LUND UNIVERSITY
School of Economics and Management

Department of Economics

Master's programme in Finance

NEKN02 Master Essay

Spring 2020

An Empirical Study: Expected Shortfall Estimation Methods for a Bank's Trading Book

Authors:
Emilia Johansson
Laura Emina Ludolphy

Supervisor:
Birger Nilsson

Date essay seminar:
June 4, 2020

Abstract

This thesis investigates methods that estimate the Expected Shortfall correctly by passing the Acerbi-Szekely (2014) backtest in both stressed and calm periods. This backtest is added to in this thesis to test against both under- and overestimation of ES. This research is relevant due to the recent shift in the Basel's Fundamental Review of the Trading Book from using Value-at-Risk as the leading risk measure to ES. The tests are performed on five hypothetical portfolios, which represent weighted asset classes and options hold in an actual bank's trading book. The tested methods are a result of the literature reviewed; namely: Normal distribution, Student's t-distribution and skewed Student's t-distribution, each with both constant volatility and GARCH(1,1), BHS and VWHS with GARCH(1,1), as a filtered form of BHS.

From a quantitative analysis, the result of this study indicates that the method with Student's t-distribution with GARCH volatility performed superior throughout both periods compared to the other methods. Secondly, the in literature supported approach of VWHS with GARCH correctly estimated ES. In particular for the calm period, the Normal distribution with constant volatility received non-rejections for four out of six tested years. An interesting result to emerge from this study is the dependence between in-sample periods and the testing year regarding the degree of similarity in the level of standard deviation, which is of great influence across all methods' test statistic results.

Keywords: Expected Shortfall, Trading Book, Student's t-distribution, GARCH(1,1), Volatility Weighted Historical Simulation

Acknowledgements

We would like to thank our supervisor Mr. Birger Nilsson for providing continuous support and valuable inputs in the fulfilment of this thesis, which is written in partial fulfilment of the degree Master of Science in Finance at Lund University School of Economics and Management.

Table of Contents

List of Tables

List of Figures

Abbreviations

VaR	Value-at-Risk
ES	Expected Shortfall
GARCH	Generalized Autoregressive Conditional Heteroskedasticity
EWMA	Exponentially Weighted Moving Average
BHS	Basic Historical Simulation
VWHS	Volatility Weighted Historical Simulation
EVT	Extreme Value Theory
CAViaR	Conditional Autoregressive Value-at-Risk
CARES	Conditional Autoregressive Expected Shortfall
GPD	Generalized Pareto Distribution
ML	Maximum Likelihood
FRTB	Fundamental Review of the Trading Book
BIS	Bank for International Settlements
BCBS	Basel Committee of Banking Supervision

1 Introduction

1.1 Background

The probability of a bad outcome and its consequence or in other words, the risk, is a major concern for the human being. Individuals, businesses, organisations, institutions and governments are all exposed to risk in various ways. The financial risk is the risk associated with financing and is particularly applicable to investors, banks and financial institutions. The banks portfolios are exposed to market risk, which is a type of financial risk, and it can be defined as "the risk of losses resulting from changes in the prices of instruments such as bonds, shares and currencies" (Bank for International Settlements, p.1, 2019a).

Value-at-Risk (VaR) and Expected Shortfall (ES) are statistical risk management metrics used to measure and evaluate the level of market risk of an investment. For the Bank for International Settlements (BIS) and the Basel Committee on Banking Supervision (BCBS), a continuing concern is to investigate and improve the standards for banks to measure the market risk of a bank's trading book and to determine the capital requirements for it. In the past and the current Basel accords, the 10-day VaR at the 99^{th} percentile one-tailed confidence level has been used as the metric to capitalize the loss of the trading book.

The 2008 financial crisis has highlighted the relevance of effective risk measures for banks by regulations. In the light of the financial crisis, the BCBS assessed the current market risk framework and recognized deficiencies related to the framework and its properties, which became apparent when it was exposed to the adverse price movements during the financial crisis. In May 2012, the BCBS proposed the *Fundamental Review of the Trading Book* (FRTB), which is the new market risk framework to strengthen the regulatory capital standards (Bank for International Settlements, 2012). One of the amendments related to the introduction of the FRTB is the shift from VaR to ES as the standardised measure of risk to determine the capital requirements of the trading book. In the FRTB, the BCBS supports the shift by the findings in the academic literature about VaR's disfavouring effect on diversification and ignorance of tail events beyond the 99^{th} percentile of the loss distribution.

The new quantitative standards for calculating the capital requirements by using Expected Shortfall, as clarified by the BCBS in the publication *Minimum Capital Requirements for Market Risk* from January 2016, is that the ES is to be computed on 10-trading day horizon at the 97.5th percentile one-tailed confidence level.

1.2 Research Problem

The change towards a new market risk framework, namely the Fundamental Review of the Trading Book suggested by the Basel Committee on Banking Supervision, prompts banks to change from Value-at-Risk to Expected Shortfall for measuring the market risk to subsequently determine the capital requirements. The remaining issue is how to estimate ES correctly such that it is not underestimated by the implemented method. As VaR used to be the standardized leading risk measure, a considerable amount of literature and studies have grown around its estimation methods, such as distributional assumptions and volatility modelling. Thus far, the research around ES estimation methods is significantly less expanded. Existing research has demonstrated that distributions accounting for excess kurtosis for instance the Student's t-distribution, volatility models representing market conditions including volatility clustering such as GARCH or EWMA and extreme loss models that can capture rare events such as Extreme Value Theory (EVT) give correct estimates for VaR.

To start with, as regards to the tail risk, ES is a better risk measurement since it considers the losses beyond VaR (Yamai & Yoshiba, 2005). Regarding the estimation methods for ES, similar starting points as for VaR have been found. Two decades ago, McNeil and Frey (2000) showed that disregarding the stylised facts of financial data leads to poor ES estimates and consequently they suggested to consider a leptokurtic distribution assuming a Student's t-distribution instead of the traditional Gaussian Normal distribution. Up to now, research comparing several ES estimations among different asset classes highlighted the factors that are associated with well-performing methods, which are the skewed Student's t-distribution as well as the conditional parametric models with the implementation of the GARCH(1,1) and the filtered versions of Historical Simulations (FHS) methods are superior methods to estimate ES, for all types of asset classes (Righi & Ceretta, 2015). Furthermore, it is now well established from several studies that confirmed the effectiveness of EVT is the ES estimation procedure (McNeil & Frey, 2000; Harmantzis, Miao & Chien, 2006; Chinhamu, Huang, Huang & Chikobvu, 2015 & Bah, Munga'tu & Waititu, 2016, EVT in combination with a GARCH (1,1) (McNeil & Frey, 2000 & Bah, Munga'tu & Waititu, 2016), and approximated with the

Generalized Pareto Distribution (GPD). Another important concern is to evaluate VaR and ES estimate by adopting a backtesting methodology. The backtesting method which has been widely employed in the literature is a backtesting method developed by McNeil and Frey (2000), which requires to perform Monte Carlo simulations. More recently, Acerbi and Szekely (2014) proposed three backtesting methods and one of the tests can be used to backtest ES directly. Therefore, it is more convenient as compared to performing a Monte Carlo simulation.

Continuous effort and research advancements regarding stylised facts in financial data increase the need of revising what is currently believed to be accurate methods. Estimating ES correctly is a key aspect for banks as otherwise their market risk and thus, capital requirements are either under- or overestimated. An underestimated ES evaluates the overall risk in the trading book portfolio as too low, which is not an adequate property of a risk measure and can result in a capital charge by Basel. This occurs because in that case banks do not fulfil the requirements set by Basel and thus, will be punished for it. On the other hand, while Basel solely is concerned with underestimation of risk, banks also do not want to overestimate ES as this would consequently result in a higher capital reserve to hedge against the risk of the trading book, which would be inefficient capital allocation. Both under- and overestimation should be avoided and thus, estimating ES using an estimation method yielding an accurate ES estimate is of major concern.

As discussed before, to date, there has been little quantitative analysis of ES estimation methods, especially on its estimation on a bank's trading book instead of single indices as implemented by previous studies. Hence, an attempt to estimate ES using loss observations of several indices, representing an all-in portfolio consisting of different asset classes and an equity option, will broaden the scientific knowledge around the ES estimation methods and create a basis for further research. The importance and originality of this study are that is provides new insights into which methods estimate ES on a bank's trading book under the in Basel determined confidence level correctly, which is evaluated by backtesting it against both under- and overestimation and thus, arrive at appropriate recommendations for banks to implement.

1.3 Statement of Research Objective

In the light of the FRTB from Basel, the objective of this thesis is to identify the estimation methods that estimates the Expected Shortall on a bank's trading book correctly for both stressed and calm periods, such that it is neither under- nor overestimated according to the "second backtest" by Acerbi and Szekely (2014).

It should be noted that a correct method in this paper refers to the estimation method that is not rejected in the backtest and thus, approximates ES correctly. This is the only criteria and other criterions such as level of difficulty and speed of implementation, level of difficulty of explaining the method, etcetera, do not determine a correct method in this study. Throughout the dissertation, the term "methods" will refer to the different methods that are tested, for instance parametric versus non-parametric approaches, distributional assumptions and volatility modelling as well as any further in literature or through the tests as relevant established factors of influence when estimating ES. The empirical analysis is performed from the perspective of a U.S. bank's trading book. Consequently, the data consist of eight U.S. Dollar denominated indices, which are aggregated into five different portfolios to represent the trading book of a U.S. bank and the analysis is conducted on daily log-loss observations obtained from the daily closing prices of the indices. The ES is estimated one day ahead at the 97.5^{th} percentile one-tailed confidence level in accordance with the new Basel standards, namely Basel III and FRTB. The ES is estimated using three years of data for a "stressed period" and a "calm period". The stressed period refers to the years 2008, 2009 and 2011 while the calm period is 2014, 2017 and 2019.

As a result of the literature review, the methods tested are the Gaussian Normal distribution, Student's t distribution and skewed Student's t-distribution with both GARCH(1,1) volatility and constant volatility, Basic Historical Simulation (BHS) and Volatility Weighted Historical Simulation (VWHS) with GARCH (1,1). The second backtest by Acerbi and Szekely's (2014) is used to evaluate the ES estimates. This test usually solely tests against underestimation; however, as the relevance of not overestimating ES for banks has been explained before, this backtest is expanded to test both sides.

1.4 Research Questions

Main Research Question:

Which methods estimate the Expected Shortfall on a bank's trading book correctly for both stressed and calm periods, such that it is neither under- nor overestimated according to the "second backtest" by Acerbi and Szekely?

Sub-Research Questions:

- SQ 1: Using existing studies, which, in literature established, methods estimate ES correctly?
- SQ 2: How can ES be backtested in a consistent way in order to account for under- and overestimations of ES according to the "second backtest" by Acerbi and Szekely?
- SQ 3: What is an appropriate representative portfolio for a bank's trading book?
- SQ 4: Which methods estimate ES correctly resulting from the tests applied in this research?
- SQ 5: To which extent do methods estimate ES similarly or differently in stressed versus calm periods?

1.5 Delimitations

There are several limitations as regards to the scope of the thesis and in which extent previously used methods can be applied in terms of the limited time frame. It is beyond the scope of this study to examine and provide a comprehensive review of all possible estimation and backtesting methods. In this thesis, the methods that are considered most relevant through established literature are applied and tested. This area leaves room for further research. Due to the timely restriction put on this thesis, the EVT approach is left for further research to determine whether it estimates ES correctly and how it performs in comparison to the methods applied in this research.

This study is unable to encompass the entire complexity of constructing a representative portfolio of what banks hold in their trading book. It is not practical to build a close to reality portfolio as this differs between banks and changes at times. This was counteracted as good as possible by constructing a representative trading book as this study's testing portfolios with several asset classes as well as an option strategy based on the suggestions by the Basel Committee on Banking Supervision.

The liquidity-adjusted ES, originated in the FRTB, is not implemented in the thesis. The FRTB considers the risk of market illiquidity by incorporating varying liquidity horizons in the calculation of the regulatory metric for market risk (Bank for International Settlements, 2012). As a result, an equation to derive the regulatory liquidity-adjusted ES is constructed (Bank for International Settlements, 2016). This is left out of this study as it requires the calculation of the ES on various sub-portfolios since different assets classes have different liquidity horizons. Hence, this is beyond the scope of this thesis.

Moreover, the data selection is limited to indices that are introduced and initiated for trading at latest in 2000-01-01 up until 2020-04-01, are denominated in U.S. Dollars and are available on either Datastream or Bloomberg.

1.6 Outline of the Thesis

The remainder of the thesis is composed in six themed sections. In Section 2, the theoretical frameworks and literature review is discussed by starting with defining VaR and ES and showing superiority of ES over the drawbacks of VaR. The in literature as best defined methods for estimating VaR are reviewed here. Afterwards, there is a critical comparison of different ES estimation methods based on existing studies and literature. Further, there is a discussion about how to backtest ES. This part forms the basis for the choice of data and methodologies used in order to perform achieve the objective of this study. Section 3 includes the methodology, in which the data from the representative portfolio is displayed, followed by how to implement the in Section 2 introduced ES estimation methods and backtesting method. In Section 4, the results are discussed and analysed before drawing a conclusion and suggesting recommendations for areas of further research and recommendations for practice of estimating ES in Section 5 and 6, respectively.

2 Theoretical Framework & Literature Review

2.1 Value-at-Risk and Expected Shortfall

2.1.1 Definition Value-at-Risk

Value-at-Risk (VaR) is a risk measure that incorporates the loss variable of a portfolio, which can be either derived from prices or returns. The latter is beneficial as its stationarity and stochastic assumption is more realistic compared to asset prices. Linear returns are defined in the following way:

$$L\,(t) \;=\; P(\,t) \;-\; P(t-1),$$

where P is the price over the period (t, t-1).

Similarly, the log-return over the holding period is defined as follows:

$$r(t\,) \;=\; ln\frac{P(t)}{P\,(t-1)}\;.$$

Using log-returns over arithmetic, linear returns is beneficial in terms of the appliance of the additivity rule and the commonly in finance assumed Gaussian distribution properties (Ballotta & Fusai, 2017). These returns can be transformed into losses by changing the sign.

VaR refers to a minimum loss l, which requires that the probability of a future loss of the portfolio L is greater than this minimum loss is smaller or equal to $1 - \alpha$, where α refers to the confidence level ($\alpha = 0.99$ in the old Basel regulation). VaR is defined in the following equation:

$$VaR_{\alpha}(L) \;= \min\{\, l\!: Pr\,(L \;>\; l\,) \;\leq\; 1 \;-\; \alpha\,\} \qquad (1)$$

The popularity of the usage of VaR comes from its simple calculation applicable to all asset classes and straightforward interpretation (Yamai & Yoshiba, 2005). Furthermore, it used to be the leading risk measure required under Basel to establish capital obligations for banks.

There are several drawbacks when using VaR. One problem is that it solely measures the percentiles of a loss occurring while overlooking the size of a loss beyond that level of VaR. In other words, it fails to detect the tail risk of the probability distribution of future profit and losses (BIS, 2000). To give an example, there might be a 5% probability that the loss is greater than 20, thus $VaR_{0.95}$ = 20 but a 1% chance that the loss is greater than 10,000 resulting in $VaR_{0.99}$ = 10,000. Therefore, the $VaR_{0.95}$ disregards the size of a loss at a higher confidence interval. This characteristic can lead to wrong indications to utility maximizing investors that are unaware of the size of a potential loss beyond the conducted VaR, which may result in great losses.

Furthermore, VaR is not a coherent risk measure as it is not sub additive. In economic terms, this implies that this risk measure does not encourage diversification, such that the risk of combined all-in portfolio is often greater than the sum of the individual risks of each sub-portfolio. This is unrealistic and does not conform with what is experienced in practice. Moreover, VaR is sensitive to stressed market condition, in which case it is likely that VaR will underestimate the actual risk, which in return will lead to false inferences (Yamai & Yoshiba, 2005).

2.1.2 Definition Expected Shortfall

Expected Shortfall (ES) is the average loss beyond the VaR level, that is for all confidence levels that are greater or equal to α. This implies that ES takes the loss distribution of the entire right tail greater than α into calculation. Mathematically this is shown as follows:

$$ES_\alpha(L) = \frac{1}{1-\alpha}\int_\alpha^1 VaR_x(L)dx \qquad (2)$$

Comparing the disadvantageous properties of VaR, such as the silence about the size of the loss beyond the VaR level, with the definition of ES it becomes clear that ES is able to eliminate these drawbacks as it captures the losses of the entire right tail. Furthermore, ES is a coherent risk measure that supports the diversification effects and thus, moves closer to real life applications. However, potential difficulties with ES arise when considering its sample size. Yamai and Yoshiba (2005) argue that in comparison to VaR a larger sample size is required to receive a similarly accurate estimation.

2.2 Stylised Facts of Financial Data & Volatility Forecasting

The stylised facts of time series of financial return data are commonly recognized and emphasized within the research field of economics and volatility modelling. Brooks (2014) define three common stylised facts exhibited by financial data, namely *volatility clustering*, *leptokurtosis* and *leverage effect*. Volatility clustering is the tendency of the volatility in the financial data to exhibit a high volatility the next day if the volatility has been high in the previous days, or low if the volatility has been low in the previous days. Hence, both positive and negative returns appear in bunches, thus high and low volatility is clustered rather than spread out over the time period. This is reflected by information arrivals and its impact the price movements. The tendency for the volatility of the financial data to exhibit a larger increase after a large fall in the price as compared to a rise of the same magnitude is called for the leverage effect. Leptokurtosis is when the distribution of the financial data is characterised by excessive peakedness at the mean and fatter tails (or heavy tails). Hence, the returns do not follow a Normal (Gaussian) distribution. Another property of distribution return is the skewness, which indicates the symmetry of the distribution around its mean value (Brooks, 2014).

In finance, volatility is a measure of risk. Several parametric and non-parametric approaches to estimate VaR and ES require forecasting of a volatility parameter by using a volatility model. Various time series models exist for this purpose. For instance a simple model is constant volatility and more sophisticated and complex models, in particular the GARCH(1,1) and the EWMA, are appropriate to account for stylised facts of volatility in financial data. Bollerslev (1986) and Taylor (1986) independently developed the Generalised Autoregressive Conditional Heteroskedasticity, GARCH(1,1) which can model volatility and at the same time it encompasses volatility clustering, time-varying volatility and leptokurtosis. However, the model cannot account for the leverage effects (Brooks, 2014). The exponentially weighted moving average (EWMA) model captures volatility clustering by letting the most recent observations to influence the forecast of volatility more than older observations (Brooks, 2014).

