Erasmus University Rotterdam

Erasmus School of Economics

Master Econometrics & Management Science : Quantitative Finance

Master's Thesis

The Estimation of Expected Shortfall in ETF Portfolios

Author: Cor van Loon

Supervisor: S. Barendse

Co-Reader: X. Leng

Abstract

This research evaluates the merit of the multivariate TGARCH-EVT-Copula model in forecasting Value at Risk and Expected Shortfall of a portfolio of Exchange Traded Funds by comparing the model with conventional benchmarks. The Value at Risk forecasts are tested using the standard Kupiec and Christoffersen tests, whereas the Expected Shortfall forecasts are tested using the state of the art testing method of Du & Escanciano (2017). Models that pass these tests are pairwise tested using the Diebold-Mariano framework with the comparative Fissler & Ziegel (2016) scoring function. The results show that including time-varying volatility improves the performance and that the multivariate TGARCH-EVT-Gaussian copula model is superior to the other models. It is also shown that multivariate modeling in combination with extreme value theory and TGARCH filtering improves the forecasting ability of ES.

Keywords : Value at Risk, Expected Shortfall, GARCH Filtering, Extreme Value Theory, Copula, Backtesting

Acknowledgements

I would like to express my gratitude to my supervisor Sander Barendse for the constructive discussions, support and encouragement. I would also like to thank Xuan Leng for her valuable comments and feedback as co-reader of this Master's Thesis. Last but not least I would like to thank my girlfriend Rosanne who has supported me along the way and never let me down!

Rotterdam, October 2017

Cor van Loon

Contents

1 Introduction

Financial institutions and other investors have to properly manage their exposure to risk factors. The crisis of 2008 has left the impression that risk management is of great importance. With the increasing number of financial products available in the investors portfolios, it has become quite difficult to measure the magnitude of these risks. All these trends have led to a demand for a portfolio-level quantitative measure, that measures the market risk appropriately. The leading measure of this type is known as Value at Risk (VaR), this is the loss level that will not be exceeded with a certain confidence level during a certain period of time (Artzner et al., 1999). Another frequently used risk measure is known as Expected Shortfall (ES), this is the expected loss beyond a given confidence level (Acerbi et al., 2001).

The contribution of this research to the existing literature is to evaluate the merit of multivariate TGARCH-EVT-Copula models over univariate models in VaR and ES estimation of a portfolio of rather unknown financial instruments, namely Exchange Traded Funds that track financial assets, such as the S&P 500 and NASDAQ. As this class of instruments is increasingly popular in the current financial world it is worthwhile to examine the properties of VaR and ES in such portfolios. To test the correct specification of the ES forecasts the state of the art testing methodology of Du & Escanciano (2017) is implemented. Furthermore a pairwise comparison between models that passed all tests is conducted using the Diebold & Mariano (1995) framework with the comparitive scoring function of Fissler & Ziegel (2016).

This research considers a number of benchmark univariate models for the estimation of VaR and ES like the Variance-Covariance model, the Historical Simulation model and the Extreme Value Theory (EVT) model. A model that incorporates time-varying volatility in the estimation is considered, this approach is known as Filtered Historical Simulation. Also a univariate approach that combines EVT and GARCH filtering is implemented. Time-varying volatility, EVT and copulas are combined in the multivariate TGARCH-EVT-Copula model.

The results show that including time-varying volatility improves the forecasting abilitie. The results from the TGARCH(1,1) and TGARCH-EVT model gives strong evidence for the use of these models in univariate forecasting of VaR and ES. From the multivariate models the TGARCH-EVT-Gaussian model performs best overall, but also the Frank and Student model perform well at the different significance levels. Furthermore it is shown that multivariate modeling improves the forecasting ability of ES.

The focus of this research is on Expected Shortfall because the Basel Committee on Banking Supervision decided in 2012 to chose ES as replacement for VaR. Until 2012 VaR has been chosen by the Basel Committee on Banking Supervision as the benchmark of risk measurement for capital requirements, unsurprisingly VaR has received a lot of negative attention in the media with lots of attention to the weaknesses of this technique. VaR is not sub-additive, meaning that the risk of a certain portfolio of assets can be larger than the sum of the individual risks of the assets when measured with VaR, as shown in Artzner et al. (1999) and Rootzén & Klüppelberg (1999). Furthermore VaR does not account for the tail risk, because VaR does not take into account the magnitude of the losses beyond VaR. Because of these weaknesses the Basel Committee on Banking Supervision decided in 2012 to chose ES as replacement for VaR. As pointed out by Yamai & Yoshiba (2002, 2005) the major challenge in the implementation of ES as the leading measure of market risk is not the estimation techniques needed but the unavailability of simple tools to its evaluation.

In the literature a wide range of models to estimate VaR and ES are available. All these models are either parametric, semi-parametric or non-parametric. The most widely implemented non-parametric model is known as Historical Simulation (HS), however Danielsson & de Vries (2000) have shown that this model produces inaccurate VaR estimates. Scaillet (2004) proposes a non-parametric kernel estimator for the estimation of VaR and ES. The parametric models measure risk by fitting probability curves to data and inferring the VaR and ES from the fitted curve. The major drawbacks from this approach are failure of the non-normality and i.i.d. assumption as shown by Abad et al. (2014). To account for the drawbacks Mittnik & Paolella (2000) consider GARCH specifications for the volatility, whereas Fleming & Kirby (2003) used stochastic volatility and Giot & Laurent (2004) used realized volatility.

Abad et al. (2014) argue that models with time-varying high-order conditional moments are better in forecasting VaR then models with constant high-order moments. The FHS model that combines the benefits of HS with the flexibility of conditional volatility models was proposed by Barone-Adesi et al. (1999). Engle & Manganelli (2004) proposed a model for the estimation of VaR based on quantile estimation and Taylor (2008) used exponentially weighting quantile regressions to estimate VaR and ES. To properly account for the tail risk Chan & Gray (2006) applied EVT in estimating VaR whereas Gilli & Këllezi (2006) used threshold models in combination with EVT to estimate ES. McNeil & Frey (2000) combined time-varying volatility and EVT in the GARCH-EVT model.

All these models do not use the dependence structure between the assets. To account for this dependence a whole new class of models involving copulas was introduced. Cherubini & Luciano (2001) are one of the first that applied Archimedean copulas to estimate VaR, by using the historical empirical distribution in the estimation of the marginal distributions. Fortin & Kuzmics (2002) used copula theory in the estimation of VaR by using linear convex combinations of copulas. Palaro & Hotta (2006) estimated VaR by means of conditional copulas, whereas Fantazzini (2008) and Caillault & Guegan (2009)) use dynamic time-varying copulas to derive VaR and ES. Hotta et al. (2008) use copulas and extreme value theory to estimate VaR. This research follows the work of Wang et al. (2010) who combined the GARCH, extreme value and copula model in the GARCH-EVT-Copula model for estimation of VaR.

The remainder of the paper is organized as follows. Section 2 describes the data and section 3 deals with the theoretical framework of VaR and ES. Section 4 describes the methodology and section 5 gives the results. Section 6 concludes and gives directions for further research.

2 Empirical Data

2.1 Data Description

The empirical analysis uses a portfolio consisting of a special kind of assets, namely Exchange Traded Funds (ETFs). An ETF can be described as an financial security that tracks an index, a commodity, bonds or a basket of assets like an index fund. These ETFs can be traded on the exchange like a normal stock. Four different ETFs are used to build the portfolio. The first ETF is the Vanguard Total Stock Market ETF (VTI), which is known for delivering promising returns. The second is based on index investing and tracks the S&P 500, this ETF is called the SPDR S&P 500 ETF (SPY). The third is based on commodity investing and especially gold, namely the SPDR Gold Shares ETF (GLD). The last one is one of the most traded ETFs, the PowerShares QQQ ETF (QQQ), which is based on a modified market cap weighted index of 100 NASDAQ listed stocks. This ETF has a large tech exposure. Further on the abbreviations VTI, SPY, GLD and QQQ will refer to this ETFs. The data sample ranges from November 18, 2004 to May 9, 2017 for a total of 3139 daily observations. The sample starts at November 18, 2004 because on this date the GLD ETF was issued. Prices used in the analysis are closing prices on the trading days, adjusted for dividends and splits. It is assumed that trading costs are negligible. The portfolio is constructed with equal weighting. As indicated in Brooks (2014) the analysis of log returns is preferred above the analysis of raw returns, because of the time additive property of log returns. Therefore in this research returns refers to log returns. The analysis in this paper uses the losses instead of the returns. Denote R_t as the raw returns, then the raw returns can be transformed to percent log losses L_t by using the following transformation

$$L_t = -100 \log \left(\frac{R_t}{R_{t-1}} \right). \tag{1}$$

Due to the log transformation, one observation is lost, leaving a total of 3138 observations.

2.2 Descriptive statistics

From the plotted losses in Figure 1 it can be seen that quite some losses have a high magnitude, especially in the time period ranging from 2008 till 2010, indicating a high volatility at that time. In all cases around 95% of the losses fall below the 2%, with outliers up to 10%. Table 1 shows the descriptive statistics of the four ETFs and the

Table 1: **Descriptive Statistics**. Descriptive statistics of the four ETFs and the equally weighted portfolio over the sample period November 19, 2004 till May 9, 2017.

	VTI	*SPY*	*GLD*	*QQQ*	*Portfolio*
Mean	-0.0242	-0.0224	-0.0306	-0.0401	-0.0293
Standard Deviation	1.2141	1.2079	1.2227	1.2790	0.9653
Skewness	0.2481	0.1074	0.3017	0.1134	0.1358
Kurtosis	14.0309	17.8549	8.7404	10.5801	12.7657
Minimum	-12.0710	-13.5577	-10.6974	-11.4799	-8.9049
Maximum	9.8197	10.3637	9.1905	9.3824	8.1088
Jarque-Bera Test Statistic	15942	28860	4356	7519	12479

equally weighted portfolio. It can be seen that the means of the four ETFs and the portfolio are of the same magnitude, whereas the standard deviations of the ETFs are also of the same magnitude but the standard deviation of the portfolio is considerably smaller, indicating that there are diversification benefits. The skewness is in all cases negative indicating that there are mainly small profits and some extreme losses, as can be seen from the figures. All ETFs have a high kurtosis, meaning that a substantial part of the variation is caused by extreme values.

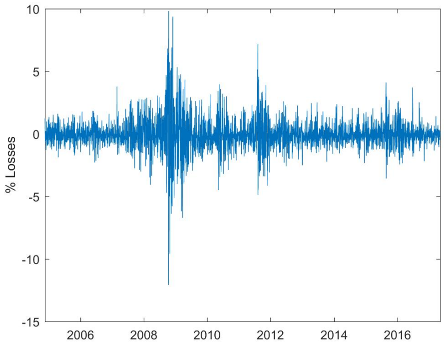
Table 2: **Cross correlations**. Cross correlations between the four ETFs over the sample period November 19, 2004 till May 9, 2017.

	VTI	*SPY*	*GLD*	*QQQ*
VTI	·	0.9881	0.0499	0.9187
SPY	0.9881	·	0.0375	0.9124
GLD	0.0499	0.0375	·	0.0069
QQQ	0.9187	0.9124	0.0069	·

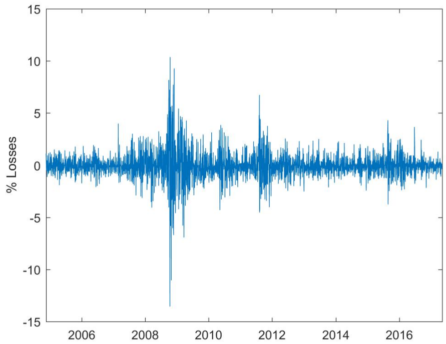
Table 2 shows the correlations between the ETFs. It is clearly seen that the VTI, SPY and QQQ ETF are highly correlated with each other, whereas the correlations of the GLD ETF with the other ETFs are quite small. Since there are high correlations between the ETFs, one can possibly benefit from this by including the dependence structure in the analysis.

2.3 Normality and i.i.d. Assumption

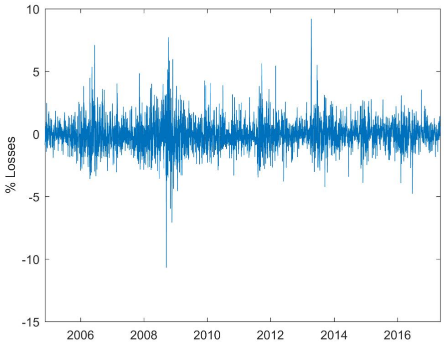
In case that the distribution of the losses is normal or Student-t distributed, a simple parametric approach can be employed to estimate the VaR and ES. To test visually for normality one can make a so called Q-Q plot, which is a graphical method for comparing distribution functions by plotting their quantiles against each other. If the data follows a normal or Student-t distribution the points in the Q-Q plot should lie on the 45-degree line, if the trend of the Q-Q plot is less steep than this line, one can conclude that the distribution on the horizontal axis is more dispersed than the one plotted on the vertical



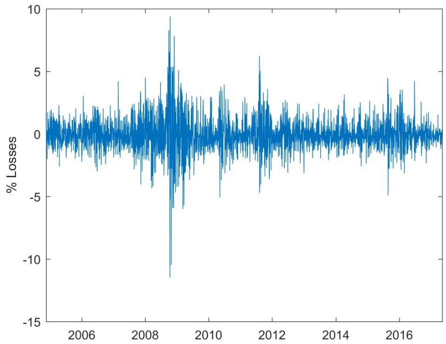
(a) Log Losses of VTI



(b) Log Losses of SPY



(c) Log Losses of GLD



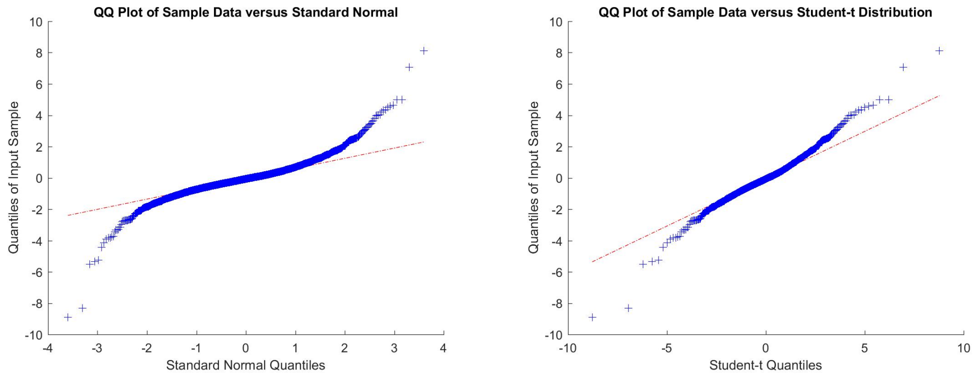
(d) Log Losses of QQQ

Figure 1: **Log Losses of the ETFS**. The plots show the log losses of the four ETFs over the sample period November 19, 2004 till May 9, 2017.

axis. It can be seen from Figure 3a that the losses do not follow a normal distribution because the tails are poorly covered, it actually indicates that a more heavy-tailed distribution should be more suitable in this context.

