Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Mathematical Sciences

Catharina Bratterud Lilleengen

Estimating and forecasting volatility of stock market returns using GARCH models – with an application to value-at-risk and expected shortfall

Master's thesis in Industrial Mathematics
Supervisor: Jacob Kooter Laading
June 2021

NTNU
Norwegian University of
Science and Technology

Catharina Bratterud Lilleengen

Estimating and forecasting volatility of stock market returns using GARCH models – with an application to value-at-risk and expected shortfall

Master's thesis in Industrial Mathematics
Supervisor: Jacob Kooter Laading
June 2021

Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Mathematical Sciences

NTNU
Norwegian University of Science and Technology

Abstract

This thesis analyses the use of different GARCH models for modelling the conditional volatility in different stock market sectors. Improving the volatility models gives better predictions in the pricing of financial assets as well as better risk management. The pandemic of COVID-19 is an important example of a volatile period in the market, and it emphasises the relevance of correctly measuring the market risk that practitioners are exposed to. Another aspect added to this debate is the green trend in the markets during recent years, i.e., both investor demand for green assets and companies desire to be seen as green. This has caused both the value and the volatility of green stocks to increase. Using both three different distributional assumptions and three conditional volatility models, we estimate the historical stock prices of the companies NEL, Tesla, Tomra, Norsk Hydro, General Motors and Aker BP in the period November 2010 to January 2021. For NEL and Norsk Hydro, the data sets are extended to include data from January 2006 until January 2021. With varying degrees of complexity, these models are able to capture the main characteristics of financial returns such as volatility clustering, leptokurtosis and asymmetry. The models are evaluated in-sample with information criteria, while the out-of-sample analysis is carried out by making volatility forecasts and applying these in value-at-risk (VaR) and expected shortfall (ES) risk measures. VaR and ES are calculated by a parametric approach using GARCH models as well as historical simulation for comparison, and their performances are evaluated with backtesting. The results document that the choice of GARCH model does not seem to be of large importance for either of the stocks. In particular, the application to VaR and ES shows that different models appear to be preferred in different scenarios. However, as the model performances are quite similar, the simplest model will be preferred due to interpretability and computational complexity. The results also show that it is necessary to investigate the estimation sample carefully, and doing this can provide a more definite answer to the choice of model. Furthermore, the choice of distribution seems to be of some importance in the application to VaR and ES, which most likely is caused by the differences in return characteristics. More non-normality in the returns calls for more complexity in the distributional assumption of the innovations. Overall, the study shows that the GARCH models investigated could be helpful in risk measurement of equities.

Sammendrag

Denne studien analyserer bruken av forskjellige GARCH-modeller for modellering av betinget volatilitet i ulike sektorer av aksjemarkedet. Forbedring av volatilitetsmodellene gir bedre prediksjoner i prissettingen av finansielle aktiva, samt bedre risikostyring. COVID-19-pandemien er et viktig eksempel på en ustabil periode i markedet, og den understreker relevansen av nøyaktig måling av markedsrisiko som utøvere utsettes for. Et annet aspekt lagt til denne debatten er den grønne trenden i markedene de siste årene, dvs. både investorens etterspørsel etter grønne eiendeler og selskapers ønsker om å bli sett på som grønne. Dette har fått både verdien og volatiliteten til grønne aksjer til å øke. Ved å bruke både tre forskjellige distribusjonsforutsetninger og tre betingede volatilitetsmodeller, estimerer vi de historiske aksjeprisene til selskapene NEL, Tesla, Tomra, Norsk Hydro, General Motors og Aker BP i perioden november 2010 til januar 2021. For NEL og Norsk Hydro er datasettene utvidet til å omfatte data fra januar 2006 til januar 2021. Med varierende grad av kompleksitet er disse modellene i stand til å fange opp de viktigste karakteristikkene til finansiell avkastning slik som gruppering av volatilitet, leptokurtose og asymmetri. Modellene blir evaluert ved hjelp av estimeringssett og informasjonskriterier, mens valideringsettet er brukt til å lage volatilitetsprognoser og anvende disse i risikomålene value-at-risk (VaR) and expected shortfall (ES). VaR og ES beregnes ved hjelp en parametrisk tilnærming ved bruk av GARCH-modeller samt historisk simulering for sammenligning, og deres prestasjon er vurdert ved hjelp av backtesting. Resultatene dokumenterer at valg av GARCH-modell ikke ser ut til å ha stor betydning for noen av aksjene. Spesielt viser anvendelsen til VaR og ES at forskjellige modeller ser ut til å være foretrekke i ulike scenarioer. Men da modellprestasjonen er ganske like, vil den enkleste modellen være å foretrekke på grunn av tolkbarhet og beregningskompleksitet. Resultatene viser også at det er nødvendig å undersøke estimeringssettet nøye, og å gjøre dette kan gi et mer bestemt svar på valg av modell. Videre ser distribusjonsvalget ut til å ha en viss betydning i anvendelsen til VaR og ES, noe som mest sannsynlig skyldes forskjellene i karakteristikker av avkastning. Mer ikke-normalitet i avkastningen krever mer kompleksitet i distribusjonsantagelsene av innovasjonene. Samlet sett viser studien at de undersøkte GARCH-modellene kan være nyttige i risikomåling av aksjer.

Preface

This thesis concludes my Master of Science in Industrial Mathematics at the Norwegian University of Science and Technology. It is written in the period January to June 2021 and represents a term workload.

I would like to thank my supervisor Jacob Kooter Laading, at the Department of Mathematical Sciences, for great support, encouragement and feedback throughout the work of this thesis. I would also like to thank DNB for providing the data sets used in this thesis.

Catharina Lilleengen
Trondheim, June 2020

Table of Contents

Chapter 1

Introduction

The stock market is one of the financial markets with the highest associated risk, and volatility is a commonly used measure for this risk. Periods of high volatility typically correspond to important events such as stock market crashes, wars, natural disasters or commodity crises. A recent event that emphasises the importance of correctly measuring the risk that market practitioners are exposed to is the sudden market fall due to the pandemic of COVID-19.

By nature, most people are risk averse, and investors would like to predict the risk of their investments in order to avoid losing capital. Investigation of the volatility structure in financial markets, and in particular the stock exchange market, has therefore been a highly debated topic. In recent years, another factor has been added to this debate. There has been an increased interest in green thinking and sustainability. People seem to be more aware of the environmental consequences of their actions and willing to prioritise accordingly financially. This has been reflected in the stock markets as people are more interested in buying green stocks.

The increased green enthusiasm has caused a rapid growth in the so-called green part of the stock market. This involves an increase in the value of green companies and an increase in the number of green companies. Particularly in the last year or so, there has been a substantial growth in the number of listed companies. This rapid growth has caused large fluctuations in certain parts of the stock market and highlights why continuous volatility modelling is needed.

2020 was a rather unusual year, and investors will remember it for the impact the COVID-19 pandemic had on the market. Global stocks experienced one of the largest declines ever, while new highs were reached by the end of the year [1]. This led to bursts of extreme volatility. Throughout the last decades, there have been numerous periods of financial distress, causing situations with extreme volatility. This has increased the need for better risk management, and as a result, financial institutions need better instruments to control their risk. For instance, the stock market crash in 1987 was the triggering event that led to the development of value-at-risk (VaR) as a risk measure. This was later extended to the conditional value-at-risk, also called the expected shortfall. These risk measures evolved from the concept of looking at percentiles of the distribution of the stock value. Given a confidence level, value-at-risk is the maximum loss on an investment during a specific time period, while the expected shortfall is the average of the losses that exceed the VaR [2].

A central part of measuring risk and obtaining precise risk measures is to generate appropriate volatility forecasts. Volatility modelling can therefore be compared to measuring the risk associated with investing in a given asset, portfolio or market. Unexpected changes in market prices lead to increased market risk. This means that the higher the volatility, the higher the associated level of risk is. Since volatility is not directly observable in the market, it has to be estimated [3]. The simplest measure for volatility is the standard deviation of returns. However, financial returns are known to exhibit certain features, called stylised facts, which are rarely captured by this measure and thus make it insufficient. As reported in the study of Bollerslev, Engle and Nelson, the stylised facts include volatility clustering, asymmetry and leptokurtosis [4]. The study of these features led to the family of Autoregressive Conditional Heteroskedasticity (ARCH) models proposed by Engle [5], which later was extended to the Generalised ARCH (GARCH) models proposed by Bollerslev [6]. Further extensions of these volatility models continue and constitute a whole family of various GARCH models.

This thesis aims to explore how different GARCH models capture the characteristics of financial returns. For this purpose, the models are tested both in-sample and out-of-sample. Backward-looking evaluation methods are used for the in-sample analysis, while the out-of-sample analysis is carried out by forecasting one-day-ahead volatilities and using these estimates in value-at-risk and expected shortfall risk measures. The risk measures are then evaluated with backtesting.

The increased interest and growth in the green part of the stock market has motivated the choice of stocks to investigate. The volatility of the green stocks NEL, Tesla and Tomra are analysed and compared to the volatility of similar non-green stocks. While the criteria for green companies are still quite imprecise, these companies are considered green as they provide products or services that are more sustainable than their competitors. The non-green stocks chosen are Norsk Hydro, General Motors and Aker BP. This thesis intends to investigate whether different GARCH models are needed for modelling volatility in different stock market sectors. The influence of a green versus non-green label is also investigated.

NEL, Tomra, Norsk Hydro and Aker BP are traded on Oslo Boers, while Tesla and General Motors are traded on the Nasdaq stock exchange and New York stock exchange, respectively. The two latter stock exchanges, which are both based in New York, are the two largest stock exchanges in the world by market capitalisation. The volume of traded stocks on these exchanges is therefore substantially larger than in the Norwegian market. The most fundamental concept in finance is the Efficient Market Hypothesis (EMH). In short, the EMH states that the price of an asset should fully reflect all available information in the market and that the prices will always immediately respond to any new information in the market [7]. As the U.S. stock market is more liquid than the Norwegian stock market, the information flow should be faster due to more participants and more frequent trading. In theory, this should lead to less leptokurtosis and more stable behaviour of the U.S. traded stocks.

When GARCH models were introduced for modelling financial time series, the error distribution was assumed to be normal. However, many studies have shown that financial time series are leptokurtic and often skewed [6, 8]. In order to capture the fat tails of return series, Bollerslev introduces using student-t distributed errors [9]. Fernandez and Steel extend this idea by allowing for skewness and propose a skew student-t distribution [10]. The specification of the error distribution is an important aspect in the modelling of volatility, and there exists a great deal of literature studying the distributional assumption. Aparicio and Estrada use time series of daily stock returns for 13 European securities markets to investigate the normal assumption [11]. This is clearly rejected, and they find overall support for the scaled t-distribution.

There also exists a great deal of literature on the GARCH models. Hansen and Lunde use the IBM stock price returns to compare 330 different GARCH models for their ability to forecast one-day-ahead volatility [12]. They report that models allowing for leverage effects could perform better than the standard GARCH(1,1) model, and the APARCH(2,2) model achieves the overall best performance.[1] Their study also documents that Gaussian distributed returns lead to better average model performance than student-t distributed returns. Similarly, the more recently published article "Does anything NOT beat the GARCH(1,1)?" by Ghalanos shows that there are few models that do not beat the standard GARCH(1,1) model with normally distributed errors [13]. He assesses the one-day out-of-sample forecasting performance using the S&P 500 index. The previous studies show neither a consensus in the literature of which GARCH model performs best nor which underlying distributional assumption the returns should have. Both the model choice and the distributional assumption will therefore be investigated for the selected stocks.

The remaining part of this thesis is organised in the following way. Chapter 2 introduces fundamental concepts of time series in addition to financial returns and their characteristics. Chapter 3 addresses the modelling of volatility. This includes an overview of GARCH models as well as

[1] The terminology applied for the GARCH models will be explained in Section 3.1.

methods for parameter estimation and model comparison. Chapter 4 introduces the measures used to assess the risk in the stock market, and Chapter 5 presents how these risk measures are calculated. Methods for backtesting the risk measures are presented in Chapter 6.

Chapter 7 gives preliminary data analyses of the stock return series. Then, Chapter 8 presents and discusses the results. To model the volatility and assess which GARCH model is most suitable for different stock market sectors, several GARCH models are fitted to each historical stock data set. From the most satisfactory models, estimated volatilities are calculated and used in the value-at-risk and expected shortfall risk measures. These risk measures are then evaluated using backtesting. Finally, a conclusion of the reported findings is given in Chapter 9.

Chapter 2

Time Series Analysis and Financial Returns

2.1 Fundamentals of Time Series Analysis

In order to understand time series analysis, it is useful to define a few basic concepts such as stationarity, autocorrelation and white noise. A stochastic process (X_t) is a sequence of random variables with indices t taken from an ordered index set $\mathcal{T}$ and defined on some given probability space. The data are commonly assumed to be equally spaced with a discrete-time observation index $t \in \mathcal{T} = \mathbb{Z}$, and a realisation of the random variable X_t will be denoted x_t. A time series model is then a specification of the joint distribution of (X_t) [14].

Often the specification of the complete distribution of (X_t) is too cumbersome, and it is common to consider the first two moments only. Let $(X_t)_{t\in\mathbb{Z}}$ be a time series with finite variance, i.e. $E(X_t^2) < \infty$. The mean function of (X_t) can then be expressed as

$$\mu_X(t) = E(X_t), \quad t \in \mathbb{Z}, \tag{2.1}$$

whereas the autocovariance function of (X_t) is given by

$$\gamma_X(s,t) = \mathrm{Cov}(X_s, X_t) = E[(X_s - \mu_X(s))(X_t - \mu_X(t))], \quad s,t \in \mathbb{Z}. \tag{2.2}$$

An important class of stochastic processes is stationary processes. The concept of stationarity relies on the fact that the behaviour of the process does not change over time, which means the process is in statistical equilibrium [15]. There are two definitions of stationarity, strict stationarity and weak stationarity [2].

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

$$(X_{t_1}, X_{t_2}, \dots, X_{t_k}) \stackrel{d}{=} (X_{t_1+h}, X_{t_2+h}, \dots, X_{t_k+h})$$

for all $t_1, \dots, t_k$, $h \in \mathbb{Z}$ and for all $k \in \mathbb{N}$. Here d means that the two sequences have equal probability distributions.

The definition states that the joint distribution of (X_t) is invariant under time shift for a strictly stationary time series. This condition is too strong for most applications, and rather than imposing conditions on all moments, we often impose conditions on the two first moments of the time series only [16].

Definition 2.2 (Weak stationarity)**.** *A stochastic process $(X_t)_{t\in\mathbb{Z}}$ is weakly stationary, also called covariance-stationary, if*

(i) $\mu_X(t) = \mu$, $t \in \mathbb{Z}$, *the mean function defined in* (2.1) *is constant and does not depend on* t.

(ii) the autocovariance function, $\gamma_X(s,t)$, *defined in* (2.2) *depends on* t *and* s *only through their difference* $h = |t-s|$.

It is important to note that for a strictly stationary process, finite second moments are not assumed. This means that strict stationarity does not imply weak stationarity unless the process has finite variance [17]. Hence, contrary to what might be expected, the weak stationarity restriction is more limiting than that of strict stationarity [18].

The second requirement in Definition 2.2 indicates that $\gamma(t-s,0) = \gamma(t,s) = \gamma(s,t) = \gamma(s-t,0)$, which means that the autocovariance function $\gamma(s,t)$ for a weakly stationary time series depends on s and t only through the lag $h = |s-t|$. Letting $s = t+h$, the notation of the autocovariance function simplifies. In other words, it can be written as a function of only one variable [19]. Hence,

$$\gamma_X(h) := \gamma_X(h,0) = \gamma_X(t+h,t), \quad \forall h \in \mathbb{Z}.$$

Observing that $\gamma(0) = \mathrm{Var}(X_t)$ for all t, it is possible to define the autocorrelation function of a weakly stationary time series [2].

Definition 2.3 (Autocorrelation function)**.** *The autocorrelation function (ACF) of a weakly stationary process $(X_t)_{t\in\mathbb{Z}}$ is the collection of autocorrelations given by*

$$\rho_X(h) = \frac{Cov(X_{t+h}, X_t)}{\sqrt{Var(X_{t+h})\, Var(X_t)}} = \frac{\gamma_X(h)}{\gamma_X(0)}, \quad \forall h \in \mathbb{Z}, \tag{2.3}$$

where $Var(X_{t+h}) = Var(X_t)$ because X_t is weakly stationary.

Investigating the ACF is an essential part of exploring the predictability and analysing time series data. For this purpose, we are interested in the autocorrelation, also called the serial correlation, $\rho_X(h)$, at lag h. This requires the sample ACF, $\hat{\rho}_X(h)$. The sample autocovariances, $\hat{\gamma}_X(h)$, of the covariance-stationary time series $(X_t)_{t\in\mathbb{Z}}$ can be calculated as

$$\hat{\gamma}_X(h) = \frac{1}{n}\sum_{t=1}^{n-h}(x_{t+h} - \bar{x})(x_t - \bar{x}),$$

where $\bar{x} = 1/n\sum_{t=1}^{n} x_t$ is the estimated sample mean and n is the number of observations [20]. Equivalent to equation (2.3), the sample ACF can be expressed as

$$\hat{\rho}_X(h) = \frac{\hat{\gamma}_X(h)}{\hat{\gamma}_X(0)}. \tag{2.4}$$

The sample ACF is used to assess whether the correlations are statistically significant at some lags, which means testing the null hypothesis that an autocorrelation coefficient is zero. The sample ACF, $\hat{\rho}(h)$, is asymptotically $N(0,1/n)$ as $n \to \infty$ for any $h \geq 1$, and the $\hat{\rho}(j)$'s are asymptotically independent [21]. Based on this, an approximate method for assessing the significance is to observe if the sample autocorrelations are outside the interval $\pm 2/\sqrt{n}$, which corresponds to a 95% confidence interval. If more than 5% of the sample ACFs are outside the interval, the null hypothesis is rejected and the autocorrelation is significant [16].

In time series analysis, more complicated processes are built up from simple ones. The simplest building block is the white noise process, which is a process with zero autocorrelation [14]. White noise is also the simplest example of a weakly stationary process.

Definition 2.4 (White noise)**.** *A stochastic process $(X_t)_{t\in\mathbb{Z}}$ is said to be white noise if it has a constant mean, constant variance and zero autocovariance at all lags. That is,*

(i) $E(X_t) = \mu$

(ii) $\gamma_X(h) = \begin{cases} \sigma^2, & h = 0 \\ 0, & h \neq 0. \end{cases}$

Usually, the mean μ is taken to be zero, and the white noise process will then be denoted $\mathrm{WN}(0, \sigma^2)$. The definition of the white noise process implies that it is weakly stationary and temporally uncorrelated. This means that the ACF will always be equal to zero, except for $h = 0$, where it is one. In addition to being temporally uncorrelated, if the process is independently and identically distributed (i.i.d.), it is a strict white noise process. This will be denoted $(X_t) \sim \mathrm{SWN}(0, \sigma^2)$. A strict white noise process is also strictly stationary [14].

Another useful concept of noise, which is common when assessing volatility and GARCH models, is a martingale-difference process. We assume that the time series $(X_t)_{t\in\mathbb{Z}}$ is adapted to some filtration $(\mathcal{F}_t)_{t\in\mathbb{Z}}$, or in the case of a discrete process, it is often referred to as the information set [22]. This contains the history of the process up to and including time t and is denoted by $\mathcal{F}_t = \sigma(\{X_s : s \leq t\})$ [2].