2.3 The Basel Standards and the Transition from VaR to ES

Since 1974, the Basel Committee of Banking Supervision (BCBS) has provided a forum for banking supervisory matters with the aim to globally improve financial stability and the banking supervision (Bank for International Settlements, 2020). The established banking regulations and standards published by BCBS are known as Basel I, Basel II and Basel III (Bank for International Settlements, 2020). In 1988, Basel I, or the *Basel Capital Accord*, was approved by the member countries and released to the banks. Later on, in 2004 the Basel I was replaced by Basel II, *the New Capital Framework*, and the three pillars made the foundation for the new framework; minimum capital requirements, supervisory review and market discipline. The financial crisis in 2008 forced the revision of Basel II as the banks liquidity and risk management were insufficient as exposed to the crisis. As a result, Basel III was approved in 2010 to enhance the capital and liquidity requirements (Bank for International Settlements, 2020). To date, the framework is still a subject for revision.

The BCBS published the Fundamental Review of the Trading Book (FRTB) in 2012 as a response to the drawbacks of the current market risk framework to propose a new framework that considers the tail risk of tail events in the measurements of market risk. VaR has been criticised in the academic literature for discouraging diversification and disregarding the tail events beyond the 99th percentile of the loss distribution, which is the confidence level used in the current framework when estimating VaR. The Bank for International Settlements (2012) supports the transition from VaR to ES by the weaknesses of the current market risk related to VaR as the risk measure. First, they support it by the fact that VaR is unable to capture the risk of tail events and such tail events may imply large losses if they occur. VAR ignores the tail risk since it does not look beyond the 99th percentile of the loss distribution. Furthermore, the shift is supported by findings in academic literature of the coherence of risk measures. VaR is not a coherent risk measure due to its non-subadditivity property. This means that VaR discourages diversification of investments. Contrariwise, the Expected Shortfall looks beyond the VaR and captures the tail risk and has the property of being sub-additive, thus it is coherent.

The new quantitative standards for calculating the capital requirements, as clarified by the BCBS in the publication *Minimum Capital Requirements for Market Risk* from January 2016 and revised version in 2019, is that the ES is to be computed on 10-trading day horizon at the 97.5th percentile one-tailed confidence level. Therefore, replacing the current metrics VaR estimated at the 99th percentile one-tailed confidence level to capitalize the loss of the trading book. Going forward, the transition from VaR to ES involves the issue of finding methods to estimate ES correctly and an accurate backtesting method for ES.

2.4 Estimating VaR and ES

2.4.1 Previous Research on VaR Estimation Methods

As indicated previously, estimating VaR correctly is still of major importance when backtesting ES and thus, the following compares existing studies of well performing VaR estimation methods. This is to give an indication of what is known about VaR methods that are well-performing in the backtest. A large body of literature has highlighted several successful methods that are described in the following.

To start with, it is widely accepted that estimation methods that are able to take rare events into account (such as EVT or fat-tailed distributions) result in more accurate estimates compared to non-fat-tailed distributions (Harmantzis, Miao & Chien, 2006). Furthermore, the advantages of a more dynamic volatility model that is able to capture volatility clustering or time-varying volatility, such as EWMA or GARCH, are well established and outperform a constant volatility model, such as the variance-covariance approach (Huang, 2000; Tang & Do, 2018). To be more specific, research has established that competitive methods include applying GARCH for volatility modelling and, in case of a parametric approach, assuming a distribution with heavy tails (excess kurtosis), skewness and asymmetry (Abad, Benito & López, 2013; Krause & Paolella, 2014). Such a distribution with these properties can be the Student's t-distribution (Lin & Shen, 2006). While Altun, Tatlidil and Ozel (2019) agree with the researchers on the benefit of excess kurtosis, they add to that by proposing a new extension to the normal distribution called alpha-skew generalized normal (ASGN), which accounts for skewed data. It is further argued that beside that newly developed distribution, also the Normal and Student's t-distribution in combination with a GARCH model deliver promising results. Regarding portfolio VaR estimations, researchers argue that Extreme Value Theory (EVT) and the non-parametric historical simulation (for which no distribution must be assumed as the empirical distribution is used) leads to a number of violation that is close to the expected number of violations and thus, perform well (Huang, 2000; Abad, Benito & López, 2013; Tang & Do, 2018).

To conclude, while some studies have demonstrated that promising VaR estimation can be achieved with EVT or Normal distribution in combination with a proven volatility model, the majority of researchers agree that a distribution that takes financial data and thus, excess kurtosis and skewness into account combined with a volatility model that is able to model time-varying volatility and volatility clustering outperforms other methods in the VaR backtest.

2.4.2 Previous Research on ES Estimation Methods

In 2000, McNeil and Frey constructed a method to estimate conditional quantiles and conditional expected shortfalls, or in other words VaR and ES. They used GARCH models to estimate the current conditional volatility and Extreme Value Theory (EVT) to model the tails of the distribution function with an application of the Generalized Pareto Distribution (GPD) approximation. When McNeil and Frey (2000) compared the results using their method with several other methods they found that a conditional approach fitted the estimation of VaR more than an unconditional approach. Another finding in the study by McNeil and Frey (2000) was using models with Normal distributions and conditional return gave poor ES estimates and consequently, the researchers suggested to take the leptokurtic property of distributions into consideration. One possible implication of this is that poor ES estimates are the consequence of disregarding the stylised facts of financial data, as discussed previously in section 2.2.

Additional research on ES estimation methods has included the EVT model. Chinhamu, Huang, Huang and Chikobvu (2015) argued that EVT has been proven successful in areas such as insurance. They showed that a GPD distribution is superior to a Gaussian or Student's t-distribution, which resulted from backtesting those distributions under the EVT approach. Harmantzis, Miao and Chien (2006) used the EVT approach to estimate ES, which together with the historical approach are concluded as methods giving accurate ES estimates compared to the Normal Gaussian distribution and the Stable Paretian methods. Almost every paper investigating ES estimations includes several methods that performed well in the studies. Therefore, Taylor (2020) tested the degree to which a combination of several well performing methods had a significant improvement of estimating ES (and in this case VaR). The conclusion is that combining is superior to using individual methods on their own. Likewise, this is supported by Bah, Munga'tu and Waititu (2016) as they found that combining EVT with a GARCH model results in a robust estimation method. It is argued that this is due to the fact that GARCH is able to account for changing volatility and volatility clustering while EVT with GPD respects the stylized financial fact of financial data, for instance fat tail of the distribution.

A profound study of several different ES estimation models, namely *A comparison of Expected Shortfall estimation models* by Righi and Ceretta, was published in 2015. Righi and Ceretta (2015) investigated 17 different ES estimations models and the models used were categorized in three types: unconditional, conditional and quantile/expectile regression-based model. The unconditional models were the Normal (Gaussian) distribution, Student's t-distribution, skewed Student's t-distribution, and Generalized Pareto distribution (GPD, used for EVT) and Historical Simulation (HS). The authors described the conditional models as "filtered versions of those unconditional models" (Righi & Ceretta, 2015, p.20). For the last category of models, the risk measure VaR and ES was estimated through quantile and expectile regressions were the dynamics of the conditional quantiles was estimated for VaR and the dynamics of the conditional expectiles was estimated for ES. Righi and Ceretta (2015) used the CAViaR models as proposed by Engle and Manganelli (2004) for VaR to model the quantiles. These models are the symmetric absolute value (SAV), the asymmetric slope (AS) and the indirect-GARCH (IG).

For ES estimation and for modelling the expectiles, the CARES approach was used for ES estimation and for modelling the expectiles, and since it is based on the CAViaR approach SAV, AS and IG is also considered in the same way. The models were estimated empirically on data of different financial assets; S&P500, DAX, FTSE100, Nikkei225, Ibovespa and Hang Seng indices representing equity investments, U.S Treasury bonds representing fixed income assets, WTI crude oil, Gold, and soybean from COMEX, NYMEX and CBOT representing commodity investments, and USD/EUR, USD/YEN, USD/GBP and BRL/USD for exchange rates. Their dataset consisted of 3198 observations for each asset from the time period 2000-01-01 – 2012-12-31 and with estimation windows equal to 250, 500 and 1000 trading days.

Righi's and Ceretta's (2015) empirical results suggest the best performing models for estimating ES are the conditional models, GARCH and filtered HS (FHS), and the authors concluded the CARES-based models as underperformers. They argue that the filtering of data is important to consider heteroscedasticity since it leads to the stylised facts of volatility clustering. Furthermore, Righi and Ceretta (2015) stressed the importance of the length of the estimation window as shorter estimation windows were concluded to give poor backtest results since information of significant relevance is left out from the estimation. Their argument of the estimation window as an important variable in ES estimation due to its effect on the overall performance of the methods conforms with this result.

A recent study by Lazar and Zhang (2019) assessed the risk of estimation models for ES based on evaluating to which extent the model must undergo correction to jointly pass ES backtests. Lazar and Zhang (2019) derived theoretical formulas for model risk, which in turn derives from the estimation and specification errors of the model. To empirically perform the test, the authors used their proposed methodology on the DJIA index from 1900-01-01 to 2017-03-05, MSCI Europe Index and MSCI World Index for equity, Bank of America Merrill Lynch US Treasury & Agency Index for bonds and the CRB Spot Index for commodity from 1986-10-31 to 2017-07-07, USD/GBP for FX and Microsoft (MSFT) shares from 1987-01-10 to 2017-20-04. According to their study, models with GARCH(1,1) residuals were found to be the models that had to undergo the smallest correction among several models with the following implementations; Historical Simulation (HS), EWMA models, Gaussian Normal distribution, Student's t-distribution, GARCH(1,1)-N, GARCH(1,1)-t, Cornish-Fisher and GARCH(1,1)-GPD. In addition, they found that $\mathrm{ES}_{0.975}$ had to undergo smaller corrections than $\mathrm{VaR}_{0.99}$ and the corrections made to the ES estimates was on average 50% smaller if corrections were made to the VaR estimates before.

2.5 Backtesting ES

In the case of ES, backtesting methods are less developed as compared to VaR where the Kupiec frequency test by Kupiec (1995) is the standard backtesting method and given to use when backtesting VaR estimates. The Kupiec test is a binomial test, where a violation occurs if the loss is larger than the VaR estimate. The observed frequency of VaR violations is compared to the predicted frequency of VaR violations.

Since the VaR has been the standard market risk measure, the Basel standard way of backtesting is conducted on VaR and by observing the number of exceptions, corresponding to the Kupiec test. Regardless of the suggested transition from using VaR to ES, the Basel Committee for the Bank for International Settlements (2019b) still recommends to backtest VaR measures at the 99^{th} and 97.5^{th} percentile confidence level, based on a sample of 250 observations and placing the estimate in one of three zones depending on the number of exceptions; green light (if zero to four exceptions), amber light (if five to nine exceptions) and red light (if ten or more exceptions). Hence, often called for the Traffic Light approach. Thus far, the BCBS has not suggested a standard backtesting method for ES.

A commonly used ES backtesting method was developed by McNeil and Frey (2000) and it tests the null hypothesis of the residuals having a mean equal to zero using a bootstrap test taking no distributional assumptions regarding the residuals. A one-sided t-test is employed to test the alternative hypothesis of a mean exceeding zero, thus implying an underestimation of the conditional ES. The backtesting method proposed by McNeil and Frey (2000) is widely adopted among the reviewed articles (Chinhamu, Huang, Huang & Chikobvu, 2015; Harmantzis, Miao & Chien, 2006; Lazar & Zhang, 2019; Righi & Ceretta, 2015; Taylor, 2020)

In 2014, the article *Backtesting Expected Shortfall,* written by Acerbi and Szekely, was published by MSCI. ES was believed to be non-backtestable after it was discovered that ES, contrary to VaR, cannot be elicited (Acerbi & Szekely, 2014). This was believed as ES takes into account the entire right tail after the VaR level and thus, the VaR estimate is of relevance for ES and its backtest. Acerbi and Szekely (2014) proved that ES can be backtested using the three methodologies they have developed and proposed in the article. The first test is backtesting ES after VaR, and it is similar to the frequently used backtesting approach developed by McNeil and Frey (2000). The third test is about estimating the ES from realized marks to backtest the whole distribution of the model. The first and the third test require Monte Carlo simulation, while the second test backtest ES estimates directly. Among these three, the first and the second model are proposed by the authors as valid backtest models. However, the first model should be used in a combination with the VaR backtest as suggested by the Basel Committee. Further, they argue that the elicitability property is to be considered when conducting an analysis of different models with the approach to select a model. Thus far, the proposed methods by Acerbi and Szekely (2014) have not been widely employed yet. Lazar and Zhang (2019), implemented the second test, Z_2, as proposed by Acerbi and Szekely (2014) among other implemented backtesting methods.

2.6 Evaluation of Previous Research

The stylised facts of time series of financial returns are emphasized in several papers written on VaR and ES estimation methods and relate to the contributions of the papers, including the studies by McNeil and Frey (2000) and Righi and Ceretta (2015). The approach by McNeil and Frey (2000), GARCH-EVT, takes into account two stylised facts of time series of financial returns. One is that the volatility of conditional return distributions is stochastic and the other is that the return distributions are leptokurtic, or in other words, the distributions are fat-tailed. McNeil's and Frey's (2000) finding as reported above is supported by the study undertaken by Righi and Ceretta in 2015 since they also found the GARCH method to be one of the best performing methods. However, the GARCH(1,1) model is not a fully idealised model to forecast volatility since it does not account for the leverage effect. Although, the GARCH(1,1) models have been widely employed in the literature, it still serves to capture some of the stylised facts of volatility in financial data. On the other hand, both McNeil and Frey (2000) and Righi and Ceretta (2015) propose other ways to consider the stylised facts. When the return distributions are leptokurtic and the tails are symmetric, McNeil and Frey (2000) suggest modelling using the t-distribution. However, Righi and Ceretta (2015) suggest an approximation with GPD since it can handle asymmetric tails, but they also find that the skewed Student's t-distribution fulfils a significant estimation function as it can take the heavy and asymmetric tails into account, and the conditional models dominate the other models used in their study.

With the transition from VaR to ES other variables than the distributional assumptions, non-parametric or parametric, and unconditional versus conditional are discussed to have an impact on the ES estimates, for instance the sample size and the length of the estimation window. Yamai and Yoshiba (2005), found that ES requires a larger sample size under market stress because a larger sample size will reduce the estimation error that becomes larger when the underlying loss distribution becomes more fat-tailed implying higher probability for large losses. Righi and Ceretta (2015) implemented three different estimation windows and argued around the estimation window as a significantly relevant variable in the ES estimation model. However, in general this variable seems to have been less considered in the literature and it might be due to that the Basel Committee's recommendation of an estimation window of 250 days has been widely accepted.

3 Data & Methodology

3.1 Data and Data Descriptive

3.1.1 Representative Trading Book Portfolio

As in indicated from previous research, equity indices, treasury bonds, exchange rates and commodities have been utilised to empirically evaluate different ES estimation methods. However, the assets or indices have been evaluated and analysed separately rather than as an all-in portfolio consisting of several different asset classes. A constructed portfolio of different assets classes is believed to bring enhanced representation of a bank's or agent's actual portfolio, and it also conforms with the hypothetical portfolio used in the *Analysis of the trading book hypothetical portfolio exercise* conducted by the Basel Committee on Banking Supervision (BCBS) published in 2014. Hence, a portfolio approach is an attempt to increase the validity of the study. An approach to construct an aggregated portfolio consisting of several different indices as a representation of a bank's trading book was implemented in this thesis. Indices were used instead of swaps or forwards to represent the actual positions hold by a bank and to include the overall pattern of these positions, which are believed to be reflected in a broad index. As seen in the portfolio exercise performed by Basel, banks make use of options to hedge their exposure. This is included into the representative portfolio of this thesis by constructing an equity index (namely S&P 500) option, which is explained in section 3.1.5.

3.1.2 Data

The data was retrieved through two financial databases from two providers, which are Thomson Reuters's Datastream and Bloomberg L.P's Bloomberg Terminal. Among many other types of data, the databases allow to retrieve time series data of indices required to achieve the research objective. Since these are two major and prominent databases, the retrieved data is subject to validation. The dataset consists of daily closing prices for equity indices, sovereign and corporate bond indices as proxies for interest rates and credit spreads, commodity index and foreign exchange (FX) index. Considering that the thesis adopts a perspective of a U.S. bank / U.S. based agent, a number of criteria was to be met for the index to be selected. The criteria for selecting the index were as follows:

- The index can be retrieved from Datastream or Bloomberg,
- it is denominated in US Dollars and,

- it has daily observations between 2000-01-01 and 2020-04-01 (approximately 250 days per year, except for year 2020).

The selected indices are presented in Table 3.1 below.

Table 3.1 Data retrieved from Datastream and Bloomberg

Asset	Name	Code	Source
Equity	S&P 500 Composite	S&PCOMP	Datastream
	S&P SmallCap 600 VALUE	SP06SVA	Datastream
Bond	US Benchmark 10 Year DS Government Index	S08729(RI)	Datastream
	S&P US Treasury Bond Index	SPBDUSB	Datastream
	S&P 500 Investment Grade Corporate Bond Index	SPUIGBD	Datastream
	Bloomberg Barclays US Corporate High Yield Total Return Index	LF98TRUU	Bloomberg
Commodity	Thomson Reuters Equal Weight Commodity Index	NYFECRB	Datastream
FX	US Dollar Index DXY	NDXYSPT	Datastream

The *S&P 500 Composite* and the *S&P SmallCap 600 Value* were used as proxies to represent the equity investments and the losses of the investments in equities. Regarding bond indices, risk factors such as time to maturity or yield are already reflected in the price and thus, it is sufficient to take the price index instead a different index for each risk factor. The bond indices presented in the Table 3.1 were used as proxies for both interest rate and credit spread. The *S&P US Treasury Bond Index* and the *US Benchmark 10 Year DS Government Index* measure the performance of the Treasury bond market and were used to approximate the interest rate. Furthermore, one investment grade bond index, the *S&P 500 Investment Grade Corporate Bond Index* and one high yield corporate bond index, the *Bloomberg Barclays US corporate High Yield Total Return Index*, were used. These two corporate bond indices measure the corporate bond market and was proxies used for the credit spread. The proxy for the position in commodities is the *Thomson Reuters Equal Weight Commodity Index*, also referred to as the *Thomson Reuters Continuous Commodity Total Return Index*. This is a broad commodity index consisting of 17 futures contracts reflecting the price movements of 17 commodities and each contract continuously preserves an equal weight of 5.88% (Thomson Reuters, 2012). The *US Dollar Index (DXY)* was used as a proxy for the foreign exchange rate positions. The DXY index is a weighted geometrically-averaged index measuring the movements of the US dollar in relation to six currencies; Euro (57.6%), Japanese Yen (13.6%), British Pound (11.9%), Canadian Dollar (9.1%), Swedish Krona (4.2%) and Swiss Franc (3.6%) (Intercontinental Exchange, 2020).

All the time-series retrieved from Datastream included non-trading days related to U.S. public holidays or other special events resulting in a closure of the exchanges. The one index retrieved from Bloomberg also included a few non-trading days. A total of 226 and 12 days of observations from Datastream and Bloomberg, respectively, were discarded from the sample due to holidays and special events. The underlying reason was that the closing price of the non-trading day was the retained closing price of the previous trading day, which eventuated into a zero return and a zero loss. The non-trading days for both stock exchanges and the bond market are New Year's Day, Martin Luther King, Jr. Day, Presidents' Day, Good Friday, Memorial Day, Independence Day, Labor Day, Thanksgiving Day and Christmas Day. Further, the bond market is closed on the Veterans Day and the Columbus Day, while the stock market remains open on these days. To maintain consistency of preventing zero losses due to non-trading, the two days were discarded regardless that observations are available for some indices at those days.