A distribution that accounts for possible leptokurticy, is the student-t distribution. As can be seen from Figure 3b this distribution approximates the distribution of the losses much better in the tails but still does not cover all aspects of the distribution of the losses. To formally test for normality of the distribution one can use the Jarque-Bera test-statistic, which tests for normality of the log losses. As can be seen from table 1 the Jarque-Bera test statistic is very large. This gives an indication that the distribution of the log losses is not normal, as was expected from the other descriptive statistics and the Q-Q plot.

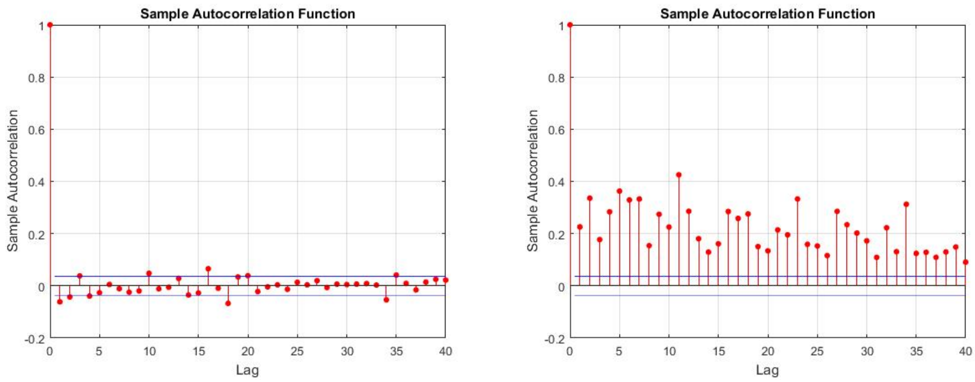
In order to apply EVT appropriately, the losses should be identically and independently distributed (i.i.d.). To graphically inspect this assumption, one can look at the auto- correlation function of the log losses and the squared log losses. From Figure 3 it can be seen that quite some auto-correlations of the log losses are significant and many of the autocorrelations of the squared log losses are significant. This strongly indicates that the losses are not i.i.d., but follow a conditional mean model with conditional het-



(a) Q-Q plot Normal Distribution (b) Q-Q plot Student-t Distribution

Figure 2: **Q-Q plots**. The plots show the Q-Q plots of the data distribution versus a normal and student-t distribution respectively

eroskedasticity. Ljung & Box (1978) and Engle (1982) tests confirm these findings.



(a) Auto-correlations of losses (b) Auto-correlations of squared losses

Figure 3: **Auto-correlation**. Auto-correlation plots of the portfolio losses and squared portfolio losses up to 40 lags.

3 Theoretical Framework

For financial risks, no definition of risk is entirely satisfactory. Depending on the context of the research one may arrive at different notions, in this research risk is defined as the quantifiable likelihood of loss. There are many different ways in which one can measure risk in one single number. In finance Value at Risk (VaR) and Expected Shortfall (ES) are the most commonly used risk measures. Section 3.1 and 3.2 provides the theoretical definitions of Value at Risk (VaR) and Expected Shortfall (ES). Section 3.3 describes the theoretical properties of VaR and ES. Section 3.4 lists the main deficiencies of VaR and in section 3.5 ES is proposed as alternative to VaR.

3.1 Definition of Value at Risk

VaR is defined as the maximum possible loss over a given holding period with a certain confidence level. In other words, VaR at the 100(1-α) percent confidence level can be defined as the 100α percentile of the loss distribution, where α is the confidence level. Given some confidence interval $\alpha \in$ (0,1), the VaR of the portfolio at the confidence level α is given by the smallest number l such that the probability that the loss L exceeds l is no larger than $1 - \alpha$. Denote Ω_{t-1} as the information set available at time $t-1$ and $F(L_t, \Omega_{t-1})$ as the conditional cumulative distribution of L_t given the information set at time $t-1$. Following the notation of Du & Escanciano (2017) and assuming a continuous distribution function F the VaR can be defined as

$$VaR_t^\alpha = \inf(L_t \in \mathbb{R} : F(L_t, \Omega_{t-1}) \geq \alpha). \tag{2}$$

Figure 4 illustrates the concept of VaR.

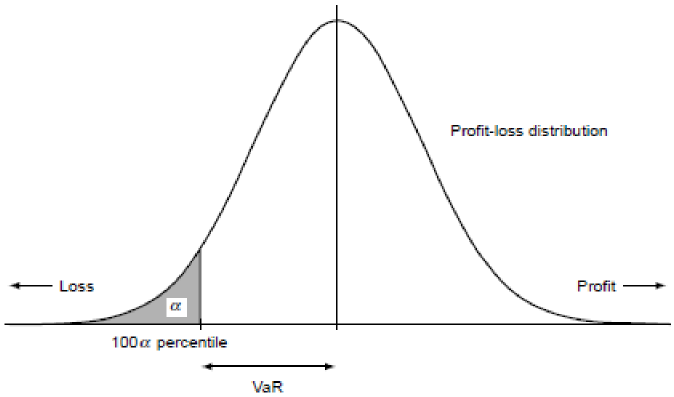


Figure 4: **Graphical illustration of VaR**. The concept of VaR illustrated for an arbitrary distribution of portfolio returns

3.2 Definition of Expected Shortfall

Artzner et al. (1997) proposed the use of ES, which is an extension of VaR. ES is the conditional expectation of a loss given that the loss is beyond the VaR level. Given some confidence interval $\alpha \in$ (0,1) the ES at the confidence level α can be calculated as follows

$$ES_t^\alpha = E[L_t | \Omega_{t-1}, L_t \geq VaR_t^\alpha]. \tag{3}$$

Assuming a continuous distribution F for the losses L_t, this can be rewritten as

$$ES_t^\alpha = \frac{1}{1-\alpha} \int_\alpha^1 VaR_t^u du \tag{4}$$

Figure 5 illustrates the concept of ES.

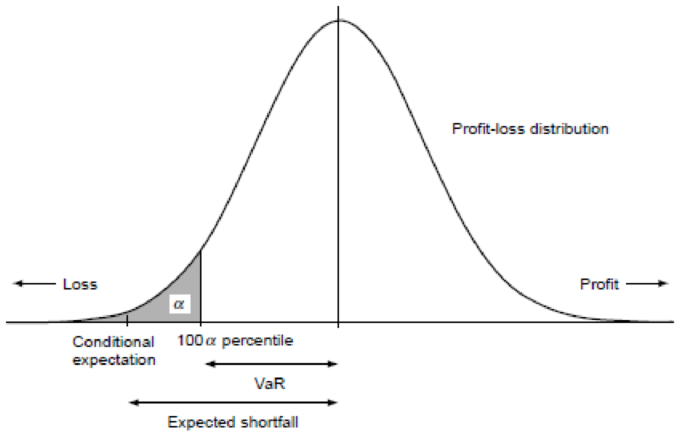


Figure 5: **Graphical illustration of ES**. The concept of ES illustrated for an arbitrary distribution of portfolio returns

3.3 Properties of Risk Measures

In the literature, several desirable properties for risk measures are examined. These properties have been proposed by Artzner et al. (1999) and Föllmer & Schied (2002). One of the most important issues involved with risk measures is that it should be interpretable, in the sense that the risk measure can be interpreted as the capital buffer needed to maintain the same level of risk. The mathematical properties of risk measures are important in understanding the trade-off between VaR and ES. The most important properties are monotonicity, positive homogeneity, translation invariance and subadditivity. See for a detailed description of these properties Appendix A.

It can be shown that VaR satisfies the properties of monotonicity, positive homogeneity and translation invariance, but not the property of subadditivity. Hence, VaR is not a coherent risk measure.

It can be verified that ES satisfies the properties of monotonicity, positive homogeneity, translation invariance and subadditivity (Acerbi & Täsche, 2002). Hence, ES is a coherent risk measure.

3.4 Criticism of VaR

In this section, the two main objections to use VaR are examined, first the failure in capturing tail risk and second the lack of subadditivity.

3.4.1 Failure in Capturing Tail Risk

VaR actually indicates the threshold loss over a given time period that will not be exceeded with a given level of confidence. For example, if the 99.9% 1-day VaR of a certain company equals 1 million then it follows that one out of 1000 days the loss will exceed this 1 million. This indicates that VaR only measures that with a certain probability the loss will exceed some threshold, but does not account for the magnitude of the loss beyond the threshold. This can be problematic, for example, if one day a firm faces a loss of 4000%, the firm will likely go bankrupt but it still satisfies the VaR requirements of one violation out of 1000 days. This deficiency of VaR can cause serious problems, see for example Basak & Shapiro (2001).

3.4.2 Lack of Subadditivity

It can be shown that VaR satisfies the properties of monotonicity, translation invariance and positive homogeneity, but that it is not subadditive and hence that it is not a coherent risk measure. This failure of subadditivity can cause that a combined portfolio can bear higher risks than the sum of the risks of the individual sub-portfolios, this contradicts the common belief that diversification lowers risk. An illustrative example of the lack of subadditivity of VaR can be found in Acerbi et al. (2001).

3.5 Expected Shortfall as Alternative Risk Measure

The lack of subadditivity and the failure to capture tail risk gives rice to the new risk measure ES. As can be seen from the definition of ES, it measures the expected loss in the tail of the profit-loss distribution. This exact definition shows that ES takes into account what happens in the tail of the distribution, overcoming the deficiency of VaR.

It has been pointed out in the literature that ES is a universal, complete and simple risk measure. VaR is still by far the most important risk measure used in financial risk management, but a change is expected. Based on the deficiencies of VaR, supervisors have proposed to use 97.5 % ES as replacement for 99 % VaR as the official risk measure. The reason behind this is that reality shows that tail risk matters and has to be incorporated into the risk measure. The debate involving this issues can be found in the Fundamental Review of the Trading Book by The Basel Committee (2013).

Due to the deficiencies of VaR, the academic and professional debate of VaR and the virtually no additional computational effort of estimating ES in comparison with VaR, we believe that using ES as a complementary risk measure to VaR is a smart move.

4 Methodology

This section describes the methodology of the paper. In section 4.1 standard univariate estimation models for VaR and ES are examined. Section 4.2 introduces the necessary theory for the multivariate copula framework. Thereafter, section 4.3 describes the multivariate TGARCH-EVT-Copula model. Section 4.4 and 4.5 describe the testing framework for VaR and ES respectively, whereas section 4.6 deals with model selection. Section 4.7 shows the estimation and forecasting framework.

4.1 Univariate Estimation models for VaR and ES

4.1.1 Variance-Covariance Model

If a parametric model with i.i.d. observations is assumed, exact formulas for VaR and ES can be derived for certain distributions of the innovations. If the losses L_t are assumed to belong to the Gaussian distribution, the formulas for VaR and ES are given as

$$\widehat{VaR}_t^{\alpha} = \hat{\mu} + \hat{\sigma}\Phi^{-1}(\alpha) \tag{5a}$$

$$\widehat{ES}_t^{\alpha} = \hat{\mu} + \hat{\sigma}\frac{\phi(\Phi^{-1}(\alpha))}{1-\alpha} \tag{5b}$$

where μ is the mean of the losses and σ the standard deviation of the losses. ϕ denotes the probability density function of the normal distribution and Φ^{-1} denotes the inverse cumulative distribution function of the normal distribution. This model is referred to as VC-N.

It is known that financial returns have fat tails, in the sense that extreme events occur more often than expected when assuming a normal distribution. To account for this possible leptokurticy, the losses L_t are assumed to belong to the Student-t distribution with $\nu > 1$ degrees of freedom. The loss distribution $L_t \sim t(\nu, \mu, \sigma)$ has moments $E(L_t) = \mu$ and $Var(L_t) = \frac{\nu\sigma^2}{\nu-2}$. The formulas for VaR and ES can then be written as

$$\widehat{VaR}_t^{\alpha} = \hat{\mu} + \hat{\sigma}t_{\hat{\nu}}^{-1}(\alpha) \tag{6a}$$

$$\widehat{ES}_t^{\alpha} = \hat{\mu} + \hat{\sigma}\frac{g_{\hat{\nu}}(t_{\hat{\nu}}^{-1}(\alpha))}{1-\alpha} \cdot \frac{\hat{\nu} + (t_{\hat{\nu}}^{-1}(\alpha))^2}{\hat{\nu}-1} \tag{6b}$$

where μ is equal to $E(L)$ which is the mean of the losses. g_ν denotes the probability density function of the Student's t-distribution with ν degrees of freedom and t_ν^{-1} denotes the inverse cumulative distribution function of the Student's t-distribution. This model is referred to as VC-S.

It is proven by McNeil et al. (2015) that for the normal distribution the shortfall-to-quantile ratio equals $ES_t^\alpha / VaR_t^\alpha = 1$ when $\alpha \to 1$. Whereas for a t-distribution with $\nu > 1$ degrees of freedom the shortfall-to-quantile ratio equals $ES_t^\alpha / VaR_t^\alpha = \frac{\nu}{\nu-1}$ when $\alpha \to 1$. This shows that for a heavy tailed distribution the difference between VaR and ES is more pronounced than for the normal distribution, indicating that the choice of distribution really matters.

4.1.2 Historical Simulation

The historical simulation model does not make any parametric assumption of the distribution of losses and is therefore widely used in practice. The foundation underling the historical simulation model is the empirical distribution function. Assume for a loss L_t, i.i.d. observations, denoted as $L_1, ..., L_n$. Then the empirical distribution function, that takes the values $L_1, ..., L_n$ with probability $\frac{1}{n}$ can be written as

$$F_n(l) = \frac{1}{n}\sum_{i=1}^{n} I(L_i \leq l), \tag{7}$$

where $I(\cdot)$ denotes the indicator function. Using asymptotic properties and the strong law of large numbers one can derive that $F_n(l)$ is a good estimator of $F_L(l)$. Because VaR is a quantile, the VaR can be estimated as $F_n^{\leftarrow}(\alpha) = \inf\{x : F(x) \geq \alpha\}$. Rank the losses $L_1, ..., L_n$ as $L_{(1)} \leq L_{(2)} \leq ... \leq L_{(n)}$, such that $L_{(i)}$ are order statistics. Following the reasoning of McNeil et al. (2015) $F_n^{\leftarrow}(\alpha)$ can be estimated as $L_{\lceil n\alpha \rceil}$. Combining all this the VaR and ES can be calculated as

$$\widehat{VaR}_t^{\alpha} = L_{\lceil n\alpha \rceil} \tag{8a}$$

$$\widehat{ES}_t^{\alpha} = \frac{\lceil n\alpha \rceil - n\alpha}{n(1-\alpha)} L_{(\lceil n\alpha \rceil)} + \frac{1}{n(1-\alpha)} \sum_{i=\lceil n\alpha \rceil + 1}^{n} L_{(i)} \tag{8b}$$

where n is the number of observations. The $\lceil \ \rceil$ symbol denotes the ceiling function. The exact derivations and underlying assumptions can be found in McNeil et al. (2015). This model is referred to as HS.