Definition 2.5 (Martingale difference)**.** *A time series $(X_t)_{t\in\mathbb{Z}}$ is known as a martingale-difference sequence with respect to the filtration $(\mathcal{F}_t)_{t\in\mathbb{Z}}$ if*

- $E(|X_t|) < \infty$,
- X_t *is* $\mathcal{F}_t$*-measurable,*
- $E[X_t|\mathcal{F}_{t-1}] = 0, \quad \forall t \in \mathbb{Z}.$

The latter property saying that given the current information, the expectation of the next value is always zero is what makes the process applicable to financial time series. The unconditional mean of a martingale difference is also zero as $E[X_t] = E[E(X_t|\mathcal{F}_{t-1})] = 0$, for all $t \in \mathbb{Z}$, and the sequence is uncorrelated over time. If the unconditional variance of the martingale difference is constant over time, then the series is also white noise [23]. The martingale difference sequence is thus a generalised white noise process.

2.1.1 Stationarity of Financial Time Series Data

Modelling time series data requires the assumption of a stationary process. An essential characteristic of a stationary process is mean reversion, meaning that the process will revert towards some fixed level mean in the long run. However, observed asset prices are generally non-stationary and therefore have to be transformed into a stationary process [20].

In order to convert a non-stationary time series into a stationary one, a technique called differencing can be applied. First-differencing means that the series is lagged one step and subtracted from the original series [24]. A plot of the time series can then be examined to check for the stationary properties given in the previous section. Further differencing can be applied if these are not present yet. Moreover, financial time series are usually exposed to exponential growth, so a logarithmic transformation is often applied to stabilise the variance.

By taking the first difference and applying the logarithmic transformation to the asset prices, the series of continuously compounded returns, the so-called log returns, can be computed [3].

Definition 2.6 (Log returns)**.** *Let S_t denote the price of a financial asset at time t. Then the logarithmic return, also called the continuously compounded return, Y_t, is given by*

$$Y_t = \log S_t - \log S_{t-1} = \log\left(\frac{S_t}{S_{t-1}}\right). \tag{2.5}$$

The log returns will hereby be referred to as the returns.

2.2 Return Characteristics

Although asset prices seem to behave randomly, their variations have some statistical properties in common. In particular, financial return time series tend to exhibit similar structures, which in financial terms are referred to as stylised facts. The most prominent features include volatility clustering, asymmetry and leptokurtosis, where the presence of the two latter leads to non-normality of returns. Two other common characteristics of financial returns are a mean close to zero and a standard deviation that tends to be substantially larger than the mean when returns are measured on a daily horizon [25].

As daily returns have very little autocorrelation, they are close to impossible to predict from their past. Squared returns, on the other hand, usually show a positive correlation with their past. The same applies to their square root, the absolute returns, which tend to be more strongly serially correlated than squared returns. These correlations indicate the presence of volatility clustering, which again implies that squared returns and absolute returns are continually changing in a partly predictable manner. Both squared returns and absolute returns are often used as proxies to model volatility. The correlations suggest that large returns tend to be followed by large returns and vice versa. In other words, if there are observed large returns the last few days, the distribution for which the next day return is taken from most likely has a large variance as well. This can be used to predict future volatility and is the idea that most volatility models are based on [25].

Already in the 1960s, Mandelbrot and Fama found evidence that daily returns tend to be leptokurtic, i.e. fat-tailed [26, 8]. This means that returns tend to have more outcomes of small and large values than suggested by the normal distribution. Closely related is another stylised fact. As volatility reacts differently to good and bad news in the market, it tends to have asymmetric responses, meaning that most returns have a skewed distribution [27]. Together these two stylised facts imply that returns are non-normal.

2.2.1 Identification of Return Characteristics

To identify the characteristics of return series, both graphical and statistical methods can be applied. A standard graphical method used to identify volatility clustering is to assess the autocorrelation function. This means checking for correlation between squared returns or absolute returns [20]. If the correlations are significant, this indicates volatility clustering in the data. Volatility clustering can also be identified using statistical methods such as the Ljung-Box test. This tests the joint significance of autocorrelation coefficients over several lags. The Ljung-Box test statistic is given as [19]

$$Q = n(n+2)\sum_{j=1}^{m}\frac{\hat{\rho}(j)^2}{n-j},$$

where $\hat{\rho}(j)$ is the sample ACF from equation (2.4), and as before, n is the number of observations. The number of lags included is denoted by m and is typically equal to 20 [16]. Under the null hypothesis of no autocorrelation, the test statistic is asymptotically chi-square distributed with m degrees of freedom. The null hypothesis is rejected if $Q > \chi^2_{1-q,m}$, or equivalently if the p-value is smaller than the significance level q.

Two prominent features of return characteristics are leptokurtosis and skewness. A distribution is called leptokurtic if the kurtosis exceeds three, which means the distribution has fatter tails than the normal distribution. An asymmetric distribution has a skewness different from zero. Negative skewness indicates a long left tail compared to the right tail, and positive skewness vice versa. A simple and intuitive test to identify the non-normality of returns is to check whether the kurtosis exceeds three and the skewness is different from zero. A more formal approach is to test whether the skewness and excess kurtosis are significantly different from zero. This could be performed with the Jarque-Bera test [28]. The test statistic is given by

$$JB = \frac{n}{6}\left(\hat{S}^2 + \frac{(\hat{K}-3)^2}{4}\right),$$

where $\hat{S}$ is the sample skewness and $\hat{K}$ is the sample kurtosis. Under the null hypothesis of normally distributed data, i.e. zero skewness and excess kurtosis, the test statistic is asymptotically chi-square distributed with two degrees of freedom. Normality is rejected if the p-value of the JB statistic is less than the significance level q.

For graphically analysing fat tails and non-normality, a qq-plot can be used. Qq-plots are used to evaluate whether a data set has a particular distribution or not. It compares the theoretical quantiles of a reference distribution to the empirical quantiles of the sample data. The qq-plot is linear if the reference distribution and empirical distribution are approximately the same. Deviation from linearity in the tails of the qq-plot indicates that the empirical distribution has fatter or thinner tails than the reference distribution. In particular, a qq-plot with an inverted S-shape indicates heavier tails for the empirical distribution than for the reference distribution [2]. Plotting a histogram of the observed returns together with a theoretical distribution can also be used to identify the distribution of the returns.

Chapter 3

Volatility modelling

In economic and financial time series, time-varying volatility, including volatility clustering, is more common than constant volatility. As volatility is considered a measure of randomness, or more specifically a measure of asset risk, modelling the randomly varying volatility is noteworthy for practitioners in the financial market. Volatility modelling can be seen as the next step in the development of more accurate modelling of asset and portfolio returns. It could also be helpful in applications such as risk management of portfolios or markets, option pricing or hedging strategies.

However, modelling volatility may be a demanding task. Volatility is not directly observable in the market and has to be estimated using observed market prices. As an example, the return after a trading day is known. However, all that can be said about the risk of that day is that if prices have had large fluctuations from that day to the next, then the risk is probably high. That is why observed market prices are used to infer the volatility [3]. A frequently applied family of volatility models is the Generalised Autoregressive Conditional Heteroskedasticity (GARCH) models. They constitute a complex framework of models which are able to cope with the stylised facts that apply for time series of returns [3]. As the name indicates, a GARCH process is autoregressive because it depends on the past values of the process. Moreover, it is conditionally heteroskedastic, which means that the conditional variance of the process is time-varying [22]. This model feature is developed to reflect the volatility clustering of financial returns.

To better understand volatility models, it is helpful to clarify the distinction between unconditional and conditional volatility. The unconditional volatility, denoted σ, is simply the standard deviation of the returns from a given period in the past. The volatility is then assumed to be constant over the entire time period considered. The conditional volatility, denoted σ_t, is defined as the volatility in a given time period conditional on what happened before [3]. As it depends on the history of the process up to that point, it will change at every point in time [22]. For instance, a forecast of the volatility at time t is typically obtained using a sample of historical observations y_t during an estimation window of length K. This can be expressed as

$$\sigma_t = \sigma(y_{t-1}, y_{t-2}, \dots, y_{t-K}),$$

where $\sigma(\cdot)$ is the method to be specified. As the forecasting horizon increases, conditional volatility models usually regress towards the unconditional volatility. This model feature is coherent with the fact that return series are usually mean reverting. However, to ensure that the unconditional volatility is defined, stationarity conditions must be imposed on the volatility models [3].

The GARCH models are in the category of conditional volatility models. The return on day t is usually modelled by

$$Y_t = \mu_t + \sigma_t Z_t, \quad t \in \mathbb{Z}. \tag{3.1}$$

If the first two conditional moments of Y_t exist, they can be expressed as

$$E[Y_t|\mathcal{F}_{t-1}] = \mu_t, \quad \mathrm{Var}[Y_t|\mathcal{F}_{t-1}] = E[Y_t^2|\mathcal{F}_{t-1}] = \sigma_t^2,$$

where $\mathcal{F}_{t-1}$ denotes the information set generated by the history of the process up to time $t-1$. Hence, μ_t is the conditional mean and σ_t^2 is the conditional variance of the process. The stochastic process (σ_t) is measurable with respect to the information set $\mathcal{F}_{t-1}$. Moreover, (Z_t) forms an i.i.d. sequence of real-valued innovations with zero mean and unit variance, i.e. $Z_t \sim$ i.i.d. $D(0,1)$ [25]. Z_t are assumed to be independent of $\mathcal{F}_{t-1}$ and $\sigma(\{Y_s : s < t\})$ [18]. Often the innovations are assumed to be normally distributed, but distributions that can capture excess kurtosis or skewness are also an option. In this thesis, the normal distribution and the ordinary and skew versions of the student-t distribution will be considered.

For return series where the mean is close to zero, the conditional mean is usually assumed to be constant and equal to zero [25]. The model in equation (3.1) then reduces to

$$Y_t = \sigma_t Z_t, \quad t \in \mathbb{Z}. \tag{3.2}$$

The data analysis presented in Chapter 7 shows that this applies to our data sets.

3.1 GARCH Models

Financial returns exhibit volatility clustering, which means that the newest observations give a better indication of whether we are in a period with high or low volatility. These observations should therefore count more than the older ones in modelling of volatility. GARCH models are based on using an optimal exponential weighting of historical observations to obtain volatility forecasts. The return on day t depends on the returns of previous days, where the recent returns typically have a higher weight than the older ones [3]. GARCH models are flexible and capture essential features of returns such as volatility clustering and leptokurtosis. Some extensions of the standard GARCH model are also able to capture asymmetry in the returns. As we will see in this section, the conditional volatility, σ_t, for a GARCH model depends on model parameters and recent return observations, whereas the unconditional volatility, σ, depends on the entire sample [3].

3.1.1 ARCH

The most elementary family of GARCH models is the Autoregressive Conditional Heteroscedasticity (ARCH) family proposed by Engle [5]. The model definition is taken from McNeil [2].

Definition 3.1 (ARCH(p) model)**.** *Let $(Z_t)_{t\in\mathbb{Z}}$ be SWN(0,1). The process $(Y_t)_{t\in\mathbb{Z}}$ is an ARCH(p) process if it is strictly stationary and if it satisfies, for all $t \in \mathbb{Z}$ and some strictly positive-valued process $(\sigma_t)_{t\in\mathbb{Z}}$, the following equations,*

$$\begin{aligned} Y_t &= \sigma_t Z_t, \\ \sigma_t^2 &= a_0 + \sum_{i=1}^{p} a_i Y_{t-i}^2, \end{aligned} \tag{3.3}$$

where $a_0 > 0$, $a_i \geq 0$ for $i = 1, \dots, p$ and p is the number of lags.

The returns are serially uncorrelated but dependent, and the dependence of Y_t on the past can be described by a quadratic function of its lagged values. The model is therefore able to capture the volatility clustering observed in financial time series. The parameters a_i need to be estimated using a numerical estimation routine such as maximum likelihood. In order to ensure a well-defined model and the conditional variance to be positive, the constraints $a_0 > 0$, $a_i \geq 0$ for $i = 1, \dots, p$, and $p \in \mathbb{N}$ are imposed.

Equation (3.3) guarantees that σ_t is $\mathcal{F}_{t-1}$-measurable. Provided that $E(|X_t|) < \infty$, we have

$$E[Y_t|\mathcal{F}_{t-1}] = E[\sigma_t Z_t|\mathcal{F}_{t-1}] = \sigma_t E[Z_t|\mathcal{F}_{t-1}] = \sigma_t E(Z_t) = 0, \tag{3.4}$$

since σ_t and Z_t are independent. This means that the ARCH process has the martingale-difference property with respect to the filtration $(\mathcal{F}_t)_{t\in\mathbb{Z}}$. Additionally, if the ARCH process is covariance-stationary, it is simply a white noise process [2].

For the ARCH process to be covariance-stationary, the unconditional variance, σ^2, must be constant over time. The derivation of the unconditional variance for an ARCH(1) model will now be elaborated, but the same procedure applies for higher orders [29]. Using the result in equation (3.4), the unconditional mean of Y_t remains zero as

$$E(Y_t) = E[E(Y_t|\mathcal{F}_{t-1})] = 0.$$

Moreover, the unconditional variance of Y_t can be achieved by

$$\begin{aligned}\mathrm{Var}(Y_t) &= E(Y_t^2) = E[E(Y_t^2|\mathcal{F}_{t-1})] = E(\sigma_t^2)\\ &= E(a_0 + a_1 Y_{t-1}^2) = a_0 + a_1 E(Y_{t-1}^2),\end{aligned}$$

since Y_t is a stationary process with $E(Y_t) = 0$ and $\mathrm{Var}(Y_t) = \mathrm{Var}(Y_{t-1}) = E(Y_{t-1}^2)$. The unconditional variance for an ARCH(1) model is thus constant and given by $\mathrm{Var}(Y_t) = a_0/(1 - a_1)$. Similarly, the unconditional variance of an ARCH(p) model is given by

$$\sigma^2 = \frac{a_0}{1 - \sum_{i=1}^{p} a_i}.$$

To ensure that the unconditional variance is defined, the restriction $a_1 < 1$ needs to be imposed for an ARCH(1) model. Equivalently, the restriction for an ARCH(p) model is $\sum_{i=1}^{p} a_i < 1$. Hence, with this condition satisfied, as (Y_t) is a martingale difference with finite time-independent variance, it is a white noise process.

3.1.2 GARCH

An extension of the ARCH process, called the generalised ARCH or GARCH process, is proposed by Bollerslev [6]. A large persistence in volatility leads to a large p in the ARCH(p) models, which means it is more parsimonious to use a GARCH model [2].

Definition 3.2 (GARCH(p, q) model)**.** *Let $(Z_t)_{t\in\mathbb{Z}}$ be SWN(0,1). The process $(X_t)_{t\in\mathbb{Z}}$ is a GARCH(p, q) process if it is strictly stationary and if it satisfies, for all $t \in \mathbb{Z}$ and some strictly positive-valued process $(\sigma_t)_{t\in\mathbb{Z}}$, the following equations,*

$$\begin{aligned}Y_t &= \sigma_t Z_t,\\ \sigma_t^2 &= a_0 + \sum_{i=1}^{p} a_i Y_{t-i}^2 + \sum_{j=1}^{q} b_j \sigma_{t-j}^2,\end{aligned} \tag{3.5}$$

where $a_0 > 0$, $a_i \geq 0$, for $i = 1, \ldots, p$, and $b_j \geq 0$ for $j = 1, \ldots, q$.

For the standard GARCH model, the conditional variance is a linear function of its own lagged conditional variance terms in addition to the lagged squared observations. This captures the feature of volatility clustering more efficiently than for the ARCH process. The constraints $a_0 > 0$, $a_i \geq 0$ for $i = 1, \ldots, p$ and $b_j \geq 0$ for $j = 1, \ldots, q$ are imposed in the parameter estimation to ensure non-negativity in the conditional volatility forecasts.

Low-order GARCH models are preferred in practice to avoid overfitting data. The most frequently used model is the GARCH(1,1) model given by

$$\sigma_t^2 = a_0 + a_1 Y_{t-1}^2 + b_1 \sigma_{t-1}^2. \tag{3.6}$$

Here the parameters a_0, a_1 and b_1 need to be estimated. The model has only three parameters but can be rewritten to an ARCH(∞) model. This demonstrates how GARCH models are more parsimonious than ARCH models and why they are often employed in practice.

The process (Y_t) is covariance-stationary if the unconditional variance exists. Using the same approach as for the ARCH models, the unconditional variance for the GARCH(1,1) model is given by

$$\sigma^2 = \frac{a_0}{1 - a_1 - b_1}.$$

Its existence requires the restriction $a_1 + b_1 < 1$ to be satisfied. Equivalent restrictions exist for the GARCH(p, q) models. If the coefficients imply non-stationarity in the variance, the GARCH models could have some undesirable properties. For instance, if $a_1 + b_1 > 1$, then a conditional variance forecast will tend to infinity as the forecast horizon increases. On the other hand, the conditional variance will revert towards a fixed mean for a stationary model [30]. These covariance-stationary restrictions will therefore be imposed in the model fitting.

There exist various extensions of the standard GARCH model, but many of them have seen limited application. Thus, in practice, two groups of extensions are used occasionally. These models account for leverage effects and power effects [3].

When positive and negative news in the market have the same effect on volatility, the returns are said to have a symmetric distribution. Classical ARCH and GARCH models work well with this assumption. However, as mentioned, the returns usually have a skewed distribution. Volatility seems to rise higher in response to negative shocks as opposed to positive shocks. This phenomenon is called the leverage effect. In order to capture this feature, it might be advantageous to use asymmetric GARCH models [31].

By studying the ACF of absolute returns and squared returns, absolute returns sometimes have stronger autocorrelations than squared returns. This feature is called the power effect. Allowing power in the volatility calculation to be flexible makes the model able to capture this feature [3].

3.1.3 Exponential GARCH

For the standard GARCH(p, q) model, equation (3.5) shows that σ_t^2 is a function of past squared values of Y_t. As the sign of the returns Y_t is not considered, the GARCH process cannot model the asymmetric effects between positive and negative returns. In other words, these models cannot capture leverage effects. The problem lies in the symmetric function x^2, and the solution is therefore to replace this with a flexible class of non-negative functions that can be asymmetric [20]. A model having this feature is the exponential GARCH (EGARCH) model developed by Nelson [32]. Our model specification is taken from Francq and Zakoian [18].

Definition 3.3 (EGARCH(p, q) model)**.** *Let $(Z_t)_{t\in\mathbb{Z}}$ be SWN(0,1). The process $(X_t)_{t\in\mathbb{Z}}$ is an EGARCH(p, q) process if it satisfies the following equations,*

$$\begin{aligned} Y_t &= \sigma_t Z_t, \\ \log \sigma_t^2 &= a_0 + \sum_{i=1}^{p} \Big(a_i Z_{t-i} + \gamma_i \big(|Z_{t-i}| - E(|Z_{t-i}|) \big) \Big) + \sum_{j=1}^{q} b_j \log(\sigma_{t-j}^2), \end{aligned} \tag{3.7}$$

where a_i and γ_i for $i = 1, \dots, p$ and b_j for $j = 1, \dots, q$ are real numbers.

Since it is the logarithm of the conditional variance being modelled, σ_t^2 will always be positive regardless of the signs of the parameters a_i and b_j. Compared to the standard GARCH model, the EGARCH model has the advantage of no restrictions on these parameters to ensure non-negative volatility forecasts. However, the constraints $-1 < \gamma_i < 1$, for $i = 1, \dots, p$, are imposed in parameter estimation. If $\gamma_i > 0$, there is a leverage effect and past negative shocks have a stronger impact on volatility than past positive shocks. On the contrary, if $\gamma_i < 0$, there is a leverage effect in the opposite direction.

