returns.

Table 8.3: Average days between extreme periods

10-day periods	S&P 500	NIKKEI 225	DAX 30
Days between extreme	168	123	137
Days between extreme (excl. continous)	261	164	197

The average length between each extreme period is between 123 and 168 trading days over the three indices, meaning that, on average, extreme periods are experienced between 1.5 and 2.0 times a year. If the continuous extreme periods are excluded when calculating the length, where one continuous extreme period is two or more such periods following each other with no normal period in-between, then the extreme periods happen 1.0 - 1.5 times a year, on average. The fact that a high proportion of the extreme periods occur together makes it so that the periods are clustered, a consequence of this is that we experience fewer extreme periods per year, but they last longer when they occur.

8.1.2 Univariate Analysis

The frequency analysis provides a quick insight into the recurrence of the 10-day periods, but tells nothing about the properties of their contents. To further enhance our understanding of what happens within each 10-day period, on average, a univariate analysis will be performed. Each index is divided into 10-day periods and the calculations are done over each period and then averaged over the whole sample.

8.1.2.1 Mean

Table 8.4: Mean indices for different market periods

	S&P 500	NIKKEI 225	DAX 30
Complete	0.03%	0.01%	0.03%
Normal	0.05%	0.03%	0.05%
Extreme	-0.24%	-0.24%	-0.25%

The overall mean of each index is lower than the mean of each respective normal period, which is due to a negative mean in the extreme periods for each index. Since the mean is the weighted average of the mean in the normal and extreme periods, the overall mean must by definition be between these two threshold. The means during extreme periods are considerably lower than the means for the normal periods. For the three indices, the mean of the extreme periods is around 5-6 fold lower than that of the normal periods. This high, negative yield indicates either a large magnitude of negative values or a number of large negative values, or both.

8.1.2.2 Standard Deviation

Table 8.5: Standard deviation indices for different market periods

	S&P 500	NIKKEI 225	DAX 30
Complete	1.09%	1.34%	1.33%
Normal	0.93%	1.16%	1.17%
Extreme	2.61%	2.68%	2.67%

Further, the extreme periods have a standard deviation which yields at a ratio of approximately 2.5 compared to normal periods, indicating that fluctuations are substantially higher for the extreme periods. As we define extreme observations based on standard deviations from the mean, this is a natural consequence. Detecting higher standard deviation for these periods is a confirmation of our own definition of an extreme period. It is, however, interesting to see that the standard deviation of the whole extreme sample is as high as it is, considering that the 10-day extreme samples also contains a majority of normal observations on average.

8.1.2.3 Range

Table 8.4: Range indices for different market periods

Range	S&P 500	NIKKEI 225	DAX 30
Normal sample	0.0277	0.0337	0.0333
Extreme sample	0.0712	0.0765	0.0773

During the extreme periods the range of the returns is noteworthy different from that of the normal periods. The range increase is between twice and three times when comparing the normal 10-day periods with the extreme ones. This is a consequence of the way the extreme sample is defined, as an increased standard deviation leads to an increase in the fluctuations of the returns, and as such the range increases, reflected through larger minimum and maximum values.

8.1.2.4 Distribution

Further, we want to investigate the internal distribution properties of each 10-day period within the normal and extreme sample. All numbers are based on calculations done on each period, which is then averaged over the sample as a whole.

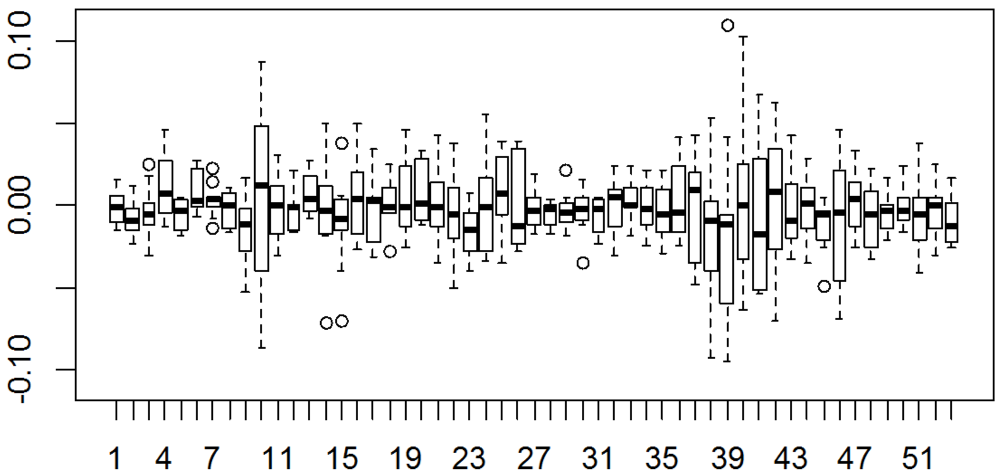
Table 8.5: Number of negative observations in each 10-day period

# Negative observations	S&P 500	NIKKEI 225	DAX 30
Normal sample	4.4632	4.4896	4.4919
Extreme sample	5.4528	5.0548	5.1077

The number of negative returns increases significantly during the extreme periods when compared to the normal 10-day periods. Out of the 10 returns the number of negative observations goes from being well under half of the returns in the normal periods, to over five for all indices during the extreme periods. Even though the extreme 10-day periods are, clearly characterised by a predominance of negative returns, in combination with a negative mean as observed earlier, it is difficult to draw conclusions on the distribution properties of both the extreme samples on their own, but also compared to the normal samples. This is a result of the differing means of the samples, making it difficult to determine where the returns are distributed.