3.1.3 Descriptive Statistics

To give an overview of the data and its properties, each asset and its descriptive statistics are displayed below in Table 3.2 (the descriptive statistics) and in Figure 3.1 (the daily log loss graphs). The y-axis is the same scale for all graphs, such that they can be comparable.

Table 3.2 Descriptive Statistics for assets for time period 2000-2020

	Minimum	**Maximum**	**Mean**	**Median**	**Std Dev**	**Skewness**	**Kurtosis**
S&PCOMP	-10.4236	12.7652	-0.0112	-0.0550	1.2530	0.4399	11.2490
SP06SVA	-8.5385	13.7096	-0.0226	-0.0673	1.5039	0.6201	8.5250
S08729	-4.0529	2.8735	-0.0096	-0.0305	0.4791	0.0719	2.8686
SPBDUSB	-1.7829	1.6938	-0.0142	-0.0200	0.2424	0.0611	3.8716
SPUIGBD	-1.8113	2.6010	-0.0013	-0.0130	0.2960	0.5821	4.8859
LF98TRUU	-2.8289	4.8147	-0.0257	-0.0438	0.3144	2.2507	34.0690
NYFECRB	-9.0120	5.8994	-0.0100	-0.0209	0.8457	0.1691	6.6034
NDXYSPT	-2.7344	3.0646	0.0006	0.0000	0.5048	0.0382	1.8134

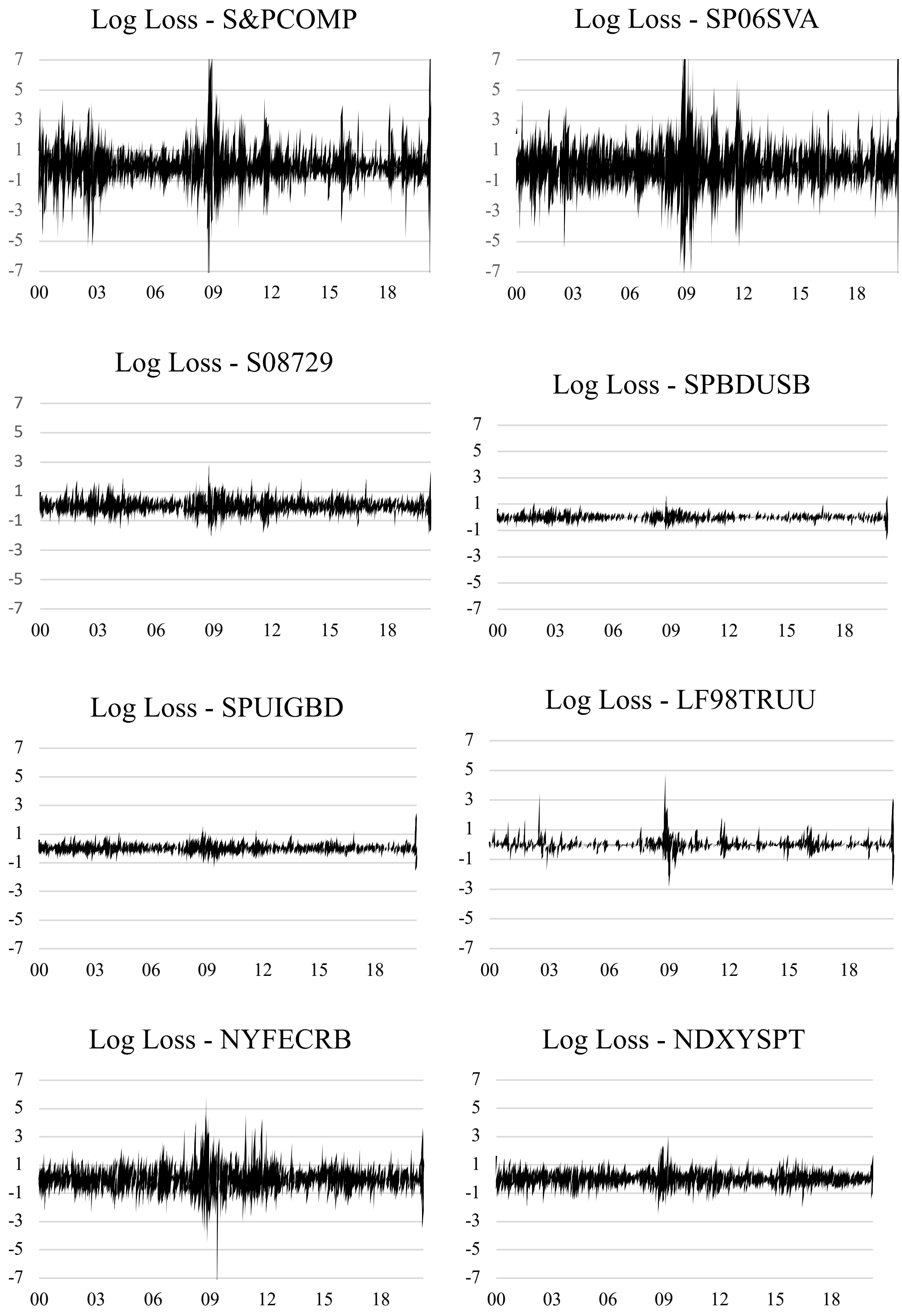

Figure 3.1 Graphs of daily log losses for the assets for time period 2000-2020

While the Bloomberg Barclays index, S&P US Treasury Bond and S&P 500 Investment Grade Corporate Bond index display less variation in absolute values of loss, the equity and commodity indices experience fluctuations on a higher scale in terms of absolute values. A common trend shown by all assets in the loss graphs is the great dispersion around 2008/ 2009, which is a consequence of the global financial crisis. Furthermore, fluctuations can be seen in 2020 during the corona crisis.

By evaluating the skewness coefficient, it becomes clear that some assets appear to be fairly symmetrically distributed (with a coefficient between 0.038 and 0.072) as it is the case for the US Benchmark 10 Year index (S08729), the S&P US Treasury Bond Index (SPBDUSB) and the US Dollar Index (NDXYSPT). The asset with the most extreme skewness is the Bloomberg Barclays index with a coefficient of 2.25. The remaining assets display skewness coefficients between 0.17 and 0.62, which means their distribution is asymmetric and positively skewed.
With the kurtosis coefficient, it is apparent that the majority of assets are not normally distributed, but rather best described by fat tails, which supports the findings from the literature review regarding stylized financial data.

The above described statistics as well as a further yearly standard deviation analysis of all indices, as seen below in Table 3.3, show that the years 2008 and 2009 have clearly above average standard deviations and thus, are considered for the stressed period. The third largest standard deviation was in 2011, which is the reason this year is also referred to for the stressed period. On the other hand, this analysis showed that the years 2014, 2017 and 2019 have the lowest standard deviations and thus, are chosen as the calm testing period for this study. For each of these mentioned testing years, the three previous years generated the input for estimating parameters or volatilities for that year. To give an example, when testing for 2008, the input from 2005-2007 was used to maximize the loglikelihood function to arrive at GARCH parameters, to calculate the constant volatility, etcetera. This period of three years as input was chosen as it is believed that among other things volatility change over time and does not stay constant. Three years is considered a medium long period, which is long enough to be able to pick up underlying trends of volatility clustering but is short enough to leave out irrelevant trends from the market, which would have an effect on the ES estimation. This three-year sample window is kept constant throughout all tests in order to ensure consistency and comparability between test results.

Table 3.3 Yearly Standard Deviation Analysis

	S&PCOMP	SP06SVA	SPUIGBD	NDXYSPT	SPBDUSB	NYFECRB	S08729	LF98TRUU	Average	Analysis	Result
2000	1.4045	1.2113	0.2379	0.5438	0.2079	0.6208	0.3757	0.2175	**0.6024**	Below	
2001	1.3642	1.3656	0.3353	0.5481	0.3163	0.6171	0.5423	0.2588	**0.6685**	Above	
2002	1.6461	1.5669	0.3025	0.4667	0.3040	0.6027	0.5229	0.3589	**0.7214**	Above	
2003	1.0788	1.1332	0.3326	0.5181	0.2893	0.6322	0.5596	0.2192	**0.5954**	Below	
2004	0.7181	1.0583	0.2835	0.5907	0.2368	0.7824	0.4556	0.1691	**0.5368**	Below	
2005	0.6512	1.0058	0.2221	0.5013	0.1792	0.6877	0.3533	0.1793	**0.4725**	Below	
2006	0.6330	1.0399	0.1985	0.4353	0.1508	0.8802	0.2939	0.1044	**0.4670**	Below	
2007	1.0007	1.2981	0.2687	0.3268	0.2147	0.7219	0.4045	0.2139	**0.5562**	Below	
2008	2.6033	2.9351	0.4637	0.7539	0.3772	1.6633	0.7275	0.7056	**1.2787**	Above	Stressed
2009	1.7230	2.4260	0.4215	0.7016	0.3425	1.3914	0.7098	0.4568	**1.0216**	Above	Stressed
2010	1.1442	1.5928	0.3151	0.5499	0.2466	0.9534	0.5586	0.2488	**0.7012**	Above	
2011	1.4725	2.0757	0.3185	0.5756	0.2496	1.0394	0.5646	0.3255	**0.8277**	Above	Stressed
2012	0.8086	1.1233	0.2229	0.4098	0.1550	0.8040	0.4168	0.1674	**0.5135**	Below	
2013	0.6984	0.8868	0.2522	0.4112	0.1482	0.5572	0.4185	0.1943	**0.4669**	Below	
2014	0.7184	0.8951	0.2003	0.3197	0.1294	0.5619	0.3289	0.2048	**0.4198**	Below	Calm
2015	0.9824	1.0117	0.2971	0.5955	0.2027	0.7271	0.4890	0.2589	**0.5706**	Below	
2016	0.8265	1.1628	0.2414	0.4771	0.2017	0.6903	0.3949	0.3286	**0.5404**	Below	
2017	0.4217	0.8079	0.1888	0.3756	0.1676	0.5394	0.3078	0.1224	**0.3664**	Below	Calm
2018	1.0826	1.0974	0.1789	0.3797	0.1623	0.6163	0.2995	0.1860	**0.5003**	Below	
2019	0.7885	1.0720	0.2327	0.2831	0.2283	0.5562	0.3910	0.1691	**0.4651**	Below	Calm
Average	**1.0883**	**1.3383**	**0.2757**	**0.4882**	**0.2255**	**0.7822**	**0.4557**	**0.2545**	**0.6146**		

3.1.4 Calculating the Losses

The daily logarithmic returns was calculated for each index in order to obtain its daily losses, which are needed to estimate ES. The log-return was calculated by the log differences where the natural logarithm of the closing price at time $t-1$ was subtracted from natural logarithm of the closing price at time t, and the losses were obtained by multiplying the log return by -$100:

$$r(t) = \ln(p_t) - \ln(p_{t-1}) \quad (3)$$

$$l(t) = -100 \cdot r(t) \quad (4)$$

3.1.5 Constructing an Option

The option is priced using the pricing formula by Black (1976) and it is defined as follows:

$$C(S_t, t) = N(d_1)S_t - N(d_2)Ke^{-r(T-t)} \tag{5}$$

$$d_1 = \frac{1}{\sigma\sqrt{T-t}}\left[\ln\left(\frac{S_t}{K}\right) + \left(r + \frac{\sigma^2}{2}\right)(T-t)\right] \tag{6}$$

$$d_2 = d_1 - \sigma\sqrt{T-t} \tag{7}$$

As presented before, for the regular indexes the loss is calculated based on a fixed amount of money that was invested the previous day compared to the index value of the following day. To be consistent with regular indexes and the option, it was decided that the call option price is calculated by using the previous day's index value as the spot price and the strike price, from which the loss to the following day is determined. This calculation yields an at-the-money option. The option price at the beginning of the day C_{t-1} and the option price at the end of the day C_t are calculated to obtain the daily returns. First the option price at the beginning of the day C_{t-1} is calculated. The time to maturity is kept constant at 30 days. The standard deviation is estimated using a rolling window of the previous month (20 days), which is then scaled to yearly volatility by multiplying with $\sqrt{252}$. The risk-free rate is assumed to be 1%. The today's closing price of the index is the spot price at the end of the day which is used to calculate C. At this point, the time to maturity is 29 days and the yearly volatility, the strike price and the risk-free rate are unchanged. Finally, the return is calculated as: $(C_t - C_{t-1})/C_{t-1}$.

As the option is a non-linear function of the underlying index, it has larger absolute deviations compared to the underlying assets and thus, to avoid extreme swings in the portfolios caused by this single index, it is scaled by multiplying with -100 just as the other asset classes; however, it receives a lower weigh to fit into the all-in portfolio. This is done as the purpose of including this option is to have a representative pattern of a non-linear option in the portfolios instead of the option dominating the portfolio losses. Therefore, as seen in Table 3.4 below, the option is included into the all-in portfolios with a weight of 1%. Consequently, this means that in total \$100*0.01 = \$1 is invested in the equity option.

3.1.6 Constructing the All-in Portfolios

As the purpose of this study is to estimate and backtest ES on a bank's trading book, the above listed asset classes need to be combined to a portfolio. To refer to Basel's portfolio exercise, they do not disclose the weights given to each asset and they constructed seven such portfolios. It is challenging to define these weights as each bank assigns them individually and they are also influenced by the negative correlation that is generally found between a bank's trading book and banking book (market risk and credit risk). If a bank faces high risk in the trading book, then it will try to reduce exposure in their loan portfolio and by doing so reducing their credit risk. This is also a matter of capital requirements and restraints for the bank imposed by regulations (Abbassi & Schmidt, 2018). Concerning the main asset classes, it can be noted that banks usually put a risk weight between 55% to 70% on equity, between 20% to 80% on commodities and up to 30% on FX. It is shown that extensive equity positions are held (PwC, 2016). In order to counteract the uncertainty that each bank holds different weights in the trading book due to several reasons, this study will consider the generally suggested weights and five different all-in portfolios are constructed to ensure that the backtesting result are valid and applicable for different banks with different books. The all-in portfolios can be found in Table 3.4.

Table 3.4 All-in portfolios with corresponding asset weights

All-in Portfolio	Equity	Equity Option	Bonds	Commodity	FX
Portfolio 1	49%	1%	5%	35%	10%
Portfolio 2	54%	1%	5%	20%	20%
Portfolio 3	59%	1%	5%	30%	5%
Portfolio 4	64%	1%	10%	20%	5%
Portfolio 5	69%	1%	10%	10%	10%

It should be noted that the two equity indices have equal weights in the portfolios, and the same holds for the four bond indices. To clarify, this means that the two equity indices have a weight of 24.5%, respectively in Portfolio 1, and each bond has a weight of 1.25% (=0.05/4) in Portfolio 1, (and in Portfolio 2), and so on.

The following table shows the descriptive statistical analysis of the five constructed all-in portfolios. It can be seen that their properties are similar to each other as they consist of the same assets with solely different weighting. Closer inspection of Table 3.5 shows that an increasing weight for equity increases the minimum and maximum data points and consequently, the standard deviation. This is not surprising as it was already identified by the previous daily log losses graphs that this asset class experience great fluctuation. Furthermore, the weights put on equity are the largest relatively to the weights of other asset classes in the portfolios.

Table 3.5 Descriptive Statistics for all-in portfolios for time period 2000-2020

All-in Portfolio	Minimum	Maximum	Mean	Median	Std. Dev.	Skewness	Kurtosis
Portfolio 1	-5.9261	8.8766	-0.0305	-0.0474	1.0245	0.4202	5.4223
Portfolio 2	-5.7872	8.8831	-0.0298	-0.0496	1.0278	0.3617	5.1734
Portfolio 3	-6.6284	9.9243	-0.0317	-0.0545	1.1353	0.4367	5.7626
Portfolio 4	-6.7024	10.0955	-0.0322	-0.0567	1.1608	0.4151	5.7637
Portfolio 5	-6.9630	10.3090	-0.0320	-0.0617	1.1939	0.3750	5.6495

3.2 Estimation Methodology

A quantitative research methodology has been undertaken since quantitative data are of importance to achieve the research objective and to answer the research question. All estimations are conducted in Excel and all formulas can be found in the Appendix (see Appendix 1: Excel Formulas and Functions). The following sections give a brief definition of each method while the focus is on how and for what reason it was implemented in this thesis.

3.2.1 Gaussian Normal Distribution

The Gaussian Normal Distribution assumes that a normal distribution generates the time series returns or losses. As previously discussed, the assumption of Normal distribution is not significantly supported in the empirical studies for VaR and ES methods, or in general for financial data. However, when estimating different VaR and ES methods the Normal distribution innovation is still useful as a reference for comparing the methods with different distributional assumptions. The VaR and the ES under the Normal distribution are estimated as:

$$VaR_{\alpha} = \mu + \sigma \cdot z_{\alpha} \tag{8}$$

$$ES_{\alpha} = \mu + \sigma \cdot \frac{f_{std}(z_{\alpha})}{1 - \alpha} \tag{9}$$

where the z_α is the quantile α of the standard normal distribution (Embrechts, Frey & McNeil, 2005). For VaR and ES, the α equals 0.975 as proposed in the FRTB by Basel. For the ES estimation, the $f_{std}(z_\alpha)$ is the probability density function of the standard Normal distribution and expressed as:

$$f_{std}(z_\alpha) = \frac{1}{\sqrt{2\pi}} exp\left(\frac{-z^2}{2}\right) \tag{10}$$

The estimation of VaR and ES was conducted using (8) and (9), by applying both the unconditional, or in other words the constant volatility, σ, through sampling the constant volatility (see section 3.2.4) and conditional volatility, σ_t, derived from implementing the GARCH(1,1) from (27). The μ is the constant mean as described in the beginning of this section.

3.2.2 Student's t-distribution

The Student's t-distribution was implemented as a distributional assumption in the VaR and ES estimations as it is capable to take the previously discussed stylised fact of financial data, namely leptokurtosis, or also called excess kurtosis (>3) and heavy tails, into account. It does so by implementing the parameter *v*, which is the degrees of freedom. The lower the degrees of freedom, the higher the excess kurtosis and the fatter the tails. Consequently, when *v* approaches infinity, it moves closer to the Normal distribution.