3.1.4 Asymmetric Power ARCH

Ding, Granger and Engle propose the asymmetric power ARCH (APARCH) model [33]. In addition to capturing leverage effects, it models σ_t^δ, where the parameter δ makes it able to capture power effects as well. Our model specification is obtained from Francq and Zakoian [18].

Definition 3.4 (APARCH(p,q) model)**.** *Let $(Z_t)_{t\in\mathbb{Z}}$ be SWN(0,1). The process $(X_t)_{t\in\mathbb{Z}}$ is an APARCH(p,q) process if it satisfies the following equations,*

$$\begin{aligned} Y_t &= \sigma_t Z_t, \\ \sigma_t^\delta &= a_0 + \sum_{i=1}^{p} a_i(|Y_{t-i}| - \gamma_i Y_{t-i})^\delta + \sum_{j=1}^{q} b_j \sigma_{t-j}^\delta, \end{aligned} \tag{3.8}$$

where $a_0 > 0$, $\delta > 0$, $a_i \geq 0$, $-1 < \gamma_i < 1$ for $i = 1, \ldots, p$ and $b_j \geq 0$ for $j = 1, \ldots, q$.

The leverage coefficients γ_i have the same interpretation as for the EGARCH model. If $\gamma_i \neq 0$, the model fits leverage effects, and if $\delta \neq 2$, the model fits power effects. An interesting fact to note about the APARCH model is that it nests several other models. For instance, if $\delta = 2$ and $\gamma_i = 0$, it leads to the standard GARCH process, and additionally, if $b_j = 0$ it leads to the ARCH process.

3.2 Estimation of GARCH Models

The marginal distribution of the i.i.d. innovations Z_t with mean zero and unit variance is often assumed to be the standard normal distribution, i.e. $Z_t \sim N(0,1)$. However, returns tend to have probability distributions with heavier tails than suggested by a normal GARCH model [3]. Even though normal GARCH models are able to accommodate for some of the non-normality in daily returns, it is often observed that they cannot capture it all. This means that normally distributed innovations could lead to underestimation.

The most dominant deviations from normality are assumed to be the fatter tails and the larger peaks around the centre of the distribution. To accommodate for some of the heavy-tailed structure in financial returns, Bollerslev introduces the student-t distribution [9]. This distribution has an additional parameter ν for the degrees of freedom, which controls the fatness of the tails. The lower the parameter ν, the fatter the tails become, and as ν goes to infinity, the distribution converges to the standard normal distribution [25]. Assuming that the innovations have the standardised student-t distribution, we can write $Z_t \sim t_\nu^*(0,1)$, where t_ν^* denotes the standardised t-distribution with ν degrees of freedom. This is equivalent to letting $Z_t = T_\nu/\sqrt{\nu/(\nu-2)}$, where T_ν is a student-t distributed variable with $\nu > 2$ degrees of freedom [21]. The scale factor multiplied by T_ν is to ensure a unit variance of Z_t, and the restriction $\nu > 2$ needs to be imposed to ensure a finite variance.

The distribution of financial returns is also known to have skewness. To account for this, Fernandez and Steel propose an extension of the student-t distribution by adding a skewness parameter [10]. Lambert and Laurent extend the work of Fernandez and Steel to the GARCH framework by reparametrizing the density as a function of the conditional mean and the conditional variance, such that the innovations have a zero mean and unit variance [34]. This means $Z_t \sim st_{\nu,\xi}^*(0,1)$, where $st_{\nu,\xi}^*$ denotes the standardised skew student-t distribution, ν the degrees of freedom and ξ the asymmetry parameter.

3.2.1 Maximum Likelihood Estimation

As there is presence of non-linearity in the models, linear regression models such as ordinary least squares cannot be used for model fitting. Instead, estimation of parameters is usually accomplished by conditional maximum likelihood. The idea behind maximum likelihood estimation (MLE) is to construct a likelihood function based on the data sample and the assumed distribution of the data sample. The parameters are then obtained by maximising the log-likelihood function.

The joint density of an i.i.d. sample $x = (x_1, \ldots, x_n)$ is the product of the density function $f(\cdot)$ at each observation x_t. To determine the parameters given our data sample, we consider the joint likelihood function

$$L(\theta|x) = \prod_{t=1}^{n} f(x_t;\theta), \tag{3.9}$$

where the density function and variable x have parameters θ. In practice, it is easier to work with a sum rather than the product of the densities. This is achieved by taking the natural logarithm of the likelihood function. The logarithm is a monotonic increasing function such that the values that maximise the likelihood will be the same as those that maximise the log-likelihood function [35]. The maximum likelihood estimator of θ is then obtained by maximising the log-likelihood function given by

$$\log L(\theta|x) = \sum_{t=1}^{n} \log f(x_t;\theta).$$

Normal distribution

With the assumption of normally distributed innovations, the model in equation (3.2) is given by

$$Y_t = \sigma_t Z_t, \quad Z_t \sim \text{i.i.d. } N(0,1).$$

The standardised normal distribution is defined as

$$f_{n^*}(z_t;\theta) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z_t^2}{2}\right). \tag{3.10}$$

Conditionally on the information set $\mathcal{F}_{t-1}$, the scale-family of the normal distribution, f_n, with zero mean and time-dependent variance σ_t can be written as [18]

$$f_n(y_t;\theta) = \frac{1}{\sigma_t} f_{n^*}\left(\frac{y_t}{\sigma_t};\theta\right) = \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{y_t^2}{2\sigma_t^2}\right). \tag{3.11}$$

Inserting this into the joint likelihood function (3.9) yields

$$L(\theta|y) = \prod_{t=1}^{n} f_n(y_t;\theta) = \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{y_t^2}{2\sigma_t^2}\right).$$

The log-likelihood function can then be expressed as

$$\log L(\theta|y) = \sum_{t=1}^{n}\left[-\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma_t^2) - \frac{1}{2}\frac{y_t^2}{\sigma_t^2}\right] = -\frac{1}{2}\sum_{t=1}^{n}\left[\log(\sigma_t^2) + \frac{y_t^2}{\sigma_t^2}\right].$$

In the last equality we have left out the constant as this includes no parameters and will have no impact on the estimation.

In order to obtain the likelihood functions for the different GARCH models, the associated expressions of the conditional variance are replaced by σ_t^2. The conditional variance equations for the standard GARCH, EGARCH and APARCH models are given in equation (3.5), (3.7) and (3.8), respectively. For example, for the GARCH(1,1) model given in equation (3.6) with normally distributed innovations, the log-likelihood function is given by

$$\log L(\theta|y) = -\frac{1}{2}\sum_{t=1}^{n}\left[\log(a_0 + a_1 y_{t-1}^2 + b_1\hat{\sigma}_{t-1}^2) + \frac{y_t^2}{a_0 + a_1 y_{t-1}^2 + b_1\hat{\sigma}_{t-1}^2}\right].$$

After replacing σ_t^2, the likelihood function is maximised to obtain the parameters in the conditional variance equation in addition to any parameters in the density function. For instance, for a GARCH(p,q) model with normally distributed innovations, the parameter vector is $\theta = (a_0, a_1, \ldots, a_p, b_1, \ldots b_q)$. The EGARCH and APARCH models both need additional parameters γ_i, and for the APARCH model, the parameter δ has to be included as well. As the normal density does not require any parameter estimation, no additional parameters are included in the θ vector. As we will see, this is not the case for the t-distributions.

Student-t distribution

For GARCH models with an assumption of standardised student-t distributed innovations, we have the model

$$Y_t = \sigma_t Z_t, \quad Z_t \sim \text{i.i.d. } t_\nu^*(0,1),$$

where as before t_ν^* denotes the standardised student-t distribution with ν degrees of freedom. This density function is given by

$$f_{t^*}(z_t;\theta) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma(\frac{\nu}{2})\sqrt{\pi\,(\nu-2)}}\left(1+\frac{z_t^2}{(\nu-2)}\right)^{-(1+\nu)/2}, \quad \text{for } \nu > 2,$$

where $\Gamma(\cdot)$ is the usual gamma function. The distribution is defined such that the random variable z_t has a mean equal to zero and variance equal to one.

Following the same procedure as for the normal distribution, the scale-family of the student-t distribution is given by

$$f_t(y_t) = \frac{1}{\sigma_t} f_{t^*}\left(\frac{y_t}{\sigma_t};\theta\right) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sigma_t\Gamma(\frac{\nu}{2})\sqrt{\pi\,(\nu-2)}}\left(1+\frac{y_t^2}{\sigma_t^2(\nu-2)}\right)^{-(1+\nu)/2},$$

which gives the log-likelihood function

$$\begin{aligned}\log L(\theta|y) =& n\left[\log\Gamma\left(\frac{\nu+1}{2}\right) - \log\Gamma\left(\frac{\nu}{2}\right) - \frac{1}{2}\log\left[\pi(\nu-2)\right]\right] \\ &- \frac{1}{2}\sum_{t=1}^{n}\left[\log(\sigma_t^2) + (\nu+1)\log\left(1+\frac{1}{\nu-2}\frac{y_t^2}{\sigma_t^2}\right)\right].\end{aligned}$$

Estimation of the student-t density requires the estimation of the parameter ν. The θ vector to be maximised therefore includes this parameter as well.

Skew Student-t distribution

With the assumption of standardised skew student-t distributed innovations, we have the model

$$Y_t = \sigma_t Z_t, \quad Z_t \sim \text{i.i.d. } st_{\nu,\xi}^*(0,1),$$

where $st_{\nu,\xi}^*$ denotes the standardised skew student-t distribution with mean zero and unit variance. Its density is given by

$$f_{st^*}(z_t;\theta) = \begin{cases} \frac{2}{\xi+\frac{1}{\xi}} s\left[g_\nu(\xi(sz_t+m))\right] & \text{if } z_t < -\frac{m}{s} \\ \frac{2}{\xi+\frac{1}{\xi}} s\left[g_\nu((sz_t+m)/\xi)\right] & \text{if } z_t \geq -\frac{m}{s}, \end{cases}$$

$$m = \frac{\Gamma\left(\frac{\nu-1}{2}\right)\sqrt{\nu-2}}{\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)}\left(\xi-\frac{1}{\xi}\right) \quad \text{and} \quad s^2 = \left(\xi^2+\frac{1}{\xi^2}-1\right)-m^2,$$

where $\nu > 2$, $\xi > 0$ and $g_\nu(\cdot)$ denotes the student-t density function. The parameters m and s^2 are the mean and variance of the non-standardised skew student-t distribution. ξ is the asymmetry parameter and is defined such that the probability masses above and below the mean are given by $\frac{P(z\geq 0|\xi)}{P(z<0|\xi)} = \xi^2$. The distribution is equal to the symmetric student-t density when ξ is equal to one.

The log-likelihood function for the skew student-t distribution is derived by Laurent and Lambert and can be expressed as [34]

$$\begin{aligned}\log L(\theta|y) =& n\left[\log\Gamma\left(\frac{\nu+1}{2}\right) - \log\Gamma\left(\frac{\nu}{2}\right) - \frac{1}{2}\log\left[\pi(\nu-2)\right] + \log\left(\frac{2}{\xi+\frac{1}{\xi}}\right) + \log(s)\right] \\ &- \frac{1}{2}\sum_{t=1}^{n}\left[\log(\sigma_t^2) + (\nu+1)\log\left[1+\left(1+\frac{(sz_t+m)^2}{\nu-2}\right)\xi^{-2I_t}\right]\right],\end{aligned}$$

$$I_t = \begin{cases} 1 & \text{if } z_t \geq -\frac{m}{s} \\ -1 & \text{if } z_t < -\frac{m}{s}, \end{cases}$$

where $z_t = y_t/\sigma_t$. As the skew student-t density requires estimating the parameters ν and ξ, these are included in the θ vector.

Estimation

From the log-likelihood functions given above, it can be observed that a distributional assumption other than the normal produces more complicated expressions. Thus, except for the simplest cases, it is not possible to obtain an analytical solution to the maximisation problem. In other words, we cannot find a closed-form solution of the ML estimators. The log-likelihood function therefore needs to be numerically maximised by computer algorithms called solvers. In practice, solvers typically use iterative algorithms to obtain parameter estimates by minimising the negative log-likelihood function [3]. This is the equivalent optimisation problem to maximising the log-likelihood function.

However, it could be a time-consuming and challenging task to obtain numerically stable results. For the GARCH(1,1) model, this usually works fine, but when the complexity of the model increases, this is not always the case. As an example, the APARCH model is usually harder to fit. Some standard techniques to reduce estimation problems are setting random starting values, increasing the data set, or trying a different solver. As for the APARCH model, it is necessary to have a large data sample for estimation, and in general, GARCH models should have an estimation sample of at least 1000 observations [25].

For fitting the volatility models, the open-source software package R is utilised. In particular, the "rugarch" package developed by Ghalanos is used [36]. This is a flexible package for univariate GARCH modelling, and it allows for the specification of the solver used in estimation. Each solver has its advantages and disadvantages, such that a model could potentially converge with one solver and not converge with another. The option to specify the solver "hybrid" is a good choice as this rotates among the available solvers until the model converges. As an example, "solnp" is the first solver that R tries. This uses the augmented Lagrange multiplier method [37].[2]

3.3 Model Comparison

The selection of lag parameters, distributional assumption and the type of GARCH model gives numerous available models. Some criteria to select the best one are therefore required. Goodness-of-fit tests are in-sample tests that measure the models' ability to capture the characteristics of the estimation sample. Post-sample prediction tests, also called out-of-sample tests, evaluate the accuracy of the predictions provided by the models. The choice of model should depend on the intended use, and in this study, the aim is to produce reasonable forecasts to understand market risk [22]. The model selection should therefore be based on the out-of-sample forecasting performance. However, given the nearly infinite number of available models, the in-sample performance is used as a filter to reduce the number of models in the out-of-sample analysis. Additionally, unless the in-sample performance of the models is satisfactory, we will have problems justifying a good out-of-sample performance.

For evaluating the in-sample fit of the models, there are several statistical methods available. For example, the significance of the estimated parameters can be examined, or the residuals can be plotted and analysed. The residuals should be i.i.d. and have the assumed conditional distribution when the volatility models have a good fit. Likelihood ratio tests are also an option to compare different models. However, adding more parameters to the model tends to increase the maximum of the likelihood function. A penalty factor-based criterion is therefore desired when comparing models of different orders p and q. This is to balance the model complexity with the model fit of the data and avoid choosing the best model as the one with the highest number of parameters [38]. An example of a penalty-based criterion is information criteria.

3.3.1 Likelihood Ratio Tests

The likelihood ratio (LR) statistic for two estimated models is given by

$$LR = -2(\log L_R(\hat{\theta}_0) - \log L_U(\hat{\theta})),$$

where $\hat{\theta}$ denotes the estimated parameter vector. The statistic compares the maximised log-likelihood, $\log L$, for the restricted and the unrestricted model, denoted by indices R and U,

[2]For further information about the "rugarch" package, the reader is referred to the reference manual [36].

respectively. The null hypothesis assumes the restricted model to be correct, so $\hat{\theta}_0$ is the maximum likelihood estimator under H_0, while $\hat{\theta}$ is the ML estimator without any restrictions [38]. The test statistic is asymptotically chi-square distributed with τ degrees of freedom, where τ is the number of restrictions.

A drawback with the likelihood ratio tests is that they only exist for nested models. A model is nested if it is possible to obtain the nested model by restricting one or more parameters in the unrestricted model. As mentioned, the ARCH and GARCH models are nested in the APARCH model, but EGARCH is not. This means that the LR test cannot be used to compare the in-sample fit of the APARCH and EGARCH models. However, the LR test will be important in the out-of-sample model evaluation described in Chapter 6.

3.3.2 Information Criteria

A frequently used approach for comparing models of different orders p and q is to use information criteria. The Akaike information criterion (AIC) is given by

$$AIC = -\frac{2\log L}{n} + \frac{2j}{n}.$$

Here $\log L$ denotes the maximum value of the log-likelihood function, j is the number of estimated parameters in the model, and n is the sample size. The best model is the one with the lowest AIC value, which corresponds to a better model fit. However, AIC tends to overestimate the number of parameters in the model, which can have a significant effect on small samples. Hence, the corrected AIC (AICc) is a better choice [20]. AICc is defined as

$$AICc = AIC + \frac{2j(j+1)}{n(n-j-1)}.$$

The best model is the one with the lowest AICc value. This tends to be a more parsimonious model than the one chosen by the AIC.

An alternative criterion is the Bayesian information criterion (BIC). In short, this criterion maximises the a posteriori probability of the model, assuming that the a priori probabilities are the same for all models. It can be shown that this is equivalent to minimising the criterion

$$BIC = -\frac{2\log L}{n} + \frac{\log(n)j}{n}.$$

The BIC tends to penalise the number of parameters more than AIC and thus tends to choose a more parsimonious model. Moreover, several simulation studies have shown that AICc tends to choose a more correct lag order in smaller samples, whereas BIC works better in larger samples [16].

An advantage of information-based criteria is that they can be used for models that are not nested. As the EGARCH model is not nested in the APARCH model, the three presented criteria will be used for the in-sample model evaluation in this study. Furthermore, we use AICc and BIC as the preferred criteria since they both choose more parsimonious models than AIC.

Chapter 4

Risk Measures

Along with the increase of financial markets and the appearance of more complex financial instruments, the sources of risk have increased. Over the years, these sources of risk have been gradually better understood and reported. The three major types of risk are market risk, credit risk and operational risk [39]. Market risk is the possibility of experiencing a loss on a financial investment due to movements in market prices [40]. Market risk can be further divided into equity risk, interest rate risk and commodity risk among others, and as already been stated, the topic of this study is equity risk.

In order to keep financial institutions and their liability holders financially stable, financial services is a closely regulated economic sector. In other words, banks are required to keep a minimum amount of capital in order to safeguard against e.g. market risks. These regulations are primarily driven by the Basel Committee on Banking Supervision (BCBS) [41]. BCBS is a multi-lateral organisation led by the central bank governors of the member countries with the purpose of ensuring global financial stability. Following recent financial crises and their unfavourable impact on the market, regulators have been proposing more strict monitoring of banks to increase the probability of surviving in extreme market conditions. The BCBS has therefore established a series of international standards for bank regulation, including the accords on capital adequacy known as the Basel I, Basel II and Basel III [42]. These are revised frameworks that are inclusive of improvements of capital charges of market risks and market risks under stressed market conditions. In order to calculate the capital charges, risk measures are needed.

According to Danielsson, a risk measure is defined as a mathematical method for computing risk, whereas a risk measurement is a number that captures risk obtained by applying data to a risk measure [3]. To better understand the notion of risk, Artzner, Delbaen, Eber and Heath introduce a unified framework for analysing, constructing and implementing risk measures [43]. They present four desirable properties in which a risk measure should fulfil in order to be coherent.

Definition 4.1 (Coherent risk measure). *Let $\mathcal{G}$ be the set of all risks. A risk measure ρ is a mapping from $\mathcal{G}$ to $\mathbb{R}$. Using $\rho(X)$ to denote the risk measure for a set of outcomes X, the risk measure is called coherent if it satisfies all of the following properties:*

(i) (Subadditivity) For all $X, Y \in \mathcal{G}$, $\rho(X+Y) \leq \rho(X) + \rho(Y)$. When adding two portfolios together the total risk cannot be worse than the sum of the two individual risks, a demonstration of the diversification principle.

(ii) (Translation invariance) For all $X \in \mathcal{G}$ and all real numbers c, $\rho(X+c) = \rho(X) - c$. Adding an amount c of cash to the portfolio, this acts as insurance and the risk will be lower.