To further investigate the distribution of the extreme periods, 10-day periods will be plotted in a so called box plot. The plot (figure 8.6) display some of the main characteristics of the extreme sample of the S&P 500. The plots for NIKKEI and DAX can be found in appendix figure 1 and 2, but contain the same characteristics as the S&P and are therefore not included here. The box-and-whisker plot shows the distribution of the returns. The inter-quartile range – from the lower to upper quartiles, i.e. the middle 50% of the data, is displayed in the clear boxes. The middle quartile, or median is shown as the dark line within the clear boxes. The top and bottom 25% of the data are shown by the whiskers, the dotted lines on both sides of the boxes. The maximum and minimum, excluding any outliers are shown as the end of the whiskers. Any outliers, data more or less than 3/2 times of the upper and lower quartiles respectively, are marked as clear dots.

Figure 8.6: Box plot S&P 500 (excl. Black Monday)



As seen in the plot, the extreme 10-day periods of the S&P 500 vary to a relatively large degree throughout the dataset. The average of the lower and upper whisker values is -3.2% and 3.1% respectively, meaning that excluding the outliers these are the average minimum and maximum values. The minimum and maximum values indicate that the range is approximately 0.9% less when taking the outliers into account, which would suggest that a few large observations increase the range over the whole sample. Yet, the middle 50% of the extreme returns, i.e. the observations covered by the boxes, are around two to three times larger than the respective numbers for the normal sample, which suggests, also when excluding outliers from the data for both the normal and extreme sample, that the observations in the extreme sample shows characteristics of exactly that, extremeness. Even though the plot, as such, suggests that the extreme sample is in fact extreme, when compared to similar numbers from the normal sample, it can be observed that several of the extreme periods are characterised by small fluctuations, namely small boxes without any outliers. This could, as touched upon earlier in the chapter, be a consequence of the market over the last year. As the mean and standard deviation over the last year is used to define the extremeness of the returns, a year marked by small fluctuations and a low mean would indicate that a 10-day period with large movements compared to the last year would be flagged as extreme, but when compared to the rest of the 40-year long sample such a period could be "normal". This dynamic could be a potential flaw in the definition of the extreme sample done so far, and as such, one to avoid when we later look at alternative ways of defining extreme periods.

8.1.2.5 Conclusion

The observations above, when analysing the concrete differences between the extreme and normal sample, indicate a clear difference between the two subsamples. The extreme returns show significantly different properties than the normal returns. However, we want to further confirm that the properties of the extreme sample are in fact displaying extreme characteristics, in particular by using a more standardised measure.

8.1.3 Bivariate Correlation Analysis

Previously in this chapter we have performed a univariate analysis to compare statistical properties in normal and extreme market periods, in an attempt to improve our understanding of the difference in dynamics during these periods. In addition to this, we seek to confirm the extremeness of the extreme sample to verify the method of defining the returns as extreme.

In the models applied earlier in the thesis a portfolio consisting of three different indices has been applied. A portfolio of assets introduces an important factor to consider when looking into the properties of the extreme periods. Correlation could potentially interfere and amplify the results of our modelling. Correlation between assets is of great importance for the volatility of a portfolio, due to the effect of diversification. The assets should have a correlation less than one, in order to reduce the standard deviation of the portfolio. As the correlation approaches a value of one, the standard deviation increases. Studies suggest that correlation between assets increases during extreme market periods, as businesses is highly interconnected, and one extreme result on the gain or loss side could potentially affect others and so on.[33]

If true, this change in degrees of connection between the indices, means that there is not a constant correlation throughout the dataset, but rather it should be possible to observe an increase when comparing the correlation between the indices over normal and extreme periods. To explore the correlation mechanisms among the three indices in our data sample, we have chosen to investigate the correlation in three scenarios for each index; *normal & normal*, *normal & extreme,* and *extreme & extreme*. Each correlation coefficient is calculated depending on the state of the two indices, normal or extreme.