4.1.3 Filtered Historical Simulation using GARCH Models

In order to account for empirical regularities in the portfolio losses, the losses are modelled using the AR-(T)GARCH model with continuous innovations

$$L_t = \mu + \phi L_{t-1} + \sigma_t z_t \qquad z_t \sim t(\nu) \tag{9}$$

where the z_t are student-t distributed with ν degrees of freedom, mean zero and unit variance. Denote $\epsilon_t = \sigma_t z_t$, which is the unpredictable component of the losses. To account for specific dynamics in the variance the following two specifications are considered:

1. GARCH(p,q) specification of Bollerslev (1986).

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 \tag{10}$$

2. TGARCH(p,q) specification of Glosten et al. (1993) and Rabemananjara & Zakoian (1993).

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} (\alpha_i + \gamma_i I(\epsilon_{t-i} < 0))\epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 \tag{11}$$

Note that the frequently used Risk Metrics approach of Morgan/Reuters (1996) is a special case of equation (10), with the restrictions $\omega = 0$, $\alpha_1 + \beta_1 = 1$ and $p = q = 1$, furthermore the parameter β_1 is not estimated but set equal to 0.94. Since GARCH models are incapable of separating out asymmetric information, the TGARCH model is included which allows the effects of good and bad news to be different on the variance, through the additional indicator term. For the sake of completeness, the optimal lag lengths will be chosen along the Akaike Information Criterion with the restriction that the maximum lag length may not exceed 5, in other words $p, q \leq 5$, although in practical applications almost always a (T)GARCH(1,1) specification suffices. The following steps are needed to compute the values for VaR and ES in this setting.

First the GARCH parameters have to be estimated by optimizing the log likelihood function, this log likelihood can be optimized using Quasi Maximum Likelihood Estimation (QMLE). With the estimated parameters, the past variances $\hat{\sigma}_t^2$ for $t \leq n$ can be estimated. After the estimation of the variances, the innovations of the GARCH process can be estimated as

$$\hat{\epsilon}_t = \frac{L_t - \hat{\mu} - \hat{\phi} L_{t-1}}{\hat{\sigma}_t} \tag{12}$$

Then the VaR of the innovations can be calculated as the HS estimate of the empirical distribution of $\hat{\epsilon}_t$. Combining all this and following Gao & Song (2008) the VaR of the FHS model can be calculated as

$$\widehat{VaR}_t^{\alpha} = \hat{\mu} + \hat{\phi} L_{t-1} + \hat{\sigma}_t \hat{\epsilon}_{\alpha} \tag{13}$$

where $\hat{\epsilon}_{\alpha}$ is the HS estimator of VaR of the innovations. The ES of the FHS model can be calculated as

$$\widehat{ES}_t^{\alpha} = \hat{\mu} + \hat{\phi} L_{t-1} + \hat{\sigma}_t \hat{\delta}_{\alpha} \tag{14}$$

where δ_{α} can be estimated as

$$\hat{\delta}_{\alpha} = \frac{\sum_{t=1}^{n} \hat{\epsilon}_t I(\hat{\epsilon}_t > \hat{\epsilon}_{\alpha})}{\sum_{t=1}^{n} I(\hat{\epsilon}_t > \hat{\epsilon}_{\alpha})}. \tag{15}$$

Barone-Adesi et al. (1999, 2002) and Pritsker (2001) have shown that this type of historical simulation performs rather well. These models are referred to as AR(1)-GARCH(p,q) and AR(1)-TGARCH(p,q) respectively.

4.1.4 Extreme Value Theory

Extreme Value Theory can be applied to develop models to describe the extremal behavior of financial risk factors. In fact, this kind of models is useful in modeling the tail of the distribution of financial risk factors. Reality often shows that risk factors are heavy tailed in comparison with normal distributions. The observations in the tail of the distribution are of extreme importance because these are the ones of interest.

There are mainly two methods available for modeling extreme values. The first method looks at the events in the data that exceeds a high threshold and fits a distribution on these exceedances. It is shown by Balkema & de Haan (1974) that the limiting distribution over the threshold is a Generalized Pareto Distribution (GPD). The second method divides the data into consecutive blocks and focuses on the series of maxima/minima in these blocks. The asymptotic distribution of these series of maxima/minima converges to the Generalized Extreme Value Distribution as proven by Embrechts et al. (1997). McNeil et al. (2015) argue that the threshold exceedance methods are preferable over the block maxima/minima models because these models make more efficient use of the often limited data on extreme outcomes. Based on this evidence in favor of threshold models, this research proceeds with the so-called Peak-over-Threshold (POT) method in combination with the GPD.

The POT method considers the distribution of exceedances over a certain threshold. Because of the importance of extreme values it is important to estimate the distribution function of values of x above a certain threshold u. Define $L_1, ..., L_n$ as a sequence of i.i.d. random variables from some unknown distribution function F and define u as some threshold. Define a sequence $Y_1, ..., Y_n$ with $Y_i = L_i - u$, effectively meaning that the Y measures the excess above the threshold. The excess distribution above the threshold u can than be written as a conditional probability

$$F_u(y) = P(L - u | L > u) \tag{16}$$

$$F_u(y) = \begin{cases} \frac{F(u+y)-F(u)}{1-F(u)}, & \text{if } y \geq 0, \\ 0, & \text{if } y < 0. \end{cases} \tag{17}$$

The largest proportion of the L_i variables lie below u and therefore there should be no problem while estimating F in this interval. However, the estimation of F_u can be rather difficult because in general only a few observations belong to this area. For this purpose EVT is extremely important, because it gives expressions for the conditional excess distribution. For sufficiently large threshold u, Balkema & de Haan (1974) and Pickands (1975) states that it is possible for a large class of distributions F to find a function H and κ such that

$$\lim_{u \to x_0} \sup_{0 \leq y \leq x_0 - u} |F_u(y) - H_{\xi,\kappa}(y)| = 0 \tag{18}$$

where x_0 is the right endpoint of F and $H_{\xi,\kappa}$ denotes the GPD which can be written as

$$H_{\xi,\kappa}(y) = \begin{cases} 1 - (1 + \frac{\xi y}{\kappa})^{\frac{-1}{\xi}}, & \text{if } \xi \neq 0, \\ 1 - exp(-\frac{y}{\kappa}), & \text{if } \xi = 0, \end{cases} \tag{19}$$

where it holds that $y \geq 0$ for $\xi \geq 0$. In other words the function $F_u(y)$ is well approximated by the function $H_{\xi,\kappa}$. The shape of the distribution is determined by the parameter ξ and the scale is determined by κ. The shape parameter ξ is also known as the tail index, which indicates the heaviness of the tail. A larger ξ, indicates a heavier tail, only distributions with $\xi \geq 0$ are suitable for modeling financial returns (Gilli & Këllezi, 2006).

From the theory it is known that the threshold u should be high, because the theorem of Balkema & de Haan (1974) and Pickands (1975) holds for sufficiently large u, but on the other hand if u gets too high, to few observations are left in the tail. So far there is no optimal choice for the selection of this threshold u. Two graphical methods appear in the literature, namely the sample mean excess plot and the Hill (1975) plot. However applying these approaches at every time point is time consuming and not feasible. **?** show using Monte Carlo experiments that the parameter estimates are robust to the threshold value if this threshold is between 5% and 13%. Carol (2008) states that using a threshold value of 10% leaves enough observations in the tail to accurately estimate the parameters of the GPD using Maximum Likelihood Estimation (MLE). Following their lead a threshold of 10% is used, therefore u is then equal to the 90% order statistic.

Several methods can be used to estimate the parameters of the GPD, namely MLE, Method of Moments, Method of Probability-Weighted Moments and the Elemental Percentile Method. For an extensive overview and discussion of these methods see Hosking & Wallis (1987), Grimshaw (1993) and Castillo & Hadi (1997). This research uses the MLE approach in the same way as is done in Gilli & Këllezi (2006). The log-likelihood function can be written as

$$L(\xi, v|y) = \begin{cases} -n\log(\kappa) - (\frac{1}{\xi}+1)\sum_{i=1}^{n}\log(1+\frac{\xi}{\kappa}y_i) & \text{if } \xi \neq 0, \\ -n\log(\kappa) - \frac{1}{\kappa}\sum_{i=1}^{n} y_i & \text{if } \xi = 0. \end{cases} \tag{20}$$

Maximizing the log-likelihood with respect to the parameters results in estimated parameter values. Using this estimated parameters we find the following analytical expressions for VaR and ES

$$\widehat{VaR}_t^{\alpha} = u + \frac{\hat{\kappa}}{\hat{\xi}}\left((\frac{n}{N_u}(1-\alpha))^{-\hat{\xi}} - 1 \right) \tag{21a}$$

$$\widehat{ES}_t^{\alpha} = \frac{\widehat{VaR}_t^{\alpha}}{1-\hat{\xi}} + \frac{\hat{\kappa}-\hat{\xi}u}{1-\hat{\xi}} \tag{21b}$$

where n is the total number of observations and N_u the number of observations above the threshold u. See for a detailed derivation of these formulas Gilli & Këllezi (2006). This model is referred to as EVT.

4.1.5 TGARCH-EVT Model

The use of normal distributions for the losses results in underestimated tails and disregards excess kurtosis and skewness. The benchmark models assume constant volatility over time, which is obviously not true given the empirical evidence of volatility clustering. VaR and ES estimates deal only with the tail of the distribution and the afore mentioned EVT approach focuses on these tails. However, applying EVT on the tails requires that the observations are i.i.d., which is obviously not the case given the volatility clustering. To overcome these shortcomings McNeil & Frey (2000) propose a model that first estimates a GARCH model and then uses the framework of EVT on the filtered residuals. These filtered residuals are approximately i.i.d., when the GARCH specification is valid. In this research this approach is implemented using an AR(1)-TGARCH(1,1) model to capture the serial dependence of the losses and using a GPD distribution for the tails of the filtered residuals of the AR(1)-TGARCH(1,1) model. The estimation algorithm can be found in appendix B.

4.2 Copulas

In practical portfolio applications the dependence structure between the different assets is assumed to be linear. The dependence is mostly measured by means of correlation, but Embrechts et al. (1999) proves that correlation on its own is not sufficient to describe the dependence structure when distributions are not normal or elliptical. A better approach to deal with the dependence structure is the use of copulas. A copula can be described as the function that forms multivariate distribution functions from one-dimensional distribution functions. Copulas express dependence on a so-called quantile scale, which is extremely usefull for describing the dependence of extreme outcomes.

Another advantage of the use of copulas is that it is flexible in the sense that it allows to focus on two different levels. The first focus is on fitting marginal distributions to each asset and the second focus is on fitting the dependence distribution of the marginal distributions. This copula model provides a much more flexible framework in describing the dependence structure.

In this research, five possible copulas are considered, namely the Clayton, Frank, Gumbel, Gaussian and Student copula. It is the goal of the research to choose the copula that delivers the best VaR and ES forecasts, rather than choosing the copula that best fits the data. Therefore no goodness of fit tests as in Genest et al. (2009) and Hofert et al. (2012b) are performed to select the best copula. The fit of the copula is assessed using the backtests for VaR and ES. In this research time-varying marginals are used in combination with conditional copulas.

This section begins with describing the basic concepts of copulas and proceeds with five special cases of copulas. The section concludes with the estimation algorithms for the copulas and a general note on sampling from the copulas.

4.2.1 Basic Concepts

A d-dimensional copula is the distribution function C on $[0,1]^d$ with standard uniform marginal distributions. Denote $C(u) = C(u_1, ..., u_d)$ as the multivariate distributions functions that are copulas. The dependence between real-valued random variables $X_1, ..., X_d$

can be completely described by their joint distribution function as follows

$$F(x_1, ..., x_d|\Omega_{t-1}) = P[X_1 \leq x_1, ..., X_d \leq x_d|\Omega_{t-1}]. \tag{22}$$

Splitting F into a part that describes the dependence structure and a part that describes the marginal behaviour, has led to the use of copulas. The joint distribution function C of $(F_1(X_1|\Omega_{t-1}), ..., F_d(X_d|\Omega_{t-1}))'$ is called the copula of the random vector $(X_1, ..., X_d)'$ or the multivariate distribution F. The multivariate distribution F can then be written as

$$\begin{aligned} F(x_1, ..., x_d|\Omega_{t-1}) &= P[X_1 \leq x_1, ..., X_d \leq x_d|\Omega_{t-1}] \\ &= P[F_1^{-1}(U_1|\Omega_{t-1}) \leq x_1, ..., F_d^{-1}(U_d|\Omega_{t-1}) \leq x_d|\Omega_{t-1}] \\ &= P[U_1 \leq F_1(x_1|\Omega_{t-1}), ..., U_d \leq F_d(x_d|\Omega_{t-1})|\Omega_{t-1}] \\ &= C(F_1(x_1|\Omega_{t-1}), ..., F_d(x_d|\Omega_{t-1})) \end{aligned} \tag{23}$$

Equation (23) can be derived from the Sklar (1959) theorem. It can be seen as the representation of the joint distribution function F in terms of the copula C and the marginal distributions $F_1, ..., F_d$. The density function $c(u_1, ..., u_d|\Omega_{t-1})$ of the copula $C(u_1, ..., u_d|\Omega_{t-1})$ can be denoted as

$$c(u_1, ..., u_d|\Omega_{t-1}) = \frac{\partial C(u_1, ..., u_d|\Omega_{t-1})}{\partial u_1 \ldots u_d} \tag{24}$$

Following Cherubini et al. (2004) the relation between the copula and the density function of F, denoted by f is given by the canonical copula representation. The canonical copula representation can be written as

$$f(x_1, ..., x_d|\Omega_{t-1}) = c(F_1(x_1|\Omega_{t-1}), ..., F_d(x_d|\Omega_{t-1})) \prod_{j=1}^{d} f_j(x_j|\Omega_{t-1}) \tag{25}$$

where f_j are the marginal densities, which can be derived as

$$f_j = \frac{dF_j(x_j|\Omega_{t-1})}{dx_j}. \tag{26}$$

In the next two sections some specific types of copulas are listed and explained.

4.2.2 Archimedean Copulas

A d-variate Archimedean copula can be defined as the following copula function

$$C(u_1, ..., u_d) = \varphi^{-1}(\varphi(u_1) + ... + \varphi(u_d)) \tag{27}$$

where $\varphi(u)$ is known as the generator function of the copula. The generator should be twice differentiable, strictly decreasing and convex. The inverse generator function denoted by φ^{-1} should be monotonic on the interval zero to infinity. In this research three different one-parameter Archimedean copulas are considered, namely Clayton, Frank and Gumbel.

1. **The Clayton Copula**
   The generator function of the Clayton copula is $\varphi(t) = (1+\theta t)^{\frac{-1}{\theta}}$, which results in the following Clayton d-dimensional copula

$$C_\theta^C = (u_1^{-\theta} + ... + u_d^{-\theta} + 1 - d)^{\frac{-1}{\theta}}. \quad (28)$$

2. **The Frank Copula**
   The generator function of the Frank copula is given by

$$\varphi(t) = \log\left(\frac{e^{-\theta t} - 1}{e^{-\theta} - 1}\right). \quad (29)$$

   Using the generator the Frank d-dimensional copula can be written as

$$C_\theta^F = -\frac{1}{\theta}\log\left(1 + \frac{\prod_{i=1}^d (e^{\theta u_i} - 1)}{(e^{-\theta} - 1)^{d-1}}\right). \quad (30)$$

   It can be proven that this only holds for $d \geq 3$.

3. **The Gumbel Copula**
   The generator function of the Gumbel copula can be written as $\varphi(t) = e^{-t^{1/\theta}}$. Using this generator the d-dimensional Gumbel copula can be written as

$$C_\theta^G = exp[-(-\log(u_1)^\theta + ... + -\log(u_d)^\theta)^{1/\theta}]. \quad (31)$$

4.2.3 Non-Archimedean Copulas

1. **The Gaussian Copula**
   The Gaussian copula, denoted as C_R^d, of a d-dimensional multivariate normal distribution with correlation matrix R_G can be written as

$$C_R^{GA} = \Phi_R^d(\Phi^{-1}(u_1), ..., \Phi^{-1}(u_d)) \quad (32)$$

   where Φ_R^d denotes the joint distribution of the d-variate standard normal distribution with correlation matrix R and Φ^{-1} denotes its inverse.