(iii) (Positive homogeneity) For all $\lambda \geq 0$ and all $X \in \mathcal{G}$, $\rho(\lambda X) = \lambda\rho(X)$. If the portfolio is increased by a factor, the risk will increase proportionally.

(iv) (Monotonicity) For all $X, Y \in \mathcal{G}$ with $X \leq Y$, $\rho(Y) \leq \rho(X)$. If a portfolio X has better values than portfolio Y under all scenarios, then the risk of X will be greater.

The simplest risk measure is volatility. This is often measured as either the standard deviation or the variance of the returns. However, this means that volatility is a sufficient risk measure only when financial returns are normally distributed. That is because all statistical properties of the normal distribution are expressed through the mean and variance, which could be the same for different assets. This implies that each asset would have the same risk profile, which clearly is not the case. Thus, the assumption about normally distributed returns is known to be violated quite often. Hence, using volatility as a risk measure can induce misleading conclusions [3].

Below, two of the most used risk measures in practice, value-at-risk and expected shortfall, are described. As they can be applied to all types of risk and securities, including complex portfolios, both are popular risk measures [20].

4.1 Value-at-Risk

One of the major crashes in financial markets occurred in 1987. As a reaction to this, Dennis Weatherstone, the former CEO of J.P. Morgan, formulated a measure of risk which later was named the value-at-risk (VaR) [44, 45]. It soon became the most frequently used risk measure in the financial industry. In short, it is a single summary statistic that measures the loss on a portfolio resulting from market movements [3].

Definition 4.2 (Value-at-risk)**.** *Given a time horizon T and a confidence level $1-\alpha$, where $\alpha \in (0,1)$, the probability that the loss on a trading portfolio exceeds or equals VaR in T is α. Equivalently, the probability of the loss being lower than VaR in T is $1-\alpha$.*

Using the notation $\mathrm{VaR}(\alpha)$ to indicate the dependence of VaR on α, VaR can be defined in more mathematical terms. Let the random variable associated with the profit and loss, or simply the return, on an investment be denoted $\mathcal{R}$ with a realisation $x_{\mathcal{R}}$. For any return distribution, $\mathrm{VaR}(\alpha)$ satisfies [21]

$$\mathrm{VaR}(\alpha) = -\inf\{x_{\mathcal{R}} \in \mathbb{R} : P(\mathcal{R} \leq x_{\mathcal{R}}) \geq \alpha\},$$

which is the negative α-quantile of the return distribution. For a continuous return distribution, $\mathrm{VaR}(\alpha)$ is given by [20]

$$P(\mathcal{R} \leq -\mathrm{VaR}(\alpha)) = \alpha. \tag{4.1}$$

VaR is a positive number, and the minus sign is used since losses are considered. Hence, VaR is the probability of losses exceeding, which means being more negative than the negative VaR. This is illustrated in Figure 4.1, which plots a standard normal profit and loss distribution. The coloured areas correspond to the 1% and 5% percentiles.

For calculation of VaR, there are three decisions to be made. The choice of confidence level, the time horizon and the profit and loss distribution. The degree of confidence is typically set to 95% or 99%, equivalent to the percentile levels $\alpha = 5\%$ and 1%, respectively [3]. In market risk management, the time horizon is usually one or ten days [2]. A one-day time horizon will be used in this study. This corresponds to trying to forecast the daily profit and loss variation in market risk portfolios. The final decision is to determine the profit and loss probability distribution. Multiple distributions have been proposed in the literature over the last couple of decades, see e.g. Nelson [32] or Rigby and Stasinopoulos [46]. Common choices, however, are the standard normal distribution or the scaled t-distribution [2]. As previously stated, the normal distribution in addition to the student-t and skewed student-t distributions will be used in this study. As we will see in Chapter 5, there also exist non-parametric methods for computing VaR that do not make an assumption about the return distribution.

A disadvantage with VaR is that it does not always satisfy the subadditivity property of a coherent risk measure. As a consequence, it might erroneously discourage diversification. However, with the assumption of normally distributed returns, VaR is proportional to volatility, which is subadditive, and VaR is therefore coherent under this assumption. Furthermore, as Danielsson, Jorgensen, Mandira, Samorodnitsky and de Vries discovered in their study, it is only if the loss distribution has very fat tails that VaR does not satisfy the subadditivity property, and this is rarely the case [47]. In particular, if the degrees of freedom in a student-t distribution is greater than two, the subadditivity property is satisfied.

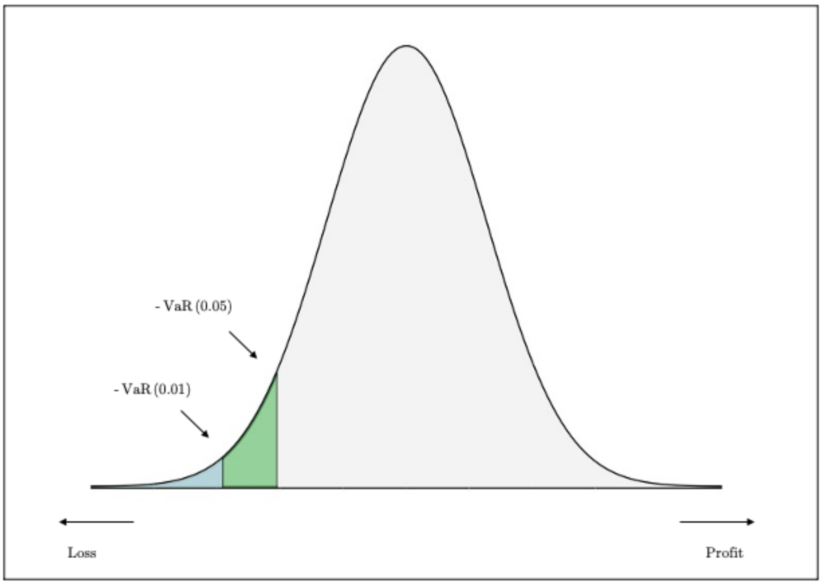


Figure 4.1: Standard normal profit and loss probability density and VaR.

Another disadvantage with VaR is that it is more concerned with the percentage of losses that exceeds VaR than the magnitude of those losses. This means that VaR only gives the worst loss that happens in $(1-\alpha)\%$ of the time and tells nothing about what happens in the remaining $\alpha\%$ of the time [48]. Additionally, VaR is known to perform poorly in stressed periods. This became particularly evident during the global financial crisis in 2007-2009 [45].

The practical advantages of VaR, such as its simplicity, ease of implementation and broad applicability, are known to outweigh its theoretical weaknesses. VaR has therefore remained the most widely used risk measure since its introduction in 1994 [47]. In terms of regulatory requirements, value-at-risk has been the official measure of market risk since the 1996 Amendment to the Basel I Accord [49]. However, with the Basel III Accord, BCBS proposes to replace value-at-risk with expected shortfall as the standard measure for market risk, partly due to its inferior performance in stressed periods [50]. Banks are currently implementing this change as a part of the final rules of Basel 4 [51].

4.2 Expected Shortfall

As stated above, VaR is more concerned with the percentage of losses that exceeds VaR than the magnitude of those losses. However, it is the extreme losses that most likely will cause financial distress. Risk measures that consider the magnitude of large losses, as well as their probability of occurring, are therefore sometimes more desired [25]. The most common alternative risk measure is the expected shortfall (ES), also known as the conditional value-at-risk [52]. Contrary to VaR, it summarises the entire tail risk of an investment.

Definition 4.3 (Expected Shortfall)**.** *Assuming that the profit and loss, $\mathcal{R}$, has a continuous distribution, the expected shortfall is the conditional expectation of the loss, given that the loss exceeds VaR. Formally,*

$$ES(\alpha) = -E[\mathcal{R}|\mathcal{R} \leq -VaR(\alpha)]. \tag{4.2}$$

For any profit and loss function $\mathcal{R}$, continuous or not, the expected shortfall can be defined as

$$\mathrm{ES}(\alpha) = \frac{1}{\alpha}\int_0^\alpha \mathrm{VaR}(\mathrm{u})du. \tag{4.3}$$

Hence, the expected shortfall is the average of VaR(u) over all percentile levels u that are less than or equal to α [20]. A comparison of the 5% VaR and ES for the normal distribution is given in Figure 4.2.

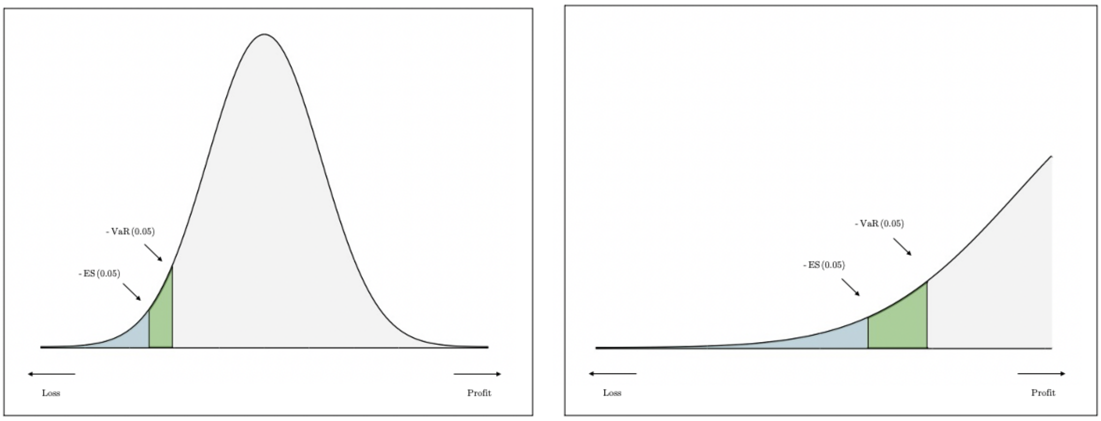


Figure 4.2: Standard normal profit and loss probability density and 5% VaR and ES.

ES is a coherent risk measure, and it also provides more information than VaR about the size of the potential loss. However, ES has several deficiencies, and the three most relevant will be given here. Firstly, ES is measured with more uncertainty than VaR as the estimation errors of VaR are incorporated in the estimation of ES. Secondly, ES is more challenging to evaluate with backtesting than VaR. The reason is that backtesting of ES requires estimates of the tail expectation to compare with the forecast of ES. VaR can be compared with actual observations, but ES can only be compared with the output from a model. Thus, ES gives more information about the risk, but it requires a larger data sample than VaR in backtesting [3]. Thirdly, the backtesting of ES can be quite complex. This will be further discussed in Chapter 6.2.

Even though VaR, especially in crisis periods, has shown to be a worse risk measure than ES, it is still commonly used due to the severe problems of backtesting ES. Using these risk measures together will give a more complete description of risk. By this, we mean that they capture both the percentile and the size of the loss. Thus, both measures will be applied in this study.

Chapter 5

Implementation of Risk Measures

Value-at-risk and expected shortfall are mainly estimated by either a non-parametric or a parametric approach. In contrast to the non-parametric approach, the parametric approach includes making an assumption about the underlying distribution of the returns, which usually calls for parameter estimation. The parametric approach is commonly known as the variance-covariance method, and the simplest non-parametric approach is called historical simulation (HS). As before, we assume that the mean of the returns is zero. For daily returns, this can be done without losing significant forecasting power in volatility modelling. The same applies to VaR and ES forecasts using a short time horizon such as ten days or less [3]. The two first sections of this chapter present the historical simulation and variance-covariance method for implementing VaR and ES forecasts. The final section describes how to use the conditional volatility models presented in Chapter 3 in risk forecasting.

5.1 Historical Simulation

Historical simulation is presumably the simplest method to forecast risk, and it is easy to grasp also for people without in-depth knowledge of risk management and statistics. It relies on the assumption that the market is persistent, and that history repeats itself. Instead of assuming a distribution for the returns, it uses the empirical distribution of the data to compute the risk forecasts. In other words, no statistical models are assumed, nor are any parameter estimates required. Compared to the parametric methods, HS has the advantage of being less prone to outliers and estimation errors. However, as each historical observation is weighted the same, HS is slow to incorporate structural changes in the asset risk. With presence of structural breaks this could be a crucial disadvantage. Although, in the absence of structural breaks, HS tends to outperform alternative methods [3].

5.1.1 Value-at-Risk

The VaR at confidence level $1 - \alpha$ is the negative $(\alpha \times n)$th value in the sorted return vector multiplied by the value of the asset. Here n is the size of the return vector and is called the window size. In practice, the number $(\alpha \times n)$ must be an integer. A common method to ensure this is to discard a few of the first observations such that n modulo 100 is zero [3].

5.1.2 Expected Shortfall

Obtaining ES forecasts by historical simulation requires the calculation of VaR as a first step. ES is then calculated by taking the mean of all losses exceeding VaR. Unless n is large, the mean is taken from a very small sample when the percentile level α is reasonable. This will not give a realistic estimate, and the method therefore requires a large sample size. As an example, a 1% VaR estimate is the $(0.01 \times n)$th smallest observation in the sample. A minimum of 10 observations is recommended, which implies a sample size of 1 000 observations for a 1% ES estimate [3].

5.1.3 Importance of Window Size

In VaR and ES forecasting, it is crucial to be aware of the importance of the window size, which is the number of observations to include in the empirical distribution of the returns. Changing the window size could potentially change the forecasts. For instance, a smaller window size leads to larger movements in the VaR forecast as the more extreme observations are expected to move by a larger amount from day to day than observations closer to the centre of the return distribution. A larger window size has the advantage of being less sensitive to outliers, while the disadvantage is that it takes a longer time for the VaR forecast to adjust to structural changes in risk. However, as the value equalling or exceeding VaR is expected to change only in one out of every $1/\alpha$ observations, the VaR forecasts are expected to be constant most of the time. As a rule of thumb, the minimum recommended sample size for HS is $3/\alpha$, and typically it is in the range between 250 and 1000 [3, 25].

5.2 Variance-Covariance Method

Contrary to the non-parametric approach, the parametric approach makes an assumption about the underlying distribution of the returns. As stated in Section 3.2, the normal, student-t and skew student-t distribution will be considered in this paper. Additionally, the parametric approach is more complex than HS as it includes statistical models and parameter estimation. This could be a disadvantage as the estimation error and model risk could be a problem. However, some parametric models can, at the expense of increased complexity, provide better forecasts.

In this section, the derivation of value-at-risk and expected shortfall will be explained using the parametric approach. This includes finding explicit expressions of VaR and ES for each of the distributional assumptions stated above. The first step of the parametric approach is to find the covariance matrix, or in the case of a simple asset, estimate the volatility [3]. This will be done using the GARCH models presented in Chapter 3. For now, it is assumed that volatility is time-independent, and the time subscript of σ will be left out.

5.2.1 Value-at-Risk

Consider holding one unit of an asset. The current portfolio value is then S_t, where S_t is the asset price at time t. To derive an expression for VaR using continuously compounded returns, we use the probability from equation (4.1) and insert the returns from equation (2.5). This gives

$$
\begin{aligned}
\alpha &= P(S_t - S_{t-1} \leq -\mathrm{VaR}(\alpha)) \\
&= P(S_{t-1}(e^{Y_t} - 1) \leq -\mathrm{VaR}(\alpha)) \\
&= P\left(\frac{Y_t}{\sigma} \leq \log\left(1 - \frac{\mathrm{VaR}(\alpha)}{S_{t-1}}\right)\frac{1}{\sigma}\right),
\end{aligned}
$$

since $-\mathrm{VaR}(\alpha)/S_{t-1} \leq 1$. Denote the cumulative distribution of the standardised returns (Y_t/σ) by $F_Y(\cdot)$ with inverse distribution $F_Y^{-1}(\cdot)$. Hence, VaR can be expressed as

$$
\mathrm{VaR}(\alpha) = -\left(\exp(\sigma F_Y^{-1}(\alpha)) - 1\right) S_{t-1}.
$$

For small values of $F_Y^{-1}(\alpha)\sigma$ this can be approximated to

$$
\mathrm{VaR}(\alpha) \approx -\sigma F_Y^{-1}(\alpha) S_{t-1}.
$$

In the following, the error distribution of the returns will be specified. For simplicity, the value S_{t-1} is assumed to be 1 NOK.

VaR with normally distributed innovations

If the innovations are assumed to be standard normally distributed, then VaR can be calculated as

$$\mathrm{VaR}(\alpha) = -\sigma\Phi^{-1}(\alpha). \tag{5.1}$$

Here $\Phi(\cdot)$ is the cumulative distribution function of a standard normal variable, and σ is the volatility forecast of the next day. A calculation of a 1% VaR estimate for an asset of value 1 NOK gives $\mathrm{VaR}(0.01) = -\sigma\Phi^{-1}(0.01) = 2.3263\sigma$. Notice that $\Phi^{-1}(\alpha)$ is always negative for $\alpha < 0.5$, so the minus sign in front ensures that VaR is a positive number.

VaR with student-t distributed innovations

For innovations assumed to be student-t distributed, scaled to have unit variance, the VaR can be calculated as

$$\begin{aligned} \mathrm{VaR}(\alpha) &= -\sigma t_\nu^{*-1}(\alpha), \\ t_\nu^{*-1}(\alpha) &= \sqrt{\frac{\nu-2}{\nu}}\, t_\nu^{-1}(\alpha). \end{aligned} \tag{5.2}$$

Here $t_\nu^*(\cdot)$ and $t_\nu(\cdot)$ denote the cumulative distribution functions of the standardised student-t distribution and the student-t distribution, respectively, and ν denotes the degrees of freedom [2]. As an example, assume $\alpha = 0.01$ and $\nu = 4$. Hence,

$$\mathrm{VaR}(0.01) = -\sigma\, t_4^{-1}(0.01)\sqrt{\frac{4-2}{4}} = -\sigma\frac{-3.746947}{\sqrt{2}} = 2.6495\sigma$$

for a single asset of value 1 NOK.

VaR with skew student-t distributed innovations

By assuming skew student-t distributed innovations, scaled to have unit variance, VaR can be calculated as

$$\begin{aligned} \mathrm{VaR}(\alpha) &= -\sigma st_{\nu,\xi}^{*-1}(\alpha), \\ st_{\nu,\xi}^{*-1}(\alpha) &= \frac{st_{\nu,\xi}^{-1}(\alpha) - m}{s}, \\ m &= \frac{\Gamma\left(\frac{\nu-1}{2}\right)\sqrt{\nu-2}}{\sqrt{\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left(\xi - \frac{1}{\xi}\right), \qquad s^2 = \left(\xi^2 + \frac{1}{\xi^2} - 1\right) - m^2. \end{aligned}$$

Here $st_{\nu,\xi}^*(\cdot)$ and $st_{\nu,\xi}(\cdot)$ denote the cumulative distribution functions of the standardised skew student-t distribution and the skew student-t distribution, respectively. Moreover, ν and ξ are the degrees of freedom and skewness parameter, respectively. Let $\alpha = 0.01$, $\nu = 4$ and $\xi = 0.9$. For a single asset of value 1 NOK, the 1% VaR is

$$\mathrm{VaR}(0.01) = -\sigma\frac{st_{4,0.9}^{-1}(0.01) - m}{s} = -\sigma\frac{-3.03510 - (-0.14928)}{1.01108} = 2.8542\sigma.$$

5.2.2 Expected Shortfall

Compared to the derivation of value-at-risk, deriving an expression for the expected shortfall is more complicated. The VaR first has to be obtained, and then the conditional expectation in equation (4.2) is calculated. As before, we assume holding one unit of an asset with value 1 NOK.