Table 8.6: Correlation between indices and market states

SP500 & DAX30	Normal	Extreme
Normal	0.4270	N/A
Extreme	0.2630	0.6073

SP500 & NIKKEI225	Normal	Extreme
Normal	0.1072	N/A
Extreme	0.0599	0.1475

DAX30 & NIKKEI225	Normal	Extreme
Normal	0.2323	N/A
Extreme	0.2748	0.4161

The tables display a trend where the correlation is found to be considerably higher during extreme periods, compared to normal periods, yielding approximately at a ratio of 1.5 higher. Restricted to this data sample and the definition of extreme periods, this indicates that the correlation is in fact higher for extreme periods than for normal periods. The intention of choosing three worldwide indices in our portfolio was to diversify as much as possible. Indices are already diversified to a great extent, containing a wide spectre of assets. However, assets have a common factor of

[33] See [20] and [21]

geography. To address this issue, three indices from different places in the world was chosen, anticipating to diversify even more. The effect of diversification, however, is faded in these periods and thus it makes portfolio returns more extreme, than if correlation was normal. Correlation dynamics are important in explaining why extreme returns are not cancelled out by diversification. The fact that the correlation in the dataset is not constant over time, could have affected the dynamics in the portfolio returns used earlier between the normal and extreme periods. There could potentially be an acceleration of extremeness in the portfolio returns during the extreme periods, as the returns move together to a higher degree than they do over the rest of the dataset. Finally, as there is a clear increase in correlation between the indices during the previously defined extreme periods, this serves as a further confirmation of the actual extremeness of the extreme sample.

8.1.4 Conclusion

Compared to the normal sample, the extreme sample certainly seem to display the characteristics it originally was intended to have. The definition of an extreme period used to divide the returns into samples is reflected in the characteristics of the returns. As such, it is possible to explain the extreme returns in the dataset as a whole based on the extreme sample. In addition, as the sample is now confirmed to be extreme, per definition, it can as a result be used to investigate the periods leading up to each 10-day period, as pre-extreme.

As the current extreme sample is defined retrospectively, we lack a method in identifying the extreme returns in a forward-looking manner without being all-knowing about the future of the market. Therefore, the identified dynamics of the extreme periods found in this chapter serves as a useful tool when investigating the probability of identifying the extreme periods before they happen.

8.2 Dynamics Before Extreme Periods

In this chapter we will investigate the possibility to predict when to switch from a normal model to an extreme model only based on pre-extreme periods, in other words we want to use the information in the returns to predict their extremeness. This could then be an alternative to the method used in the Autonomous model, where three observations were counted which, during the 10 latest days, violated being +/- two standard deviations from the mean, before it switched from normal to extreme modelling. The downside of this method is the time lag, as switching of models is triggered after the extreme observations have in fact occurred. We will, therefore, analyse further the time leading up to the extreme periods, as we have defined them earlier, to see whether the data in the returns prior to the extreme period can indicate in any way the extreme observations coming up.

The interval before the previously defined extreme periods are analysed over the 30 trading days leading up to each extreme period in 10-day intervals. Through this we seek to identify any trend in the key statistics used earlier, to analyse the extreme periods itself.

8.2.1 Mean

When comparing the 1-10 days (-10 days) before an extreme period with the 11-20 (-20 days) and 21-30 days (-30 days) before, there appears to be a clear trend in the mean when moving towards the extreme period itself.

Table 8.7: Mean before extreme periods

	S&P 500	NIKKEI 225	DAX 30
Mean -10 days	-0.0025	-0.0028	-0.002843
Mean -20 days	-0.0018	-0.0020	-0.002095
Mean -30 days	-0.0006	-0.0008	-0.000317

As showed earlier the mean is around five times lower than the average of the normal periods. In the 30 days leading up to an extreme period on average, the mean, moves from slightly negative, but still lower than the mean of the normal sample, to a 4 to 5-fold decrease from here. Over the last 10-day period before an extreme period, the mean is only slightly higher than that of the extreme 10-day period itself.

8.2.2 Standard Deviation

Table 8.8: Standard deviation before extreme periods

	S&P 500	NIKKEI 225	DAX 30
Standard deviation - 10 days	0.0161	0.0155	0.0170
Standard deviation - 20 days	0.0135	0.0139	0.0138
Standard deviation - 30 days	0.0122	0.0112	0.0118

The table above displays an increasing deviation from 1.1 - 1.2% in the five to six weeks (30 trading days) leading up to the extreme period, to between 1.6-1.7% in the last 10-day period before the extreme period itself. As shown earlier, the standard deviation during the extreme period is even higher, around 2.2-2.4% over the three indices. The increased standard deviation in the weeks before the extreme periods shows that the market is more volatile, with larger fluctuations in the returns on a day-to-day basis, before the actual extreme observations occur. As the extreme periods themselves are defined by returns with large movements from the mean, measured in standard deviations, it is expected that the data around these periods will display some of the same characteristics as shown in table 8.8. However, it is not given that the returns before the extreme observations themselves show such a clear trend, but the fact that they do could prove useful when trying to predict when the extreme observations occur.

8.2.3 Distribution

Another clear trend in the data is the increase of negative returns compared to positive, as the extreme period is approaching.

Table 8.9: Number of negative returns before extreme periods

	S&P 500	NIKKEI 225	DAX 30
# Negative returns -10 days	5.1887	4.9589	5.0000
# Negative returns -20 days	4.9057	4.8630	5.1077
# Negative returns -30 days	4.6981	4.3836	4.8615

As seen earlier the number of negative returns during the extreme periods is higher than the corresponding number during the normal periods, with 5.1-5.5 negative returns on average observed during the extreme periods, to approximately 4.5 in the normal sample. As the extreme period comes closer, the number of negative returns seems to follow this trend, increasing from 4.4 - 4.9 to 5.0 - 5.2 over the three indices. Once again it is observable, as was the case for both the mean and standard deviation, that an extreme period is approaching, in terms of the statistical properties of the returns gradually getting closer to the data observed during the 10-day extreme period itself.