2. **The Student-t Copula**
   The Student-t copula of a d-variate t-distribution with ν degrees of freedom and

correlation matrix R_S can be written as

$$C^t_{\nu,R} = t^d_{\nu,R}(t^{-1}_\nu(u_1), ..., t^{-1}_\nu(u_d)) \tag{33}$$

where $t^d_{\nu,R}$ denotes the joint distribution of the d-variate student-t distributions and t^{-1} its inverse.

4.2.4 Estimation Framework

Estimation of the copulas can be done using Maximum Likelihood estimation (MLE), Inference for Margins (IFM) or Canonical Maximum Likelihood (CML). The MLE method estimates jointly the parameters of the marginal distributions and the copula, whereas IFM estimates first the parameters of the univariate marginal distributions and thereafter the parameters of the copula. The CML method uses the idea that the copula parameters can be estimated without specifying the marginal distributions, the marginal distributions in this method are calculated using the empirical distribution function. Exact details on these methods can be found in Bouye et al. (2000) and Romano (2002). In the framework of this thesis, the marginals are estimated using kernel estimation and extreme value theory, this leads to the choice of MLE for only the copulas in this framework. With the estimated $\hat{\theta}_1$, containing the parameters of the marginals, the parameters of the copula denoted by θ_2 can be estimated by solving

$$\hat{\theta}_2 = \arg\max_{\theta_2} \sum_{t=1}^{n} \log c(F_1(L_{1,t}|\Omega_{t-1}), ..., F_d(L_{d,t}|\Omega_{t-1}); \theta_2, \hat{\theta}_1). \tag{34}$$

For the estimation of high dimensional Archimedean copulas see for example Hofert et al. (2012a).

Once the parameters are estimated, sampling from the copula can be easily done using the conditional sampling method. In the case of the Archimedean copulas, this can be computationally intensive in high dimensional cases, to remedy this issue one can use the sampling method of Cherubini et al. (2004). However, as the dimensionality in this research is limited, the conditional sampling method suffices.

4.3 TGARCH-EVT-Copula Model

The use of normal distributions for the losses results in underestimated tails and disregards excess kurtosis and skewness. VaR and ES estimates deal only with the tail of the distribution and the afore mentioned EVT approach focuses on these tails. However, applying EVT on the tails requires that the observations are i.i.d., which is obviously not the case given the volatility clustering. To overcome these shortcomings McNeil & Frey (2000) propose a model that first estimates a GARCH model and then use the framework of EVT on the filtered residuals. It is shown in the literature that to describe the dependence structure of financial assets only using correlation as a dependence measure is not enough. To remedy this issue, the use of copulas was proposed. In this extension of existing models, the AR(1)-TGARCH(1,1) model and EVT approach are combined with copulas, in order to capture the dependence structure between the assets appropriately. This approach consists of the following steps.

1. Estimate an AR(1)-TGARCH(1,1) model for each ETF separately

$$L_{t,j} = \mu_j + \phi L_{t-1,j} + \sigma_{t,j}\epsilon_{t,j} \qquad \epsilon_{t,j} \sim t(\nu) \tag{35a}$$

$$\sigma^2_{t,j} = \omega + (\alpha_1 + \gamma_1 I(\epsilon_{t-1,j} < 0))\epsilon^2_{t-1,j} + \beta_1\sigma^2_{t-1,j} \tag{35b}$$

2. With the estimated parameters filter out the standardized residuals

$$\hat{\epsilon}_{t,j} = \frac{L_{t,j} - \hat{\mu}_j - \hat{\phi}L_{t-1,j}}{\hat{\sigma}_{t,j}} \tag{36}$$

3. Estimate the marginal distributions of the standardized residuals $\epsilon_{t,j}$ of the ETFs using a semi parametric approach. It is known that the majority of the observations lie in the center, therefore the estimation of the center should not be too hard. The estimation of the tails of the distribution can be problematic as a limited number of observations lie in the tail. Therefore the interior of the marginal distributions is estimated using a Gaussian kernel function and the tails are estimated using extreme value theory by means of the GPD. Denote $\epsilon_{1,j}, ..., \epsilon_{n,j}$ as an i.i.d. random sample for a continuous density function f. Then the kernel density estimator of f can be defined as

$$\hat{f}_j(x;h) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\sqrt{2\pi h}}e^{\frac{-(x-\epsilon_{i,j})^2}{2h^2}}. \tag{37}$$

4. Transform the obtained standardized residuals ϵ_j to uniform variables by plugging them in their respective marginal distributions F_j. The continuity of the piecewise marginals is ensured by using interpolation to combine the tails with the center, this results in approximate piecewise continuous marginals. The uniform variables are of the form $U_j = F_j(\epsilon_j)$ for $1 \leq j \leq 4$.

5. Fit a copula using the MLE estimation method on the transformed marginal distributions U_j for $1 \leq j \leq 4$.

6. Using the estimated parameters of the copula, one can simulate from the copula. In this multivariate setting N simulations are performed, resulting in a Nx4 matrix M of simulated marginal values. Denote column j as M_j for $1 \leq j \leq 4$ which contains the simulated marginal uniform values for the losses of ETF j.

7. The simulated values are on a uniform scale, to rescale them to the original scale

one can use the inverse marginal functions F_j^{-1}, this results in $W_j = F_j^{-1}(M_j)$ for $1 \leq j \leq 4$.

8. To reintroduce this autocorrelation and heteroskedasticity, transform the simulated values W_j to log losses using the AR(1)-TGARCH(1,1) specification as in equation (35a) and (35b). This results in the following simulated log losses, denoted as $R_{t,j}^i$

$$R_{t,j}^i = \hat{\mu}_j + \hat{\phi} R_{t-1,j}^i + \hat{\sigma}_{t,j} W_{t,j}^i \tag{38}$$

9. The simulated log losses of the ETFS are denoted by R_j^i for $1 \leq j \leq 4$. To transform them to portfolio losses of an equally weighted portfolio, use the transformation $\frac{1}{4}\sum_{j=1}^{4} R_{t,j}^i$ for $1 \leq i \leq N$. Denote these portfolio losses as L_{t+1}^i.

10. The VaR and ES estimates of the TGARCH-EVT-Copula approach can be calculated as

$$\widehat{VaR}_t^{\alpha} = L_{t+1}^{\lceil N\alpha \rceil} \tag{39a}$$

$$\widehat{ES}_t^{\alpha} = \frac{\lceil N\alpha \rceil - N\alpha}{N(1-\alpha)} L_{t+1}^{(\lceil N\alpha \rceil)} + \frac{1}{N(1-\alpha)} \sum_{i=\lceil N\alpha \rceil + 1}^{N} L_{t+1}^{(i)} \tag{39b}$$

where $L_{t+1}^{(i)}$ are order statistics and the $\lceil\ \rceil$ symbol denotes the ceiling function.

4.4 Model Evaluation VaR

To test if the prediction of future risks is accurate, one should backtest the forecasts. The proportion of exceptions should match the confidence level, backtesting tests of this type are known as tests of unconditional coverage. For a good VaR model not only the proportion of exceptions should match the confidence level but at the same time, the exceptions should be evenly spread over time, in the sense that they are independent of each other. Clustering of exceptions indicates that the model does not accurately capture changes in market conditions, like volatility and correlations. Backtesting tests of this type are conditional coverage tests. The unconditional coverage is tested using the Kupiec (1995) proportion of failures test and the conditional coverage is tested using the Christoffersen (1998) independence test. For notational convenience denote $1 - \alpha$ as $\tilde{\alpha}$.

4.4.1 Unconditional Coverage Test

Define the following indicator function which measures the exceptions

$$h_t^{\alpha} = I(L_{t+1} > VaR_t^{\alpha}). \tag{40}$$

In an ideal situation, the VaR model produces a number of exceptions in line with the used confidence level. If the number of exceptions is less than the confidence level, the risk is overestimated. On the other hand, if the number of exceptions is more than the confidence level, the risk is underestimated. Define T as the total number of observations, $T_1 = \sum_{t=1}^{T} h_t^{\alpha}$ as the number of exceptions and $T_0 = T - T_1$ as the number of non-exceptions. Assuming independent exceptions, Kupiec's test can be used to determine whether the number of exceptions is consistent with the confidence level. The null hypothesis can be written as

$$H_0 : E[h_t^{\alpha}] = \tilde{\alpha}$$

The likelihood ratio test of correct unconditional coverage can be computed as

$$LR_{uc} = -2\log\left(\frac{(\tilde{\alpha})^{T_0}(1-\tilde{\alpha})^{T_1}}{(\tilde{\pi})^{T_0}(1-\tilde{\pi})^{T_1}}\right) \sim \chi^2(1) \tag{41}$$

where $\tilde{\pi}$ is the maximum likelihood estimate of the proportion. This test statistic is asymptotically χ^2 distributed with one degree of freedom. If LR_{uc} is greater than the critical value, H_0 is rejected and the model is classified as inaccurate. Note that this test corresponds to testing that the sequence $\{h_t^{\alpha} - \tilde{\alpha}\}_{t=1}^{T}$ has zero mean in the Du & Escanciano (2017) framework.

4.4.2 Conditional Coverage Test

Good VaR estimates should have independence, in the sense that VaR exceptions should be spread out over the sample and not come in clusters. This boils down to testing the following null hypothesis

$$H_0 : E[h_t^{\alpha} | \Omega_{t-1}] = E[h_t^{\alpha}].$$

Independence is tested against the specific alternative of a first-order Markov chain for h_t^{α}, with transition probability matrix

$$\Pi_1 = \begin{bmatrix} 1-\pi_{01} & \pi_{01} \\ 1-\pi_{11} & \pi_{11} \end{bmatrix} \tag{42}$$

where $\pi_{ij} = P[h_t^{\alpha} = j | h_{t-1}^{\alpha} = i]$. The null hypothesis of independence implies that $\pi_{01} = \pi_{11} = \pi_2$. Define T_{ij} as the number of observations such that $h_t^{\alpha} = j$ and $h_{t-1}^{\alpha} = i$. The likelihood ratio test of independence can be computed as

$$LR_{ind} = -2\log\left(\frac{(1-\pi_2)^{T_{00}+T_{11}}(\pi_2)^{T_{01}+T_{11}}}{(1-\pi_{01})^{T_{00}}(\pi_{01})^{T_{01}}(1-\pi_{11})^{T_{10}}(\pi_{11})^{T_{11}}}\right) \sim \chi^2(1) \tag{43}$$

where π_{01}, π_{11} can be calculated using maximum likelihood. The maximum likelihood estimate of π_2 is equal to $(T_{01} + T_{11})/T$. The test statistic is asymptotically χ^2 distributed with one degree of freedom. If LR_{ind} is greater than the critical value, H_0 is rejected and the model is classified as inaccurate. Note that this test corresponds to testing that the sequence $\{h_t^\alpha\}_{t=1}^T$ is uncorrelated in the Du & Escanciano (2017) framework.

4.4.3 Joint Conditional and Unconditional Coverage Test

Good VaR estimates should not only are independent but also have good coverage. The null hypothesis for this joint test can be written as

$$H_0 : E[h_t^\alpha|\Omega_{t-1}] = E[h_t^\alpha] = \tilde{\alpha}$$

The likelihood ratio test statistic can be written as

$$LR_{cc} = LR_{uc} + LR_{ind} \sim \chi^2(2) \tag{44}$$

which is asymptotically χ^2 distributed with two degrees of freedom. Exact derivations and underlying assumptions of this test can be found in Christoffersen (1998). Note that this test corresponds to testing that the sequence $\{h_t^\alpha - \tilde{\alpha}\}_{t=1}^T$ is uncorrelated in the Du & Escanciano (2017) framework.

4.5 Model Evaluation ES

To test if the prediction of future risks is accurate, one should backtest the forecasts. Whereas the testing of VaR is straightforward, testing ES appears to be rather difficult. Gneiting (2011) showed that ES on its own lacks a mathematical property called elicitability which makes it difficult to backtest using scoring functions. Fissler & Ziegel (2016) showed that ES is jointly elicitable with VaR, but their approach was aimed on model selection and not on testing correct specification. Following this work, many researchers and professionals were convinced that it was not possible to properly backtest ES at all.

However recent research conducted by Du & Escanciano (2017) shows that backtesting ES does not necessarily depend on elicitability. They show that testing can be done using the property that cumulative violations form a class of Martingale Difference Sequences. This cumulative violation process accumulates all violations in the right tail, just like the ES accumulates the VaR in the right tail. Du & Escanciano (2017) propose unconditional and conditional tests based on this cumulative violation process, these tests for ES are comparable with the ones proposed in section 4.4 for VaR.

4.5.1 Unconditional Backtest

Unlike VaR, which only contains information on one quantile level, ES contains information from the whole right tail, by integrating all VaRs from α to 1. This results in the following equation for ES

$$ES_t^\alpha = \frac{1}{\tilde{\alpha}} \int_\alpha^1 VaR_t^u du. \tag{45}$$

Following Du & Escanciano (2017) the correct specification of ES can be tested by using the integral of the indicator function of violations h_t, in the literature known as the cumulative violation process, by replacing VaR_t^α with h_t^α in equation (45)

$$H_t^\alpha = \frac{1}{\tilde{\alpha}} \int_\alpha^1 h_t^u du. \tag{46}$$

Since this h_t^α has mean u, it can be derived using Fubini's theorem that H_t^α has mean $\tilde{\alpha}/2$. Following the notation of Du & Escanciano (2017) u_t can be defined as some function $F_t(L_t, \Omega_{t-1})$. Using the fact that $h_t^u = I(u_t \leq u)$ the following can be derived

$$H_t^\alpha = \frac{1}{\tilde{\alpha}} \int_\alpha^1 I(u_t \leq u) du \tag{47a}$$

$$= \frac{1}{\tilde{\alpha}} (\tilde{\alpha} - u_t) I(u_t \leq \tilde{\alpha}). \tag{47b}$$

Like violations, cumulative violations are distribution free, in the sense that the sequence $\{u_t\}_{t=1}^T$ consists of independent and identically distributed $U[0,1]$ variables, see for example Rosenblat (1952) and Berkowitz (2001) for further details. As pointed out by Du & Escanciano (2017) cumulative violations contain information about the tail risk in the sense that when a violation occurs, the cumulative violation measures how far the actual value of L_t lies from its quantile through the term $\tilde{\alpha} - u_t$. The variables $\{u_t\}_{t=1}^T$ are generally unknown since the distribution of F_t is unknown. In practice often a parametric conditional distribution $F_t(\cdot, \Omega_{t-1}, \theta)$ is specified, where θ is estimated before estimating VaR and ES. With this estimated parameters the percentiles u_t can be calculated

$$\hat{u}_t = F_t(L_t, \hat{\Omega}_{t-1}, \hat{\theta}). \tag{48}$$

This $\hat{u}_t$ can be obtained be simulating M times from the conditional predictive distribution, ranking this generated random variates in ascending order and calculating the number of observations larger than the actual observed loss. Denote this number as S then the term $\hat{u}_t$ can be approximated by $1 - S/M$. The estimated cumulative violations can then be calculated as

$$\hat{H}_t^\alpha = \frac{1}{\tilde{\alpha}} (\tilde{\alpha} - \hat{u}_t) I(\hat{u}_t \leq \tilde{\alpha}). \tag{49}$$

Testing for correct unconditional specification of ES can be done by testing whether the sequence $\{H_t^\alpha - \tilde{\alpha}/2\}_{t=1}^T$ has zero mean, this boils down to testing the following null hypothesis

$$H_0 : E(H_t^\alpha) = \tilde{\alpha}/2. \tag{50}$$

Straightforward calculations show that the variance of H_t^α is equal to $\tilde{\alpha}(1/3-\tilde{\alpha}/4)$, therefore a simple t-test statistic is as follows

$$U_{ES} = \frac{\sqrt{T}(\bar{H}^\alpha - \tilde{\alpha}/2)}{\sqrt{\tilde{\alpha}(1/3-\tilde{\alpha}/4)}} \sim N(0,1) \tag{51}$$

where $\bar{H}^\alpha = \frac{1}{T}\sum_{t=1}^{T}\hat{H}_t^\alpha$. Following the reasoning of Du & Escanciano (2017) the limiting distribution of U_{ES} converges to a standard normal distribution when $T/n \to 0$, $T \to \infty$ and $n \to \infty$.