ES with normally distributed innovations

When the innovations are assumed to be standard normally distributed, the expected shortfall can be expressed as [2]

$$\mathrm{ES}(\alpha) = \sigma\frac{\phi(\Phi^{-1}(\alpha))}{\alpha}.$$

Here $\phi(\cdot)$ is the density of a standard normal variable, and as before, $\Phi(\cdot)$ is the cumulative distribution function. The derivation of this expression is given in Appendix A.1.

ES with student t-distributed innovations

When the innovations are assumed to be student-t distributed, scaled to have unit variance with $\nu > 2$ degrees of freedom, the expected shortfall can be calculated as [2]

$$\mathrm{ES}(\alpha) = \sigma\sqrt{\frac{\nu-2}{\nu}}\frac{g_\nu(t_\nu^{-1}(\alpha))}{\alpha}\left(\frac{\nu+(t_\nu^{-1}(\alpha))^2}{\nu-1}\right).$$

Here $g_\nu(\cdot)$ denotes the density and $t_\nu(\cdot)$ the cumulative distribution function of a student-t distributed variable. The derivation of this expression can be found in Appendix A.2.

ES with skew student t-distributed innovations

To calculate the expected shortfall when the innovations are assumed to have a scaled skew student-t distribution, we have to use numerical integration of equation (4.3).

5.3 VaR and ES with Time-Dependent Volatility

The volatility has until now been assumed independent of time. In the following, an explanation of how to incorporate the time-varying volatility into the forecasting of VaR and ES will be elaborated. The first step is to calculate the one-step-ahead volatility forecast $\hat{\sigma}_{t+1}$. Using the GARCH models stated in Section 3.1 to produce this forecast requires parameter estimation. As an example, consider finding the one-day-ahead volatility forecast $\hat{\sigma}_{t+1}$ for the standard GARCH(1,1) model given in equation (3.6). This can be achieved by using the previous volatility estimate $\hat{\sigma}_t$ and the fitted parameter vector. Thus, $\hat{\sigma}_{t+1}$ can be found iteratively as

$$\hat{\sigma}_{t+1} = \sqrt{\hat{a}_0 + \hat{a}_1 y_t^2 + \hat{b}_1\hat{\sigma}_t^2},$$

where $\hat{a}_0, \hat{a}_1$ and $\hat{b}_1$ denote estimated coefficients. For each volatility forecast, the parameters have to be re-estimated. The volatility σ_1 of the first day is set to be the standard deviation of the entire data sample. The same procedure can then be applied to all GARCH models using their respective formulas for conditional volatility. Finally, the obtained estimate for $\hat{\sigma}_{t+1}$ can be exchanged with the time-independent volatility σ in the given formulas for VaR and ES. We will use the notation $\mathrm{VaR}_t(\alpha)$ and $\mathrm{ES}_t(\alpha)$ to denote the time-varying VaR and ES forecasts, respectively.

Chapter 6

Evaluating Risk Measures with Backtesting

In order to assess individual risk models, evaluation criteria like testing for parameter significance or information criteria can be applied. These criteria do not, however, properly evaluate the risk forecasting power of the models. To compare and choose among different risk models, a standard procedure is to use out-of-sample quantitative methods known as backtests. This involves comparing ex ante risk measure forecasts, either VaR or ES, from a particular model with ex post realised returns, i.e. historical observations. Backtesting can be seen as a final diagnostic check on the volatility models developed, complementing the previously discussed diagnostics [25]. This chapter illustrates the concept of backtesting by applying it to VaR, and then backtests for both VaR and ES will be explained.

6.1 Backtesting Value-at-Risk

As VaR violations are infrequent events, assessing the accuracy of VaR forecasts by future model performance would require a long time period. For this reason, backtesting evaluates VaR forecasts by looking at how a model performs over a period in the past. For understanding the concept of backtesting, it is essential to distinguish between the estimation window and the testing window [3].

Definition 6.1 (Estimation window). *The estimation window, K, is the number of observations used to make a risk forecast.*

Definition 6.2 (Testing window). *The testing window, T, is the days for which a risk forecast is made.*

The definitions stated above will be important in illustrating how the backtesting procedure applies to VaR. Let N be the total sample size equal to the sum of the estimation window K and the testing window T. The sample from the estimation window of the first K days is used to make a VaR forecast at day $K+1$. The estimation window is then moved one day to obtain a risk forecast at day $K+2$. This continues until a VaR forecast at day N is obtained, which means a total of T risk forecasts. Figure 6.1 illustrates how this rolling window method works. After obtaining the VaR forecasts, they can be compared to the already known daily returns. In particular, if the actual return on a day exceeds the VaR forecast made for that day, then a VaR violation is said to have occurred [3].

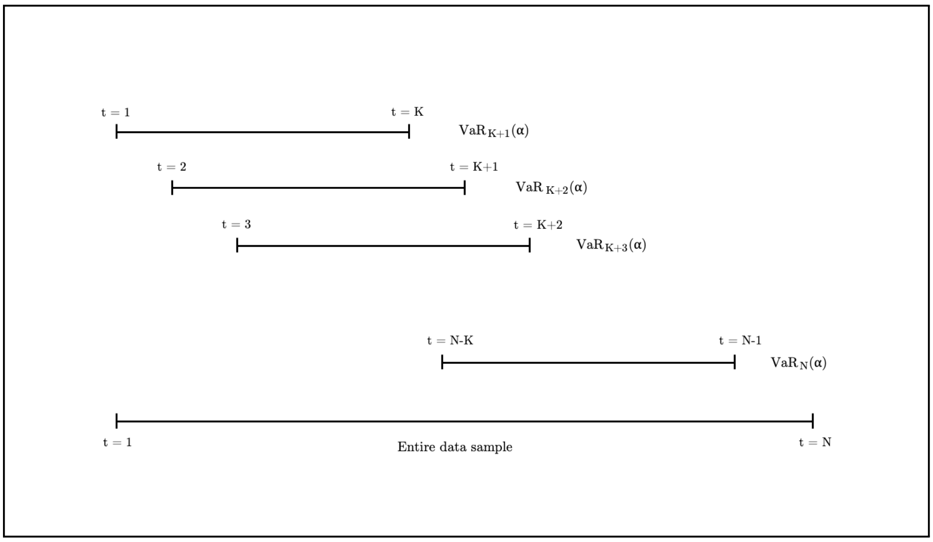


Figure 6.1: Rolling window forecasts of VaR.

In most VaR backtests, the violations over time are assumed to be an i.i.d. Bernoulli distributed sequence (η_t) [3].

Definition 6.3 (VaR violation)**.** *A VaR violation, also called a VaR exceedance, is an event such that*

$$\eta_t = \begin{cases} 1 & if \;\; Y_t < -VaR_t(\alpha) \\ 0 & if \;\; Y_t \geq -VaR_t(\alpha), \end{cases}$$

where Y_t is return and $VaR_t(\alpha)$ is the VaR estimate made at time $t-1$ for time t.

If the VaR model is accurate and the sequence (η_t) follows an i.i.d. Bernoulli process, η_t is equal to 1 with probability α because the probability of a VaR violation is α. Similarly, η_t is equal to 0 with probability $1-\alpha$. The total number of violations is then a sum of independent Bernoulli distributed variables and will follow a binomial distribution with parameters $p = \alpha$ and $n = T$ when T forecasts of VaR are made. The expected number of VaR violations is then αT.

There are different statistical tests to apply for judging the quality of VaR forecasts. A relatively simple and intuitive evaluation method is to report the relative number of violations, called the violation ratio. Let T_1 be the number of violations and T_0 the number of days without violations [3]. These numbers are obtained by

$$\begin{aligned} T_1 &= \sum_t \eta_t, \\ T_0 &= T - T_1. \end{aligned}$$

Definition 6.4 (Violation ratio)**.** *The violation ratio (VR) is*

$$VR = \frac{T_1}{\alpha T}.$$

Hence, the violation ratio is the observed number of violations divided by the expected number of violations. The VaR model underestimates risk if the violation ratio is larger than 1 and overestimates if the ratio is smaller than 1. Some statistical errors are expected, and values different from 1 could be statistically significant. As a rule of thumb, if $VR \in [0.8, 1.2]$, the forecast is good, and the model is imprecise if $VR < 0.5$ or $VR > 1.5$. However, with increasing testing window, both bounds get narrower [3].

6.1.1 Parameter Choice and Estimation

In order to evaluate risk forecasts by backtesting, the estimation window and the testing window must be decided. The length of the estimation window is highly dependent on the VaR model and the confidence level. As stated in Section 5.1.3, HS needs at least $K = 300$ for 1% VaR forecasting, i.e. a minimum of $3/\alpha$ observations, and GARCH models need even more. Similarly, the size of the testing window is highly dependent on the chosen confidence level. In particular, the testing window has to increase with more extreme confidence levels. The reason is that VaR violations are infrequent events, and for 1% VaR forecasts, a violation is expected to occur only every 100 days. The sample size of violations could potentially become quite small, and a large testing window T is required to guarantee reliable results [3].

Backtesting is based on the assumption that no changes in the portfolio are made when estimating one-day VaR. In reality, actual realised returns might reflect various fee incomes, revenue and transaction costs from intra-day trading and re-balancing. This implies that the returns used in backtesting should represent a portfolio without changes [3]. The data sets applied in this paper are in this sense static, and no adjustments have to be made for this study.

6.1.2 Significance of Backtesting VaR

For assessing the significance of backtesting VaR, more sophisticated methods than just the given rule of thumb should be applied. A good VaR model should satisfy two properties, the unconditional coverage property and the independence property.

For a 1% VaR backtest, a VaR violation is expected in 1% of the time. If the observed number of violations is more frequent, the VaR model systematically underestimates risk at the 1% level. The unconditional coverage property ensures that the theoretical confidence level α is similar to the empirical probability of violation. In other words, the number of VaR violations is close enough to the expected value. This property can be tested by the unconditional coverage test proposed by Kupiec [53].

Despite the model's performance of coverage rate, the risk of significant losses will be considerably higher if VaR violations tend to occur in a rapid successive manner. The independence property requires two successive observations in the sequence (η_t) to be independent of each other [3]. For instance, a violation today should not influence the likelihood of observing a violation tomorrow. For testing this property, Christoffersen proposes the independence test [54]. Christoffersen also proposes a joint test called the conditional coverage test to test the unconditional coverage and independence properties simultaneously.

When the unconditional coverage and independence properties are satisfied, the η_t's are i.i.d. Bernoulli distributed variables with probability α. For ease of notation, consider the time index set $T = \{K+1, \ldots, N\}$ for the days where a VaR forecast is made.

6.1.3 Unconditional Coverage Test

The unconditional coverage test by Kupiec is used to assess whether the fraction of violations obtained from a particular VaR model, π, is significantly different from the promised fraction α [53]. The null hypothesis for VaR violations is

$$H_0 : \eta_t \sim B(\alpha),$$

where $B(\cdot)$ denotes the Bernoulli distribution. The unrestricted likelihood function can be written in terms of the estimated probability $\hat{\alpha}$ as

$$L_U(\hat{\alpha}) = \prod_{t=1}^{T} (1-\hat{\alpha})^{1-\eta_t} \hat{\alpha}^{\eta_t} = (1-\hat{\alpha})^{T_0} \hat{\alpha}^{T_1},$$

where as before, T_0 denotes the number of days without violations and T_1 is the number of violations. The probability α can be estimated by $\hat{\alpha} = \frac{T_1}{T}$, which is the observed fraction of violations.

Under H_0, $\pi = \alpha$ such that the restricted likelihood function can be written as $L_R(\alpha) = L_U(\alpha)$, with the probability $\hat{\alpha}$ restricted to α. The likelihood ratio statistic can then be expressed as

$$LR_{uc} = -2(\log L_R(\alpha) - \log L_U(\hat{\alpha})) \overset{\text{asymptotic}}{\sim} \chi_1^2.$$

The null hypothesis is rejected if $LR_{uc} > \chi^2_{1-q,1}$ using a $q\%$ significance level, or equivalently if the p-value is less than q.

6.1.4 Independence Test

To assess whether violations cluster, Christoffersen propose the independence test [54]. On two consecutive days, the probabilities for a new violation or no violation can be expressed as

$$p_{ij} = P(\eta_t = i | \eta_{t-1} = j),$$

where $i, j = 0$ or 1. For instance, p_{11} is the probability of a violation today conditional on a violation yesterday, and p_{10} is the probability of no violation today when there was a violation yesterday. As the violation sequence (η_t) is assumed to be dependent over time with one lag, it can be described as a first-order Markov sequence with transition probability matrix defined as

$$\Pi_1 = \begin{pmatrix} 1 - p_{01} & p_{01} \\ 1 - p_{11} & p_{11} \end{pmatrix}.$$

Using the given transition matrix, the unrestricted likelihood function can be written as

$$L_U(\Pi_1) = (1 - p_{01})^{T_{00}} p_{01}^{T_{01}} (1 - p_{11})^{T_{10}} p_{11}^{T_{11}},$$

where T_{ij} is the number of observations where j follows i. Maximising $L_U(\Pi_1)$ gives the maximum likelihood estimate

$$\hat{\Pi}_1 = \begin{pmatrix} \frac{T_{00}}{T_{00}+T_{01}} & \frac{T_{01}}{T_{00}+T_{01}} \\ \frac{T_{10}}{T_{10}+T_{11}} & \frac{T_{11}}{T_{10}+T_{11}} \end{pmatrix}.$$

The null hypothesis assumes no clustering, which means that the probability of a violation today is independent of any violations yesterday. This yields, $p_{01} = p_{11} = p$. The estimated probability, $\hat{p}$, can be computed as

$$\hat{p} = \frac{T_{01} + T_{11}}{T_{00} + T_{10} + T_{01} + T_{11}},$$

and the estimated transition matrix can be reduced to

$$\hat{\Pi}_0 = \begin{pmatrix} 1 - \hat{p} & \hat{p} \\ 1 - \hat{p} & \hat{p} \end{pmatrix}.$$

The restricted likelihood function can then be expressed as

$$L_R(\hat{\Pi}_0) = (1 - \hat{p})^{T_{00}+T_{10}} \hat{p}^{T_{01}+T_{11}},$$

which gives the likelihood ratio test statistic

$$LR_{ind} = -2(\log L_R(\hat{\Pi}_0) - \log L_U(\hat{\Pi}_1)) \overset{\text{asymptotic}}{\sim} \chi_1^2.$$

The null hypothesis is rejected if the p-value of the statistic is less than the significance level q.

A crucial drawback of this test is that it only considers violations that occur in pairs. It will not detect violations that occur with two or more days between, although this implies that the independence property is not satisfied. For instance, approaches such as HS, which are not very responsive to markets with changing volatility, often do not satisfy the independence property. Unfortunately, the described test will not always be able to detect clustering for HS as the clustering is not sequential but still significant.

6.1.5 Joint Conditional Coverage Test

Christoffersen also developed a joint test to check for coverage and independence called the conditional coverage test [54]. The test statistic can be calculated by simply adding the two individual test statistics. This yields,

$$LR_{cc} = LR_{uc} + LR_{ind} \sim \chi_2^2.$$

Under H_0, both LR_{uc} and LR_{ind} are asymptotically chi-square distributed with one degree of freedom, which means that their sum is chi-square distributed with two degrees of freedom. Intuitively, testing both properties at the same time would be the best. However, if the VaR model only satisfies one of the properties, the joint test has less power to reject the model. If the joint test does not reject the model, then the individual tests should be performed.

6.2 Backtesting Expected Shortfall

In the evaluation of forecasts, we want to compare forecasted estimates with observed data. For this purpose, a scoring function $s(x, y)$ could be useful. In simple terms, a scoring function is a loss function used to evaluate a prediction of some measure of a probability distribution. It evaluates the performance of the forecasts x given the observations y. A statistical measure used for assessing the accuracy of a model is called elicitable if it allows forecasted estimates and observed data to be compared [45].

Gneiting shows that in contrast to VaR, ES is not elicitable [55]. He proves that a necessary condition for the elicitability of a risk measure is the existence of convex level sets and disproves their existence for ES. This means that it is not possible to find a scoring function $s(x, y)$ such that the forecasted ES x of the true ES y can be obtained as the x that minimises $s(x, y)$ [56]. In other words, we cannot find a strictly consistent scoring function for the expected shortfall. The paper by Gneiting led to a widespread belief that it is not possible to backtest ES at all [57].

However, not all interpreted Gneiting's findings as evidence for ES not being backtestable. Emmer, Kratz and Tasche show that ES is conditionally elicitable for continuous distributions with finite means [58]. They also conclude that even though ES is not elicitable, it is the most suitable risk measure. Furthermore, Fissler and Ziegel show that ES is jointly elicitable with VaR, and Acerbi and Szekely argue that elicitability has to do with model selection and not with model testing [59, 60]. After these studies were published, there emerged a consensus among practitioners that the problem of comparing the forecasting performance of different methods, which requires the elicitability property, is distinct from the problem of comparing ex ante risk measure forecasts with ex post realised returns, i.e. backtestability [61]. This means that even though ES is non-elicitable it should be possible to find a backtest that does not exploit the property of elicitability and makes ES backtestable.

In addition to the lack of the desired elicitability property, as an expectation rather than a single quantile is considered, ES is more challenging to backtest than VaR. The accuracy of ES backtests is much lower than for VaR backtests because a formal test of ES would be a joint test of the accuracy of VaR and the expectation beyond VaR. This means the errors in estimating VaR also have to be taken into account. The reliability of an ES backtest procedure is therefore expected to be much lower than for the procedure of a VaR backtest. As such, backtesting ES also requires a lot more observations than backtesting VaR does. This is part of the reason VaR is more frequently used as a risk measure. However, after the change from VaR to ES in the regulatory capital requirements for market risk, an extensive literature on the backtesting of ES has emerged. In the following, some popular backtest will be elaborated.

The notation used in equation (4.2) for the expression of ES was conditional on the loss exceeding VaR but not conditional on time. We now consider that VaR and ES are estimated at time $t-1$ and used to forecast the tail of the return distribution from time $t-1$ to t. For the expected shortfall at confidence level $1-\alpha$ estimated for time t, we consider the following notation

$$\mathrm{ES}_t(\alpha) = -E[Y_t | Y_t < -\mathrm{VaR}_t(\alpha)], \tag{6.1}$$

where $\mathrm{VaR}_t(\alpha)$ is the value-at-risk estimated for time t. For ease of notation we denote the negative

VaR and ES estimates by $\hat{v}_t(\alpha)$ and $\hat{e}_t(\alpha)$, respectively.

6.2.1 Exceedance Residuals Test

The exceedance residuals (ER) backtest proposed by McNeil and Frey is one of the oldest backtests developed for ES and is still one of the most frequently used [62]. This test considers the difference between the realised returns and the negative estimate of the expected shortfall, conditional on the returns exceeding the VaR. Thus, the exceedance residuals can be written as

$$\boldsymbol{\varepsilon} = \{\varepsilon_t : \; t \in T, \; Y_t < \hat{v}_t(\alpha)\},$$

where

$$\varepsilon_t = Y_t - \hat{e}_t(\alpha)). \tag{6.2}$$

Under the assumption that the process dynamics and the expected shortfall are estimated correctly, these residuals should behave like an i.i.d. sample with zero mean. The null hypothesis of ε_t having mean zero is tested against the alternative hypothesis that the residuals have a mean greater than zero, which is equivalent to the expected shortfall being underestimated. The choice of a one-sided hypothesis test is argued by the fact that underestimation is more critical than overestimation. The regulators want to avoid underestimation such that the capital reserves are not too small, while market practitioners do not want to endanger the existence of their business. The null hypothesis and the alternative hypothesis can be expressed as

$$H_0 : \mu_\varepsilon = 0 \quad \text{vs.} \quad H_1 : \mu_\varepsilon > 0, \tag{6.3}$$

where μ_ε is the mean of the exceedance residuals.