8.2.4 Range

When investigating the range of the returns over the 30 days, a trend is found, corresponding well to the increased standard deviation.

Table 8.10: Range before extreme periods

	S&P 500	NIKKEI 225	DAX 30
Range -10 days	0.0516	0.0499	0.0524
Range -20 days	0.0438	0.0448	0.0427
Range -30 days	0.0389	0.0359	0.0373

The ranges grow consistently as the extreme period is closing in, increasing roughly 1.5 times from trading day 30-21 to trading day 10-1. It seems as higher standard deviations in pre-extreme periods are directly reflected in the range for the same periods. This increased range could be caused by negative extreme values only, but as seen in table 8.11 this is not the case. Both minimum values and maximum values becomes higher in absolute terms, as extreme periods are approaching, indicating that the market dynamics become more unstable for both negative and positive returns.

Table 8.11: Min and max returns before extreme periods

	S&P 500	NIKKEI 225	DAX 30
Min -10 days	-0.0280	-0.0278	-0.0285
Min -20 days	-0.0238	-0.0241	-0.0233
Min -30 days	-0.0200	-0.0201	-0.0191
Max -10 days	0.0236	0.0221	0.0239
Max -20 days	0.0201	0.0207	0.0193
Max -30 days	0.0189	0.0158	0.0183

The larger fluctuations in returns observed through the increased standard deviation naturally leads to greater variation in the minimum and maximum values. The increased range, minimum and maximum values, although increasing over the 30 days, are still far lower than the corresponding observed values during the extreme period. As seen before, the range is between 7% and 8% during the 10-day extreme periods on average. Given the observed higher standard deviation during these periods, this is natural. As such, there seems to be a consistency in the observations of both range and standard deviation during and before the extreme periods.

The key takeaway from the above numbers is that extreme periods are not only getting closer with respect to time, but also in trend. It seems that the closer we get to an extreme 10-day period, the more we approximate the actual returns found in it. This could, if we are able to isolate these effects from the noise of the day-to-day volatility in the market mean that we, prospectively could define when an extreme period is coming. However, as these effects are observed as a gradual approximation to the extreme period, it could be difficult to pick up on the changes on a day-to-day basis, as there are no definite signs that an extreme period is actually approaching, rather a trend in the data. It is this change in trend, moving from a normal period towards an extreme, we will try to seek out when defining the parameters to determine if an extreme period is approaching.

8.2.5 Switching Procedure

An important factor when establishing parameters to predict the extreme periods before they happen, in order to switch from the normal to extreme model, is to avoid "false alarms", meaning that models should not be switched if there turns out to be no basis to do so. To use the extreme model in a normal state would imply an overestimation of the risk. We will therefore try to minimise the probability of switching to extreme modelling without cause, by using two different parameters to decide whether an extreme period is approaching, only shifting model when both statements are true.

We have seen that there is a significant predominance of negative returns in the two weeks before the extreme periods, on average. Further, there is an increased range in the returns over the same 10 trading days, as there are larger minimum and maximum values in the observations, indicating a higher standard deviation at the same time. As both the number of negative returns and the increased range of returns are clearly observed trends and entirely separate observations, which both are indications on their own, the two are suited as parameters when choosing if the extreme model should be used.

The number of negative returns have proven to be significantly higher when approaching an extreme period than over the rest of the dataset. The higher number of red trading days reflects both the negative mean and skewness of the returns over the 10-day period. Increased range reflects the greater fluctuations in the returns, while also accounting for the magnitude of the returns being large, not only as a factor of the mean, but being large enough to be "extreme" measured on their own merit. This is a potential mitigation of including extreme periods which are only a result of the market over the last year, rather than taking into account its extremeness in a larger historic perspective, as observed in the box-plot earlier.

Using the number of negative returns as selection criteria when choosing between normal and extreme modelling, could potentially affect the results by implicitly excluding extreme positive return periods, not containing 5 or more negative returns. However, as shown in this chapter, pre-extreme periods for the extreme sample is largely characterised by a magnitude of negative returns and thus we argue that this is an appropriate constraint to include.

8.2.6 Conclusion

Combined, the number of negative returns and increased range size, provide a robust measure of extremeness approaching. The two constraints together exclude periods of negative consolidation in the market, where several marginally negative returns occur, by ensuring that the range of the returns in the 10-day period has to be large enough. Further, the two combined also increases the probability of accurate results through ensuring that the range of the returns is not affected by one or two outliers increasing the range, in including the criterion of a predominance of negative numbers, proved to be the case when approaching an extreme period. In addition, periods where only one of the two is true is defined as normal. This ensures that the number of periods where the extreme model is applied is not too high, leading to a smaller number of false alarms.

With the parameters to notify if an extreme period is approaching in place, the setup of the Autonomous model from chapter 6 will be modified to switch prospectively in the next section.