4.5.2 Conditional Backtest

Define the lag-j autocovariance and autocorrelation of H_t^α for $j \geq 0$ as

$$\gamma_j = Cov(H_t^\alpha, H_{t-j}^\alpha) \quad and \quad \rho_j = \frac{\gamma_j}{\gamma_0}. \tag{52}$$

In this framework $\{H_t^\alpha\}_{t=1}^T$ is unobservable therefore the estimated $\hat{H}_t^\alpha$ is used instead of H_t^α, resulting in the following equations

$$\hat{\gamma}_{Tj} = \frac{1}{T-j}\sum_{t=1+j}^{T}(\hat{H}_t^\alpha - \tilde{\alpha}/2)(\hat{H}_{t-j}^\alpha - \tilde{\alpha}/2) \quad and \quad \hat{\rho}_{Tj} = \frac{\hat{\gamma}_{Tj}}{\hat{\gamma}_{T0}}. \tag{53}$$

Testing for correct conditional specification can be done by testing whether the sequence $\{H_t^\alpha - \tilde{\alpha}/2\}_{t=1}^T$ is uncorrelated. This boils down to testing the following null hypothesis

$$H_0 : E(H_t^\alpha - \tilde{\alpha}/2|\Omega_{t-1}) = 0. \tag{54}$$

Under H_0 all ρ_j for $j \geq 1$ should be zero. This can be tested using the Box-Pierce test statistic

$$C_{ES}^m = T\sum_{j=1}^{m}\hat{\rho}_{Tj}^2 \sim \chi^2(m) \tag{55}$$

Following the reasoning of Du & Escanciano (2017) the limiting distribution of C_{ES}^m converges to a chi-squared distribution with m degrees of freedom when $T/n \to 0$, $T \to \infty$ and $n \to \infty$.

4.6 Model Selection

The choice of an appropriate risk measure for risk management that is easily backtestable appears to be difficult. In light of this debate, Weber (2006) and Gneiting (2011) have shown that ES on its own is not elicitable. Especially this means that there is no strictly consistent scoring function S such that for any random variable L_t with finite mean the following holds

$$ES_t^\alpha(L) = \arg\min E[S_{E,t}(e_t, L_t)]. \tag{56}$$

Note that V and E denote VaR and ES respectively whereas e_t and v_t denote a realization of VaR and ES respectively. It can been verified that such strictly consistent scoring functions for VaR can be constructed and are of the form

$$S_{V,t} = (h_t^\alpha - \tilde{\alpha})(G(L_t) - G(v_t)) \tag{57}$$

where G is a strictly increasing function. However it is shown by Fissler & Ziegel (2016) that ES is elicitable of higher order in the sense that the pair $(VaR_t^\alpha, ES_t^\alpha)$ is jointly elicitable. This leads to the use of comparative tests. This implies that there exists a scoring function such that the following holds

$$(VaR_t^\alpha(L_t), ES_t^\alpha(L_t)) = \arg\min E[S_{V,E,t}(v_t, e_t, L_t)]. \tag{58}$$

Fissler & Ziegel (2016) propose the following choice of $S_{V,E,t}$

$$S_{V,E,t}(v_t, e_t, L_t) = (h_t^\alpha - \tilde{\alpha})(G_1(L_t) - G_1(v_t)) + \frac{1}{\tilde{\alpha}} G_2(e_t) h_t^\alpha (L_t - v_t) + G_2(e_t)(e_t - v_t) - \mathcal{G}_2(e_t) \tag{59}$$

with G_1 and G_2 strictly increasing continuously differentiable functions and $\mathcal{G}_2^{'} = G_2$. The first part of the equation shows that VaR on itself is elicitable and the second part shows that the remaining part cannot be split in a part only depending on e and a part only depending on v, this illustrates the statement that ES is not elicitable on its own. Fissler et al. (2015) propose the choice for G_1 and G_2 as $G_1(v) = v$ and $G_2(e) = exp(e)$. Define the loss differential D_t as follows

$$D_t = S_{V,E,t}^i - S_{V,E,t}^j \tag{60}$$

where $S_{V,E,t}^i$ and $S_{V,E,t}^j$ denote the scoring function for model i and j respectively. Fissler & Ziegel (2016) state that for model verification elicitability is not necessary, but that it is extremely important for model ranking. Because the strictly consistent scoring functions allow for comparison between models. In the special case of model selection one wants to do a pairwise comparison between model i and model j. This leads to the following null hypothesis

$$H_0 : E[S_{V,E,t}^i] - E[S_{V,E,t}^j] = 0 \tag{61}$$

where $E[S_{V,E,t}^i]$ and $E[S_{V,E,t}^j]$ are approximated with their sample means, which are equal to $\bar{S}_{V,E}^i = \frac{1}{T}\sum_{t=1}^T S_{V,E,t}^i(v_t, e_t, L_t)$ and $\bar{S}_{V,E}^j = \frac{1}{T}\sum_{t=1}^T S_{V,E,t}^j(v_t, e_t, L_t)$ respectively. This leads to the following alternative hypothesis

$$H_a : E[S_{V,E,t}^i] - E[S_{V,E,t}^j] < 0. \tag{62}$$

All possible pairs i, j are tested to avoid a selection bias, in the sense that selecting the best model beforehand is unfair towards other models. Using the scoring function in equation

(59) one can derive the following Diebold & Mariano (1995) based statistic

$$DM = \frac{\bar{S}^i_{V,E} - \bar{S}^j_{V,E}}{\sqrt{\frac{\tilde{\sigma}^2_D}{T}}} \sim N(0,1) \tag{63}$$

where $\tilde{\sigma}^2_D$ is the Newey-West estimate of the asymptotic variance. Diebold & Mariano (1995) show that the test statistic under the null hypothesis is asymptotically standard normal when $T \to \infty$ and $n \to \infty$.

4.7 Forecasting Framework

For the empirical application of this paper, a number of $m = 100$ in sample observations is used for the HS and CV models. For all other models a number of $m = 2500$ in sample observations is used. The size of the backtesting sample is in both cases equal to $T = 638$. The paper is based on a moving window estimation scheme with m observations. The focus of this paper is on estimating one-day ahead VaR and ES. This framework results in 638 estimated values for VaR and ES. The statistical significance of the VaR and ES estimates is assessed with the tests described in the previous sections.

5 Empirical Results

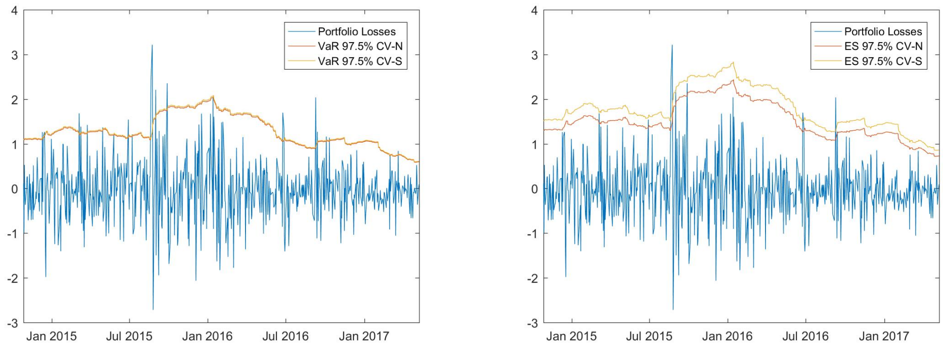
This section discusses the empirical results of the paper. In section 5.1 the results of the univariate estimation models are given. Section 5.2 describes the empirical findings for the multivariate TGARCH-EVT-Copula model. Section 5.3 concludes with ranking the models that passed all tests.

5.1 Univariate Estimation models

All benchmark models are univariate, indicating that at the beginning of the analysis the losses of the four ETFs are converted to portfolio losses by means of an equally weighted portfolio. Hence for the univariate models, the dependence structure between the ETFs is not important, since an equally weighted portfolio is used. In this section with benchmark models, all models are first used separately and in the following section the models are combined in the TGARCH-EVT-Copula model.

5.1.1 Benchmark Models

It can be seen from Figure 6 that the VaR forecasts of the CV model for a normal and *t*-distribution are almost the same, although the forecasts of the *t*-distribution are somewhat higher. The differences between the ES forecasts are more visible, it can be seen that the forecasts from the *t*-distribution model are higher, this is not surprising because this distribution accounts for fat tails and this is exactly what ES is about, the tail of the distribution. The fluctuating line is not surprising because the CV model only uses the first two moments of the distributions. These moments are estimated using 100 in-sample observations with a moving window, therefore these estimates of the moments do vary over time. These models take into account the current state of the financial world, for example around July, 2015 when a high loss enters the estimation sample, the estimates of VaR and ES become higher!



(a) VaR Variance-Covariance model

(b) ES Variance-Covariance model

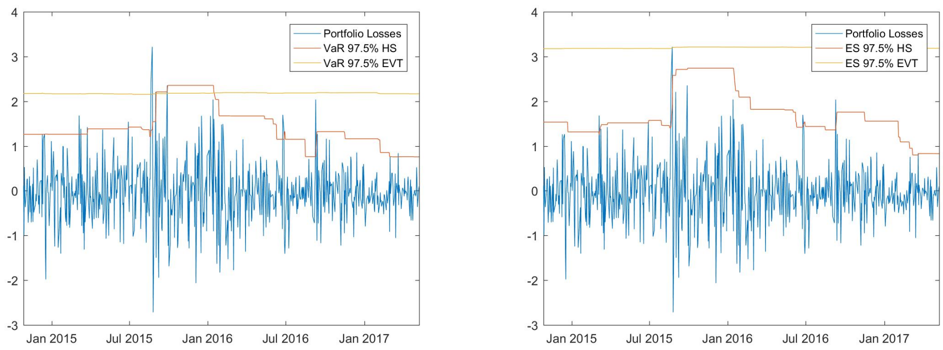
Figure 6: **VaR and ES plots of CV Model**. The plots show the 97.5% VaR and ES of the Variance-Covariance Model.

In Figure 7 the VaR and ES forecasts of the HS and EVT model are shown. It can be seen that the forecasts for both VaR and ES of the EVT model are rather constant. This is not surprising because this model considers the tail of the distribution and hence when no additional large loss occurs the tail remains unaltered. Although the other i.i.d. models in this research are estimated with $m = 100$ in sample observations, the EVT model uses $m = 2500$ observations, because of the fact that approximately 200 observations should lie in the tail (McNeil & Frey, 2000). The estimated tail index parameter is equal to 0.1550 and therefore positive as can be expected for financial heavy tailed data and the estimated scale factor is also positive and equal to 0.6949.

After the high peak around July 2015 the HS forecasts of the VaR and ES increase, this is obviously caused by the change of the tail in the sense that extreme observations are added. The HS forecasts tend to follow the the direction of the portfolio losses, although the forecasts are clearly discontinuous as can be seen in Figure 7.

5.1.2 Time-varying Volatility Models

As can be seen from the losses in Figure 7 the volatility seems to change over time, therefore this research considers the AR(1)-GARCH(p,q) and AR(1)-TGARCH(p,q) specification to deal with the time-varying volatility. To select the order of lags in the specifications, the AIC and BIC values are compared to each other. Using these values as a measure, the optimal lag lengths are chosen to be $p = q = 1$. Consequently, this research considers an AR(1)-GARCH(1,1) and AR(1)-TGARCH(1,1) model.



(a) VaR HS and EVT model

(b) ES HS and EVT model

Figure 7: **VaR and ES plots of HS and EVT model**. The plots show the 97.5% VaR and ES of the HS and EVT models.

The average parameters, average standard errors and average t statistics of these models over the 638 models can be found in Table 3. It may appear a bit strange that the estimated mean μ is negative, but this is due to the fact that losses are considered instead of returns. The parameters of the AR(1) model are in both cases quite similar to each other, which indicates that the mean difference between the two models lies in the variance specification, which of course was expected. Also, the estimated β_1 and the estimated degrees of freedom are comparable for the two specifications.

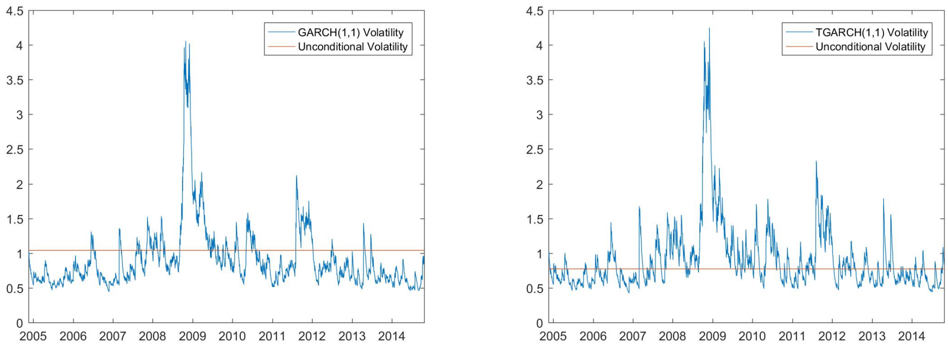
The difference lies in the parameters α_1 and γ_1, which measures the effects of shocks to the variance. Although the AR(1)-TGARCH(1,1) model allows the effects of positive and negative shocks to be different, the average impact $\alpha_1 + \frac{\gamma_1}{2}$ of shocks is roughly the same. All parameters in both models are significant except of the ϕ parameter, which has a remarkably high standard error in comparison with the parameter values.

Table 3: **Estimated parameters**. Estimated parameters of the GARCH and TGARCH model. The models are estimated using a moving window with 2500 observations, this results in 638 estimated models. The shown values are the average parameters in the top row, the average standard errors between brackets and the average t-statistics of the 638 estimated models in the bottom row.