For the ES estimation method, the probability density function of the standardized Student's t-distribution is required, which is expressed as:

$$f_{std}^{*}(t_{\alpha,v}) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{v \cdot \pi} \cdot \Gamma\left(\frac{v}{2}\right)} \cdot \left[1 + \frac{1}{v} \cdot t_{\alpha,v}^{2}\right]^{-\frac{(v+1)}{2}} \tag{11}$$

where $t_{\alpha,v}$ is the α-quantile for the distribution (Laurent & Bauwens, 2002). The *v* denotes the degrees of freedom and is expressed

$$v = \frac{4k - 6}{k - 3} \tag{12}$$

The (12) is the solution to the expression for kurtosis, *k*, which is:

$$k = 3 + \frac{6}{v - 4} \tag{13}$$

Following these equations, VaR and ES under the Student's t-distribution for each day in the out of sample period were estimated as defined by Embrechts, Frey and McNeil (2005):

$$VaR_\alpha = \mu + \sigma \cdot \sqrt{\frac{v - 2}{v}} \cdot t_{\alpha,v} \tag{14}$$

$$ES_\alpha = \mu + \sigma \cdot \sqrt{\frac{v-2}{v}} \cdot t_{\alpha,v} \cdot \frac{f_{std}^*(t_{\alpha,v})}{1 - \alpha} \cdot \left(\frac{v + t_{\alpha,v}^2}{v - 1}\right) \tag{15}$$

In the above cases, σ refers to the regular unconditional variance; however, if it is replaced by a time-varying model such as GARCH(1,1), the approximations for VaR and ES for the Student's t-distribution change to the following expressions:

$$VaR_\alpha = \mu + \sigma_t \cdot \sqrt{\frac{v - 2}{v}} \cdot t_{\alpha,v} \tag{16}$$

$$ES_\alpha = \mu + \sigma_t \cdot \sqrt{\frac{v-2}{v}} \cdot t_{\alpha,v} \cdot \frac{f_{std}^*(t_{\alpha,v})}{1 - \alpha} \cdot \left(\frac{v + t_{\alpha,v}^2}{v - 1}\right) \tag{17}$$

3.2.3 Skewed Student's t-distribution

As resulted from the Literature Review discussed by several authors, the skewed Student's t-distribution is useful for modelling data that fits the stylised financial facts. While the symmetric Student's t-distribution considers the heavy tails and excess kurtosis compared to the Normal distribution, this extension also accounts for the skewness and thus, the asymmetry as compared to the symmetrical Student's t-distribution (Jones & Faddy, 2002).

In a study by Bauwens and Laurent (2002) it becomes apparent that a symmetric Student's t-distribution is particularly applicable when modelling stock return data. Combining this distribution with a GARCH volatility model is superior to other methods.

To start with, μ in the skewed Student's t-distribution refers to the mean of the underlying data, y_t, and σ is the standard deviation of y_t. Further, it is assumed that z_t is an independently and identically distributed process with $E(z_t) = 0$ and $Var(z_t) = 1$. Following this definition, the underlying data, y_t, can be defined as:

$$y_t = \mu + \sigma z_t \tag{18}$$

Solving (18) for the z_t:

$$z_t = \frac{(y_t - \mu)}{\sigma} \tag{19}$$

The authors assume that z_t follows a skewed Student's distribution and thus, introduce the parameters ξ and v, which measure skewness and kurtosis, respectively. While v remains the same as for the symmetric Student's t-distribution and thus, (12), the parameter ξ takes on a value below or above 1 to be asymmetric (if $\xi = 1$ the distribution is the symmetric Student's t-distribution). Specifically, $\xi > 1$ means right skewness and $\xi < 1$ means left skewness. The following loglikelihood equation in Bauwens and Laurent (2002) needs to be maximized in order to find the parameters v and ξ:

$$\begin{gathered} ln\ L_t(\mu, \sigma, \xi, v) = \\ = \ln\left(\frac{2}{\xi + \frac{1}{\xi}}\right) + ln\ \Gamma\left(\frac{v+1}{2}\right) - \frac{1}{2}\ln\big(\pi(v-2)\big) - ln\Gamma\left(\frac{v}{2}\right) \\ + \ln\left(\frac{s}{\sigma}\right) - \frac{1}{2}(v+1)\ln\left[1 + \frac{(sz_t + m)^2 \xi^{-2I_t}}{v-2}\right] \end{gathered} \tag{20}$$

where:

$$m(\xi, v) = \frac{\Gamma\left(\frac{v-1}{2}\right)\sqrt{v-2}}{\sqrt{\pi}\Gamma\left(\frac{v}{2}\right)}\left(\xi - \frac{1}{\xi}\right) \tag{21}$$

$$s^2(\xi, v) = \left(\xi^2 + \frac{1}{\xi^2} - 1\right) - m^2 \tag{22}$$

$$I_t = \begin{cases} 1\ if\ z_t \geq -\frac{m}{s} \\ -1\ if\ z_t < -\frac{m}{s} \end{cases} \tag{23}$$

In the above case, σ is assumed to be constant, but can be replaced with a time-varying volatility, which would then be expressed as σ_t in (19) and (20). The conditional volatility σ_t is derived from the GARCH(1,1) estimation as described below. Consequently, the parameters ξ and v was replaced with the values from the ML estimation allowing for GARCH(1,1).

To estimate VaR and ES under the assumption of the skewed Student's t-distribution, the quantile function is calculated according to the following definition given by Laurent (2002):

$$t_{\alpha,v,\xi} = \begin{cases} \dfrac{\frac{1}{\xi} t_{\alpha,v}\left(\frac{\alpha}{2}(1+\xi^2)\right) - m}{s} & if\ \alpha < \dfrac{1}{1+\xi^2} \\ \dfrac{-\xi t_{\alpha,v}\left(\frac{1+\alpha}{2}(1+\xi^{-2})\right) - m}{s} & if\ \alpha \geq \dfrac{1}{1+\xi^2} \end{cases} \tag{24}$$

where $t_{\alpha,v,\xi}$ is the quantile function of the Skewed student's t-distribution, $t_{\alpha,v}$ is the quantile function for the Student's t-distribution (see above for the Student's t-distribution).
The function is estimated at seven different quantile levels, $\alpha = 0.999, 0.995, 0.991, 0.987, 0.982, 0.979, 0.975$.

The VaR under the skewed Student's t-distribution was then estimated using the $t_{\alpha,v,\xi}$ as the input in accordance with the following expression for VaR:

$$VaR_\alpha = \mu + \sigma \cdot t_{\alpha,v,\xi} \ . \tag{25}$$

The ES for each day in the estimation period is simply the average of the VaR estimates at the same day and can be described as

$$ES_\alpha = \text{average}(VaR_{0.975}, VaR_{0.979}, \dots, VaR_{0.999}) \ . \tag{26}$$

In the same manner as for the Normal distribution and for the Student's t-distribution, the σ which is the constant volatility in (25) is replaced by the estimated GARCH(1,1) estimation of volatility (see section 3.2.5 below) to obtain the method estimates assuming skewed Student's t-distribution with the application of conditional volatility.

3.2.4 Constant Mean and Volatility

In the formulas for VaR and the ES the μ denotes the constant mean and σ denotes the constant volatility, and these are the sample average and standard deviation, respectively. The constant mean and volatility are estimated using the rolling window sampling technique. The window is fixed to the number of days in the in-sample period. By moving one day ahead, the observation the day before is included and the oldest observation is excluded from the window. The process proceeds until the end of the sample period has been reached.

The constant mean applies to the VaR and ES estimations with the Normal distribution, the Student's t-distribution and the skewed Student's t-distribution, for both the constant volatility and the GARCH volatility, in order to maintain consistency in the treatment of the methods. The Normal distribution with constant volatility is the only method that does not require estimations of parameter and therefore, the μ is not estimated with the maximum likelihood as for the other methods.

3.2.5 GARCH(1,1) Parameter Estimation

The GARCH(1,1) model, as developed independently by Bollerslev (1986) and Taylor (1986), was implemented in order to account for the stylised facts, specifically volatility clustering and time varying volatility. In the GARCH(1,1) model, the conditional variance depends on the past conditional variances, σ_{t-1}^2, and the conditional variance on past squared errors, η_{t-1}^2 . Thus, the definition of the conditional variance is given by Brooks (2014), and here defined with the parameters ω, α and β:

$$\sigma_t^2 = \omega + \alpha\eta_{t-1}^2 + \beta\sigma_{t-1}^2 \, . \tag{27}$$

The parameters α and β determine the dynamics of the volatility while the restrictions put on the parameters ω, α and β that non can be negative ensure that the conditional variance, σ_t^2, does not become negative.

The model's mean equation is defined as follows:

$$r_t = \mu + \eta_t, \tag{28}$$

where r_t is the log return of the underlying data, μ is the average of it while η_t are the residuals. The above-mentioned parameters as well as μ had to be defined. This was done by performing maximum loglikelihood (ML) estimation by maximising the sum of the loglikelihood functions of each day t in the in-sample period, where $t = 1, 2, \dots T$ and T is between 747 - 750 depending on number of trading days of each year. This is a mathematical tool to estimate the most likely parameters that will generate the known observations (which in this case are the all-in trading portfolios) and thus, those optimised parameters are such that they can explain and fit the given data best. Here, it should be noted that the μ in (28) is estimated by the ML estimation and not the rolling window which applies to the μ in the VaR and ES expressions.

The function to perform the ML estimation under the Normal distribution is as follows (Peters, 2001):

$$ln\, L_T\, (\mu, \omega, \alpha, \beta) = \sum_{t=1}^{T} \left(-\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln(\sigma_t^2) - \frac{\eta_t^2}{2\sigma_t^2} \right) \tag{29}$$

For the Student's t-distribution, the parameter for the degrees of freedom v is needed in the loglikelihood function, as given by Peters (2001), as this distribution account for excess kurtosis:

$$\begin{aligned} ln\, L_T\, (\mu, v, \omega, \alpha, \beta) &= \ln\left[\Gamma\left(\frac{v+1}{2}\right)\right] - \ln\left[\Gamma\left(\frac{v}{2}\right)\right] - 0.5\ln[\pi(v-2)] \\ &- 0.5\sum_{t=1}^{T}\left[ln\sigma_t^2 + (1+v)\ln\left(1 + \frac{z_t^2}{v-2}\right)\right] \end{aligned} \tag{30}$$

Here the degrees of freedom is $2 < v \leq \infty$ and it is estimated according to (12). Moreover, for the skewed Student's t-distribution both kurtosis and skewness are required and therefore, the loglikelihood function is as follows:

$$ln\ L_t(\mu, \xi, v, \omega, \alpha, \beta) =$$
$$= \ln\left(\frac{2}{\xi + \frac{1}{\xi}}\right) + ln\ \Gamma\left(\frac{v+1}{2}\right) - \frac{1}{2}\ln(\pi(v-2)) - ln\Gamma\left(\frac{v}{2}\right) \tag{31}$$
$$+ \ln\left(\frac{s}{\sigma_t}\right) - \frac{1}{2}(v+1)\ln\left[1 + \frac{(sz_t + m)^2 \xi^{-2I_t}}{v-2}\right]$$

where z_t is defined in (19) and I_t in (23), as defined in Bauwens and Laurent (2002).
In the following, the estimation procedure is explained. To start with, initial values for the four parameters are set. To start with, μ is initially defined as the average of the underlying log-return data, which mostly generated a value close to 0. The remaining three parameters are set in line with the restrictions that none of them can be negative. As the estimation is sensitive to the initial parameters, values are chosen that are expected to be close to the actual parameter values after the optimisation: $\mu \approx 0$, $\omega = 0.01$, $\alpha = 0.05$ and $\beta = 0.9$. The parameters μ, ω, α and β are estimated each year using a window of three years prior to the upcoming year. Consequently, the parameters used for modelling the GARCH in 2008 are estimated using the loss observations from 2005-2007, adding up to 750 in-sample observations in total. The loss observations from 2006-2008 (also a total of 750 observations) are used to estimate the parameters for 2009 and this process is followed for the years 2011, 2014, 2017 and 2019.

For the estimation procedure of GARCH assuming Normal distribution innovations, the first step was to define the residual η_t. According to (28) the residual is $\eta_t = r_t - \mu$. Afterwards, the next step was to calculate the conditional variance according to (27) with the initially set parameter values. As the conditional variance cannot be defined for the first observation out of each sample when t = 1, the unconditional variance $Var(r_t)$, the sample variance of the three years sample input, was used. Afterwards, the loglikelihood function as in (29) was maximized. To simplify, the first step was to define the loglikelihood value for each period with:

$$ln\ L_t\ (\mu, \omega, \alpha, \beta) = -0.5\ln(\sigma_t^2) - \frac{\eta_t^2}{2\sigma_t^2} \tag{32}$$

This was followed by summing the values for each day in the sample period. This sum was maximized by changing the four parameters and allowing for the restrictions that ω, α and β

are non-negative. After the maximization was conducted the final parameter values were reported and eventually used in the GARCH(1,1) model to capture the volatility.
The GARCH estimation assuming the symmetric Student's t-distribution was performed by first calculating the residual η_t in the same manner as for the normal distribution. Afterwards, z_t was calculated as in (19) before calculating the conditional variance as described above according to (27). The loglikelihood function in (30) was calculated for each day in the sample period. The sum of all the loglikelihood values in the sample period was then maximised to obtain the GARCH parameters ω, α, β and v.

Finally, the GARCH estimation assuming the skewed Student's t-distribution was conducted. The η_t and z_t were estimated in the same way as described for the symmetric Student's t-distribution. This was followed by defining I_t as described in (23), where m and s were defined as (21) and (22). Again, the next step was to calculate the loglikelihood function for each day in the sample period and summing all the loglikelihood values, as in (31). This sum was maximised in order to obtain the parameter estimates of v (degrees of freedom) and ξ (the skewness parameter) and ω, α, β of the GARCH model.

3.2.6 Historical Simulation

The Historical Simulation (HS) approach is a non-parametric approach and is independent of distributional assumptions. It takes the actual empirical distribution of the underlying data and thus, is able to navigate certain drawbacks when choosing a distribution that is unable to fit financial data. The sampling technique used to estimate VaR and ES with HS is the rolling window or moving window. The window contains the same number of loss observations as the length of the window (in terms of days of the in-sample periods in this case) to estimate the VaR and ES the day following the window. The window is fixed to the number of days in the in-sample period, so when it is moving one day ahead the observation from the previous day is included in the window and the oldest loss observation in the window is excluded from it.

i. Basic Historical Simulation

For the simplest form of the non-parametric approach, namely the basic historical simulation (BHS), VaR is estimated as follows:

$$VaR_\alpha(L) = (1 - \alpha)N + 1 \quad (33)$$

where N is the number of sample observations. Consequently, ES is approximated by taking the average of the losses that are greater than the VaR estimate.

ii. Volatility Weighted Historical Simulation

The Volatility Weighted Historical Simulation (VWHS) implements BHS, but on the rescaled losses for which each loss receives the same probability as suggested by Hull and White (1998). The intuition behind this is volatility clustering, meaning that if volatility is higher than average at time t, it is likely that it is also higher at time t+1 and vice versa. For that reason, this method can model and account for events that are currently undergoing in the market.

The Volatility Weighted Historical Simulation (VWHS) implements BHS, but on the rescaled losses for which each loss receives the same probability. The intuition for this is volatility clustering, meaning that if volatility is higher than average at time t, it is likely that it is also higher at time $t + 1$ and vice versa. For that reason, this method can model and account for events that are currently undergoing in the market. The rescaled losses are defined as follows for T losses:

$$\begin{aligned} l_1^* &= \frac{\sigma_{T+1}}{\sigma_1} l_1 \\ l_2^* &= \frac{\sigma_{T+1}}{\sigma_2} l_2 \\ &\vdots \\ l_T^* &= \frac{\sigma_{T+1}}{\sigma_T} l_T \end{aligned} \tag{34}$$

Four years (the three in-sample years and the out-of-sample year) of daily rescaled losses are obtained for each month, January – December, where σ_{T+1} denotes the conditional volatility at the first trading day of the month in the out-of-sample period. This means, 12 columns with rescaled losses are conducted for each portfolio. Afterwards, the VaR and ES are estimated by applying the BHS as described above, but in this case on the rescaled losses. The procedure is described in detail below.

To estimate the VaR and ES with the VWHS, a rolling window of three years consisting of the rescaled losses is implemented. The first VaR is estimated using the out-of-sample (three years) rescaled losses from January, and the second VaR estimate is estimated using the three years of rescaled losses from January plus the first loss in the out of sample period minus the first loss in the in-sample, maintaining the same length of the window. When the VaR estimates for all days in January is calculated, the first VaR estimate in February is calculated using the three years of rescaled loss from February, the rolling window switches month from January to February and moves one step ahead. Hence, the rolling window includes (and excludes) the out of sample (and the in-sample) days in January. The same procedure follows for each month of the out of sample period.

3.3 Backtesting

After completing the estimations of ES, backtesting was conducted in order to evaluate the performance of the methods used to estimate ES. Backtesting is the widest accepted method to determine how accurate the method applied on the historical data forecasts the actual outcome. As previously mentioned, Acerbi and Szekely (2014) developed three methods for backtesting ES. Here, the second test proposed by Acerbi and Szekely (2014) was employed to backtest the ES estimates. This method has a number of adequate features compared to the first and the third method. Firstly, it is testing ES directly. Secondly, the test only requires the predicted (one day ahead) $ES_{\alpha,t}\ (L)$ and the magnitude $L_t I_t$ of the loss if a VaR exception occurs at day *t*. Thirdly, it is characterized by a noticeable stability of the critical values over different distributions, or shapes of tails (Acerbi & Szekely, 2014). The authors highlight this as advantageous since it implies no necessity to note daily predictions of the distribution. In other words, a Monte Carlo simulation of the distributions of the test statistics is not required to obtain critical values. However, a Monte Carlo simulation is required for the first and the third test, thus the main disadvantage of these two tests is based on this requirement. Hence, the tests are not directly applicable in the same extent as the second test.

For simplicity, the hypotheses as stated by Acerbi and Szekely (2014) are rewritten to the following:

$H_0 \quad : \quad ES^P_{\alpha,t} = ES^F_{\alpha,t}$, ES is correctly estimated for each day t

$H_1 \quad : \quad ES^P_{\alpha,t} < ES^F_{\alpha,t}$, ES is underestimated for at least one day *t*.

Furthermore, as this thesis set out to also test against overestimation, which is equally bad for banks as underestimation is for Basel, the following alternative hypothesis results from it:

$H_1 \quad : \quad ES^P_{\alpha,t} > ES^F_{\alpha,t}$, ES is overestimated for at least one day *t*.

The P denotes the actual distribution of losses of the model/method used to estimate ES, while the F denotes the unknown and true distribution of losses.