8.3 Pre-Extreme Switching of Model

The value-at-risk and expected shortfall is now modelled on a basis of the switch between normal and extreme, i.e. student's t-GARCH and conditional POT, happening before the actual extreme observations occur. The model does its calculations on a day-to-day basis, meaning it estimates when to switch based on the properties of the last 10-day period each day. When the requirements of switching to extreme modelling is fulfilled, the model uses extreme modelling the next trading day. In theory this should mean that the model better takes into account the characteristics of the extreme observations as it then, theoretically, applies the correct model to all normal and extreme returns. However, we must expect the model to falsely predict extreme returns approaching, and vice versa, model actual extreme returns as normal. This is only natural, as the parameters cannot correctly take into account every extreme case without being too tightly defined. The criteria for switching to the extreme modelling of VaR and ES is set to be a range greater than 2.5% and over five negative returns observed over the last 10 trading days.

The definition of an extreme period, and as a consequence a normal period, is the same as applied earlier in the thesis. The reasoning for using the same approach in deciding the normal and extreme subsamples is twofold; first, if the subsamples are defined by using the pre-extreme parameters, it is not certain whether the actual extreme observations are modelled, as a forward-looking measure is used, rather than a retrospective measure deciding the extremes. As such, the risk of modelling the distribution parameters and then later VaR and ES on the wrong returns is present, since the pre-extreme forward-looking definition in no way can hit the extreme returns as accurately as the original, retrospective definition. Second, we want consistency in the models, and using pre-extreme parameters would mean we would not be able to compare the different models. It is the application of the VaR and ES models that is important, as the successfulness of the different approaches is measured based on their respective results. Further, we have seen in chapter 8.1 that the original definition of the extreme sample is a good fit, and as such it is the switch between the different VaR and ES models that lacks in hit rate.

8.3.1 Autonomous Model

The Autonomous model is still realistic in terms of actual usage, only using information from the past in determining the choice of how to model the next day. As such the trade-off compared to the earlier applied, all-knowing, models is clear, and a worsened estimation of the risk compared to these is expected. Therefore, it only makes sense to compare the results of this approach to the Autonomous model presented in chapter 6, where the switch between normal and extreme VaR calculations were done after three extreme observations were observed. As a consequence, we expect, predicting extreme observations in advance, to improve the accuracy of the model – given that actual prediction is succeeding.

8.3.1.1 Original Sample

When applying this procedure on the original sample, using pre-extreme parameters to decide when to switch between normal and extreme modelling, the Autonomous model yields the following results for the extreme periods:

Table 8.12: Violation rates using pre-extreme switching procedure [34]

	Extreme sample			
	VaR-	VaR+	ES-	ES+
Violations	27	15	13	7
Violation rate	4.58%	2.54%	2.20%	1.19%
Sample size	590	590	590	590

[34]Appendix table 7 and 8: Violation rates for the complete and normal sample.

The violation rate of the VaR on both the loss and gain side is quite high. This means the model underestimates the risk quite severely, as we have far more violations than the wanted 1% level. However, compared to the earlier presented Autonomous model, which switches from normal to extreme retrospectively, the results have improved. Here, the VaR violation rate was 5.4% and 2.9% for the left and right side of the distribution respectively. Although small, the fact that the results have improved could suggest that extreme observations are modelled more correctly now, by predicting when these returns occur before they do. The continued high number of violations suggest that modelling VaR when lacking the answer to when the extreme observations actually occur is problematic. The expected shortfall numbers show the same trend as the VaR estimates when compared to the earlier tested Autonomous model. The ES violation rates are down for both the loss and gain side, compared to the corresponding figures of 2.7% and 1.5% found earlier, when using the retrospective switching procedure in the Autonomous model.

8.3.1.2 Out-of-Sample

To further strengthen the robustness of the results, we apply the same model and methodology on the out-of-sample previously presented. The logic behind using this sample is the same as earlier, we wish to test the model on a dataset in which the model is not perfectly fitted on. As the pre-extreme information is based on information in the original sample, the results above is somewhat biased. Therefore, by using an out-of-sample we can either be able to further generalise the results above, or have to limit the results to the particular data they are fitted on.

Table 8.13: Violation rates using pre-extreme switching procedure [35]

	Extreme sample			
	VaR-	VaR+	ES-	ES+
Violations	19	3	7	1
Violation rate	4.42%	0.70%	1.63%	0.23%
Sample size	430	430	430	430

The VaR out-of-sample estimates showcase approximately the same trend as for the original sample on the loss side. The VaR+ has a much lower violation rate, indicating that the model overestimates the VaR on the gains side. Once again it makes sense to compare these results with the Autonomous out-of-sample numbers found earlier, where the VaR for the left and right side were 5.4% and 1.6% respectively. This

[35] Appendix table 9 and 10: Violation rates for the complete and normal sample.

means that the difference between the two models is very similar to the differences found for the original sample. Through applying the new model, the VaR estimates are improved on both sides of the distribution as the numbers are closer to the preferred 1% violation rate. The expected shortfall figures suggest the same trend in the violation rates.