	μ	ϕ	ω	α_1	β_1	γ_1	ν
GARCH(1,1)	-0.0767	-0.0416	0.0150	0.1055	0.8827	·	6.2917
	(0.0142)	(0.0214)	(0.0044)	(0.0154)	(0.0149)	·	(0.8006)
	-5.3826	-1.9480	3.4255	6.8699	59.2612	·	7.8588
TGARCH(1,1)	0.0573	-0.0401	0.0168	0.1684	0.8919	-0.1656	6.5429
	(0.0143)	(0.0211)	(0.0037)	(0.0136)	(0.0210)	(0.0241)	(0.8466)
	-4.0067	-1.9002	4.5822	12.3821	42.5662	-6.8775	7.7286

Both models are on average covariance stationary, note that the necessary condition for this is $\alpha_1 + \beta_1 < 1$ for the GARCH(1,1) model and $\alpha_1 + \beta_1 + \frac{\gamma_1}{2} < 1$ for the TGARCH(1,1) model. To show the effect of time-varying volatility, recall that the unconditional variance of a GARCH(1,1) model is equal to $\frac{\omega}{1-\alpha_1-\beta_1}$ and the unconditional variance of a

TGARCH(1,1) model is equal to $\frac{\omega}{1-\alpha_1-\beta_1-\frac{\gamma_1}{2}}$. The resulting volatilities estimated over the first 2500 observations are shown in Figure 8. It can be seen that the estimated volatilities are not constant over time but fluctuate around the unconditional volatility. As the Great Depression hits the financial market around 2008 a large volatility shock is observed in both models. Both volatility series show the same behavior, although the TGARCH(1,1) volatilities are larger for negative shocks due to the asymmetric shock component in its specification.



(a) Volatility GARCH(1,1) Model (b) Volatility TGARCH(1,1) Model

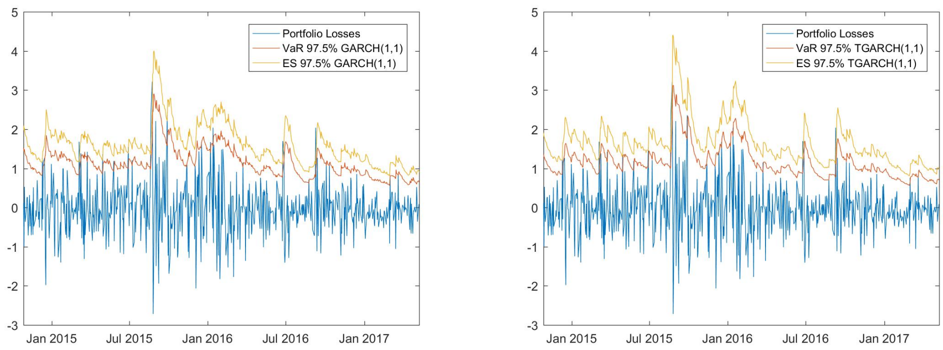
Figure 8: **Volatility plots of the GARCH(1,1) and TGARCH(1,1) models**. The plots show the estimated and unconditional volatilities of the GARCH(1,1) and TGARCH(1,1) model over the sample period November 19, 2004 till October 24, 2014.

The plots of the estimated VaR and ES of the GARCH(1,1) and TGARCH(1,1) model can be found in Figure 9. Where the other models deliver forecasts that are actually quite constant, the GARCH(1,1) and TGARCH(1,1) models deliver forecasts that move in the direction of the portfolio losses. Large shocks in losses are reflected in large shocks to the estimates of VaR and ES. The estimates of ES are higher then the estimates of VaR which is not surprising because ES accounts for the average loss in the tail. It can be seen that the estimates of ES and VaR of the GARCH(1,1) and TGARCH(1,1) model are quite similar, however at some dates the TGARCH(1,1) estimates are higher due to the asymmetric news component.

To apply EVT correctly, EVT and GARCH are combined in the TGARCH-EVT model. Using this model also time-varying volatility can be introduced in the model. The resulting VaR and ES estimates are shown in Figure 10. It is clearly seen that introducing time-varying volatility and GARCH filtering improved the estimates of VaR and ES. Where the estimated VaR and ES of the normal GPD are nearly constant, the estimates of the TGARCH-EVT model fluctuate in the direction of the portfolio losses.

5.1.3 Statistical Results

In order to do statistical interference on the models the tests as described in section 4.4 and 4.5 are conducted. The summary statistics for these tests for $\alpha = 97.5\%$ and $\alpha = 99\%$



(a) VaR and ES forecasts GARCH(1,1) (b) VaR and ES forecasts TGARCH(1,1)

Figure 9: **VaR and ES plots of the GARCH(1,1) and TGARCH(1,1) models**. The plots show the estimated 97.5% VaR and ES of the GARCH(1,1) and TGARCH(1,1) models over the sample period November 19, 2014 till May 9, 2017.

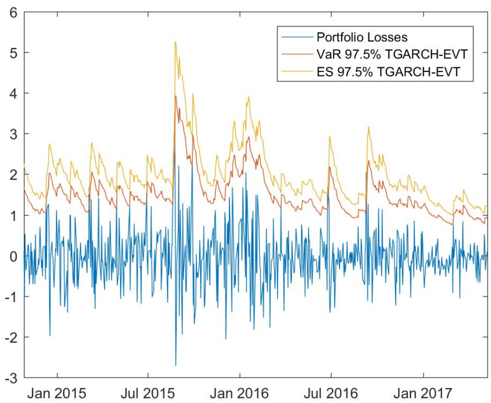


Figure 10: **VaR and ES plot of the TGARCH-EVT model**. The plots show the estimated 97.5% VaR and ES of the TGARCH-EVT model over the sample period November 19, 2004 till May 9, 2017.

can be found in Table 4 and Table 5 respectively.

It can be seen that the number of violations of the EVT model is much lower than expected. The p-value for the LR_{IND} statistic in the case of $\alpha = 99\%$ for the EVT seems somewhat high, but this is due to the fact that there is only one violation. Therefore the violations are not correlated over time and hence the p-value approaches one.

Lower exceptions then expected indicate that the models overestimate the risk associated with this portfolio. Hence the corresponding VaR forecasts are too high and the actual losses are lower than the VaR forecasts. One can argue that this outcome is preferable because the risk is limited in this case. However, in risk management, VaR determines the amount of capital that is needed as buffer to back losses. If the risk is overestimated, too much capital is reserved as a buffer and hence firms loose possible profits, that could result from investing the capital.

The test of correct unconditional coverage of VaR is passed by four models at signifi-

Table 4: **Backtesting results for $\alpha = 97.5\%$.** The table shows the estimated proportion of exceptions and the respective p-values belonging to the test statistics. Test statistics for ES are obtained by simulating $M = 10000$ times from the predictive distribution and by setting $m = 5$ as in Du & Escanciano (2017). Tests are conducted on a 5% significance level.

	VC-N	VC-S	*HS*	*GARCH(1,1)*	*TGARCH(1,1)*	*EVT*	*TGARCH-EVT*
$\hat{\pi}$	3.61%	3.29%	2.35%	3.92%	3.29%	0.63%	1.72%
LR_{UC}	0.0932	0.2216	0.8078	0.0338	0.2216	0.0003	0.0128
LR_{IND}	0.0135	0.0534	0.0043	0.0310	0.4154	0.0150	0.0128
LR_{CC}	0.0116	0.0733	0.0165	0.0103	0.3401	0.0000	0.0186
U_{ES}	0.0142	0.0036	0.1979	0.1645	0.0692	0.0005	0.0159
C_{ES}^5	0.0001	0.0000	0.0000	0.0370	0.5932	0.0000	0.9271

Table 5: **Backtesting results for $\alpha = 99\%$.** The table shows the estimated proportion of exceptions and the respective p-values belonging to the test statistics. Test statistics for ES are obtained by simulating $M = 10000$ times from the predictive distribution and by setting $m = 5$ as in Du & Escanciano (2017). Tests are conducted on a 5% significance level.

	VC-N	*VC-S*	*HS*	*GARCH(1,1)*	*TGARCH(1,1)*	*EVT*	*TGARCH-EVT*
$\hat{\pi}$	1.72%	1.25%	1.41%	1.72%	2.04%	0.16%	0.63%
LR_{UC}	0.0956	0.5352	0.3262	0.0956	0.0209	0.0077	0.3092
LR_{IND}	0.0004	0.0908	0.1246	0.0128	0.0295	0.9987	0.9899
LR_{CC}	0.0005	0.1975	0.1899	0.0112	0.0065	0.0287	0.5963
U_{ES}	0.0079	0.0307	0.0211	0.3492	0.1840	0.0281	0.0703
C_{ES}^5	0.0000	0.0000	0.0049	1.0000	0.9999	0.0000	0.7623

cance level 97.5% and five models at the 99% significance level. The test of independence of the VaR violations is passed by the VC-S and TGARCH(1,1) model at the 97.5% significance level and by the VC-S, HS and TGARCH-EVT model at the 99% level while ignoring the high p-values for the EVT model. The test of correct conditional coverage is only passed by the VC-S and TGARCH(1,1) model at 97.5% level and the VC-S, HS and TGARCH-EVT model at 99% level.

The unconditional test of ES is only passed by three models at the 97.5% level and by three at the 99% level. The test of uncorrelated ES forecasts is only passed by the TGARCH(1,1) model and the TGARCH-EVT model at the 97.5% level. On the 99% level only the GARCH based models pass the test. The results suggest that the VC-S and HS model are a good fit for forecasting VaR but are poor in ES forecasting.

When including time-varying volatility in the models the forecasting ability of ES increases, in the sense that from the GARCH based models only the TGARCH-EVT model at the 97.5% does not pass the test of correct unconditional coverage. This results show that by including the current financial state in the analysis, the models are more capable to incorporate the tail behavior. It can be seen that the GARCH(1,1) and TGARCH(1,1) model slightly underestimate the risk, however, the GARCH based models are most promising according to these results.

Overall it can be concluded that by including time-varying volatility in the models, the forecasting ability increases. As the Basel Committee recommends to change from $VaR^{99\%}$ as risk measure towards $ES^{97.5\%}$, the results from the TGARCH(1,1) and TGARCH-EVT model gives strong evidence for the use of these models in univariate forecasting of VaR and ES. These unconditional and conditional backtests only judge the forecasting abilities of the models individually, but cannot be used for model comparison. Pairwise model comparison can be done using the Diebold & Mariano (1995) framework with the Fissler & Ziegel (2016) scoring function and will be done in section 5.3.

5.2 TGARCH-EVT-Copula Model

5.2.1 Marginal Distribution Modeling

The aforementioned models are all univariate models in the sense that the analysis is done on the portfolio losses rather than on the ETF losses separately. This section deals with the multivariate copula model that takes the dependence structure of the ETFs into account. Therefore for each ETF separately an AR(1)-TGARCH(1,1) model is estimated. The resulting average parameters, average standard errors and average t statistics can be found in Table 6.

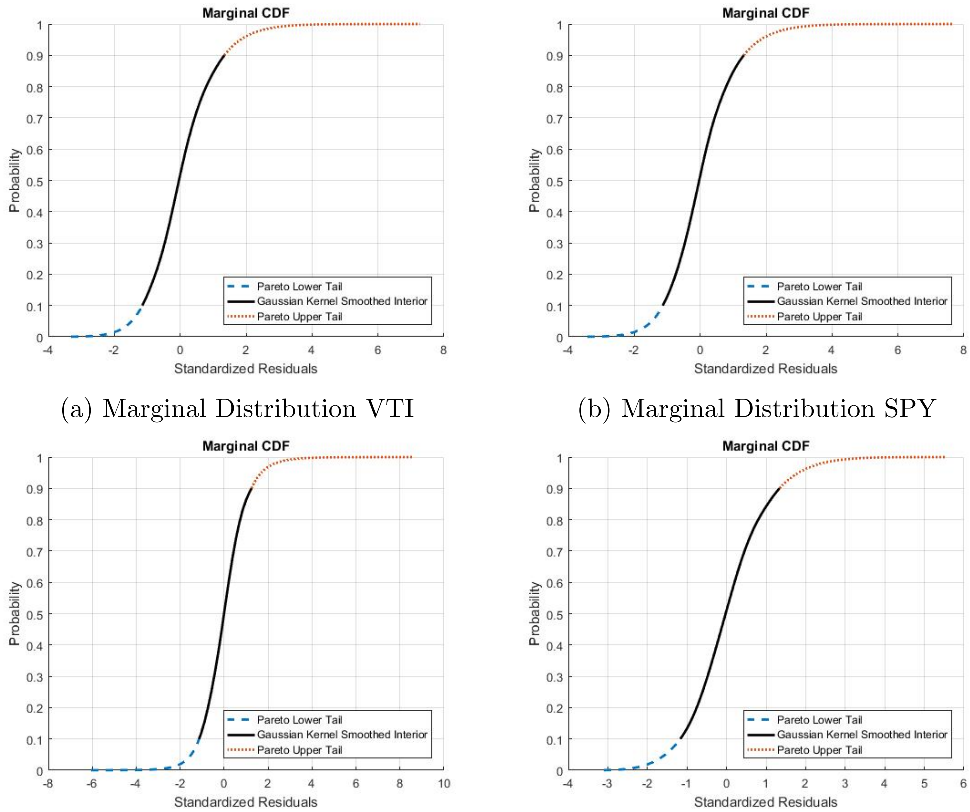
Table 6: **Estimated parameters**. Estimated parameters of the AR(1)-TGARCH(1,1) model for each ETF separately. The models are estimated using a moving window with 2500 observations, this results in 638 estimated models. The shown values are the average parameters in the top row, the average standard errors between brackets and the average t-statistics of the 638 estimated models in the bottom row.

	μ	ϕ	ω	α_1	β_1	γ_1	ν
	-0.0587	-0.0267	0.0223	0.2170	0.8727	-0.2170	6.5669
VTI	(0.0160)	(0.0219)	(0.0044)	(0.0155)	(0.0262)	(0.0284)	(0.9384)
	-3.6651	-1.2183	5.0109	14.0378	33.2672	-7.6320	6.9980
	-0.0585	-0.0425	0.0217	0.2255	0.8698	-0.2255	6.1506
SPY	(0.0153)	(0.0217)	(0.0043)	(0.0152)	(0.0269)	(0.0293)	(0.8418)
	-3.8123	-1.9559	5.0568	14.8118	32.2951	-7.7022	7.3063
	-0.0560	-0.0403	0.0100	0.0291	0.9535	0.0281	4.9986
GLD	(0.0199)	(0.0197)	(0.0045)	(0.0084)	(0.0078)	0.0128	(0.5264)
	-2.8080	-2.0401	2.2285	3.4749	122.7416	2.1974	9.4951
	-0.0832	-0.0109	0.0369	0.2097	0.8700	-0.2097	7.0392
QQQ	(0.0190)	(0.0217)	(0.0074)	(0.0161)	(0.0264)	0.0283	(1.1711)
	-4.3876	-0.5017	5.0164	13.0318	32.9081	-7.4088	6.0109

The resulting parameters are quite normal for financial returns, the term $\alpha_1 + \beta_1 + \frac{\gamma_1}{2}$ is close to one for such returns and all models are covariance stationary. The AR parameters of the four ETFs separately are of the same magnitude as the AR parameters of the portfolio. In both cases the ϕ parameter is not significant, although the parameter of the GLD ETF is significant. It can be seen that there is large estimation error in forecasting this particular AR parameter.

The parameters of the TGARCH specification for each ETF correspond closely to the ones obtained for the portfolio, except for the GLD ETF. The estimated α_1 and γ_1 are substantially smaller than the ones of the other ETFs and the portfolio, this gives an indication that the news component is not important for the variance of this ETF, whereas it was important for the portfolio. The higher estimated β_1 for the GLD ETF indicates that the variance of this ETF can be mostly explained by the previous period variance. For the other parameters of the ETFs the same behavior as the portfolio is observed, the average effect of news shocks is approximately the same. The resulting standardized residuals from the AR(1)-TGARCH(1,1) models are then used as input for the GPD model.