The test statistics Z_2, as defined by Acerbi and Szekely, is defined in (35) where $T(1-\alpha)$ is the expected number of violation days. If the ES is correctly estimated by the method the expected Z_2 equals zero under the null hypothesis, since under the null the test statistics Z_2 is equal to zero.

$$Z_2 = -\frac{1}{T(1-\alpha)} \sum_{t=1}^{t=T} \frac{L_t I_t}{ES_{\alpha,t}(L_t)} + 1 \tag{35}$$

The following equations (36), (37), (38) and (39) depicts how the expected Z_2 equals zero under the null hypothesis. In equation (36), the definition of ES is rewritten and the L_t represents the stochastic loss at day *t*. I_t represents the indicator function which, as defined in (37), takes on the value one if a VaR violation occurs at day *t* and zero otherwise.

$$ES_{\alpha,t}(L_t) = E\big[L_t | L_t > VaR_{\alpha,t}(L_t)\big] = \frac{E[L_t I_t]}{E[I_t]} = \frac{E[L_t I_t]}{1-\alpha} \tag{36}$$

$$I_t = \begin{cases} 1 \; if \; L_t > VaR_{\alpha,t}(L_t) \\ 0 \; if \; L_t \leq VaR_{\alpha,t}(L_t) \end{cases} \tag{37}$$

The equation in (38) follows from (36) and is applied in the equation in (39).

$$ES_{\alpha,t}(L_t) = \frac{E[L_t I_t]}{1-\alpha} \tag{38}$$

$$E_{H_0}[Z_2] = -E\left[\frac{1}{T}\sum_{t=1}^{t=T}\frac{L_t I_t}{(1-\alpha)ES_{\alpha,t}(L_t)} + 1\right] =$$

$$= \frac{1}{T}\sum_{t=1}^{t=T}\frac{E[L_t I_t]}{1-\alpha}\cdot\frac{1}{ES_{\alpha,t}(L_t)} \tag{39}$$

$$= -\frac{1}{T}\cdot T + 1 =$$

$$= -1 + 1 = 0$$

To conclude, the expected Z_2 under the null and the alternative hypotheses for under- and overestimation, respectively, is as follows:

$$E_{H_0}[Z_2] = 0\,, \quad E_{H_1}[Z_2] < 0, \quad E_{H1}[Z_2] > 0\,.$$

As previously stated above, Acerbi and Szekely (2014) found that the critical values for Z_2 showed noticeable stability over different distributions. The authors suggest that if the significance threshold is 5% and 0.01% the critical values are Z_2^{crit} = -0.7 and Z_2^{crit} = -1.80, respectively, for underestimation.
This implies that the null hypothesis is rejected if the test statistics, Z_2, is below -0.7 or -1.80 for the significance level 5% and 0.01%, respectively, due to underestimation. Furthermore, Acerbi and Szekely (2014) suggest that the aforementioned critical values perfectly apply for a Basel traffic-light approach.

This backtesting method described above is applicable to test against underestimation as required by the Basel Committee; however, for banks overestimation of risk is just as critical. For this reason, the Acerbi and Szekely test is added to by the upper test statistics to reject a method when it is overestimating ES. The critical values for overestimation are +0.59 for the 5% significance level and +0.93 for the 0.01% significance level. These were obtained with a standard Monte-Carlo simulation of Z-statistic under the null hypothesis. See Acerbi and Szekely (2014) for details on this simulation.

In this thesis a traffic-light approach is employed to resemble the Basel Committee's proposed traffic-light mechanisms for backtesting, and the method is rejected if it does not pass the 5% significance level. After identifying the test statistics, Z_2, each method is assigned the colour "green", "amber" or "red" and is either rejected or not rejected. To not reject the method, the

Z_2 must have a value between -0.7 and 0.59, thus the method will be assigned a green colour since it is neither underestimated (from the perspective of the Basel Committee) nor overestimated (from the perspective of banks). This is in accordance with the critical values suggested by Acerbi and Szekely (2014).

For the significance level 0.01% the lower critical value is -1.8 for the normal distribution, according to Acerbi and Szekely (2014). However, the critical value for the significance level 0.01% is rather unstable for different distributions as compared to the critical value at the 5% significance level. According to their result, this means that the critical value will become more negative when the distribution exhibits a kurtosis larger than three (kurtosis > 3, or when v becomes smaller) which is the case here. Therefore, the critical values for both sides will be considered as approximate critical values, or approximate boundaries of the red zones. The values used as approximated critical values for the 0.01% significance level are -1.8 and +0.93. This is a conservative approach since the actual critical values will be larger when the distribution exhibits excess kurtosis, thus the actual critical values will be more negative (positive) for underestimation (overestimation) than the approximated critical values. Hence, methods that have Z_2 < -1.8 or Z_2 > 0.93 will be rejected and placed in the red zones as suggested by Basel's traffic light system. The approximated upper critical value for the 0.01% level is chosen in such way it corresponds to the lower critical level of the 0.01% level. This is because limited sources exist for upper critical values suited for the implementation of the backtest by Acerbi and Szekely (2014). This is a solid solution in the attempt to represent the perspective of banks in the estimation of ES.

Therefore, in the case the Z_2 is between -0.70 and -1.8 or between +0.59 and +0.93 it is rejected since it requires further improvements to not underestimate or overestimate the ES, thus it receives the amber light. As already mentioned, methods fall in the approximated red zone if the methods have a Z_2 below -1.8 or above +0.93 since it underestimates and overestimates the ES. This rejection procedure is summarized in Table 3.6.

Table 3.6 Summary Rejection and Traffic Light Result Procedure based on Test Statistics

Test Statistics	**Not reject/ Reject**	**Traffic Light**	**Reason**
>-0.70 and <0.59	Not reject	Green	Correctly estimated
-0.70 to -1.80 and +0.59 to +0.93	Reject	Amber	Slight under- or overestimation
<-1.80 and >+0.93	Reject	Red	Severe under- or overestimation

4 Results & Analysis

Based on the discussed methods, the tests are performed, and ES estimates are obtained for these several methods for the five representative trading portfolios for the stressed periods, namely the years 2008, 2009 and 2011, as well as the calm period, which is the years 2014, 2017 and 2019. This section first presents the estimated parameters from both constant volatility and GARCH (1,1) and secondly the backtest test statistics (Z_2) and traffic light system results of the various methods for the two periods are analysed. Furthermore, the five portfolios are evaluated, and the performance of the methods are analysed year by year. Finally, an overall summary of the results and the analysis is given.

4.1 Estimated Parameters

To start with, the GARCH parameters μ, ω, α and β for the three distributions are estimated with maximum likelihood. For the Student's t-distribution the parameter ν and for the skewed Student's t-distribution the parameters ν and ξ are also included in the maximum likelihood estimation, with both constant and GARCH volatility, as explained in the methodology in the previous chapter. The GARCH parameters are derived for each separate representative portfolio in each period for the three different distributions, respectively. In Table 4.1 and 4.2, the parameters ν and ξ estimated for constant volatility are presented together with the parameter μ for Student's t and skewed Student's t-distribution.

Table 4.1 Estimated Parameters for Stressed Period – Constant Volatility

		2008		2009		2011	
	Portfolio	**Student**	**skewed Student**	**Student**	**skewed Student**	**Student**	**skewed Student**
μ	1	-0.0569	-0.0525	-0.0397	0.0015	-0.0605	-0.0328
	2	-0.0508	-0.0435	-0.0336	0.0067	-0.0482	-0.0239
	3	-0.0581	-0.0518	-0.0425	0.0044	-0.0632	-0.0292
	4	-0.0549	-0.0466	-0.0402	0.0077	-0.0583	-0.0235
	5	-0.0519	-0.0405	-0.0365	0.0129	-0.0506	-0.0176
v	1	10.6039	10.4265	3.3142	3.3337	4.0853	4.1079
	2	8.9679	8.9442	3.2847	3.2976	4.1212	4.1330
	3	9.3832	9.2850	3.2011	3.2170	3.9821	3.9994
	4	8.6782	8.6262	3.1547	3.1657	3.9563	3.9668
	5	8.3673	8.3519	3.1502	3.1566	3.9674	3.9727
ξ	1		1.0338		1.0855		1.0498
	2		1.0442		1.0764		1.0431
	3		1.0391		1.0832		1.0526
	4		1.0446		1.0783		1.0513
	5		1.0558		1.0752		1.0471

Table 4.2 Estimated Parameters for Calm Period – Constant Volatility

		2014		2017		2019	
	Portfolio	**Student**	**skewed Student**	**Student**	**skewed Student**	**Student**	**skewed Student**
μ	1	-0.0662	-0.0473	-0.0262	-0.0247	-0.0529	-0.0511
	2	-0.0682	-0.0533	-0.0364	-0.0332	-0.0547	-0.0514
	3	-0.0769	-0.0543	-0.0308	-0.0277	-0.0597	-0.0558
	4	-0.0821	-0.0603	-0.0364	-0.0324	-0.0627	-0.0580
	5	-0.0854	-0.0658	-0.0436	-0.0381	-0.0653	-0.0590
v	1	4.6847	4.6829	8.6307	8.6400	5.2815	5.2966
	2	4.6624	4.6538	7.8314	7.8480	4.8939	4.9132
	3	4.5128	4.5154	8.5030	8.5343	4.9760	5.0021
	4	4.4611	4.4630	8.1713	8.2192	4.7287	4.7557
	5	4.4225	4.4236	7.8086	7.8718	4.5616	4.5909
ξ	1		1.0574		1.0092		1.0075
	2		1.0459		1.0189		1.0128
	3		1.0599		1.0177		1.0136
	4		1.0560		1.0222		1.0161
	5		1.0504		1.0289		1.0202

The degrees of freedom parameter ν ranges between 3.1052 and 10.6039 for the symmetrical Student's t-distribution and between 3.1566 and 10.4265 for the skewed Student's t-distribution in all years. The highest values of the parameter ν are the ones estimated for 2008 and 2017 and the lowest are observed for 2009. For the skewed Student's distribution, the skewness parameter is ξ >1 for all five portfolios for all six years, which implies asymmetry in the loss distributions. It means that the loss distributions exhibit right (positive) skewness, thus it has more probability in the right tail. This is an indication of deviation from a symmetrical distribution. A lower degrees of freedom parameter indicates excess kurtosis and fatter tails, so the lower degrees of freedom parameters estimated for 2009 by using the years 2006-2008 are very likely a consequence of the financial crisis and its influence on the loss distribution resulting in heavy tails, thus more probability for a higher loss in the tails. Furthermore, the skewness parameters as estimated for 2009 are further away from ξ =1 in absolute terms in relation to the skewness parameters for the other years, which indicates greater right (positive) skewness and more probability in the right tail. Contrariwise, the estimated degrees of freedom parameters for 2017 are higher as compared to those of the portfolios in 2009. This implies that the relatively stable years 2014-2016 exhibits a loss distribution with non-heavy tails. Taking the skewness parameter into consideration, the average value of the skewness parameters in 2017 is the second lowest (after the average value in 2019) relatively the average of the skewness parameters in the other years. Hence, one can conclude that these loss distributions are normally distributed. To conclude, in the case of estimating the parameters using constant volatility there are indications of higher values of the skewness parameters and lower values of the degrees of freedom parameters when the period used for estimating the parameters involves more volatile years and vice versa.

In Table 4.3 and 4.4 the ν and ξ parameters are presented with the GARCH parameters μ, ω, α and β, as obtained from the ML estimation with conditional volatility, thus GARCH application, for all three distributions.

Table 4.3 Estimated Parameters for Stressed Period – GARCH Volatility

		2008			2009			2011		
	Portfolio	**Normal**	**Student**	**skewed Student**	**Normal**	**Student**	**skewed Student**	**Normal**	**Student**	**skewed Student**
μ	1	-0.0615	-0.0684	-0.0649	-0.0529	-0.0598	-0.0489	-0.0920	-0.0956	-0.0876
	2	-0.0523	-0.061	-0.0548	-0.0451	-0.0545	-0.0413	-0.0936	-0.0955	-0.0907
	3	-0.0627	-0.0709	-0.0659	-0.0552	-0.0648	-0.0512	-0.0977	-0.1033	-0.0933
	4	-0.0579	-0.0672	-0.0607	-0.0527	-0.0638	-0.0487	-0.0994	-0.1045	-0.0955
	5	-0.0525	-0.0643	-0.0549	-0.0483	-0.0621	-0.0452	-0.1024	-0.1063	-0.0999
ω	1	0.0328	0.0265	0.0258	0.0546	0.0261	0.0264	0.0652	0.0300	0.0300
	2	0.0297	0.0182	0.0173	0.0410	0.0169	0.0176	0.0533	0.0232	0.0235
	3	0.0376	0.0263	0.0254	0.0566	0.0266	0.0270	0.0708	0.0323	0.0324
	4	0.0361	0.0217	0.0208	0.0478	0.0211	0.0218	0.0654	0.0291	0.0294
	5	0.0368	0.0186	0.0176	0.0390	0.0151	0.0159	0.0581	0.0254	0.0259
α	1	0.0423	0.034	0.0336	0.1605	0.0943	0.0925	0.1718	0.0886	0.0876
	2	0.0445	0.0312	0.0305	0.1388	0.0843	0.0832	0.1749	0.0881	0.0875
	3	0.0494	0.0354	0.0349	0.1551	0.0930	0.0913	0.1740	0.0896	0.0886
	4	0.0502	0.0335	0.0328	0.1416	0.0868	0.0853	0.1749	0.0891	0.0883
	5	0.0524	0.0331	0.0324	0.1285	0.0802	0.0790	0.1764	0.0885	0.0879
β	1	0.9538	0.9247	0.9261	0.8973	0.8870	0.8874	0.8990	0.8988	0.8994
	2	0.9554	0.9701	0.9712	0.9150	0.9523	0.9521	0.8994	0.9493	0.9494
	3	0.9514	0.9304	0.9321	0.9040	0.8931	0.8932	0.8996	0.8995	0.9000
	4	0.953	0.9398	0.9417	0.9151	0.9051	0.9049	0.9004	0.9011	0.9013
	5	0.9528	0.9457	0.9474	0.9261	0.9175	0.9170	0.9015	0.9031	0.9031
v	1		11.3698	11.2172		8.3145	8.5550		8.6709	8.7751
	2		9.8411	9.863		7.4079	7.6618		11.2432	11.3932
	3		10.5682	10.4942		7.7715	8.0095		9.2707	9.4247
	4		9.8427	9.8443		7.1182	7.3518		10.5652	10.7937
	5		9.4656	9.5636		6.7701	7.0368		12.7556	13.1329
ξ	1			1.0309			1.0744			1.0373
	2			1.0431			1.0693			1.0268
	3			1.0363			1.0756			1.0430
	4			1.0425			1.0693			1.0401
	5			1.0553			1.0673			1.0332

Table 4.4 Estimated Parameters for Calm Period – GARCH Volatility

		2014			2017			2019		
	Portfolio	**Normal**	**Student**	**skewed Student**	**Normal**	**Student**	**skewed Student**	**Normal**	**Student**	**skewed Student**
μ	1	-0.0727	-0.0693	-0.0566	-0.0405	-0.0389	-0.0354	-0.0777	-0.0667	-0.0687
	2	-0.0869	-0.0737	-0.0640	-0.0524	-0.0512	-0.0455	-0.0819	-0.0674	-0.0678
	3	-0.0863	-0.0807	-0.0660	-0.0434	-0.0437	-0.0385	-0.0852	-0.0730	-0.0738
	4	-0.0949	-0.0874	-0.0736	-0.0490	-0.0500	-0.0438	-0.0898	-0.0753	-0.0752
	5	-0.1017	-0.0933	-0.0813	-0.0562	-0.0580	-0.0500	-0.0931	-0.0774	-0.0755
ω	1	0.0486	0.0099	0.01	0.2246	0.0910	0.0882	0.2103	0.0667	0.0671
	2	0.1246	0.0193	0.0193	0.2303	0.0990	0.0940	0.2212	0.0843	0.0844
	3	0.0940	0.0159	0.0159	0.2646	0.1076	0.1031	0.2245	0.0717	0.0718
	4	0.1279	0.0243	0.0242	0.2805	0.1169	0.1109	0.2241	0.0768	0.0768
	5	0.1487	0.0378	0.0387	0.2957	0.1271	0.1188	0.2327	0.0872	0.0867
α	1	0.1098	0.0415	0.0415	0.2350	0.1262	0.1255	0.2610	0.1207	0.1217
	2	0.2021	0.0514	0.0512	0.2605	0.1437	0.1434	0.2940	0.1397	0.1399
	3	0.1585	0.0486	0.0483	0.2372	0.1289	0.1279	0.2811	0.1279	0.1283
	4	0.1958	0.0572	0.0568	0.2461	0.1358	0.1349	0.3018	0.1373	0.1373
	5	0.2153	0.0708	0.0717	0.2584	0.1443	0.1440	0.3201	0.1465	0.1456
β	1	0.9194	0.9482	0.9480	0.7113	0.7380	0.7428	0.7035	0.7775	0.7762
	2	0.8293	0.9628	0.9629	0.6971	0.8435	0.8480	0.6808	0.8559	0.8557
	3	0.8798	0.9377	0.9377	0.7100	0.7350	0.7414	0.7104	0.7810	0.7806
	4	0.8472	0.9221	0.9222	0.7008	0.7210	0.7289	0.7064	0.7692	0.7693
	5	0.829	0.8974	0.8953	0.6948	0.7086	0.7182	0.7000	0.7524	0.7536
v	1		6.5985	6.6072		10.5511	10.6265		6.2589	6.2276
	2		6.6197	6.5943		9.8315	9.9419		5.9046	5.9007
	3		6.7507	6.7810		10.3593	10.4984		6.0223	6.0121
	4		6.8655	6.8895		10.0129	10.2024		5.7985	5.7991
	5		6.9906	7.0126		9.6800	9.9172		5.6865	5.7006
ξ	1			1.0621			1.0331			0.9891
	2			1.0474			1.0541			0.9979
	3			1.0665			1.0450			0.9961
	4			1.0620			1.0543			1.0003
	5			1.0554			1.0698			1.0088

The explanatory power of the of the GARCH parameters from the ML estimation should be noted with caution as they are solely the parameter estimations; however, comparing in particular the alpha and beta parameters for the different periods, it becomes apparent that the betas are lower in the calm period compared to the stressed period while the alpha parameters is slightly higher in the calm period when evaluating with the stressed period. This may mean that the past conditional variances have a larger influence in the stressed period while the variances of the squared errors are slightly more crucial for the calm period. This in turn can be interpreted as that the aspect of volatility clustering and time-varying volatility is more dominant in a financially distressed period than in the calm period. This is intuitive as such periods are characterised by above average volatility, meaning that high volatility at time t influences the volatility at time $t + 1$ to also be high. This dependence becomes clear in a higher beta value while in the calm period the volatility is more constant over time, which in turn proceeds into a lower beta.

The degrees of freedom parameter v ranges between 5.6865 and 12.7556 for the symmetrical Student's t-distribution and 5.7006 and 13.1329 for the skewed Student's t-distribution for all years. The highest observed values of the parameter v are the ones estimated for 2008, 2011 and 2017 the lowest are observed for 2019. The skewness parameter is $\xi > 1$ for all five portfolios for all six years, except for Portfolio 1, 2 and 3 in 2019 where $\xi < 1$, which implies some left (negative) skewness in the loss distribution. However, this difference is fairly small, and the skewness parameter is closer to $\xi \approx 1$, implying no asymmetry of the distributions in this period. An interesting result is that the degrees of freedom parameters when allowing for conditional volatility are relatively higher than the parameters observed in Tables 4.1 and 4.2.