Due to the one-sided hypothesis test, the ER test tends to favour models that overestimate the ES. The test also heavily relies on the VaR estimates, which is an obvious deficiency. The ES backtest could be misleading if the conditional volatility model is incorrectly specified or the distributional assumption of the innovations is wrong. Another drawback of the exceedance residuals test concerns the sample size. Only the observations corresponding to a VaR violation are used, and too few data points could lead to less reliable results.

The hypothesis test given in (6.3) can easily be performed with the bootstrapping procedure of Efron and Tibshirani [63]. We want to test if the residuals ε_t, which are distributed according to some distribution function F, have mean $\mu_\varepsilon = 0$ as assumed in our null hypothesis. The test statistic is given as

$$\mathit{\Theta}_{ER}(\varepsilon) = \frac{\hat{\mu}_\varepsilon - \mu_\varepsilon}{\hat{\sigma}_\varepsilon / \sqrt{T_1}},$$

where $\hat{\mu}_\varepsilon$ and $\hat{\sigma}_\varepsilon$ are the sample mean and standard deviation based on the sample of size T_1, where T_1 is the number of VaR violations in a particular model.

The empirical distribution $\hat{F}$ needs to be translated to have the desired mean as under H_0, which means for ε_t to have mean zero. To sample data from under the null hypothesis, we use the shifted residuals $\tilde{\varepsilon}_i = \varepsilon_i - \hat{\mu}_\varepsilon + \mu_\varepsilon$, for $i = 1, \ldots, T_1$. Then we draw B bootstrap samples $\tilde{\varepsilon}_1^*, \ldots, \tilde{\varepsilon}_{T_1}^*$ with replacement from $\tilde{\varepsilon}_1, \ldots, \tilde{\varepsilon}_{T_1}$. For each bootstrap sample we compute the statistic $\mathit{\Theta}_{ER}(\tilde{\varepsilon}^*) = \frac{\hat{\mu}_\varepsilon^* - \mu_\varepsilon}{\hat{\sigma}_\varepsilon^* / \sqrt{T_1}}$. The significance level then corresponds to the proportion of samples for which $\mathit{\Theta}_{ER}(\tilde{\varepsilon}^*)$ is higher than $\mathit{\Theta}_{ER}(\varepsilon)$.

There is also an option to use the standardised residuals instead of the raw residuals in equation (6.2). This would require volatility forecasts as additional information since the standardised residuals are given by the raw residuals ε_t divided by the volatility forecasts σ_t. McNeil and Frey report that both residuals give similar results when using a bootstrap test.

6.2.2 V-test

Embrechts, Kaufmann and Patie introduce the measure $V = (|V_1| + |V_2|)/2$ for backtesting ES [64]. The first measure, V_1, is the conditional average of the difference between the observed return Y_t

and the negative of the estimated expected shortfall, $\hat{e}_t(\alpha)$. This is equivalent to calculating the average on days when the actual returns exceed the VaR estimate. Thus,

$$V_1 = \frac{\sum_{t=1}^{T} (Y_t - \hat{e}_t(\alpha)) I_{\{Y_t < \hat{v}_t(\alpha)\}}}{\sum_{t=1}^{T} I_{\{Y_t < \hat{v}_t(\alpha)\}}},$$

where $I_{\{\cdot\}}$ is the indicator function and takes the value one if the argument is true and zero otherwise. A value of V_1 close to zero indicates a good model. Since only values below the negative VaR are considered, this statistic strongly depends on the VaR estimates. This is a crucial shortcoming of the V_1 measure. Depending on the performance of the VaR estimator, one could in fact take the mean of a subsample which is very different from $\alpha\%$ of the values, meaning a subsample different from αT.

To correct for the shortcomings of the V_1 measure, the second measure Embrechts, Kaufmann and Patie introduce, V_2, replaces the VaR estimates with an empirical α-quantile. The statistic is defined as

$$V_2 = \frac{\sum_{t=1}^{T} d_t I_{\{d_t < d^\alpha\}}}{\sum_{t=1}^{T} I_{\{d_t < d^\alpha\}}},$$

where $d_t = Y_t - \hat{e}_t(\alpha)$ and d^α is the empirical α-quantile of $(d_t)_{1 \le t \le T}$. As $\hat{e}_t(\alpha)$ is an estimate with confidence level $1-\alpha$, it is expected that d_t is negative in less than one out of $1/\alpha$ cases. A good ES model leads to a V_2 statistic close to zero. The initially described V statistic thus measures how well the forecasted ES fits the real data. As before, the statistic should be close to zero if the model is good. However, Embrechts, Kaufmann and Patie do not explicitly state how close to zero the measure should be for the forecasts to be acceptable.

6.2.3 Conditional Calibration Backtests

Nolde and Ziegel introduce conditional calibration (CC) backtests based on strict identification functions of the respective functional [65]. As ES is jointly elicitable with VaR, the two functionals are put together into the two-dimensional functional (VaR_t, ES_t). For simplicity of the notation, we leave out the dependence on α as it is a fixed quantity. For the pair VaR and ES at confidence level $1-\alpha$, Nolde and Ziegel choose the strict 2×1 identification function [66]

$$V(Y_t, v_t, e_t) = \begin{pmatrix} \alpha - I_{\{Y_t < v_t\}} \\ e_t - v_t + I_{\{Y_t < v_t)\}}(v_t - Y_t)/\alpha \end{pmatrix},$$

where the expectation is zero if and only if the true VaR and ES are inserted. The forecasts $\hat{v}_t$ and $\hat{e}_t$ are tested with the hypothesis

$$H_0 : E\Big[V(Y_t, \hat{v}_t, \hat{e}_t) | \mathcal{F}_{t-1}\Big] \ge 0 \quad \text{vs.} \quad H_1 : E\Big[V(Y_t, \hat{v}_t, \hat{e}_t) | \mathcal{F}_{t-1}\Big] < 0,$$

component-wise for all $t = \{1, \ldots, T\}$, where $\mathcal{F}_{t-1}$ denotes the information up to time t. The requirement $E[V(Y_t, \hat{v}_t, \hat{e}_t)|\mathcal{F}_{t-1}] = 0$ is equivalent to testing $E[\boldsymbol{h_t'} V(Y_t, \hat{v}_t, \hat{e}_t)] = 0$ for all $\mathcal{F}_{t-1}$-measurable $\mathbb{R}^2$-valued functions $\boldsymbol{h_t}$. Considering that this is infeasible, Nolde and Ziegel suggest using an $\mathcal{F}_{t-1}$-measurable sequence of $l \times 2$-matrices of test functions $\boldsymbol{h_t}$ to create the Wald-type test statistic

$$W_{CC} = T \left(\frac{1}{T} \sum_{t=1}^{T} \boldsymbol{h_t} V(Y_t, \hat{v}_t, \hat{e}_t) \right)' \hat{\Omega}_T^{-1} \left(\frac{1}{n} \sum_{t=1}^{T} \boldsymbol{h_t} V(Y_t, \hat{v}_t, \hat{e}_t) \right).$$

Here

$$\hat{\Omega}_T = \frac{1}{T} \sum_{t=1}^{T} (\boldsymbol{h_t} V(Y_t, \hat{v}_t, \hat{e}_t)) (\boldsymbol{h_t} V(Y_t, \hat{v}_t, \hat{e}_t))'$$

is a consistent estimator of the covariance of the l-dimensional vector $\boldsymbol{h_t} V(Y_t, \hat{v}_t, \hat{e}_t)$. Under the null hypothesis, the test statistic W_{CC} is asymptotically chi-square distributed with l degrees of freedom.

Two versions of the conditional calibration test are proposed. The first uses no additional information besides the risk forecasts and is named the simple CC test. The second requires volatility forecasts in addition to risk forecasts and is called the general CC test. For the simple CC test, Nolde and Ziegel propose the test function $\boldsymbol{h_t} = \boldsymbol{I_2}$, the identity matrix. For the one-sided general CC test, they propose to use

$$\boldsymbol{h_t} = \begin{pmatrix} 1 & |\hat{v}_t| & 0 & 0 \\ 0 & 0 & 1 & \hat{\sigma}_t^{-1} \end{pmatrix}',$$

where $\hat{\sigma}_t$ is a forecast of the volatility. Similarly to the standardised ER test, the general CC test is a backtest for both the VaR, ES and volatility.

6.2.4 Regression-Based Backtests

Bayer and Dimitriadis propose three regression-based ES backtests, which test whether a series of ES forecasts is correctly specified relative to the series of returns (Y_t) for $t = \{1, \ldots, T\}$ [67]. This is performed by regressing the returns on the ES forecasts by using a regression equation designed for the functional ES,

$$Y_t = \beta_0 + \beta_1 \hat{e}_t + u_t^e, \tag{6.4}$$

where $\hat{e}_t$ is the negative ES forecasts, u_t^e are the errors $Y_t - \hat{e}_t$ and $\mathrm{ES}(u_t^e | \mathcal{F}_{t-1}) = 0$ almost surely. As before, for simplicity of the notation, we drop the dependence of α. Estimating the parameters β stand-alone is not possible as there do not exist strictly consistent loss and identification functions for the expected shortfall, meaning that it is not elicitable [55]. Based on the seminal work of Fissler and Ziegel, who introduce joint loss and identification functions for the VaR and ES [59], Dimitriadis and Bayer, Patton and Barendse propose a joint regression framework for the estimation of the joint regression parameters (γ, β) [66, 68, 69]. In particular, they propose a second auxiliary quantile regression equation given by

$$Y_t = \gamma_0 + \gamma_1 \hat{v}_t + u_t^v, \tag{6.5}$$

where $\hat{v}_t$ denotes negative VaR forecasts, u_t^v is the error term for the value-at-risk estimation and $\mathrm{VaR}(u_t^v | \mathcal{F}_{t-1}) = 0$ almost surely. The first test Bayer and Dimitriadis propose, called the Auxiliary ESR backtest, uses auxiliary VaR forecasts as the explanatory variable in the quantile equation (6.5), but only tests the parameters β associated with the expected shortfall. The second test, called the Strict ESR backtest, uses the ES forecasts as the explanatory variable in both the quantile and the ES regression equations (6.5) and (6.4), and as before, only tests the ES specific parameters β. The Auxiliary ESR backtest has the disadvantage of requiring VaR forecasts as additional input parameters, whereas the Strict ESR test only requires ES forecasts. For both these tests, when the ES forecasts are accurate, the intercept and slope parameters should be equal to 0 and 1, respectively. The backtests are performed with a two-sided hypothesis test as we do not know how the slope and intercept parameters will be influenced by underestimated or overestimated ES forecasts, respectively.

However, to avoid keeping too large capital reserves, market participants want to moderately underestimate the risk. The regulatory authorities, on the other hand, want to avoid underestimation of the risk as this could lead to insolvency of financial institutions. This justifies the choice of a one-sided backtest. Bayer and Dimitriadis therefore propose a third test, named the Intercept ESR backtest. This is based on regressing the forecast errors, $Y_t - \hat{e}_t$, on an intercept term only, meaning it has a fixed slope parameter equal to one. Thus,

$$Y_t - \hat{e}_t = \beta_0 + u_t^e,$$

where as before, $\mathrm{ES}(u_t^e | \mathcal{F}_{t-1}) = 0$ almost surely. The estimation of β_0 is accomplished by computing the empirical expected shortfall of the forecast errors. A one-sided hypothesis can be defined as

$$H_0 : \beta_0 \geq 0 \quad \text{vs.} \quad H_1 : \beta_0 < 0,$$

which can be tested using the bootstrap procedure of Efron and Tibshirani, similar to that described for the exceedance residuals test in Section 6.2.1 [63]. The t-statistic can be expressed as

$$\Theta_{ESR} = \frac{\hat{\beta}_0 - \beta_0}{\sqrt{\hat{\Sigma}_{22}/T}},$$

where $\hat{\Sigma}_{22}$ is the lower diagonal element of the asymptotic covariance matrix from the joint estimation of the sample quantile and expected shortfall [66]. The covariance matrix, Σ, has the following components,

$$\begin{aligned}
\Sigma_{11} &= \frac{\alpha(1-\alpha)}{f_Y^2(v(\alpha))}, \\
\Sigma_{12} &= \Sigma_{21} = (1-\alpha)\frac{v(\alpha)-e(\alpha)}{f_Y(v(\alpha))}, \\
\Sigma_{22} &= \frac{1}{\alpha}\mathrm{Var}(Y-(-v(\alpha))|Y \leq -v(\alpha)) + \frac{1-\alpha}{\alpha}(v(\alpha)-e(\alpha))^2,
\end{aligned}$$

where $v(\alpha)$ and $e(\alpha)$ represent the value-at-risk and expected shortfall under H_0. The two first elements of Σ represent the quantile and the covariance, respectively. The test statistic Θ_{ESR} only requires the variance for the expected shortfall, Σ_{22}, and we will therefore not go into detail on $f_Y(\cdot)$.

We draw B bootstrap samples from the errors $u_t^e = Y_t - \hat{e}_t$ and name them $U^{(b)}$. The test statistics are then calculated as

$$\Theta_{ESR}^{(b)} = \frac{ES(U^{(b)}) - \beta_0}{\sqrt{\hat{\Sigma}_{22}^{(b)}/T}},$$

where $\hat{\Sigma}_{22}^{(b)}$ is the estimate of the asymptotic covariance matrix of the bootstrap set $U^{(b)}$ [70]. The p-value is calculated as the proportion of the B bootstrap test statistics that are larger than or equal to the t-statistic from the data sample, Θ_{ESR}.

Bayer and Dimitriadis report that their ES Regression-based backtests are more powerful than the backtests of McNeil and Frey [62] and Nolde and Ziegel [65] in almost all simulation designs studied. Moreover, the Strict ESR and Intercept ESR tests proposed by Bayer and Dimitriadis are the first backtests that only require ES forecasts as input variables and not any additional information. These tests are particularly important as the Basel Committee regulations now only requires financial institutions to report their ES forecasts [71, 72].

6.2.5 Selection of Evaluation Methods

A problem with backtests where more information than solely the ES forecasts are used is that rejecting the null hypothesis does not necessarily imply that the ES forecasts are wrong. A rejection of the null hypothesis means that some component of the input parameter is wrong [65]. However, ES depends on the VaR, and sensible ES forecasts are based on correctly specified VaR forecasts. Prior to the emerging literature on ES backtests, the expected shortfall was backtested through the value-at-risk at the same confidence level. If the VaR forecasts were rejected, then the ES forecasts would most likely be inaccurate as well [71]. However, acceptance of VaR forecasts does not guarantee acceptance of ES forecasts and pure ES backtests are therefore desired.

The previously described backtests of McNeil and Frey [62] and Nolde and Ziegel [65] have versions that only require VaR forecasts in addition to ES forecasts, and not additionally the volatility [62, 65, 73], the cumulative violation process [58, 74, 75] or the entire return distribution [60, 76, 77]. The Strict ESR and Intercept ESR backtests proposed by Bayer and Dimitriadis require no additional information other than the ES forecasts. Unless they are used for internal purposes, the backtests that require additional private information are not applicable for regulatory requirements as this information is not reported by the financial institutions [71, 72]. Additionally, the Intercept ESR test allows for one-sided hypothesis tests such that the regulators can avoid underestimation.

Conceptually, there exist several methods for backtesting ES. However, in addition to the aforementioned requirements, a substantial part of these tests are either impractical to implement computationally, require an unreasonable amount of available data, or both of these criteria apply. As the intercept ESR test does not require any extra information, allows for one-sided hypothesis testing and has a reasonable running time when implemented, only this will be considered in addition to the three VaR backtests. In the following, we denote the Intercept ESR test by simply the ESR test.

Chapter 7

Preliminary Data Analysis

In both the scientific and the public community, a common consensus around the effects of climate change has emerged over the last decades. Particularly in recent years, the interest in environmental issues such as sustainability, pollution and waste reduction has increased. An example of this is the Paris Agreement adopted in December 2015, an international agreement with the common goal of limiting global warming. More sustainable policies have led to changes in corporate activities. Both investors and consumers are today far more conscious of where they are placing their capital. Practitioners are for instance more interested in investing in so-called ESG stocks, which stands for Environmental, Social and corporate Governance stocks. As the name indicates, in addition to the desired earnings, the ethical action towards the environment is an essential part of the government for these companies. Thus, it is of interest to investigate whether this green focus has made any changes in certain parts of the stock market.

A distinct partition of the green and non-green stocks can be quite difficult as there is still no robust and unambiguous criterion for this. For instance, the ESG-score of a company can be quite ambiguous. Consider Tesla, which in public is known to have a green image. The company has set a new standard for the electric car industry and is defined as an ESG stock. However, there have been raised questions whether Tesla in fact should be an ESG stock. The company has made a significant contribution to battery technology and made electric cars the new in. The environmental part of the question is therefore highly earned. The social part, on the other hand, including labour treatment, is not fulfilled. The governance part is not fulfilled either, as buying Tesla stocks mainly adds to the wealth of the leader Elon Musk [78]. Making a clear distinction of the market into green and non-green stocks could therefore be challenging.

The data sets studied in this paper consist of stock prices from the companies NEL, Tesla, Tomra, Norsk Hydro, General Motors and Aker BP. The three former have an operation or produce products with more green and sustainable focus than compared to many of their competitors. In other words, they perform well on the environmental part of the ESG score. They will therefore be referred to as green stocks, and the three latter as the non-green stocks. The green stocks operate in three different sectors of the market. The non-green stocks have been chosen for their similarity to the industrial sectors of the green stocks, however, they are not perfect substitutes.

NEL, Tomra, Norsk Hydro and Aker BP are Norwegian companies traded on Oslo Boers, while Tesla and General Motors are U.S.-based and traded on the Nasdaq stock exchange and New York stock exchange, respectively. Stocks are only traded on so-called trading days, which does not include weekends or holidays. For simplicity, the data are assumed to be equally spaced even though this is an approximation of reality [20]. The size of the data sets is thus equivalent to the number of trading days in the given time period. In particular, the green data include historical closing prices of the NEL, Tesla and Tomra stocks from 17.11.2010 to 21.01.2021. The same applies to the non-green data. We also include data of the global financial crisis from 01.01.2006 for the NEL and Norsk Hydro stocks. [3]

[3] The data sets used in this study are provided by DNB.

This chapter aims to study the characteristics of the stock returns and identify the most important features. In order to obtain stationary time series, the stock prices are logarithmically transformed and differenced one lag as described in Section 2.1.1. The first two sections of this chapter study the green and non-green stocks in the period 17.11.2010 to 21.01.2021 separately. The final section gives a short analysis of the data sets from 01.01.2006 to 31.12.2010 for the NEL and Norsk Hydro stocks.

7.1 Green Stocks

The left panel of Figure 7.1 shows the historical stock prices of NEL, Tesla and Tomra from November 2010 to January 2021. For NEL we see a decline in the firm value from 2010 to 2014 and a very rapid increase starting in 2019 and continuing into 2021. For Tesla and Tomra, there can be seen a similar increase in firm value starting in 2020 and 2018, respectively. The stock market is commonly known to have a slow and steady increase with sudden rapid drops. In other words, there are usually not observed as rapid increases as drops. This means that the market tends to be negatively skewed. However, in recent years the opposite has been observed. There have been rapid increases in stock prices and generally more volatile periods in the market. The price dynamics of these green stocks are a good example of this feature.