A crucial part of copula theory is the identification of the marginal distribution of the assets. This research uses a semi-parametric approach for this purpose. The resulting semi-parametric marginal distributions over the first 2500 observations are shown in Figure 11, the three separate parts of the marginals are clearly visible. It can be seen that 10% of the observations is reserved for the upper and lower tail respectively. Because the focus is on losses we are interested in the upper tail, it can be seen that indeed all ETFs are heavily tailed.



(a) Marginal Distribution VTI

(b) Marginal Distribution SPY

(c) Marginal Distribution GLD

(d) Marginal Distribution QQQ

Figure 11: **Semi-Parametric Marginals**. The plots show the marginal distributions of the four ETFs over the sample period November 19, 2004 till October 24, 2014.

5.2.2 Dependence Structure Modeling

The second step in the copula framework is the choice of a certain type of copula. The fit of the copula is assessed using the backtests for VaR and ES. The estimated parameters of these copulas are shown in Table 7. It can be seen from the estimated correlation matrices that the ETFs are in fact heavily correlated. For example, VTI and GLD have a correlation of 0.9848 and VTI and QQQ have also a high correlation of 0.8955, this shows clearly that there is a dependence structure between these ETFs.

Table 7: **Estimated Copula parameters**. Estimated parameters of the copulas. The θ corresponds to the parameter in the Archimedean copulas. The R_G and R_S are the correlation matrices of the Gaussian and Student copula respectively. The ν corresponds to the degree of freedom of the Student copula. The copulas are estimated using a moving window with 2500 observations, this results in 638 estimated models. The shown parameters are the average parameters of the 638 estimated models.

	Clayton	Frank	Gumbel	Gaussian	Student
θ	0.7977	3.4506	1.4699	·	·
R	·	·	·	R_G	R_S
ν	·	·	·	·	4.8217

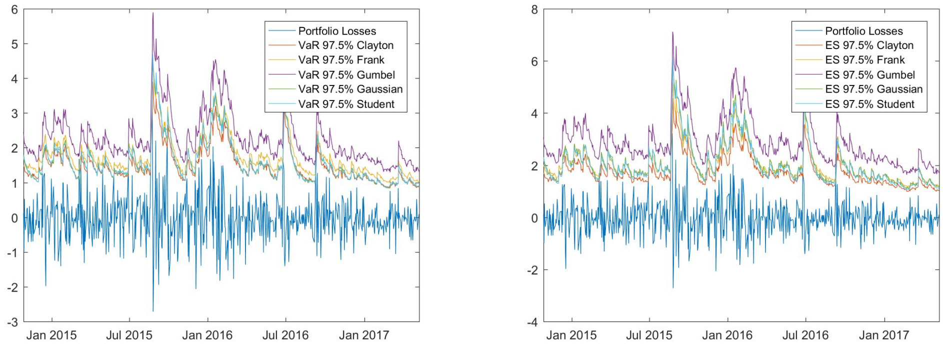
$$R_G = \begin{bmatrix} \cdot & 0.9848 & 0.0962 & 0.8955 \\ 0.9848 & \cdot & 0.0854 & 0.8905 \\ 0.0962 & 0.0854 & \cdot & 0.0483 \\ 0.8955 & 0.8905 & 0.0483 & \cdot \end{bmatrix} \qquad R_S = \begin{bmatrix} \cdot & 0.9895 & 0.0868 & 0.8990 \\ 0.9895 & \cdot & 0.0771 & 0.8932 \\ 0.0868 & 0.0771 & \cdot & 0.0480 \\ 0.8990 & 0.8932 & 0.0480 & \cdot \end{bmatrix}$$

The estimated correlation matrices correspond closely to the sample correlations in Table 2, although these correlations are somewhat higher. The difference in correlations is the largest for the GLD ETF. The results show that the imposed dependence structure by means of copulas, results in a higher correlation. Therefore the univariate models disregard this dependence structure by using equally weighting beforehand.

The Frank copula has the highest degree of dependence and the Clayton the lowest degree of dependence. The lowest degree of dependence of the Clayton copula was expected because this copula exhibits lower tail dependence, but the focus in this case is on the upper tail, therefore the estimated strength of dependence is lower. On the other hand the Gumbel copula exhibits right tail dependence and therefore the estimated strength of dependence is larger. The Frank copula exhibits neither left nor right tail dependence and is therefore symmetric. The estimated degree of freedom of the Student copula is of the expected magnitude, a degree of freedom of around 5 is expected for such financial data.

The plots of the VaR and ES of the five copulas are shown in Figure 12. From Figure 12a is clearly seen that the Gumbel model delivers the highest forecasts of VaR and that the Clayton model delivers the lowest forecasts of VaR. For the forecasts of ES essentially the same pattern is observed, the Gumbel model has the highest forecasts and the Clayton the lowest. This can be explained from the insight that the Clayton copula exhibits lower tail dependence, therefore dependence in the upper tail is less emphasized, this results in lower estimated ES and VaR.

On the other hand the Gumbel copula exhibits upper tail dependence and therefore the upper tail is more emphasized, this results in higher forecasts. The Gaussian and Frank copula have neither lower nor upper tail dependence and lie therefore in between the other models. There seems to be some kind of ranking between the models, the forecasts of Frank, Gaussian and Student lie between the forecasts of Clayton and Gumbel. Loosely speaking it can be concluded that the Clayton model underestimates the risk, whereas the Gumbel model overestimates the risk, however this is just arguing from intuition. For statistical inference on the performance the backtests are used.



(a) VaR forecasts

(b) ES forecasts

Figure 12: **VaR and ES plots of the five TGARCH-EVT-Copula models**. The plots show the 97.5% VaR and ES over the sample period November 19, 2014 till October May 9, 2017.

5.2.3 Statistical Results

The expected number of exceptions is around 16 at the 97.5% significance level and around 6 at the 99% significance level. It can be clearly seen that the Clayton model under estimates the risk as it has for both significance levels to many exceptions. On the other hand, the Gumbel model seems to over estimate the risk as there are to few exceptions. From the firms perspective neither of the two is desirable because if the risk is under estimated the firm is at risk, but on the other hand, if the risk is over estimated too much capital is needed to retain the capital buffer. This immediately leads to missed profits, therefore over- and underestimation of risk are both dangerous.

The tests of correct conditional and unconditional coverage are passed by three models at the 97.5% level and by two models at the 99% level. Out of the ten estimated models, only two do not pass the test of correct conditional coverage, this clearly indicates that including the dependence structure in the estimation improves the independence of the violations in the sense that the violations are now evenly spread over the sample.

Three models at the 97.5% level and four models at the 99% level pass the tests of

Table 8: **Copula Backtesting Results for** $\alpha = 97.5\%$. The table shows the estimated proportion of exceptions and the respective p-values belonging to the test statistics. Test statistics for ES are obtained by simulating $M = 10000$ times from the predictive distribution and by setting $m = 5$ as in Du & Escanciano (2017). Tests are conducted on a 5% significance level.

	Clayton	*Frank*	*Gumbel*	*Gaussian*	*Student*
$\hat{\pi}$	3.45%	1.72%	0.78%	1.41%	1.41%
LR_{UC}	0.1465	0.1840	0.0012	0.0552	0.0552
LR_{IND}	0.5994	0.0128	0.9859	0.9658	0.9658
LR_{CC}	0.3035	0.0186	0.0053	0.1589	0.1589
U_{ES}	0.0464	0.3304	0.0103	0.3304	0.1840
C_{ES}^5	0.0195	0.7885	0.6466	1.0000	0.9999

Table 9: **Copula Backtesting Results for** $\alpha = 99\%$. The table shows the estimated proportion of exceptions and the respective p-values belonging to the test statistics. Test statistics for ES are obtained by simulating $M = 10000$ times from the predictive distribution and by setting $m = 5$ as in Du & Escanciano (2017). Tests are conducted on a 5% significance level.

	Clayton	*Frank*	*Gumbel*	*Gaussian*	*Student*
$\hat{\pi}$	1.88%	0.78%	0.16%	0.47%	0.31%
LR_{UC}	0.0463	0.5862	0.0077	0.1336	0.0416
LR_{IND}	0.0197	0.9859	0.9987	0.9935	0.9964
LR_{CC}	0.0091	0.8496	0.0287	0.3245	0.1255
U_{ES}	0.0389	0.7515	0.0746	0.3149	0.1980
C_{ES}^5	0.8251	0.9999	0.8712	1.0000	1.0000

unconditional and conditional coverage of ES. This clearly indicates that including the dependence structure by means of copulas improved the forecasts of ES. The Clayton model is the only model that does not pass the backtests for ES at the 97.5% level and only passes the unconditional backtest at the 99% level.

Judging from these results the Gaussian and Student models deliver the best forecasts at the 97.5% level and the Frank and Gaussian model deliver the best forecasts at the 99% level. Overall the Gaussian copula seems a good choice for estimating the VaR and ES of these ETFs.

When comparing the results of the multivariate models with the univariate models it can be seen that the multivariate models are better suited for forecasting VaR then the univariate benchmark models. However the GARCH based univariate models perform also well in forecasting VaR. The multivariate models are better suited for forecasting ES then the univariate models, this can be seen from the fact that almost all multivariate models pass both the unconditional and conditional backtest, whereas this is certainly not the case for the univariate models. Clearly the combination of including time-varying volatility and including the dependence structure improved the forecasting ability of ES.

5.3 Model selection

In this section, the univariate models that passed the unconditional and conditional backtests of VaR and ES are ranked with the multivariate models that passed those tests using the Diebold & Mariano (1995) framework with the Fissler & Ziegel (2016) scoring function. This way of testing prohibits bad models to be chosen over good models (Sarma et al., 2003).

For example when a model structurally delivers too high values of VaR and ES, the model selection procedure will be unfair because the score of this model will be too low. This can be seen from the fact that with little exceptions, the part in equation (59) with h_t^α cancels out most of the times, therefore the score of the model is biased downwards. To avoid this biasness this two stage selection procedure is used.

From the univariate models only the TGARCH(1,1) model passes all tests and from the multivariate models only the Gaussian and Student models passed all tests at the 97.5% level. Further, from the univariate models, only the TGARCH-EVT model passes all tests and from the multivariate models, only the Frank and Gaussian model at the 99% level. The results of these rankings can be found in Table 10 and 11.

It is interesting to see that the models that exhibit lower or upper tail dependence are not among the models that passed all backtests. This indicates that in this particular application with ETFs there is no gain in the heavier weights in the tails. This effect may stem from the fact that the model with lower tail dependence tends to underestimate the risk whereas the model with upper tail dependence tends to overestimate the risk. It appears that models that are symmetric find a kind of balance in this and deliver therefore better VaR and ES forecasts.

Table 10: **Model selection means and p-values for $\alpha = 97.5\%$.** The table shows the means of the loss differential of the Diebold-Mariano test statistic for model i (left column) versus model j (upper row) and their respective p-values. Only univariate and multivariate models that passed all backtests are considered. Tests are conducted on a 5% significance level.

	TGARCH(1,1)	*Gaussian*	*Student*
TGARCH(1,1)	·	1.2605 (0.9996)	0.7779 (0.9577)
Gaussian	-1.2605 (0.0004)	·	-0.4826 (0.0011)
Student	-0.7779 (0.0423)	0.4826 (0.9989)	·

Table 11: **Model selection means and p-values for $\alpha = 99\%$.** The table shows the means of the loss differential of the Diebold-Mariano test statistic for model i (left column) versus model j (upper row) and their respective p-values. Only univariate and multivariate models that passed all backtests are considered. Tests are conducted on a 5% significance level.

	TGARCH-EVT	*Frank*	*Gaussian*
TGARCH-EVT	·	3.6453 (0.9782)	2.4166 (1.0000)
Frank	-3.6453 (0.0218)	·	-1.2287 (0.1811)
Gaussian	-2.4166 (0.0000)	1.2287 (0.8189)	·

At the 97.5% level both the Student and the Gaussian model have a significantly lower mean score, indicating that they both outperform the TGARCH(1,1) model. This clearly indicates that multivariate modeling in combination with EVT and TGARCH filtering significantly improves the forecasting ability of the models. It can be seen that the Gaussian model outperforms the Student model, because of the significant negative mean score. This clearly indicates that the multivariate models perform better and that of the multivariate models the Gaussian model outperforms the Student model. Following the results of the other backtests and the high estimated correlation matrix in the copula model, we deem the Gaussian model as the best performing model at the 97.5% significance level.

At the 99% significance level both the Frank and the Gaussian model outperform the TGARCH-EVT model, as can be seen from the negative significant means in the test results. The difference in means between the two multivariate models is not significant and therefore these models have equal forecasting ability. This may be due to the fact that both models have neither lower nor upper tail dependence.

Overall it is clearly shown that by including an a-symmetric news component the forecasting ability improves. Summarizing at the 97.5% significance it seems that the Gaussian model performs the best, whereas at the 99% significance level both the Frank and Gaussian model are the best.

6 Conclusion

The vast majority of firms use Value at Risk (VaR) as the risk measure for their risk management. This paper highlights that VaR is not a coherent risk measure because of its lack of subadditivity. Furthermore, it is also mentioned that VaR does not account for tail risk. Therefore Expected Shortfall (ES) is proposed as the replacement for VaR. The losses of four ETFs are used to demonstrate the value of ES over VaR.

The research starts with reviewing univariate estimation models like the Variance-Covariance (VC) model, the Historical Simulation (HS) model, the Filtered Historical Simulation (FHS) model and the Extreme Value Theory (EVT) model. Furthermore, the FHS and EVT are combined in a new approach, the TGARCH-EVT model. It is shown using the standard tests of Kupiec (1995) and Christoffersen (1998) for VaR and the state of the art testing method of Du & Escanciano (2017) for ES that the VC, HS and EVT approaches are outperformed by the univariate GARCH based models. It is clearly shown that including time-varying volatility in the estimation improves the forecasting abilities of the models.

Because of the lack of a dependence structure in the univariate models, a model based on GARCH filtering, EVT and copulas is implemented, the so called TGARCH-EVT-Copula model. It is clearly shown that multivariate modeling improves the forecasting abilities for both VaR and ES, however, the effect for VaR is smaller than for ES. The

Gaussian TGARCH-EVT-Copula model delivers the best results over the two significance levels. Whereas in the past testing ES appeared to be a major challenge, the testing framework of Du & Escanciano (2017) shows that these tests can be done using the cumulative violations process. The results prove that the multivariate models are better suited to forecast ES.

The backtests can only be used for testing model specification and not for model selection. Therefore, based on the result that VaR and ES are jointly elicitable the models that passed all backtests are ranked using the Diebold & Mariano (1995) framework with the Fissler & Ziegel (2016) scoring function. It is shown that the Gaussian TGARCH-EVT-Copula model perform best at the 97.5% significance level. The Frank and Gaussian TGARCH-EVT-Copula model have equal predictive ability at the 99% significance level, but both outperform the univariate TGARCH-EVT model.

Based on the recommendation of the Basel Committee on Banking Supervision to substitute $ES^{97.5\%}$ as alternative for $VaR^{99\%}$, we conclude that the Gaussian TGARCH-EVT-Copula is the best choice in forecasting VaR and ES because this is the only model that passed all tests at both significance levels.