Similar to the pattern displayed among the parameters in Tables 4.1 and 4.2, there is a tendency for higher values of skewness parameters to be accompanied by lower values of the degrees of freedom parameters. Moreover, the degrees of freedom parameters in the calm period are lower than in the stress period, which also aligns with the conclusion above. Again, the interpretations of the parameters should be considered with cautions since this is solely a trend among relatively few estimated periods.

4.2 Backtest Test Statistics and Traffic Light System Results

In the following tables, Table 4.5 and 4.6, the test statistics from the backtests as well as the traffic light results for each labelled method for both stressed and calm years, respectively, are shown. These tables serve as an overview of the general test results, which are discussed in detail in the following sub-parts. The discussion is structured method by method to arrive at a meaningful sub-conclusion about the performance of the particular method for the chosen years.

Table 4.5 Backtest Z Test Statistics and Traffic Light System Results for Stressed Period

	Portfolio	ES-N	ES-N-GARCH	ES-t	ES-t-GARCH	ES-Skew t	ES-Skew t-GARCH	ES-BHS	ES-VWHS-GARCH
2008	1	-6.6511	-0.3360	-6.0923	-2.6079	-3.3189	-0.5862	-5.2114	-3.3441
		Reject	Not reject	Reject	Reject	Reject	Not reject	Reject	Reject
	2	-6.7712	0.0070	-6.1032	-2.3527	-2.6859	0.0108	-4.9670	-2.9872
		Reject	Not reject	Reject	Reject	Reject	Not reject	Reject	Reject
	3	-6.8420	-0.1454	-6.1817	-2.3772	-3.1380	-0.2440	-5.0825	-3.0981
		Reject	Not reject	Reject	Reject	Reject	Not reject	Reject	Reject
	4	-6.8193	0.0213	-6.1954	-2.3475	-2.9304	-0.0032	-5.0842	-2.8921
		Reject	Not reject	Reject	Reject	Reject	Not reject	Reject	Reject
	5	-7.0541	-0.0696	-6.3179	-2.0987	-2.8784	0.0995	-5.4967	-2.5156
		Reject	Not reject	Reject	Reject	Reject	Not reject	Reject	Reject
2009	1	-0.7650	1.0000	-0.4709	-0.4188	0.9850	0.8898	-0.1597	0.2816
		Reject	Reject	Not reject	Not reject	Reject	Reject	Not reject	Not reject
	2	-0.4003	1.0000	-0.1630	-0.0919	0.9858	1.0000	-0.0327	0.5738
		Not reject	Reject	Not reject	Not reject	Reject	Reject	Not reject	Not reject
	3	-0.7579	1.0000	-0.5539	-0.3021	0.9684	0.9450	-0.0543	0.4175
		Reject	Reject	Not reject	Not reject	Reject	Reject	Not reject	Not reject
	4	-0.7160	1.0000	-0.5109	-0.0960	0.9687	1.0000	-0.0440	0.4298
		Reject	Reject	Not reject	Not reject	Reject	Reject	Not reject	Not reject
	5	-0.5365	1.0000	-0.3645	0.2889	0.9684	0.9819	0.0712	0.7108
		Not reject	Reject	Not reject	Not reject	Reject	Reject	Not reject	Reject
2011	1	0.1365	0.6465	0.2789	-0.0518	0.9197	0.6268	0.5500	-0.0900
		Not reject	Reject	Not reject	Not reject	Reject	Reject	Not reject	Not reject
	2	0.1307	0.6483	0.2707	-0.0638	0.9155	0.4981	0.4323	-0.4914
		Not reject	Reject	Not reject	Not reject	Reject	Not reject	Not reject	Not reject
	3	-0.0051	0.6447	0.6088	-0.1651	0.9675	0.6009	0.5462	-0.2843
		Not reject	Reject	Reject	Not reject	Reject	Reject	Not reject	Not reject
	4	0.1162	0.6463	0.4667	-0.3265	0.9668	0.5535	0.4255	-0.3454
		Not reject	Reject	Not reject	Not reject	Reject	Not reject	Not reject	Not reject
	5	0.1062	0.6490	0.3216	-0.3789	0.9492	0.4058	0.3071	-0.6963
		Not reject	Reject	Not reject	Not reject	Reject	Not reject	Not reject	Not reject

Table 4.6 Backtest Z Test Statistics and Traffic Light System Results for Calm Period

	Portfolio	ES-N	ES-N-GARCH	ES-t	ES-t-GARCH	ES-Skew t	ES-Skew t-GARCH	ES-BHS	ES-VWHS-GARCH
2014	1	0.5492	1.0000	0.8354	-0.0644	1.0000	0.8488	0.8319	-0.3488
		Not reject	Reject	Reject	Not reject	Reject	Reject	Reject	Not reject
	2	-0.0206	1.0000	0.8282	0.0656	0.9844	0.8885	0.3326	-0.4270
		Not reject	Reject	Reject	Not reject	Reject	Reject	Not reject	Not reject
	3	0.5511	1.0000	0.8416	0.0568	1.0000	0.9087	0.7122	-0.2481
		Not reject	Reject	Reject	Not reject	Reject	Reject	Reject	Not reject
	4	0.4061	1.0000	0.7169	0.0527	1.0000	0.9087	0.7020	-0.2854
		Not reject	Reject	Reject	Not reject	Reject	Reject	Reject	Not reject
	5	0.1123	1.0000	0.5864	0.0506	1.0000	0.8461	0.4504	-0.3218
		Not reject	Reject	Not reject	Not reject	Reject	Reject	Not reject	Not reject
2017	1	0.5461	0.8468	0.7005	0.4685	0.9407	0.6887	0.5828	0.5251
		Not reject	Reject	Reject	Not reject	Reject	Reject	Not reject	Not reject
	2	0.3881	0.7128	0.5514	0.4336	0.9644	0.7040	0.5693	0.3854
		Not reject	Reject	Not reject	Not reject	Reject	Reject	Not reject	Not reject
	3	0.5458	0.8539	0.7039	0.4719	0.9623	0.7694	0.7116	0.5275
		Not reject	Reject	Reject	Not reject	Reject	Reject	Reject	Not reject
	4	0.4026	0.7245	0.5774	0.3458	0.9642	0.7288	0.5870	0.3999
		Not reject	Reject	Not reject	Not reject	Reject	Reject	Not reject	Not reject
	5	0.3874	0.7216	0.4271	0.3182	0.9635	0.7551	0.4585	0.3961
		Not reject	Reject	Not reject	Not reject	Reject	Reject	Not reject	Not reject
2019	1	-0.3684	0.7181	-0.2002	0.2065	0.7024	0.7506	-0.1295	-0.6466
		Not reject	Reject	Not reject	Not reject	Reject	Reject	Not reject	Not reject
	2	-0.3763	0.5717	-0.2603	0.3064	0.7823	0.8511	-0.1171	-0.6510
		Not reject	Not reject	Not reject	Not reject	Reject	Reject	Not reject	Not reject
	3	-0.3794	0.5803	-0.2746	0.2076	0.6596	0.7460	-0.1166	-0.6530
		Not reject	Not reject	Not reject	Not reject	Reject	Reject	Not reject	Not reject
	4	-0.3733	0.4419	-0.2347	0.3056	0.8024	0.8325	0.0169	-0.7618
		Not reject	Not reject	Not reject	Not reject	Reject	Reject	Not reject	Reject
	5	-0.5072	0.4340	-0.0951	0.3055	0.7638	0.8304	0.0183	-0.8774
		Not reject	Not reject	Not reject	Not reject	Reject	Reject	Not reject	Reject

4.2.1 Backtest of ES Estimates with Normal Distribution

To start with the ES estimation under Normal distribution with constant volatility for the stressed years, it can be seen from the test statistics for 2008 that ES is underestimated substantially and thus, the method is rejected. For 2009, the estimates for three out of five portfolios are rejected at the 5% significance level and thereby received the amber traffic light, which means that further improvements of the method are required. Consequently, the method does not underestimate ES as greatly as compared to 2008 estimates. For 2011, all five portfolio test statistics are not rejected and thus, awarded with the green light. An explanation for these results is that the three years sample period input for 2008 are generated from a rather calm period (as the assets in the portfolios in 2005-2007 experienced a below average standard deviation), thus these input values forecast a calm period. For this reason, each ES estimate underestimates the risk in 2008 as the actual risk in that year was higher due to the start of the global financial crisis. The same explanation holds for the result of ES estimates for 2009, with the difference that 2008 is an input year. The ES estimates in 2009 are influenced by a variation of volatilities, both low and high, and therefore, the method does not underestimate ES as much as in 2008. This leads to an improvement in the ES estimates for 2009. The results for the last stressed year, 2011, are not rejected as the sample period input comes from the stressed years 2008 and 2009 (and the calm year 2010) and thus, deliver accurate forecasts for that period, which is displayed in non-rejected test statistics. To be more specific, the variation in the volatility is lower and thereby the method can deliver more accurate ES estimates.

From the test statistics in Table 4.6 for the Normal distribution with constant volatility, it can be seen that this method gives better ES estimates for the calm period in comparison to the previously discussed stressed period. More interestingly, ES is correctly estimated for all portfolios in all the three calm years. Therefore, the test statistics are within the green light boundaries, which means the method does neither underestimates nor overestimates the ES. One possible explanation for this might be that the in-sample data's standard deviation is similar enough to the testing years' deviation, such that the Normal distribution with constant volatility is able to pick up the relatively constant standard deviation from the in-sample years to estimate ES correctly. Furthermore, the method with Normal distribution might perform better in the calm years due to less probability of big losses occurring, meaning less excess kurtosis (as discussed in the estimated degree of freedom parameter analysis in the previous section).

As a sub-conclusion, the Normal distribution with constant volatility as an ES estimation method performs well for the calm period, which is surprising as neither the distribution nor the constant volatility was found as a well performing method in this thesis's literature review; however, an intuitive explanation for this performance is provided above. Furthermore, in particular from the backtesting results from the stressed period it emerges that there possibly is a dependence between the correctness of the ES estimate and the level of similarity of its sample period's standard deviation to the testing year. This is a preliminary finding and requires further investigation by other methods' results to become reliable as a finding.

Evaluating Table 4.5 for the Normal distribution in combination with GARCH volatility, the test statistics indicate no rejection for 2008 and consequently lies within the green light boundaries, while both for 2009 and 2011 all ES estimates are overestimated and given the red and amber light, respectively. Comparing the result for 2008 under normal distribution with constant volatility versus GARCH volatility, it appears that the implementation of GARCH volatility accounts for the recent volatility changes. The constant volatility in that period is characterised by calm movements, GARCH proves to be able to account for volatility clustering that starts at the end of 2007 and beginning of 2008 and thus, it is able to predict ES for 2008 accurately. Implying this property of volatility clustering, the ES are overestimated for 2009 and 2011 as the greatest volatility is observed in 2008 and thus, GARCH forecasts a similar level of volatility for the other two stressed years. However, the overestimation of the ES in 2011 is not as severe as in 2009. As already discussed previously in this section, there is less fluctuations in volatility in the years used to forecast the volatility for 2011, which results in more accurate ES estimations compared to 2009 (yet still slightly overestimated). Overall, evaluating this result for the stressed years, the dependence of both the length and experienced standard deviation of the in-sample period seems to be relevant. For this study, assuming the estimates for 2011 were predicted with only two years data input from 2009 and 2010, this method might have been able to estimate ES correctly as these years have similar levels of standard deviations compared to 2011.

By a closer inspection of Table 4.6 for the method with Normal distribution and GARCH volatility, it becomes apparent that the ES estimates are rejected at the 0.01% and 5% significance level for 2014 and 2017, respectively, due to overestimation while the test statistics indicate correct estimations for 2019 for four out of five portfolios. However, the rejected test statistics for 2017 are in the amber zone, thus the methods are not rejected at the 0.01% statistical level since the test statistics are not above the approximated critical value +0.93.
In 2014, another great overestimation (in terms of test statistic values and received traffic lights) happens, which can be explained by the input years that reflect the stressed year 2011. The GARCH model takes the above average high volatility from this year into account and thus, overestimates the below average volatility and thus, the risk in the testing year 2014. The result for 2017 can be evaluated when looking closer at the standard volatility. This year reports the lowest volatility out of the entire sample period, which causes the overestimation. Regardless of the fact that the three years input for that year are also calm periods, they are still experiencing higher volatility than 2017 and thus, the overestimation occurs. In 2019, the ES is correctly estimated at the 5% significance level for four out of five portfolios and thus, the test statistics are within the green light boundaries. This can be explained by the three years input data which is similar to that of the year 2019, based on the standard deviation. Furthermore, taking into account the extremely calm year 2017, the test statistics for 2019 are closer to

overestimation than to underestimation. Here the opposite affect occurs compared to 2008 results with Normal distribution and constant volatility, which are underestimated due to low standard deviations from the in-sample data. This further suggests a clear link between the in-sample periods and the ES estimations' correctness being applicable in both directions (over- and underestimation).

To conclude the result for the Normal distribution with GARCH volatility, the method does not follow a clear trend to what extent it estimates ES correctly consistently; however, it should be noted that there is a strong dependence between the volatility observed in the sample years with the volatility experienced with the testing years. If these volatilities are similar, GARCH is able predict ES correctly. If the sample period volatility is greater than the volatility in the testing year, the test statistics indicate overestimations. On the other hand, if the sample period volatility is mostly below average, the GARCH model is still able to react quickly and take volatility clustering and increased volatility into account as seen for 2008. Overall, this suggests that GARCH is sensitive to the input volatilities and tends to overestimate the ES rather than underestimating it.

4.2.2 Backtest of ES Estimates with Student's t-distribution

Table 4.5 illustrated the breakdown of test statistics and traffic light system results for the Student's t-distribution. Starting with this distribution in combination with constant volatility and in the stressed period, it is apparent that the results are similar to those under Normal distribution applying the same volatility modelling. The test statistics for the ES estimates in 2008 are also in the red zone because they are more negative than the approximated critical value -1.8, thus the ES is underestimated by the method. This can be explained in the same manner as for the method with the Normal distribution and constant volatility. In this year there is a slight improvement when using Student's t-distribution as compared to the Normal distribution, such as the test statistic values are less negative in relation to the test statistics for the Normal distribution with constant volatility.

As seen in Table 4.1, the degrees of freedom parameters v differ significantly between 2008 and 2009. In year 2009, each portfolio had a much lower degrees of freedom parameter than in 2008, implying that the loss distributions exhibit leptokurtosis, such that it has heavy tails. Therefore, the Student's distribution improves the ES estimation for this year. Thus, the trend as found in the previous research of Student's t-distribution being the superior distribution can be supported here preliminary. Furthermore, it also becomes obvious that the results are influenced to a great extent by the volatility model applied and less by the distributional assumptions.

Moving on to the results of the same method for the calm period. In Table 4.6, it can be recognised that the ES results are similar, but less often correctly estimated, to those under Normal distribution with constant volatility. If ES estimates and its test statistics are rejected, it is due to overestimation. For this reason, the same explanation as for the Normal distribution in that period holds. The more surprising observation is that under Normal distribution with constant volatility more estimates are not rejected compared to when using Student's t-distribution, which does not confirm the outcome from the literature discussed and from this method's stressed period results found in this study. However, the method does not overestimate ES in such way it is in the red zone for overestimation, thus it is not rejected at the 0.01% significance level.

In summary, these results suggest and confirm further that the outcome of the backtest is greatly influenced by the volatility model applied, rather than by the underlying distribution that is assumed. This can be seen in the fact that estimates conducted with the same volatility model but with different distributional assumptions perform similarly.

The test statistics of the method using Student's t-distribution in combination with forecasted volatility using the GARCH model for the stressed period, as shown in Table 4.5, display an opposite trend compared to those ES estimates under the Normal distribution and with GARCH volatility for the same period. Further, if this method is rejected it is due to underestimation while for Normal distribution using GARCH the reason of rejection is overestimation. Regardless of this opposition, it appears that this is a good performing method combination as 10 out of 15 estimates in the stressed period are not rejected. Thus, it can be argued that this method combinations perform superior while both methods in combination with other distributions and volatility models estimate ES less often correct.

The estimated degrees of freedom parameters for 2009, as seen in Table 4.3, are also lower than those of 2008, but at the same time higher than the parameters estimated with constant volatility. The lower values of the degrees of freedom parameters indicate more leptokurtosis, thus heavy tails. Therefore, it is intuitive that the ES estimations were improved with the assumption of the

Student’s t-distribution since it can accommodate kurtosis larger than three, and thus account for the heavy tails.

If now moving on to the same combination in the method for the calm period, it is apparent that while 2014 and 2017 display opposite results when comparing to the ES estimates under Normal distribution with GARCH, the backtest outcomes for 2019 are in agreement. In total, for this period 15 out of 15 test statistics are correct and thus, are not rejected and receive the green light. This further supports the suggestion found for the stressed period that this method combination works well.

Taken together, the results for the method with Student’s t-distribution with constant volatility display similar patterns as for Normal distribution with constant volatility while The results in this section indicate that the method combination of Student’s t-distribution with GARCH estimates ES correctly for both stressed and calm period more often than other methods.

4.2.3 Backtest of ES Estimates with Skewed Student’s t-distribution

In this section the results of the ES estimation using the Skewed Student’s t-distribution with both constant volatility and GARCH for the stressed and calm period as seen in Table 4.5 and 4.6, are discussed.

For the stressed period using constant volatility, all estimates for the three tested years are rejected with mostly the red light (two exceptions for the amber light in 2011). In 2008, this is due to severe underestimation, similar to the Normal and Student’s t-distribution using constant volatility. For 2009 and 2011 the red light is given due to overestimation, which shows the constant volatility is not able to adjust for less stressed period when having a very stressed year’s volatility as input. This result indicates the in-sample and out-of-sample data dependence.

Nevertheless, it further appears that this particular distribution is less appropriate in fitting the data compared to Normal and regular Student’s t-distribution for the stressed period. Remarkably, all ES estimates in the stressed period are either rejected at the 0.01% significance with a red light or not rejected at the 0.01% level but still overestimated and thus, within the amber zone. Given this outcome, the method estimates ES poorly and it has a poor performance when compared to the other methods.

Turning to the calm period estimates with this distribution using constant volatility, it is noted that also for this period the methods using the Normal and Student's t-distribution using the same volatility model outperform the method with skew Student's t-distribution for every year in the calm period. In 2014 and 2017, the method in discussion is rejected for all years at the 0.01% significance level, as it overestimates the ES and thereby, it is placed in the red zone.

To conclude regarding the performance of the skew Student's t-distribution with constant volatility, it is found that the method more often overestimates the ES compared to other distributions with the same volatility modelling, which suggest that the issue arises from the skewed Student's t-distribution being unable to fit the data as the symmetric Student's t-distribution. A possible explanation might be that the positive skewness puts too much weight to the right of the distribution, which consequently leads to a too high ES estimate and thus, overestimation. The difference between existing studies' suggestion of an outperformance of the skewed Student's t-distribution and this study's result is surprising, especially when evaluating the skewness parameters being above zero and thus, displaying asymmetry, both for assets individually and the combined portfolios, as reported in Table 3.2 and 3.4.