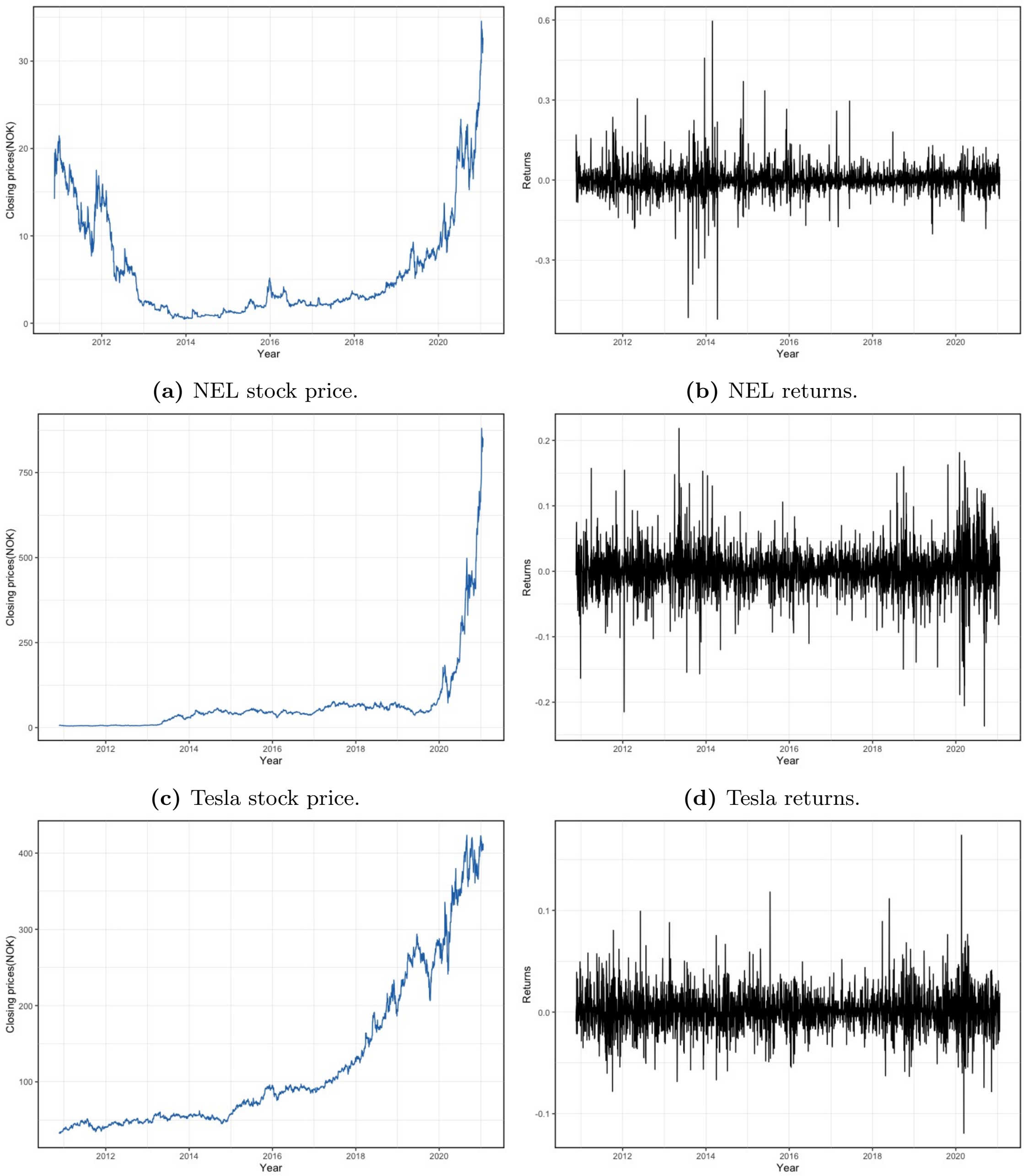
The plotted returns in the right panel of Figure 7.1 shows that the time series are mean reverting to zero and that the variance is roughly constant. For comparison, Table 7.1 summarises some descriptive statistics for the historical returns from 17.11.2010 to 21.01.2021 and shows that NEL returns have a daily mean of 0.031% and a daily standard deviation of 5.182%. A mean close to zero is therefore confirmed, and the daily standard deviation is significantly larger than the mean. The same applies to the Tesla and Tomra returns. However, we notice that Tesla has the highest reported mean. Additionally, we notice that NEL has the highest standard deviation on average, which implies the highest volatility.

Table 7.1: Descriptive statistics for the stock return series from 17.11.2010 to 21.01.2021. NH and GM are acronyms for Norsk Hydro and General Motors, respectively.

	NEL	Tesla	Tomra	NH	GM	Aker BP
Sample size	2656	2656	2656	2656	2656	2656
Mean	0.031%	0.187%	0.095%	0.004%	0.020%	0.085%
Standard Deviation	5.182%	3.428%	1.918%	1.972%	2.058%	2.856%
Min	-52.189%	-23.652%	-11.925%	-12.856%	-19.023%	-33.219%
Max	59.585%	21.838%	17.401%	11.689%	18.185%	42.934%
Skewness	0.504	-0.033	0.518	-0.115	-0.180	1.112
Kurtosis	22.721	6.371	5.618	3.431	10.418	31.123
Jarque Bera Test (p-value)	0.000	0.000	0.000	0.000	0.000	0.000
Ljung-Box Test Returns, 20 lags (p-value)	0.369	0.427	0.003	0.848	0.000	0.000
Ljung-Box Test Squared Returns, 20 lags (p-value)	0.000	0.000	0.000	0.000	0.000	0.000

Even though the plotted returns appear to be stationary and fluctuating around zero, some large positive and negative values can be observed. For NEL this happens during the second half of 2013 and the first half of 2014, while for Tomra, this occurs in 2020 in particular. Moreover, for Tesla large returns are observed in both these time periods. The plots clearly show periods of high and low volatility, which confirms the presence of volatility clustering.

The ACF plots of returns and squared returns in Figure 7.2 show that most of the correlation of returns lies inside the 95% confidence interval, while the correlation of squared returns lies outside the confidence interval and is significant. This is clear evidence of volatility clustering of the returns. The fact that there are few significant serial correlations in the ACF of the returns indicate uncorrelated return series, and the significant correlations in the ACF of squared returns imply dependent time series. This is common for time series of returns and is the feature that a univariate volatility model is designed to capture [29].



(a) NEL stock price. **(b)** NEL returns.

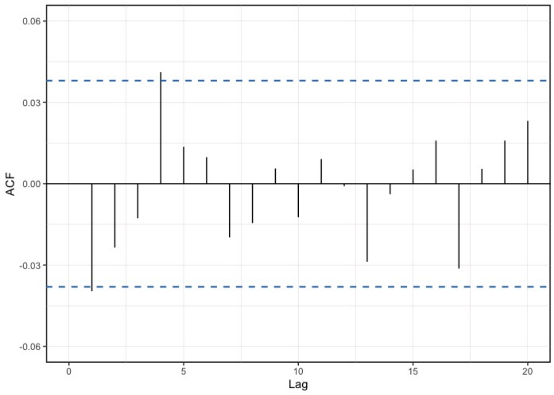
(c) Tesla stock price. **(d)** Tesla returns.

(e) Tomra stock price. **(f)** Tomra returns.

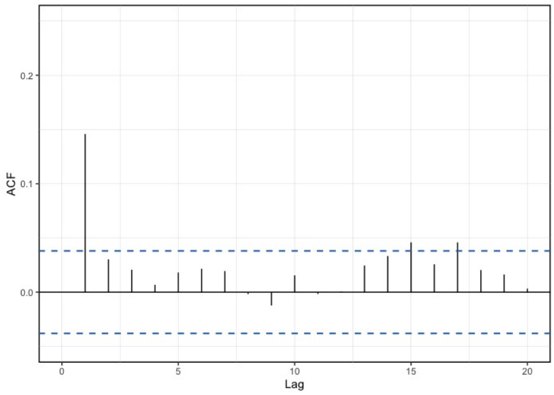
Figure 7.1: NEL, Tesla and Tomra closing prices and returns from 17.11.2010 to 21.01.2021.

More formally, the Ljung-Box test given in Table 7.1 shows that the null hypotheses of no volatility clustering are rejected as the p-values are below a significance level of 5% for squared returns. This implies that the squared returns are predictable and can be modelled. Moreover, for NEL and Tesla, we confirm that the returns are uncorrelated as the null hypotheses of white noise series are not rejected. For Tomra, however, the null hypothesis is rejected. But by restricting the null hypothesis to be weak white noise rather than white noise, the returns are uncorrelated. This is sufficient for the returns to be considered unpredictable [79].

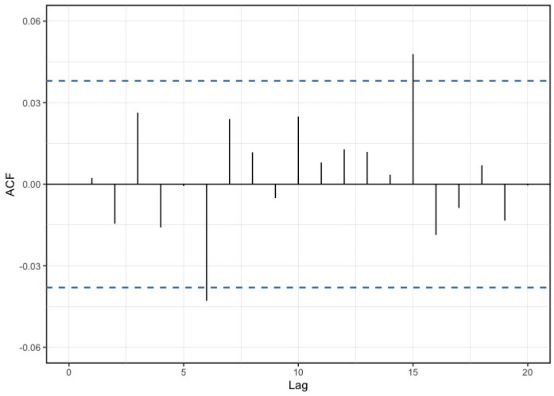
The non-normality of NEL returns can be assessed by viewing the qq-plot in Figure 7.3a. An inverted S-shape can easily be detected and indicates heavy tails for the return series. The same applies to Tesla and Tomra returns in Figure 7.3c and 7.3e, respectively. However, there seems to be less fatness in the tails of the two latter stocks as the inverted S-shape is less evident. The histograms of the returns in the right panel of Figure 7.3 confirm the leptokurtic feature as the data lie outside the blue normal curves in the centre and the tails of the distributions. The heavy-tailed student-t distribution with four degrees of freedom, indicated by the green curves, can be seen to capture the fat tails of the distributions more efficiently. We also notice the positive skewed empirical distribution of the NEL and Tomra returns from the histograms as the right tails are longer than the left tails. For the Tesla returns, the left tail are a bit longer than the right tail, indicating a slightly negative skewed distribution.



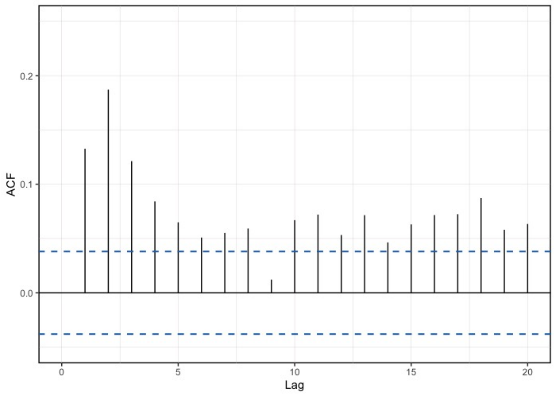
(a) ACF of NEL returns.



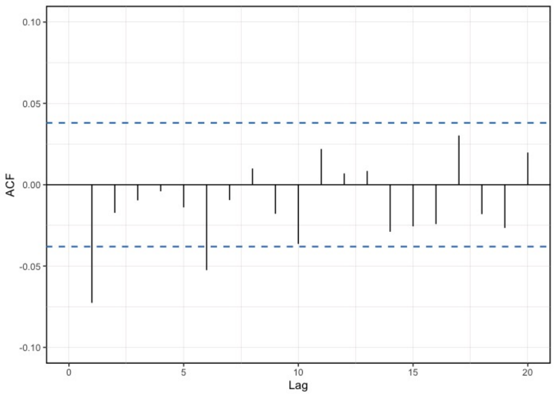
(b) ACF of squared NEL returns.



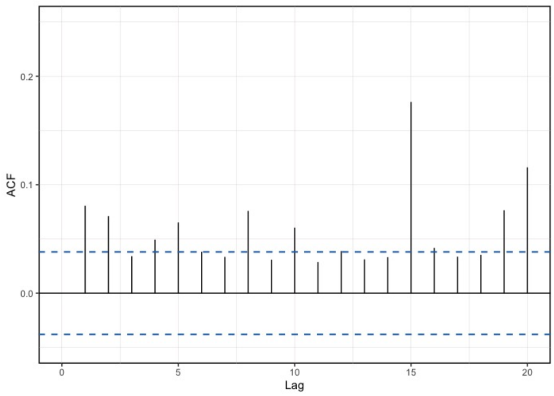
(c) ACF of Tesla returns.



(d) ACF of squared Tesla returns.



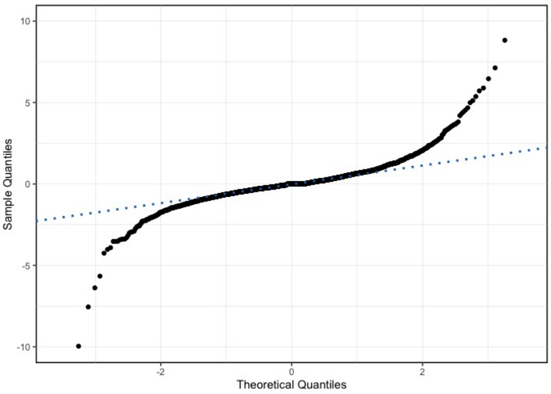
(e) ACF of Tomra returns.



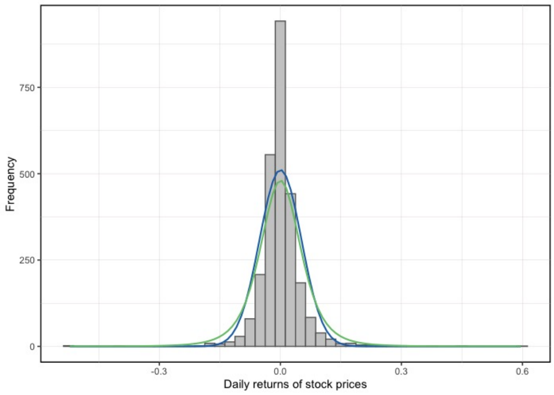
(f) ACF of squared Tomra returns.

Figure 7.2: Sample ACF of returns and squared returns associated to the daily closing values of the NEL, Tesla and Tomra stocks from 17.11.2010 to 21.01.2021. The horizontal dashed lines denote the two standard error limits of the sample ACF.

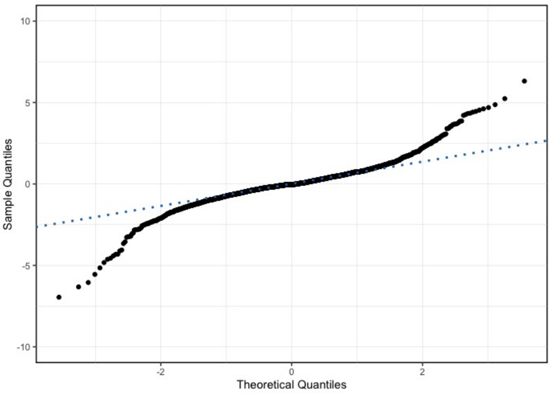
The reported kurtosises in Table 7.1 are larger than three for all stocks and coincide with the graphical evidence of heavy tails. As the U.S. stock exchanges are more liquid than the Norwegian market, we would expect the kurtosis of NEL and Tomra to be larger than for Tesla. This is not the case for Tomra. Furthermore, the table shows a positive skewness of 0.504 and 0.518 for NEL and Tomra, respectively, while Tesla has a slightly negative skewness of -0.033. This indicates an asymmetric distribution for the returns. The market seems to experience some changes as contrary to the expectation of negative skewness, both NEL and Tomra have positive skewness. More formally, the non-normality of the data can be verified by the reported Jarque Bera Test in Table 7.1. The p-values are below a significance level of 5%, and the null hypotheses of normally distributed returns are thus rejected.



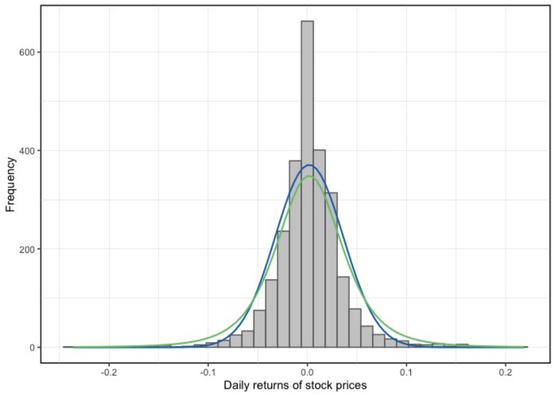
(a) Normal qq-plot for NEL returns.



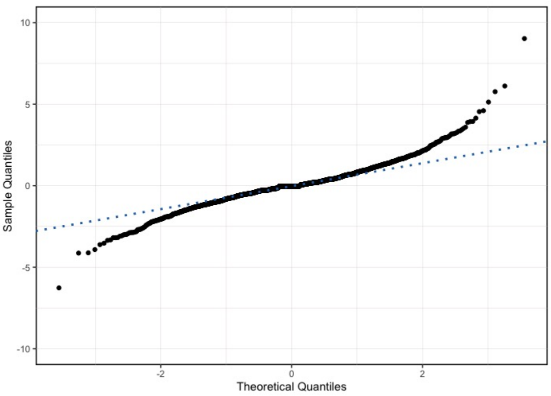
(b) Histogram for NEL returns.



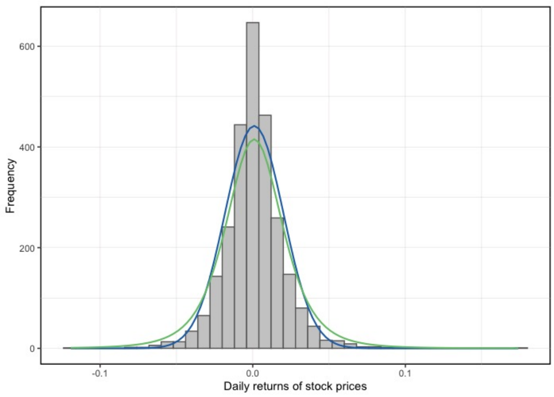
(c) Normal qq-plot for Tesla returns.



(d) Histogram for Tesla returns.



(e) Normal qq-plot for Tomra returns.

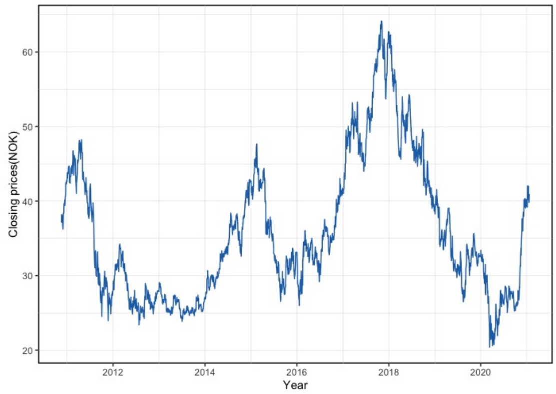


(f) Histogram for Tomra returns

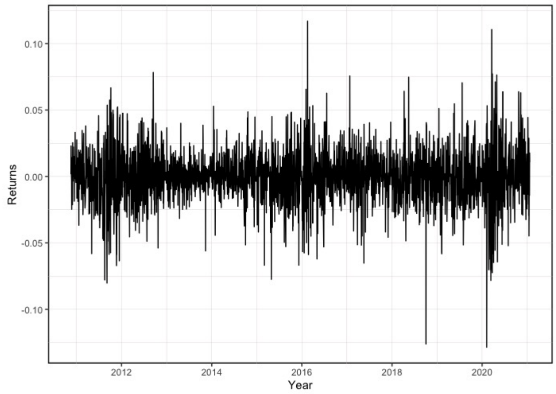
Figure 7.3: Normal qq-plot and histogram for NEL, Tesla and Tomra stock returns from 17.11.2010 to 21.01.2021. The blue line in the histograms indicates the normal distribution, while the green line indicates a student-t distribution with four degrees of freedom.