The fact that the model performs as similar on the out-of-sample as the original sample, showcasing the same positive trend when moving to the pre-extreme parameter model versus the original Autonomous model, suggests the definition applied to predict the extreme observations might work. The results indicate an improvement of estimates for both VaR and ES, when switching based on the prediction of extreme observations coming up, and the fact that this is the case for the two independent samples supports this.

As we saw in subchapter 8.1, between one-third and one-fourth of the extreme periods are continuous, as a result we would expect to miss out on modelling quite a few extreme observations correctly when switching retrospectively. Then, given that we are able to predict the extreme observations better than the number of returns we miss out on due to switching too late, we would expect the new Autonomous model, using pre-extreme parameters, to produce better estimates than the old model. As we have seen above, this seems to be the case, and as such we can conclude that the methodology and parameters chosen to predict the extreme observations are somewhat accurate. However, the violation rates are still high, especially for the loss side of the distribution, meaning that the prediction, and as a result, switching of models between normal and extreme is off, in particular from the almost perfect results obtained when we know the timing of the observations. Still, this is expected due to the difficult nature of the properties of returns, and especially the extreme observations. We would not expect a model to perfectly hit the 1% violation rate when it is based on realistic terms, in the meaning that it is estimated as a model that knows nothing about the next day it will model. As such, we should get the same results if we on a day-to-day basis today apply the model on the next day and saves the result continuously. If we were able to get a perfect hit rate then, the model would be perfect as a risk management tool, which is an unrealistic thought.

8.4 Concluding Remarks

The extreme period investigation in this chapter has given insight into the properties of the returns defined as extreme throughout the thesis. The findings seem to support the actual extremeness of these returns when compared to the dataset as a whole. Further, we were able to incorporate the use of information obtained about the properties of the dynamics during the extreme periods with an analysis of the dynamics before the extreme periods. Together they serve as a way to predict the extreme observations with a certain degree of success compared to the earlier applied retrospective method.

For a portfolio manager, the high failure rate of the somewhat realistic Autonomous model is worrying. We wanted to improve the failure rate of the model found in chapter 6 by trying to predict the timing of the extreme observations in the sample. We were able to lower the failure rate by about 0.5 percentage points in the original sample, but the violations for the VaR is still about 4.6% for the loss side of the distribution, meaning that the underestimation of risk in the model will lead to higher than expected losses.

The difficulties in both modelling and predicting extreme returns is no surprise. They are extreme and causes severe losses for a portfolio manager for a reason – they cannot be anticipated and when they eventually happen they hit hard. A final violation rate of approximately 5% for VaR and 2% for ES on the loss side of the distribution suggests that using the Autonomous model will lead to no different results.

Final Remarks

Our thesis has tested a Gaussian Basel in order to see how well it estimates risk, through the performance of value-at-risk and expected shortfall. A well-diversified portfolio is set, consisting of three geographically spread stock indices. Moreover, the portfolio is found to consist of several periods with persistent increased volatility, and is as a result divided into a normal and an extreme sample, to identify how the Basel framework performs under different market conditions.

The most noticeable errors of the Basel framework are during extreme market periods, where risk level estimates exceed the 99% VaR and ES too often. Implementation of ES should intentionally provide more information about tail events, however the output is only as good as the input; if the underlying historical data and Gaussian distribution give little information or underestimates probability of tail events, ES will not give adequate estimates of risk exposure. The underestimation of risk is a result of the bad fit between the empirical data and the assumption of a Gaussian distribution. This results in a higher frequency of extreme returns in the dataset, compared to the corresponding theoretical probability of extreme returns derived by a Gaussian distribution. It is further found, that the use of one year equally weighted mean and standard deviation in the Basel framework leads to slow risk measures not capable of incorporating new market information sufficiently quick, such that the current state of the market is not included in the estimation to a large enough degree. The result is an overestimation of risk during less volatile periods, while the risk is underestimated when markets turn more volatile.

The market characteristics of the normal and extreme samples are found to differ. The shape and distribution of both subsamples is different from a Gaussian distribution, and normality is rejected on a statistical significant level. A student’s t-distribution is found to be a good fit for the normal sample, as it takes into account the heavy tails and negative skewness it consists of. Further, extreme value theory and a generalised Pareto distribution is deemed to be able to describe the extremeness of the extreme sample by explaining the large returns found in the tails of the subsample accordingly. Moreover, both samples are characterised by volatility clustering, which suggests that the variance of the returns is conditional of time. This is dealt with by the implementation of a volatility model – GARCH, to increase the adaptability to the current state of the market.

The choice of distributions and implementation of conditional volatility modelling results in a student's t-GARCH model for the normal sample. For the extreme sample, extreme value theory is employed through a conditional peak-over-threshold model. Both are found to provide clear improvements to the Basel framework estimation of VaR and ES. The two models are, in contrast to Basel, able to incorporate and adapt to new market information quickly, while at the same time incorporate the tail risk found to be present in both subsamples.