Turning now to the backtest of the skewed Student's t-distribution with GARCH volatility for the stressed period. The results appear to follow a similar pattern to the method under Normal distribution with GARCH volatility; however, this method combination performs slightly better than for the Normal distribution. On the other hand, GARCH under this distribution compared to the Student's t-distribution do not appear to resemble in the same direction, while the latter displays a slight outperformance over this method. When skewed Student's t-distribution with GARCH is rejected it is because the method overestimates the risk estimates, which confirms the trend discussed previously for this distribution and thus, emerges as a common issue for the skewed Student's t-distribution. However, comparing the test statistics of the method with the skewed Student's t-distribution with constant volatility and GARCH volatility, it is apparent that the GARCH model can accommodate volatility clustering.

Regarding this method combination's result for the calm period, one can observe from Table 4.6 the test statistics appear to follow a different pattern to the GARCH under Student distribution. This distribution with GARCH performs better compared to the skewed Student's t-distribution with GARCH, as the latter does not estimate ES correctly for any portfolio in the three years. The results are rejected due to overestimation more often when compared to Normal and Student's t-distribution applying GARCH. While the overestimation in 2017 is partly a common problem due to the extremely low volatility in that year and the in comparison higher volatility in the sample period 2014-2016, also the estimates for 2019 are overestimated (received the amber light), which is not rejected by other GARCH implying methods.

To conclude the evaluation of the skewed Student's t-distribution, for both constant and time-varying GARCH volatility, this distribution overestimates ES more often than other distributions in both the stressed and calm period. Comparing the performance between the two periods, it can be summarized that it estimates ES for the stressed period better than for the calm years.

4.2.4 Backtest of ES Estimates with BHS and VWHS

The following discusses the last section of Table 4.5 and 4.6, which is the backtesting result for the methods BHS and VWHS with GARCH volatility. To start with, BHS in the stressed period naturally follows a similar result pattern as the Normal and Student's t-distribution with constant volatility for this period, as BHS also applies constant volatility. A difference is that this method estimates ES correctly more often in 2009 compared to the Normal distribution, which underestimated ES in that year. For 2008, it underestimates the forecasts due to the same reason being the calm input years for a stressed out-of-sample year. It is worth noting, while this method also receives the red light due to underestimation for all estimates in 2008, the underestimation is less severe in absolute test statistic values compared to the other two mentioned methods. The overall result for BHS in the stressed method displays that this method estimates ES well with less and lower underestimation compared to Normal and symmetrical Student's t-distribution. This suggests that a method that does not fit the data into a certain distribution can be beneficial. This result contradicts the findings in previous findings. In the study by Righi and Ceretta (2015), the historical simulation did not give accurate ES estimates.

Moving on Table 4.6 to discuss the BHS results for the calm period, it is apparent that this ES method receives solely the green and amber light for its ES estimations. The reason for rejecting in 2014 and 2017 is overestimation within the amber zone, meaning the overestimation is not severe, but some improvements to the method are required. Further, as found in many test results, BHS performs similar to other methods assuming constant volatility, which is in an indicator that it is due to the dependence between in-sample periods' and out-of-sample period's volatility.

To conclude, the BHS seems to give more accurate ES estimates in the calm years than it does for the stressed years. This suggests that especially in calm, but also in stressed years it can be beneficial to model the empirical distribution of the underlying data instead of forcing it into a certain type of distribution, in particular when forced into the skewed Student's t-distribution. Moreover, it is notable that BHS does not underestimate ES in the stressed period as severe as other comparable methods; however, it is also more likely to (slightly) overestimate ES in calmer periods.

Lastly, the VWHS using GARCH for the stressed period performs similar to other GARCH implying model, in particular to the method with Student's t-distribution and GARCH volatility. In 2008 the estimates are rejected at the 5% and 0.01% statistical level due to underestimation, and thus placed in the red zone. This is the common issue found in most methods due to the sudden financial distressed experienced in that year. Moreover, this method delivers accurate estimates and thus, a good overall performance for the stressed period, which supports the finding from BHS that stressed period's data can be modelled well in an empirical distribution.

The same method also performs superior in the calm period as it is solely rejected for two portfolios due to underestimation in 2019. This is explained by the in- and out-of-sample dependence again as the in-sample period is from an extremely calm period (2017) while the out of sample testing year is an average calm year. However, these underestimations receive the amber traffic light, which is a sign for solely slight underestimation.

To conclude, VWHS in combination with GARCH is one of the better performing methods in terms of estimating ES correctly for both the stressed and the calm period. This result is also supported in previous studies. It conforms with the findings of Righi and Ceretta (2015), which showed that the filtered Historical Simulation performed well in their study. Nevertheless, it also suffers from dependence of input data and its volatility like most other methods also experience.

4.2.5 Portfolio Performance

In this thesis, the number of backtests of ES estimation performed results in a total of 240 test statistics, 48 for each portfolio. Table 4.7 presents the number of test statistics in each zone between the five trading portfolios; thus, it reveals the difference between each portfolio in regard to the result of the ES estimation methods. As can be seen from the table, the ES on the Portfolios 2 and 5 are estimated correctly by the estimation methods 28 and 27 times, respectively, such that the ES is not underestimated or overestimated, and therefore the methods have not been rejected at the 5% significance level and were placed within the green zone. In other words, using these two portfolios the ES has been estimated correctly most times by the methods. Interestingly, there are no similarities between these two portfolios in terms of weights in the assets. Portfolio 1 and 3 have fewest ES methods in the green zone and the greatest number of methods in the amber zone. These two portfolios have the greatest weights in commodity, 35% and 30%, respectively. As seen in Table 3.3, the commodity index has the highest standard deviation behind the two equity indices, which may explain larger deviations in Portfolio 1 and 3 comparably to the other portfolios. However, as discussed in section 3.1.6 and Table 3.5, the portfolios with the largest minimum and maximum and thus, standard deviation are the ones with the highest equity weights, which are portfolio 4 and 5. Therefore, the fact that portfolio 1 and 3 are estimated correctly and received a green traffic light slightly less often compared to the other portfolios might be more coincidentally rather than a meaningful finding.

Table 4.7 Cumulative Number of Test Statistics in each Traffic Light Zones for Each Portfolio over all tested Years

Zone	Portfolio 1	Portfolio 2	Portfolio 3	Portfolio 4	Portfolio 5
Green Zone	23	28	22	25	27
Amber Zone	14	8	13	10	8
Red Zone	11	12	13	13	13

While the portfolio performance analysis showed that the weights distribution between asset classes is not relevant for the performance of ES methods, this finding validates the design of this study and increases its applicability to banks. Assuming the actual bank's trading book is composed of similar asset classes (as suggested by Basel), the actual asset weights implied by a bank are unknown and differ between banks; however, this is shown irrelevant as methods give similar results (in terms of rejection and non-rejection) to all five tested portfolios. Therefore, this finding is reliable as it was tested for five portfolios for in total six years and thus, this study design is valid and applicable to real banks.

4.2.6 Year by Year Analysis

As one of the sub-research objectives is to determine whether methods perform differently or similarly in stressed versus calm periods, the following summarizes the main findings for this (see Table 4.8 for a detailed year by year method performance breakdown).

Table 4.8 Year by Year Method Performance Ranking

Year	Method Ranking		Number of Rejections (per method)
2008	1st	ES-N-GARCH, ES-Skew t-GARCH	0
	2nd	All other methods	5
2009	1st	ES-t, ES-t-GARCH, ES-BHS	0
	2nd	ES-VWHS-GARCH	1
	3rd	ES-N	3
2011	1st	ES-N, ES-t-GARCH, ES-BHS, ES-VWHS-GARCH	0
	2nd	ES-t	1
	3rd	ES-Skew t-GARCH	2
2014	1st	ES-N, ES-t-GARCH, ES-VWHS-GARCH	0
	2nd	ES-BHS	3
	3rd	ES-t	4
2017	1st	ES-N, ES-t-GARCH, ES-VWHS-GARCH	0
	2nd	ES-BHS	1
	3rd	ES-t	2
2019	1st	ES-N, ES-t, ES-t-GARCH, ES-BHS	0
	2nd	ES-N-GARCH	1
	3rd	ES-VWHS-GARCH	2

The following discussion is solely concerned about the methods that achieve zero rejections in the respective years to determine the methods that estimate each portfolio's ES correctly in that year and not to just determine methods that are able to estimate ES correctly most of the time or a few times only.

Firstly, it is noted that there is no clear indication that one certain method or method combination outperforms the other methods dominantly. Furthermore, the only method that was unable to estimate ES correctly without zero rejections for at least one portfolio in one year is the skewed Student's t-distribution with constant volatility. In total, the Student's t-distribution using GARCH performed the best as it was able to estimate ES correctly for all portfolios for five out of six tested years, followed by the Normal distribution using constant volatility, which achieved zero rejections for all portfolios for four out of six years.

To look closer in which periods these non-rejections occurred, it is apparent that the Normal distribution with constant volatility was able to predict ES correctly for each of the calm years and for only one stressed year. Consequently, it can be assumed that this method performs better in calm periods. Secondly, the Student's t-distribution using GARCH was able to estimate ES correctly for two years in the stressed period and for all years in the calm period. Lastly, the third most correctly estimating methods in this study is VWHS with GARCH and BHS as they have zero rejections for in total three out of the six tested years. For VWHS with GARCH two of these years are calm years and solely one stems from the stressed period, while it is the other way around for BHS (correctly estimated for two stressed and one calm year). This small difference in distribution of correct estimates of the non-parametric approaches between calm and stressed years is not significant.

To summarize the emerged findings regarding method performance in stressed versus calm period, the Student's t-distribution using GARCH performs well in both periods and while the Normal distribution with constant volatility show a (slightly) better estimation performance for the calm period.

4.3 Summary of Results and Analysis

The following summarizes the most meaningful trends identified in the method by method and year by year analysis. As the portfolio performance analysis did not result in any significance trends, this part is omitted here.

This thesis is concerned with which ES estimation method estimate the ES such that the ES is neither underestimated not overestimated both the perspective of the Basel Committee and the perspective of Banks are considered. From the result, there are more methods that would be preferred as seen from the perspective of the Basel Committee because in general the methods applied in this thesis on the portfolio leads to more overestimations than underestimations. Hence, the regulatory perspective should emphasis the application of the Normal Distribution with GARCH or the skewed Student's t-distribution for stressed periods and Normal distribution with both constant volatility and GARCH volatility, the Student's t-distribution and the VWHS model with GARCH volatility for the calm period. The banks are concerned with the overestimation since it can lead to inefficient capital allocation, therefore the banks would prefer the method with the Student's t-distribution and GARCH volatility or the VWHS with GARCH volatility.

In Table 4.9 the cumulative number of lights awarded to each method as identified by the test statistic can be found. Those results are discussed below for the perspective of both Basel and banks, meaning the method only performs well when it does not over- nor underestimate ES.

Table 4.9 Cumulative Number of Test Statistics in each Traffic Light Zones for Each Portfolio, sorted by methods used

Zone	ES-N	ES-N-GARCH	ES-t	ES-t-GARCH	ES-Skew t	ES-Skew t-GARCH	ES-BHS	ES-VWHS-GARCH
Green Zone	22	9	18	25	0	8	21	22
Amber Zone	3	11	7	0	7	18	4	3
Red Zone	5	10	5	5	23	4	5	5

Based on this quantitative analysis, the method, which estimates ES correctly more often and is within the green light boundaries, throughout all periods and portfolios is the Student's t-distribution with GARCH volatility, received the green lights 25 times. Moreover, the Normal distribution with constant volatility and the VWHS with GARCH received the green light 22 times. Therefore, these are on the second place after the Student's t-distribution with GARCH. On third place comes the BHS with 21 correct estimations and thus, green lights. The remaining methods score a considerably lower number of correct estimations and hence, are not considered well performing in this study.

While quantitively speaking and also the method-by-method and year-by-year analysis shows that there are several methods that perform similarly well in the estimation of the ES correctly and thus, there is no clear indication that there is a single method or method combination that is superior.

The correct estimations from methods such as Student's t-distribution, GARCH and VWHS are conformed with the reviewed literature and the reported importance of taking stylized facts into account (Huang, 2000; McNeil & Frey, 2000; Righi & Ceretta, 2015; Tang & Do, 2018). Moreover, especially the finding that the Normal distribution with constant volatility modelling belongs to the one of the best performing methods in this study contradicts the findings from existing research. While this distribution was only included in the tests to potentially find outperforming of other distributions (that account for stylized financial facts such as excess kurtosis and skewness), it proves to be able to estimate ES correctly. Moreover, the skewed Student's t-distribution was promising as it accounts for two important financial data facts; nevertheless, its result in particular in combination with constant volatility is rather disappointing.

Finally, the single most consistent and remarkable finding to emerge from the analysis is the strong dependence between in-sample data and out of sample and their level of standard deviation. The more similar that level is between input years and testing year, the better are the ES estimates in terms of not being under- or overestimated. This is a trend seen by almost every tested method and year. In general, this accords with earlier observations by Righi and Ceretta (2015), indicating that the estimation window itself is an important variable in estimation of and the performance of the ES estimates. This study confirms that the performance of the ES method is associated with the estimation window, and thus and important variable in ES estimation as suggested by Righi and Ceretta (2015).

This result is partly related to the findings observed in earlier studies by Yamai and Yoshiba (2005), indicating that the sample size for estimating ES must be larger than for VaR. This identifies a possibility for further research to investigate whether, given the same data as in this study, the ES estimations are correct more often when increasing (or potentially decreasing) the in-sample size from three years.

5 Conclusion

The study was designed to determine the methods that are able to correctly estimate the Expected Shortfall on a bank's trading book for both stressed and calm period, in such a way that the methods are not rejected for under- or overestimation when backtesting the Expected Shortfall estimates using the second backtest developed by Acerbi and Szekely (2014). The following answers each sub-research questions directly before drawing an overall conclusion concerning the main objective.

From previous studies, the established methods to estimate the ES are those that in different combinations make allowance for the stylised facts of financial data, namely excess kurtosis, skewness and volatility clustering. In the previous research, the methods that correctly estimated the ES are the symmetrical Student's t-distribution and the skewed Student's t-distribution, with conditional volatility from volatility models such as GARCH(1,1) or EWMA, VWHS and EVT with GARCH volatility, and the method that was commonly rejected is the one with Normal distribution innovations. In accordance with the previous research findings, this study has implemented the Normal distribution, the symmetrical and skewed Student's t-distributions and VWHS, with constant volatility and GARCH forecasted volatility, and also the BHS, in order to estimate ES. The Student's t-distribution implies the degrees of freedom parameter, which measures the excess kurtosis, and the skewed version of this distribution takes the skewness parameter into account, which allows for an asymmetric distribution. The BHS models the empirical distribution of the underlying data. Furthermore, GARCH and VWHS take volatility clustering and time-varying volatility into the estimation procedure.

A review of the existing backtesting methods for ES estimates revealed that the second backtest by Acerbi and Szekely (2014) was the most appropriate one as it tests ES directly without requiring a Monte Carlo simulation as other backtests do. This implies that a rather comprehensible and adaptable backtesting method is available for evaluation of the ES estimation methods. In order to resemble the method proposed by the Basel Committee, an extension of the backtesting method by Acerbi and Szekely (2014) was conducted in order to account for both under- and overestimation of the ES. This was made by placing the methods not rejected at the significance level 5% (test statistic above or below the critical value -0.7 and +0.59, respectively) in the green zone, the methods rejected at the 5% level and not rejected at the 0.01% significance level in the amber zones and finally, the methods with a test statistics exceeding the approximated critical values -1.8 and +0.93 for the 0.01% significance level in the red zones.

To construct a representative portfolio for a bank's trading book several asset classes, namely equity, foreign exchange, commodities, proxies for credit spread and interest rates, as well as an equity option, as suggested in Basel's *Analysis of the trading book hypothetical portfolio exercise,* were included in all-in portfolios. The asset weighting was made in accordance with the suggestions of PwC (2016); however, as the actual assets and its weights are unknown and differ between banks, in total five all-in portfolios were assembled to mitigate this uncertainty. The testing was performed on the daily log losses of these five portfolios. The results in this study showed that the weights distribution between asset classes is not relevant for the performance of ES methods, therefore this finding validates the design of this study and increases its applicability to banks.

An finding to emerge from this study's testing is that the Student's t-distribution and the skewed Student's t-distributions are only superior to the Normal distribution with GARCH, otherwise when the constant volatility is applied the method with the Normal distribution showed good performance in relation to the methods with Student's t-distributions. Furthermore, the choice of volatility model seems to have a greater impact on the performance of the ES estimate as compared to the choice of distributional assumptions.

Moreover, a quantitative analysis of the number of non-rejections showed that the Student's t-distribution with GARCH volatility received the highest number of non-rejections of the test statistics, followed by the Normal distribution with constant volatility and the VWHS with GARCH and lastly BHS. All other methods received a considerably lower number of not rejected test statistics. This study has shown that the method using the Student's t-distribution with GARCH volatility is the method that performs equally as well in the stressed and the calm periods. The Normal distribution with constant volatility gives more correct ES estimations for the calm period. Both VWHS and BHS do not display a significant trend in which period they perform better.

This thesis is concerned with both underestimation and overestimation of the ES, thereby the perspective of the Basel's Committee and the perspective of the banks are considered. This study showed that the methods that correctly estimates the ES on a bank's trading book for both the stressed and calm period, taking both under- and overestimation into consideration, are the methods with the Student's t-distribution with GARCH volatility and secondly also VWHS with GARCH volatility. Regarding estimations in calm out-of-sample periods, also the Normal distribution with constant volatility predicts ES correctly. The most unanticipated finding should be acknowledged here, which affects all methods used. It emerged that there is a strong dependence between the in-sample period's standard deviation and the backtest result for the out-of-sample testing year. It was identified that the more similar the level of deviation in these years, the more often does any method estimate ES correctly.

6 Recommendations

The study should be repeated using the same data and all-in portfolios but changing the in-sample period by length, such as one years, two years or five years, and possibly by using a rolling window for parameter updating. The in-sample period should be kept constant throughout the study to be comparable across methods and testing years. By changing the length of the in-sample period, different levels of standard deviations will be used to predict ES for the out of sample year. It should be studied whether or not an increase, or decrease, of the in-sample period's length has a consistent impact on the correctness of estimations. The issue of this dependence is an intriguing one, which could be usefully explored, such that, to give an example, a shorter in-sample period might have a more accurate level of standard deviation comparing to the testing year as they are timely more close or whether a longer in-sample period might ensure stability among ES estimates and thus, is estimating it correctly more often.

Moreover, as acknowledged in delimitations, due to a timely restriction put on this thesis, the EVT was not tested; however, it was argued to be a well-performing method by existing literature and thus, should be tested in further research that implements a similar representative portfolio of a bank's trading book.