In this research, only four assets are used to demonstrate the superiority of the multivariate models and to argue that ES can be estimated without additional computational effort. A suggestion for future research could be to extend the models to the high dimensional cases and see if the results still hold.

Furthermore, in this research simplifying assumptions are made to make the analysis feasible in the sense that it was assumed that trading costs are negligible and that the firms are buy-and-hold investors. It could be interesting for future research to include also dynamic investing and short/long investing.

Another suggestion for further research is to model the dependence structure of the assets using dynamic copulas instead of static copulas and see if the results still hold.

References

Abad, P., Benito, S., & López, C. (2014). A comprehensive review of Value at Risk methodologies. *The Spanish Review of Financial Economics*, *12*, 15-32.

Acerbi, C., Nordio, C., & Sirtori, C. (2001). *Expected Shortfall as a tool for financial risk management. working paper.* Retrieved from `http://www.gloriamundi.org/var/wps.html>`

Acerbi, C., & Täsche, D. (2002). Expected Shortfall: a natural coherent alternative to Value at Risk. *Economic notes*, *31*(2), 379-388.

Artzner, P., Delbaen, F., Eber, J. M., & Heath, D. (1997). Thinking coherently: Generalised scenarios rather than VaR should be used when calculating regulatory capital. *Risk London Risk Magazine Limited*, *10*, 68-71.

Artzner, P., Delbaen, F., Eber, J.-M., & Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, *2*(3), 203-228.

Balkema, A., & de Haan, L. (1974). Residual life time at great age. *Annals of Probability*, *2*(5), 792-804.

Barone-Adesi, G., Giannopoulos, K., & Vosper, L. (1999). VaR without Correlations for Portfolios of Derivative Securities. *Journal of Futures Markets*, *19*(5), 583–602.

Barone-Adesi, G., Giannopoulos, K., & Vosper, L. (2002). Backtesting Derivative Portfolios with Filtered Historical Simulation (FHS). *European Financial Management*, *8*(1), 31-58.

Basak, S., & Shapiro, A. (2001). Value-at-Risk Based Risk Management: Optimal Policies and Asset Prices. *The Review of Financial Studies*, *14*(2), 371-405.

Berkowitz, J. (2001). Testing density forecasts with applications to risk management. *Journal of Business and Economic Statistics*, *19*(4), 465-474.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics*, *31*(3), 307-327.

Bouye, E., Durrleman, V., Nileghbali, A., Riboulet, G., & Roncalli, T. (2000). *Copulas for Finance: a reading guide and some applications* (Manuscript). Financial Econometrics Research Center.

Brooks, C. (2014). *Introductory Econometrics for Finance* (Third ed.). Cambridge University Press.

Caillault, C., & Guegan, D. (2009). Forecasting VaR and Epected Shortfall using Dynamical Systems: A Risk Management Strategy. *Frontiers in finance and economics*, *6*(1), 26-50.

Carol, A. (2008). *Market Risk Analysis (Volume IV.) Value at Risk Models.* John Wiley and Sons.

Castillo, E., & Hadi, A. (1997). Fitting the Generalized Pareto Distribution to Data. *Journal of the American Statistical Association*, *92*(440), 1609-1620.

Chan, K. F., & Gray, P. (2006). Using Extreme Value Theory to measure Value-at-Risk for daily electricity spot prices. *International Journal of Forecasting*, *22*(2), 283-300.

Cherubini, U., & Luciano, E. (2001). Value at Risk trade-off and capital allocation with copulas. *Economic Notes*, *30*(2), 235-256.

Cherubini, U., Luciano, E., & Vecchiato, W. (2004). *Copula Methods in Finance.* John Wiley and Sons.

Christoffersen, P. F. (1998). Evaluating Interval Forecasts. *International Economic Review*, *39*(4), 841-862.

Danielsson, J., & de Vries, C. (2000). Value-at-Risk and extreme returns. *Annales d'Economie et de Statistique*, *60*, 239-270.

Delbaen, F. (1998). Working paper, coherent risk measures on general probability spaces..

Diebold, F. X., & Mariano, R. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, *13*(3), 253-263.

Du, Z., & Escanciano, J. C. (2017). Backtesting Expected Shortfall: Accounting for Tail Risk. *Management Science*, *63*(4), 940-958.

Embrechts, P., Kluppelberg, C., & Mikosch, T. (1997). Modeling Extremal Events for Insurance and Finance. *Berlin: Springer*.

Embrechts, P., McNeil, A. J., & Straumann, D. (1999). Correlation and dependence properties in risk management: Properties and pitfall. In M. Dempster (Ed.), *Risk management: Value at risk and beyond* (p. 176-223). Cambridge University Press.

Engle, R. F. (1982). Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, *50*(4), 987-1007.

Engle, R. F., & Manganelli, S. (2004). CAViar: Conditional Autoregressive Value at Risk by Regression Quantiles. *Journal of Business & Economic Statistics*, *22*(4), 367-381.

Fantazzini, D. (2008). Dynamic Copula Modelling for Value at Risk. *Frontiers in Finance and Economics*, *5*(2), 72-108.

Fissler, T., & Ziegel, J. F. (2016). Higher order elictability and Osband's principle. *The Annals of Statistics*, *44*(4), 1680-1707.

Fissler, T., Ziegel, J. F., & Gneiting, T. (2015). *Expected Shortfall is jointly elicitable with Value at Risk – Implications for Backtesting.* Retrieved from `http://arxiv.org/abs/1507.00244v2`

Fleming, J., & Kirby, C. (2003). A Closer Look at the Relation between GARCH and Stochastic Autoregressive Volatility. *Journal of Financial Econometrics*, *1*(3), 365-419.

Föllmer, H., & Schied, A. (2002). Convex measures of risk and trading constraints. *Finance and Stochastics*, *8*(4), 429-447.

Fortin, I., & Kuzmics, C. (2002). Tail dependence in stock return pairs. *International Journal of Intelligent Systems in Accounting, Finance Management*, *11*, 89-107.

Fundamental review of the trading book: A revised market risk framework. (2013). Bank for International Settlements.

Gao, F., & Song, F. (2008). Estimation risk in GARCH VaR and ES estimates. *Econometric Theory*, *24*(5), 1404-142.

Genest, C., Remillard, B., & Beaudoin, D. (2009). Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and Economics*, *44*(2), 199-213.

Gilli, M., & Këllezi, E. (2006). An application of Extreme Value Theory for Measuring Financial Risk. *Computational Economics*, *27*(1), 1-23.

Giot, J., & Laurent, S. (2004). Modelling daily Value-at-Risk using realized volatility and ARCH type models. *Journal of Empirical Finance*, *11*(3), 379-398.

Glosten, L., Jaganathan, R., & Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance*, *48*(5), 1779-1801.

Gneiting, T. (2011). Making and evaluating point forecasts. *Journal of the American Statistical Association*, *106*(494), 746-762.

Grimshaw, S. (1993). Computing the Maximum Likelihood Estimates for the Generalized Pareto Distribution to Data. *Technometrics*, *35*(2), 185-191.

Hill, B. M. (1975). A Simple General Approach to Inference About the Tail of a Distribution. *The Annals of Statistics*, *3*(5), 1163-1174.

Hofert, M., Machler, M., & McNeil, A. J. (2012a). *Estimators for Archimedean copulas in high dimensions.* Retrieved from `http://arxiv.org/pdf/1207.1708.pdf`

Hofert, M., Machler, M., & McNeil, A. J. (2012b). Likelihood inference for Archimedean copulas in high dimensions under known margins. *Journal of Multivariate Analysis*, *110*, 133-150.

Hosking, J. R. M., & Wallis, J. R. (1987). Parameter and quantile estimation for the Generalised Pareto Distribution. *Technometrics*, *29*(3), 339-349.

Hotta, L. K., Lucas, E. C., & Palaro, H. P. (2008). Estimation of VaR using Copula and Extreme Value Theory. *Multinational Finance Journal*, *12*(3/4), 205-218.

Kupiec, P. (1995). Techniques for verifying the accuracy of risk management models. *Journal of Derivatives*, *3*(2), 73-84.

Ljung, G. M., & Box, G. E. P. (1978). On a Measure of a Lack of Fit in Time Series Models. *Biometrika*, *65*(2), 297-303.

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*, *7*(3/4), 271-300.

McNeil, A. J., Frey, R., & Embrechts, P. (2015). *Quantitative Risk Management: Concepts, Techniques and Tools.* Princeton University Press.

Mittnik, S., & Paolella, M. S. (2000). Conditional density and Value-at-Risk prediction of Asian currency exchange rates. *Density Forecasting in Economics and Finance*, *19*(4), 313-333.

Morgan/Reuters, J. P. (1996). *Riskmetrics Technical Document, Part II: Statistics of*

Financial Market Returns (Fourth ed.). New York.

Palaro, H. P., & Hotta, L. K. (2006). Using Conditional Copula to Estimate Value at Risk. *Journal of Data Science*, *4*, 93-115.

Pickands, J. (1975). Statistical inference using extreme order statistics. *Annals of Statistics*, *3*(1), 119-131.

Pritsker, M. (2001). Evaluating Value at Risk Methodologies: Accuracy versus Computational Time. *Journal of Financial Services Research*, *12*, 201-242.

Rabemananjara, R., & Zakoian, J. M. (1993). Threshold ARCH models and asymmetries in volatility. *Journal of Applied Econometrics*, *8*(1), 31-49.

Romano, C. (2002). *Calibrating and simulating copula functions: an application to the Italian stock market* (Working paper No. 12/2002). CIDEM.

Rootzén, H., & Klüppelberg, C. (1999). A single number can't hedge against economic catastrophes. *Ambio*, *28*(6), 550-555.

Rosenblat, M. (1952). Remarks on a multivariate transformation. *The Annals of Mathematical Statistics*, *23*(2), 470-472.

Sarma, M., Thomas, S., & Shah, A. (2003). Selection of Value-at-Risk models. *Journal of Forecasting*, *22*(4), 337-358.

Scaillet, O. (2004). Nonparametric estimation and sensitivity analysis of Expected Shortfall. *Mathematical Finance*, *14*(1), 115-129.

Sklar, A. (1959). Fonctions de repartition a n dimensions et leurs marges. *publication Insst. Statist. Univ. Paris*, *8*, 229-231.

Tasche, D. (2002). Expected Shortfall and beyond. *Journal of Banking and Finance*, *26*(7), 1519-1533.

Taylor, J. W. (2008). Using exponentially weighted quantile regression to estimate Value at Risk and Expected Shortfall. *Journal of Financial Econometrics*, *6*(3), 382-406.

Wang, Z. R., Chen, X. H., Jin, Y. B., & Zhou, Y. J. (2010). Estimating risk of foreign exchange portfolio: Using VaR and CVaR based on GARCH–EVT-Copula model. *Physica A: Statistical Mechanics and its Applications*, *389*(21), 4918-4928.

Weber, S. (2006). Distribution-invariant risk measures, information, and dynamic consistency. *Mathematical Finance*, *16*(2), 419-441.

Yamai, Y., & Yoshiba, T. (2002). Comparative Analyses of Expected Shortfall and Value-at-Risk: Their Estimation Error, Decomposition, and Optimization. *Monetary and Economic Studies*, *20*(1), 87-121.

Yamai, Y., & Yoshiba, T. (2005). Value-at-Risk versus Expected Shortfall: A practical perspective. *Journal of Banking & Finance*, *29*(4), 997-1015.

Appendix

In this section the appendix is given. In the report references will be made to the subsections below.

A Properties of Risk Measures

Define L as the loss returns of the portfolio, l and λ as two deterministic quantities and denote $\varrho(L)$ as the risk measure of the portfolio losses L. The following gives a brief description of the six properties that a risk measure should have in order to quantify as a sufficient risk measure.

- **Monotonicity**

$$L_1 \leq L_2, \quad \text{implies that} \quad \varrho(L_1) \leq \varrho(L_2)$$

  This means that positions that lead to higher losses in every state of the world require more risk capital. Positions with $\varrho(L) \leq 0$ do not require any capital.

- **Translation Invariance**

$$\varrho(L + l) = \varrho(L) + l$$

  This property states that by adding or subtracting a deterministic quantity l to a position leading to the loss L, the capital requirements change with exactly the same amount.

- **Positive Homogeneity**

$$\varrho(\lambda L) = \lambda \varrho(L) \quad \text{for every} \quad \lambda > 0$$

  This means that multiplying the capital with a scalar, the risk also multiplies with the same scalar.

- **Subadditivity**

$$\varrho(L_1 + L_2) \leq \varrho(L_1) + \varrho(L_2)$$

  This means that diversification must actually matter, in the sense that two portfolios combined may not be riskier than the sum of the risks of the portfolios individually.

- **Convexity**

$$\varrho(\lambda L_1 + (1-\lambda)L_2) \leq \lambda \varrho(L_1) + (1-\lambda)\varrho(L_2) \quad \text{if} \quad \lambda \in [0, 1]$$

This property indicates that investing in different assets and diversification may never increase the risk, but it is possible that it reduces the risk.

- **Normalization**

$$\varrho(0) = 0$$

This obviously means that having no position immediately means having no risk.

When a risk measure is monotonic, translation invariant, positive homogeneous and subadditive it is characterized as a coherent risk measure. As proven in the literature, it is desired for a risk measure to be coherent, to qualify as decent risk measure. Subadditivity together with positive homogeneity implies convexity of the risk measure.

B Estimation Algorithm TGARCH-EVT Model

1. Estimate an AR(1)-TGARCH(1,1) model

$$L_t = \mu + \phi L_{t-1} + \sigma_t \epsilon_t \qquad \epsilon_t \sim t(\nu) \tag{64a}$$

$$\sigma_t^2 = \omega + (\alpha_1 + \gamma_1 I(\epsilon_{t-1} < 0))\epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \tag{64b}$$

2. With the estimated parameters filter out the standardized residuals

$$\hat{\epsilon}_t = \frac{L_t - \hat{\mu} - \hat{\phi} L_{t-1}}{\hat{\sigma}_t} \tag{65}$$

3. Define y as the excesses above the threshold of the residuals, where the threshold can be chosen using the mean excess and Hill plot. Estimate the following GPD distribution on these excesses

$$H_{\xi,\kappa}(y) = \begin{cases} 1 - (1 + \frac{\xi y}{\kappa})^{\frac{-1}{\xi}}, & \text{if } \xi \neq 0, \\ 1 - exp(-\frac{y}{\kappa}), & \text{if } \xi = 0, \end{cases} \tag{66}$$

4. Calculate the VaR and ES using the following GPD expressions for VaR and ES

$$\widehat{VaR}_t^{\alpha} = u + \frac{\hat{\kappa}}{\hat{\xi}} \left((\frac{n}{N_u}(1-\alpha))^{-\hat{\xi}} - 1 \right) \tag{67a}$$

$$\widehat{ES}_t^{\alpha} = \frac{\widehat{VaR}_t^{\alpha}}{1-\hat{\xi}} + \frac{\hat{\kappa} - \hat{\xi} u}{1 - \hat{\xi}} \tag{67b}$$

5. The VaR and ES estimates can be calculated using these expressions in equations (13) and (14) of the FHS approach, where the HS estimates in those formulas are replaced with the formulas in equation (67a) and (67b) respectively.

C List of Tables

D List of Figures