To move from a theoretical setup to a real life applicable approach, the two models are combined into the Autonomous model. Now, the switch between normal and extreme modelling is retrospective, and in addition the fitting of distributions is done continuously from day-to-day. As such, the assumptions of student's t-GARCH and the c-POT model are tightened. It is found that this results in less accurate estimates, in particular for the extreme returns, when compared to the results of the two theoretical models separately. Yet, the Autonomous model outperforms the Basel framework significantly both on the original sample and in an out-of-sample test on three Nordic stock indices.

As a consequence of the underestimation of risk for extreme observations, the extreme periods itself and the time leading up to them is analysed further. It is found, through the analysis of pre-extreme periods, that the switching procedure can be optimised compared to the retrospective method. By using two parameters, the number of negative returns observed and the range of the returns, over the last 10 days, qualified guesses are performed to anticipate extreme returns approaching. Even though the results of VaR and ES is found to be improving, the violation rate is still high, implying a systematic underestimation of risk in the model.

The Autonomous model is, overall, reckoned to be more applicable and appropriate for a portfolio manager to use. The risk is far more accurately estimated than the standard Basel framework for both normal and extreme periods. However, the violation rate during extreme periods for the Autonomous model is still too high, and of concern for its ability to provide sufficient coverage during non-normal market conditions. Nevertheless, for a portfolio manager, the improvement compared to the Basel framework is still directly transferable to better risk management potential by generating far fewer losses. The enhancement in risk measuring could provide a larger upside, as the model, in addition to fewer losses, is more adjusting to the current state of the market compared to the Basel framework. All in all, this results in less overestimation of risk exposure and less violations of risk limits.

Bibliography

[1] Basel Committee. "Fundamental review of the trading book: A revised market risk framework." *Consultative Document, October* (2013).

[2] Danielsson, Jon, and Casper G. De Vries. "Value-at-risk and extreme returns." *Annales d'Economie et de Statistique* (2000): 239-270.

[3] Yamai, Yasuhiro, and Toshinao Yoshiba. "Value-at-risk versus expected shortfall: A practical perspective." *Journal of Banking & Finance* 29.4 (2005): 997-1015.

[4] Casella, George, and Roger L. Berger. *Statistical inference*. Vol. 2. Pacific Grove, CA: Duxbury, 2002.

[5] Clauset, Aaron. "A brief primer on probability distributions." *Santa Fe Institute. http://tuvalu. santafe. edu/~ aaronc/courses/7000/csci7000-001_2011_L0. pdf* (2011).

[6] Basel Committee. " Amendment to the Capital Accord to incorporate market risks." *November* (2005).

[7] Basel Committee. "Guidelines for computing capital for incremental risk in the trading book." (2009).

[8] Basel Committee. " Fundamental review of the trading book." (2012).

[9] Basel Committee. "Fundamental review of the trading book: A revised market risk framework." *Consultative Document, October* (2013).

[10] Jarque, Carlos M., and Anil K. Bera. "Efficient tests for normality, homoscedasticity and serial independence of regression residuals." *Economics letters* 6.3 (1980): 255-259.

[11] Mandelbrot, Benoit B. "The variation of certain speculative prices." *Fractals and Scaling in Finance*. Springer New York, 1997. 371-418.

[12] Ljung, Greta M., and George EP Box. "On a measure of lack of fit in time series models." *Biometrika* (1978): 297-303.

[13] Brooks, Chris. "Introductory Econometrics for Finance, 2008."

[14] Danielsson, Jon, and Casper G. De Vries. "Value-at-risk and extreme returns." *Annales d'Economie et de Statistique* (2000): 239-270.

[15] Embrechts, Paul, Sidney I. Resnick, and Gennady Samorodnitsky. "Extreme value theory as a risk management tool." *North American Actuarial Journal* 3.2 (1999): 30-41.

[16] Daníelsson, Jón. *Financial risk forecasting: the theory and practice of forecasting market risk with implementation in R and Matlab*. Vol. 588. John Wiley & Sons, 2011.

[17] Levine, Damon. "Modeling tail behavior with extreme value theory." *Risk Management* 17 (2009): 14-18.

[18] Hosking, Jonathan RM, and James R. Wallis. "Parameter and quantile estimation for the generalized Pareto distribution." *Technometrics* 29.3 (1987): 339-349.

[19] McNeil, Alexander J. "Extreme value theory for risk managers." *Departement Mathematik ETH Zentrum* (1999).

[20] Longin, Francois, and Bruno Solnik. "Extreme correlation of international equity markets." *The journal of finance* 56.2 (2001): 649-676.

[21] Das, Sanjiv R., et al. "Common failings: How corporate defaults are correlated." *The Journal of Finance* 62.1 (2007): 93-117.

Appendix

List of Appendix Tables

Appendix table 1:

Akaike-, Bayesian-, Schwarz- and Hannan-Quinn Information Criterion for different p and q. Average results of the four models are presented below.