The findings of this study have a number of important implications for future practice of estimating ES correctly. The information and findings can be used to develop targeted courses of actions for when estimating ES:

- To imply an ES estimation method that proved to perform constant between periods (such as the Student's t-distribution with GARCH), when the underlying condition of the to-be-estimated period is unknown or uncertain
- To imply an ES estimation method that proved to perform better in either calm or stressed period (such as the Normal distribution with constant volatility for calm periods), when the underlying condition of the to-be-estimated period can be forecasted with certainty
- To determine an appropriate and stable in-sample period (based on length and level of standard deviation) to achieve more accurate and correct ES estimates

References

Abad, P., Benito, S. & López, C. (2013). A comprehensive review of Value at Risk methodologies, *The Spanish Review of Financial Economics*, vol. 12, pp. 15-32

Abbassi, P. & Schmidt, M. (2018). A comprehensive view on risk reporting: Evidence from supervisory data, *Journal of Financial Intermediation*, vol. 36, pp. 74-85

Acerbi, C. & Szekely, B. (2014). Backtesting Expected Shortfall: Introducing three model-independent, non-parametric back-test methodologies for Expected Shortfall, Available online, https://www.msci.com/documents/10199/22aa9922-f874-4060-b77a-0f0e267a489b [Accessed 30 March 2020]

Altun, E., Tatlidil, H. & Ozel, G. (2019). Value-at-risk estimation with new skew extension of generalized normal distribution, *Communications in Statistics – Theory and Methods*, vol. 48, pp. 3663-3681

Ballotta, L. & Fusai, G. (2017). A Gentle Introduction to Value at Risk, Available online: https://www.researchgate.net/publication/315924017_A_Gentle_Introduction_to_Value_at_Risk [Accessed 18 March 2020]

Bank for International Settlements. (2014). Analysis of the trading book hypothetical portfolio exercise, Basel Committee on Banking Supervision, Available online: https://www.bis.org/publ/bcbs288.pdf [Accessed 6 April 2020]

Bank for International Settlements. (2012). Fundamental review of the trading book, Basel Committee on Banking Supervision, Available online: https://www.bis.org/publ/bcbs219.pdf [Accessed 6 April 2020]

Bank for International Settlements. (2020). History of the Basel Committee, Available online: https://www.bis.org/bcbs/history.htm?m=3%7C14%7C573%7C76 [Accessed 6 April 2020]

Bank for International Settlements. (2016). Minimum capital requirements for market risk, Basel Committee on Banking Supervision, Available online: https://www.bis.org/bcbs/publ/d352.pdf [Accessed 6 April 2020]

Bank for International Settlements. (2019a). The market risk framework – In brief, Basel Committee on Banking Supervision, Available online: https://www.bis.org/bcbs/publ/d457_inbrief.pdf [Accessed 28 May 2020]

Bank for International Settlements. (2019b). Minimum capital requirements for market risk, Committee on Banking Supervision, Available online: https://www.bis.org/bcbs/publ/d457.pdf [Accessed 6 April 2020]

Bank for International Settlements. (2000). Stress Testing by Large Financial Institutions: Current Practise and Aggregation Issues, Available online: https://www.bis.org/publ/cgfs14.pdf [Accessed 6 April 2020]

Bauwens, L. & Laurent, S. (2002). A new class of multivariate skew densities, with application to GARCH models, Available online: https://www.researchgate.net/publication/23977564_A_new_class_of_multivariate_skew_densities_with_appplication_to_GARCH_models [Accessed 22 May 2020]

Black, F. (1976). The pricing of commodity contracts, *Journal of Financial Economics*, vol. 3, no. 1-2, pp. 167-179

Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity, *Journal of Economics*, vol. 32 no. 3, pp. 407-327

Brooks, C. (2014). Introductory Econometrics for Finance, 3rd ed, Cambridge: Cambridge University Press

Embrechts, P., Frey, R. & McNeil, A. J. (2005). *Quantitative Risk Management Concepts, techniques and tools*, Princeton University Press Princeton and Oxford.

Engle, R.F. & Manganelli, S. (2004). CAViaR: Conditional autoregressive value at risk by regression quantiles, *Journal of Business and Economic Statistics*, vol. 22, no. 4, pp. 367-381

Harmantzis, F. C., Miao, L. & Chien, Y. (2006). Empirical study of value-at-risk and expected shortfall models with heavy tails, *The Journal of Risk Finance*, vol. 6, pp. 117-135

Huang, A. (2000). A Comparison of Value at Risk Approaches and a New Method with Extreme Value Theory and Kernel Estimator, Available online: https://www.researchgate.net/publication/228708519_A_Comparison_of_Value_at_Risk_Approaches_and_a_New_Method_with_Extreme_Value_Theory_and_Kernel_Estimator [Accessed 16 March 2020]

Hull, J. & White, A. (1998). Incorporating volatility updating into the historical simulation method for VaR, *Journal of Risk*, vol. 1

Intercontinental Exchange. (2020). U.S. Dollar Index, Available online: https://www.theice.com/publicdocs/ICE_USDX_Brochure.pdf [Accessed 13 May 2020]

Jones, M. C. & Faddy, M. J. (2002). A skew extension of the t-distribution, with applications. *J. R. Statist. Soc. B*, vol. 65, pp. 159-174

Krause, J. & Paolella, M. S. (2014). A Fast, Accurate Method for Value-at-Risk and Expected Shortfall, *Econometrics*, vol. 2, pp. 98-122

Kupiec, P. (1995). Techniques for Verifying the Accuracy of Risk Management Models, *Journal of Derivatives*, vol. 3, pp. 73-84

Laurent, S. (2002). Asymmetry and fat-tails in financial time series, Thesis, Universiteit Maastricht, Available online: https://cris.maastrichtuniversity.nl/ws/portalfiles/portal/1186499/guid-4bdc3e21-ba0d-4e82-80bb-d16bafffb4f7-ASSET1.0.pdf [Accessed 29 May 2020]

Lazar, E. & Zhang, N. (2019). Model risk of expected shortfall, *Journal of Banking & Finance*, vol. 105, pp. 74-93

Lin, C.-H. & Shen, S.-S. (2006). Can the student-t distribution provide accurate value at risk?, *The Journals of Risk Finance*, vol. 7, pp. 292-300

McNeil, A. J. & Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach, *Journal of Empirical Finance*, vol. 7, no. 3–4, pp. 271–300

Peters, J.-P. (2001). Estimating and forecasting volatility of stock indices using asymmetric GARCH models and (Skewed) Student-t densities, Ecole d'Administration des Affaires, University of Liège, Available online: https://pdfs.semanticscholar.org/f445/493c46b3b4d52e958f69d502ccf8f03d51c3.pdf [Accessed 13 May 2020]

PwC. (2016). Basel IV: Revised Standardised Approach for Market Risk: Increasing risk sensitivity through the "Sensitivities-based Method". Available online: https://www.pwc.com/gx/en/advisory-services/basel-iv/basel-iv-revised-standardised-.pdf [Accessed 29 April 2020]

Righi, M. B. & Ceretta, P. S. (2015). A comparison of Expected Shortfall estimation models, *Journal of Economics and Business*, vol. 78, pp. 14-47

Tang, M.-L. & Do, T. K. (2018). In search of robust methods for multi-currency portfolio construction by value at risk, *Asia-Pacific Financial Markets*, vol. 26, pp. 107-126

Taylor, S.J. (1986). Modeling Financial Time Series. Chichester: John Wiley and Sons

Thomson Reuters. (2012). Thomson Reuters Indices: Continuous Commodity Total Return Index, Available online: https://www.thomsonreuters.com/content/dam/openweb/documents/pdf/tr-com-financial/methodology/cci-methodology.pdf [Accessed 13 May 2020]

Yamai, Y. & Yoshiba, T. (2005). Value-at-Risk Versus Expected Shortfall: A Practical Perspective, *Journal of Banking & Finance*, vol. 29, pp. 997–1015

Appendix

Appendix 1: Excel Formulas and Functions

Section	Description	Excel Function
Data Return	**Data Return**	= LN(Data_t) - LN(Data_t-1)
Descriptive Statistics	**Minimum**	= MIN(Data)
	Maximum	= MAX(Data)
	Mean	= AVERAGE(Data)
	Median	= MEDIAN(Data)
	Standard Deviation	= STDEV.S(Data)
	Skewness	= SKEW(Data)
	Kurtosis	= KURT(Data)
BS Option Pricing	**Yearly Volatility**	= SQRT(252)*STDEV.S(Data(T))
	Option price C t-1	=(NORM.DIST(d1,0,1,TRUE)*S[t])-(NORM.DIST(d2,0,1,TRUE)*K*EXP(-r*t))
Portfolio Constructi on	**Portfolio 1**	=(Equity/2)*0.45+(Proxies for IR and CS/4)*0.05+Commodity*0.35+FX*0.1+Option*0.05

	Portfolio 2	=(Equity/2)*0.5+(Proxies for IR and CS/4)*0.05+Commodity*0.2+FX*0.2+Option*0.05
	Portfolio 3	=(Equity/2)*0.52+(Proxies for IR and CS/4)*0.05+Commodity*0.3+FX*0.05+Option*0.08
	Portfolio 4	=(Equity/2)*0.55+(Proxies for IR and CS/4)*0.1+Commodity*0.2+FX*0.05+Option*0.1
	Portfolio 5	=(Equity/2)*0.65+(Proxies for IR and CS/4)*0.1+Commodity*0.1+FX*0.1+Option*0.05
Normal Distributio n GARCH	**Residual (t)**	=Data(t)-mu
	Time varying variance (t=1)	=STDEV(Data of 750 observations)
	Time varying variance (all t>1)	=SQRT(omega+alpha*residual(t-1)^2+beta*time varying variance (t-1))
	Loglikelihood	=-0.5*LN(time varying variance(t))-(residual(t)^2/(2*time varying variance(t)^2))
	Max Sum	=SUM(Loglikelihood for all t)
Student t constant volatility	**z(t)**	=(Data(t)-mu)/time varying variance(t)
	ML constant sigma	=LN(GAMMA((v+1)/2))-LN(GAMMA(v/2))-0.5*LN(PI()*(v-2))-0.5*(LN(sigma)^2)+(1+v)*LN(1+((z(t)^2)/(v-2))))
Student t GARCH	**ML GARCH**	=LN(GAMMA((v+1)/2))-LN(GAMMA(v/2))-0.5*LN(PI()*(v-2))-0.5*(LN(time varying volatility (t)^2)+(1+v)*LN(1+((z(t)^2)/(v-2))))
skewed Student t constant volatility	**I(t)**	=IF(z(t)<m/s,-1,1)

	ML constant sigma	=LN(2/(xsi+(1/xsi)))+LN(GAMMA((v+1)/2))-0.5*LN(PI()*(v-2))-LN(GAMMA(v/2))+LN(s/sigma)-0.5*(v+1)*LN(1+(((s*z(t)+m)^2*xsi^(-2*I(t)))/(v-2)))
skewed Student t GARCH	**ML GARCH**	=LN(2/(xsi+(1/xsi)))+LN(GAMMA((v+1)/2))-0.5*LN(PI()*(v-2))-LN(GAMMA(v/2))+LN(s/time varying volatility (t))-0.5*(v+1)*LN(1+(((s*z(t)+m)^2*xsi^(-2*I(t)))/(v-2)))
VWHS Estimation	**Jan.**	=GARCH SD Normal Distribution first estimate in Jan from test period
	Scaled Loss GARCH STD(t+1) Jan.	=Data*(Jan./GARCH SD(t))
VaR and ES estimations :	**Constant mean (t)**	=AVERAGE(in-sample data[rolling window])
	Constant volatility (t)	=STDEV.S(in-sample data[rolling window])
Normal Distribution	**VaR N (t) constant volatility**	=constant mean(t)+constant volatility(t)*NORM.S.INV(0.975)
	ES N (t) constant volatility	=constant mean(t)+constant volatility(t)*(NORM.S.DIST(NORM.S.INV(0.975),0)/(1-0.975))
	Violations ES N (t) constant volatility	=(Data(t)*(IF(Data(t)>VaR N (t),1,0))/ES N (t))
	VaR N GARCH (t)	=constant mean(t)+GARCH Normal time varying volatility (t)*NORM.S.INV(0.975)
	ES N GARCH (t)	=constant mean(t)+GARCH Normal time varying volatility (t)*(NORM.S.DIST(NORM.S.INV(0.975),0)/(1-0.975))
	Violations ES N GARCH (t)	=(Data(t)*(IF(Data(t)>VaR N GARCH (t),1,0))/ES N GARCH (t))
Student t	**v Student t constant volatility**	estimated by solver

	v Student t GARCH	estimated by solver
	VaR t (t) constant volatility	=constant mean (t)+SQRT(((v Student t constant volatility-2)/v Student t constant volatility))*constant volatility (t)*T.INV(0.975,v Student t constant volatility)
	ES t (t) constant volatility	=constant mean(t)+SQRT(((v Student t constant volatility-2)/v Student t constant volatility))*constant volatility (t)*T.DIST(T.INV(0.975,v Student t constant volatility),v Student t constant volatility,0)/(1-0.975)*((v Student t constant volatility+((T.INV(0.975,v Student t constant volatility))^2))/(v Student t constant volatility-1))
	Violations ES t (t) constant volatility	=(Data(t)*(IF(Data(t)>VaR t (t),1,0))/ES t (t))
	VaR t GARCH (t)	=constant mean+SQRT(((v Student t GARCH-2)/v Student t GARCH))*GARCH t time varying volatility(t)*T.INV(0.975,v Student t GARCH)
	ES t GARCH (t)	=constant mean+SQRT(((v Student t GARCH-2)/v Student t GARCH))*GARCH t time varying volatility (t)*T.DIST(T.INV(0.975,v Student t GARCH),v Student t GARCH,0)/(1-0.975)*((v Student t GARCH+((T.INV(0.975,v Student t GARCH))^2))/(v Student t GARCH-1))
	Violations ES t GARCH (t)	=(Data(t)*(IF(Data(t)>VaR t GARCH (t),1,0))/ES t GARCH (t))
Skew t constant volatility	**[v, xsi, mu] skew t constant volatility**	estimated by solver
	m skew t constant volatility	=(((GAMMA(((v skew t constant-1)/2)))*SQRT(v skew t constant-2))/((SQRT(PI()))*GAMMA(v skew t constant/2)))*(xsi skew t constant-(1/xsi skew t constant))
	s skew t constant volatility	=SQRT((xsi skew t constant^2+(1/(xsi skew t constant^2))-1)-m skew t constant^2)

	Quantile skew t (prob.) constant volatility	=(IF(prob.<(1/(1+xsi skew t constant^2)), ((1/xsi skew t constant)*T.INV((prob./2)*(1+xsi skew t constant^2),v skew t constant)-m skew t constant)/s skew t constant, ((-xsi skew t constant)*T.INV(((1-prob.)/2)*(1+(1/xsi skew t constant^-2)),v skew t constant)-m skew t constant)/s skew t constant))
	VaR skew t (t) constant volatility (prob.)	=constant mean (t)+sigma skew t constant*Quantile skew t constant(prob.)
	ES skew t (t) constant volatility	=AVERAGE(VaR skew t (t) constant volatility for all prob.)
	Violations ES skew t (t) constant volatility	=Data(t)*(IF(Data(t)>VaR skew t (t) constant (prob.),1,0))/ES skew t (t) constant
	Average Violations ES skew t (t) constant volatility	=AVERAGE(Violations ES skew t (t) constant volatility)
Skew t GARCH	**[v, xsi, mu, alpha, beta, omega] skew t GARCH**	estimated by solver
	m skew t GARCH	=(((GAMMA(((v skew t GARCH-1)/2)))*SQRT(v skew t GARCH-2))/((SQRT(PI()))*GAMMA(v skew t GARCH/2)))*(xsi skew t GARCH-(1/xsi skew t GARCH))
	s skew t GARCH	=SQRT((xsi skew t GARCH^2+(1/(xsi skew t GARCH^2))-1)-m skew t GARCH^2)
	Quantile skew t (prob.) GARCH	=(IF(prob.<(1/(1+xsi skew t GARCH^2)), ((1/xsi skew t GARCH)*T.INV((prob./2)*(1+xsi skew t GARCH^2),v skew t GARCH)-m skew t GARCH)/s skew t GARCH, ((-xsi skew t GARCH)*T.INV(((1-prob.)/2)*(1+(1/xsi skew t GARCH^-2)),v skew t GARCH)-m skew t GARCH)/s skew t GARCH))
	VaR skew t (t) GARCH (prob.)	=constant mean (t)+sigma skew t GARCH*Quantile skew t GARCH(prob.)
	ES skew t (t) GARCH	=AVERAGE(VaR skew t (t) GARCH for all prob.)

	Violation ES skew t (t) GARCH	=Data(t)*(IF(Data(t)>VaR skew t (t) GARCH (prob.),1,0))/ES skew t (t) GARCH
	Average Violation ES skew t (t) GARCH	=AVERAGE(Violation ES skew t (t) GARCH)
BHS constant volatility	**VaR BHS (t)**	=PERCENTILE.INC(750 obs. input data (t),0.975)
	ES BHS (t)	=AVERAGEIF(750 obs. Input data, CONCATENATE(">",VaR BHS(t)))
	Violations ES BHS (t)	=Data(t)*(IF(Data(t)>VaR BHS (t), 1, 0))/ES BHS (t)
VWHS GARCH	**VaR VWHS GARCH (t) (start at t=750)**	=PERCENTILE.INC(Scaled Loss GARCH STD Jan. 1: Scaled Loss GARCH STD Jan. 749),0.975)
	Violations VWHS GARCH (t)	=IF(Data(t=750)>VaR HS GARCH (t=750),1,0)
	ES VWHS GARCH (t) (start at t=750)	=AVERAGEIF(Sclaed Loss GARCH STD Jan.1:Scaled Loss GARCH STD Jan.749, CONCATENATE(">",VaR HS GARCH 750))
	Violations ES VWHS GARCH (t)	=Data(t)*(IF(Data(t)>VaR HS GARCH (t), 1, 0))/ES HS GARCH (t)
Backtest: Test Statistics	**T*(1-alpha)**	=252*(1-0.975)
	ES-N	=-(1/T*(1-alpha))*SUM(Violations ES N (t) constant volatility)+1
	ES-N-GARCH	=-(1/T*(1-alpha))*SUM(Violations ES N (t) GARCH)+1
	ES-t	=-(1/T*(1-alpha))*SUM(Violations ES t (t) constant volatility)+1

ES-t-GARCH	=-(1/T*(1-alpha))*SUM(Violations ES t (t) GARCH)+1
ES-skew t-constant	=-(1/T*(1-alpha))*SUM(Average Violations ES skew t (t) constant volatility)+1
ES-skew t-GARCH	=-(1/T*(1-alpha))*SUM(Average Violations ES skew t (t) GARCH)+1
ES-BHS	=-(1/T*(1-alpha))*SUM(Violations ES BHS(t))+1
ES-VWHS-GARCH	=-(1/T*(1-alpha))*SUM(Violations ES VWHS GARCH (t))+1