Students t-GARCH		q		
	(p,q)	1	2	3
p	1	7.0969	7.0968	7.0966
	2	7.0963	7.0967	7.0962
	3	7.0958	7.0962	7.0957

Appendix table 2:

VaR violation rates student's t-GARCH(1,1)

	Complete sample		Normal sample		Extreme sample	
	VaR-	VaR+	VaR-	VaR+	VaR-	VaR+
Violations	126	108	93	89	33	19
Violation rate	1.27%	1.09%	1.00%	0.96%	5.59%	3.22%
Sample size	9900	9900	9310	9310	590	590

Appendix table 3:

ES violation rates student's t-GARCH(1,1)

	Complete sample		Normal sample		Extreme sample	
	ES-	ES+	ES-	ES+	ES-	ES+
Violations	41	41	27	35	14	6
Violation rate	0.41%	0.41%	0.29%	0.38%	2.37%	1.02%
Sample size	9900	9900	9310	9310	590	590

Appendix table 4:

Generalised Pareto Distribution parameters for each tail of the distribution, found by maximum likelihood estimation

GPD Parameters	β	ξ	μ	q 99%
Negative returns	7.87E-01	8.23E-02	4.09E-03	3.62E+00
Positive returns	6.31E-01	4.17E-02	4.50E-03	2.84E+00

Appendix table 5:

VaR violation rates conditional-POT

	Complete sample		Normal sample		Extreme sample	
	VaR-	VaR+	VaR-	VaR+	VaR-	VaR+
Violations	38	54	29	49	9	5
Violation rate	0.38%	0.55%	0.31%	0.53%	1.53%	0.85%
Sample size	9900	9900	9310	9310	590	590

Appendix table 6:

ES violation rates conditional-POT

	Complete sample		Normal sample		Extreme sample	
	ES-	ES+	ES-	ES+	ES-	ES+
Violations	13	17	7	17	6	0
Violation rate	0.13%	0.17%	0.08%	0.18%	1.02%	0.00%
Sample size	9900	9900	9310	9310	590	590

Appendix table 7:

Original sample VaR violation rates using pre-extreme switching procedure

	Complete sample		Normal sample		Extreme sample	
	VaR-	VaR+	VaR-	VaR+	VaR-	VaR+
Violations	133	115	106	100	27	15
Violation rate	1.34%	1.16%	1.14%	1.07%	4.58%	2.54%
Sample size	9900	9900	9310	9310	590	590

Appendix table 8:

Original sample VaR violation rates using pre-extreme switching procedure

	Complete sample		Normal sample		Extreme sample	
	ES-	ES+	ES-	ES+	ES-	ES+
Violations	44	47	31	40	13	7
Violation rate	0.44%	0.47%	0.33%	0.43%	2.20%	1.19%
Sample size	9900	9900	9310	9310	590	590

Appendix table 9:

Out-of-sample VaR violation rates using pre-extreme switching procedure

	Complete sample		Normal sample		Extreme sample	
	VaR-	VaR+	VaR-	VaR+	VaR-	VaR+
Violations	92	53	73	50	19	3
Violation rate	1.42%	0.82%	1.20%	0.82%	4.42%	0.70%
Sample size	6490	6490	6070	6070	430	430

Appendix table 10:

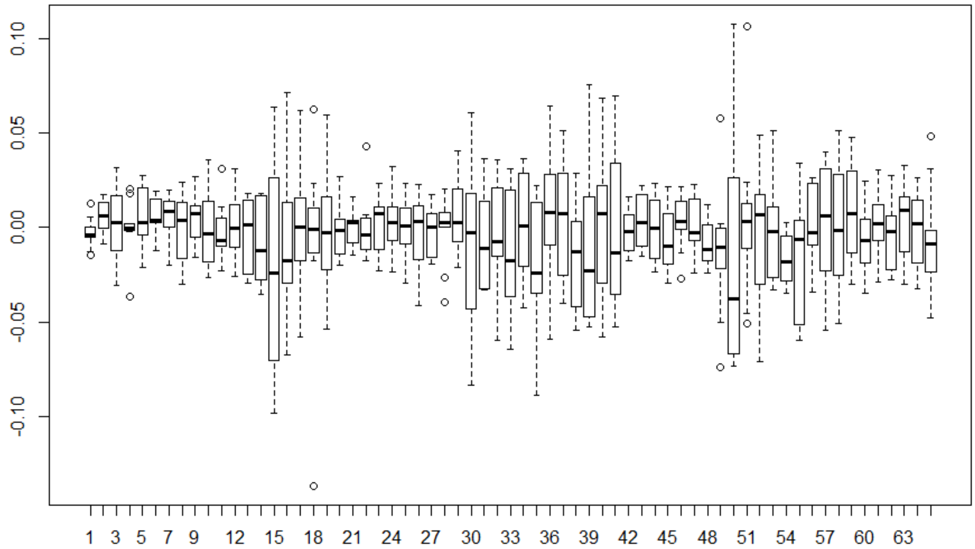
Out-of-sample ES violation rates using pre-extreme switching procedure

	Complete sample		Normal sample		Extreme sample	
	ES-	ES+	ES-	ES+	ES-	ES+
Violations	32	17	25	16	7	1
Violation rate	0.49%	0.26%	0.41%	0.26%	1.63%	0.23%
Sample size	6490	6490	6070	6070	430	430

List of Appendix Figures

Appendix figure 1:

Box plot DAX 30



Appendix figure 2:

Box plot NIKKEI 225