7.2 Non-Green Stocks

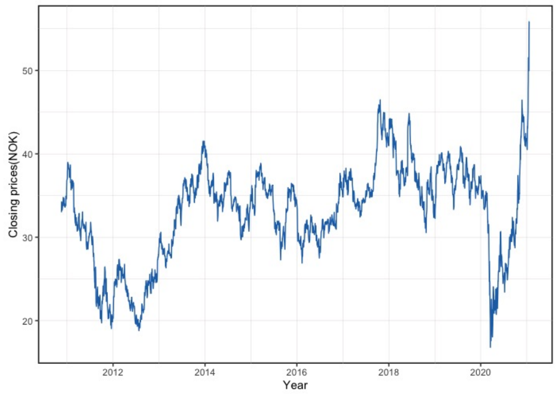
Figure 7.4 shows stock prices and returns of the companies Norsk Hydro, General Motors and Aker BP from 17.11.2010 to 21.01.2021. Compared to the green stock price dynamics in the left panel of Figure 7.1, the non-green stocks have a different price dynamic. The stock prices seem more stable with sudden bursts both up and down for Norsk Hydro and General Motors. This could explain the negative skewness of their respective returns reported in Table 7.1. However, we also notice the rapid increase in the value of General Motors after the outbreak of COVID-19 in the spring of 2020. Furthermore, Aker BP seems to have experienced more rapid increases than drops in the stock price, and the returns are thus positively skewed with a value of 1.112.



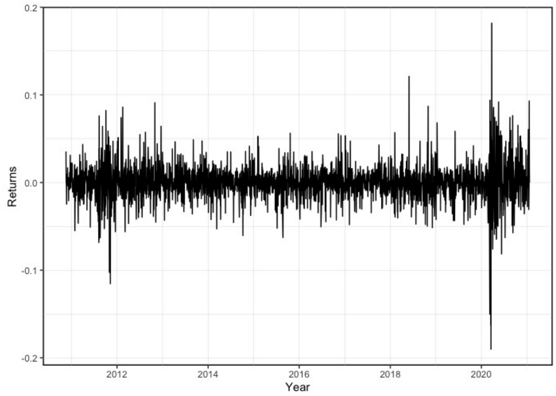
(a) Norsk Hydro stock price.



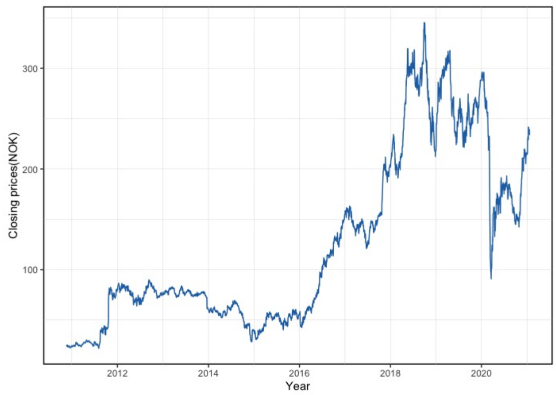
(b) Norsk Hydro returns.



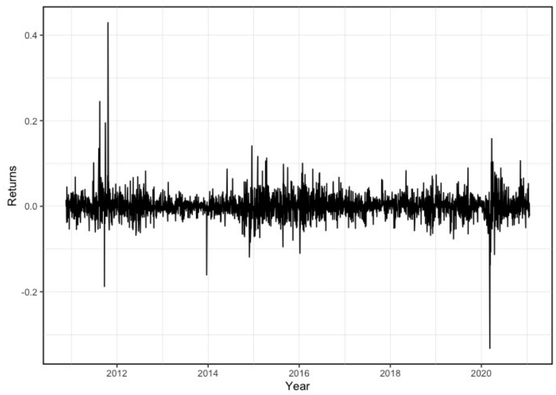
(c) General Motors stock price.



(d) General Motors returns.



(e) Aker BP stock price.

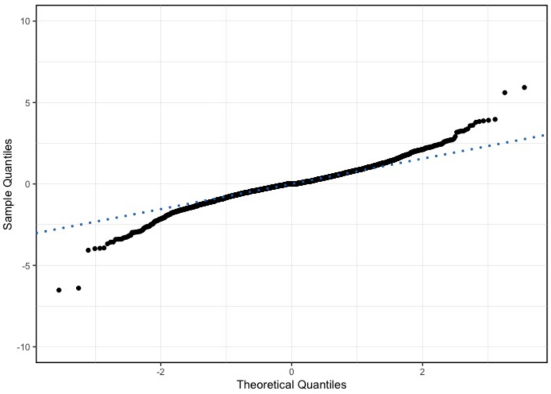


(f) Aker BP returns.

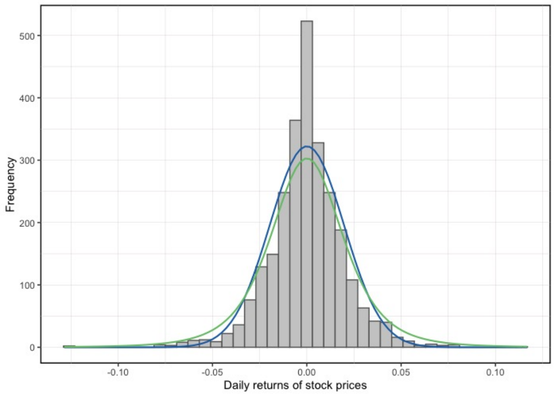
Figure 7.4: Norsk Hydro, General Motors and Aker BP closing prices and returns from 17.11.2010 to 21.01.2021.

The Jarque-Bera test shows that, similarly to the green stocks, the non-green stocks reject the null hypothesis of normality. However, a few differences in the qq-plots and histograms in Figure 7.5 can be observed. As mentioned in Section 3.2, the most dominant deviations from normality are the higher peaks around the centre of the distribution and the fatter tails. The qq-plots show that the returns of Norsk Hydro are closest to the normal line. In other words, Norsk Hydro has the least leptokurtic returns. The kurtosis also confirms this as Norsk Hydro has the smallest reported value. This feature may be caused by the fact that Norsk Hydro are traded on other exchanges, such as OTCQX International in the U.S., in addition to Oslo Boers [80]. This is a more liquid market, and we expect less leptokurtosis for stocks that are traded here. Moreover, the histogram for Aker BP clearly shows more peakedness around the centre of the distribution. General Motors also seems to experience a similar trend but to a lesser extent. This could partially explain the high kurtosis of Aker BP and General Motors reported in Table 7.1. In fact, Aker BP has the highest reported kurtosis with a value of 31.123. The histogram of Aker BP also shows a positive skewness.

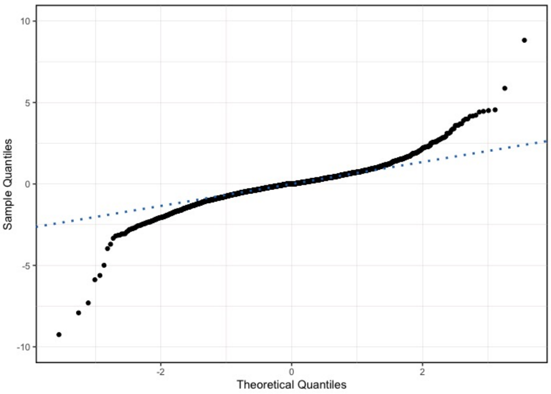
Besides the differences noted above, the non-green stocks seem to provide similar results to the green stocks when performing an identical data analysis. ACF plots of returns and squared returns for the non-green stocks are provided in Appendix B.



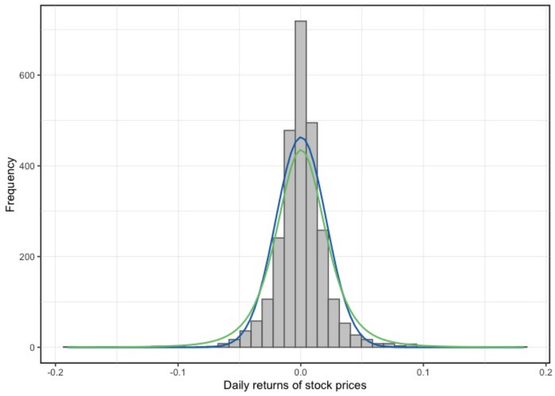
(a) Normal qq-plot for Norsk Hydro returns.



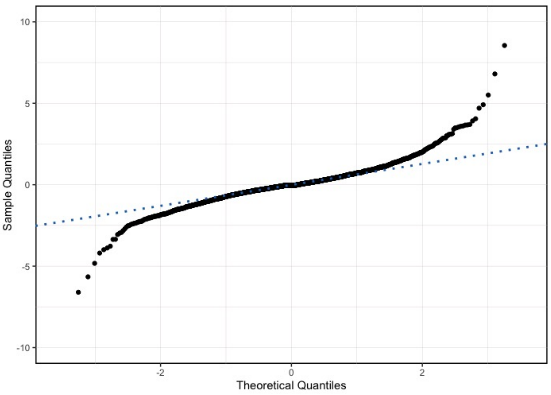
(b) Histogram for Norsk Hydro returns.



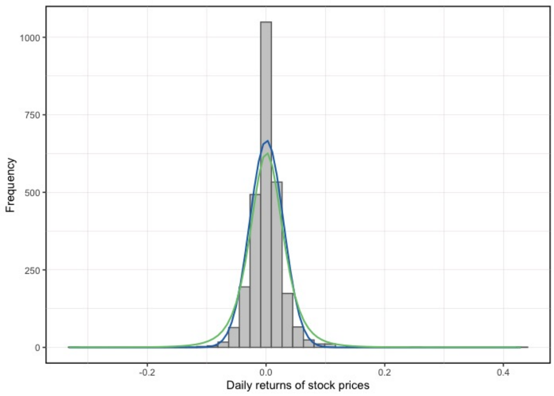
(c) Normal qq-plot for General Motors returns.



(d) Histogram for General Motors returns.



(e) Normal qq-plot for Aker BP returns.



(f) Histogram for Aker BP returns.

Figure 7.5: Normal qq-plot and histogram for Norsk Hydro, General Motors and Aker BP stock returns from 17.11.2010 to 21.01.2021. The blue line in the histograms indicates the normal distribution, while the green line indicates a student-t distribution with four degrees of freedom.

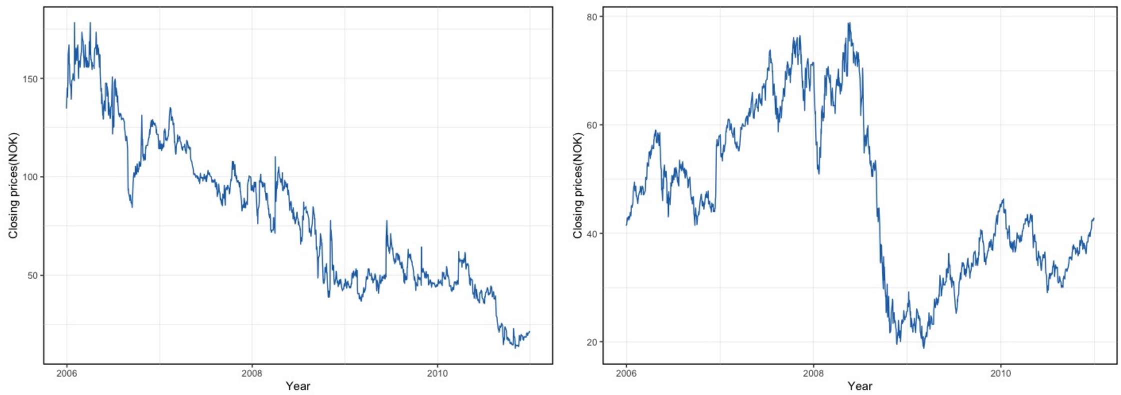
7.3 Data from the Global Financial Crisis

For investigating the effect of the estimation sample used in model fitting, we perform a short analysis of the data including the global financial crisis between mid 2007 and early 2009 for the NEL and Norsk Hydro stocks. In particular, the data sets cover the period 01.01.2006 to 31.12.2010. Table 7.2 gives some descriptive statistics for the stock return series, and Figure 7.6 plots the stock prices.

Table 7.2: Descriptive statistics for the stock return series of NEL and Norsk Hydro from 01.01.2006 to 31.12.2010.

	NEL	Norsk Hydro
Sample Size	1305	1305
Mean	-0.142%	0.002%
Standard Deviation	5.065%	3.047%
Min	-26.620%	-16.4724%
Max	47.2794%	18.764%
Skewness	2.026	-0.092
Kurtosis	20.159	5.107
Jarque Bera Test (p-value)	0.000	0.000
Ljung-Box Test Returns, 20 lags (p-value)	0.001	0.000
Ljung-Box Test Squared Returns, 20 lags (p-value)	0.000	0.000

Figure 7.6a shows a steady decrease with some minor ups and downs for the NEL stock from 2006 up to 2011. As this is the opposite of the expected behaviour of equities, it may explain the high positive skewness reported in Table 7.2. Contrary, we see a very rapid decrease in the firm value of Norsk Hydro during 2008 and the global financial crisis. We also see a steady increase in the Norsk Hydro stock for the periods 2006 to early 2008 and 2009 up to 2011. This is a more typical stock price behaviour and may explain the reported negative skewness.



(a) NEL stock price. **(b)** Norsk Hydro stock price.

Figure 7.6: NEL and Norsk Hydro closing prices from 01.01.2006 to 31.12.2010.

Compared to the data set presented in Section 7.1, the mean of NEL returns is negative and of larger magnitude with the data set from the global financial crisis. This reflects the decrease in the stock price of NEL during the financial crisis. The Ljung-Box test of the returns rejects the null hypothesis of a white noise series, but as before, the returns are uncorrelated by restricting the null hypothesis to be weak white noise rather than white noise. The same applies to the Norsk Hydro returns. The standard deviation and kurtosis of Norsk Hydro are slightly larger for the global financial crisis data. This indicates more fluctuations in the stock price of Norsk Hydro during the financial crisis. Except for these findings, a preliminary data analysis of the financial crisis data gives similar results as in the two previous sections. Further analysis using the financial crisis data can be found in Section 8.4.

Chapter 8

Results and Discussion

This chapter reports the results of fitting GARCH volatility models to the stock data sets presented in the data analysis. At first, an in-sample analysis using information criteria is outlined in Section 8.1. The superior volatility models from the in-sample analysis are then evaluated out-of-sample in Section 8.2 using value-at-risk and expected shortfall risk measures. Section 8.3 gives a comparison of models across market sectors, and the presented results will be summarised and discussed. For these analyses, the data sets presented in Section 7.1 and 7.2 are applied. Section 8.4 provides a short analysis of the effect of the estimation sample. For this analysis, we apply the data sets presented in Section 7.3.

8.1 In-sample Analysis

For ease of notation, acronyms are introduced for each of the GARCH models. These are given in Figure 8.1. The models will be referred to as the normal, student-t or skew student-t GARCH models depending on the distributional assumption. Moreover, depending on both the type of GARCH model and the distributional assumption, notations like GARCH-n, EGARCH-t or APARCH-st will also be used.

G_{n} : GARCH(1,1) with normally distributed innovations
G_{t} : GARCH(1,1) with student-t distributed innovations
G_{st} : GARCH(1,1) with skew student-t distributed innovations
E_{n} : EGARCH(1,1) with normally distributed innovations
E_{t} : EGARCH(1,1) with student-t distributed innovations
E_{st} : EGARCH(1,1) with skew student-t distributed innovations
A_{n} : APARCH(1,1) with normally distributed innovations
A_{t} : APARCH(1,1) with student-t distributed innovations
A_{st} : APARCH(1,1) with skew student-t distributed innovations

Figure 8.1: Acronyms for different GARCH models.

8.1.1 Implementation and Parameter Choices

A substantial part of the studies on GARCH volatility models reports that models with lag parameters $p = q = 1$ usually perform the best [81]. Moreover, when fitting a statistical model, we prefer a parsimonious model to avoid overfitting. Only models which have parameters $p = 1, 2$ and $q = 1, 2$ will therefore be considered in this thesis. For each model, the innovations are assumed to be normal, student-t and skew student-t distributed.

The last 1000 observations, equivalent to approximately the last four years of data, are reserved for validation and backtesting from each stock data set. Since the return series include 2656 daily observations, each GARCH model is estimated with 1656 observations. In particular, the in-sample analysis is performed on the data from 17.11.2010 to 23.03.2017, while the data from 24.03.2017 to 21.01.2021 are used for backtesting.

8.1.2 In-sample Results

Tables C.1-C.3 in Appendix C report information criteria results for NEL stock returns in the period 17.11.2010 to 23.03.2017. They contain the results for all possible combinations of the given parameters for each of the three GARCH models discussed in Chapter 3.1. The models that perform best, i.e. have the lowest information criteria, are highlighted with bold numbers. The same applies for Tesla, Tomra, Norsk Hydro, General Motors and Aker BP in Tables C.4-C.6, C.7-C.9, C.10-C.12, C.13-C.15 and C.16-C.18, respectively. From these tables it is evident that within each type of GARCH model, AICc chooses the models with parameters $p = q = 1$ and student-t distributed innovations. This applies for all stocks, except for NEL, where the skew student-t distributed innovations seem to be preferred. Notice also that for some data sets, if AIC is used as the model criterion instead of AICc, then as expected, less parsimonious models would be preferred.

Using BIC as model criterion leads to approximately the same conclusions. One exception can be noticed for the Tomra stock where the GARCH(1,2) and EGARCH(2,1) models with student-t distributed innovations seem to be slightly better than the other models within their model type. However, the differences within models of different lag orders are smaller than compared to using a different distributional assumption. Based on these results, only GARCH models with parameters $p = q = 1$ will be considered further in this thesis. Hereby they will be referred to as the GARCH, EGARCH and APARCH models, respectively.

When comparing the BIC of all three GARCH models on the same data set, the EGARCH model appears to be the most optimal model for all stocks. This can be seen in Table 8.1 where the models with the lowest AICc and BIC from the tables in Appendix C are reported. Using AICc for model comparison, the standard GARCH model seems to be slightly better than the EGARCH model for NEL, Tesla, Tomra and General Motors. For Norsk Hydro the GARCH and EGARCH models seem equivalent, and for Aker BP the EGARCH model is somewhat better. However, the model performances are quite similar.

Table 8.1 also displays estimated coefficients with corresponding p-values for the models that perform best according to the AICc. The p-values are associated with the hypothesis of whether the coefficient estimates are equal to zero or not. The parameter a_1 tells us how volatility reacts to new information in the market, and b_1 how much volatility remembers from the past [82]. Both parameters are positive and significant for nearly all models in all stocks. This indicates presence of GARCH effects in the stock returns, which implies heteroscedasticity, i.e. non-constant volatility, in the stock market. Moreover, for nearly all models the sum of a_1 and b_1 is approximately one, and b_1 is close to 0.9. This indicates that volatility shocks are persistent. For daily returns, a large value of σ_{t-1}^2 will then be followed by a large value of σ_t^2, or equivalently, large changes in the returns will be followed by large changes and vice versa. This confirms the volatility clustering found in the data analysis in Chapter 7. Brooks suggests that a reasonable explanation for this behaviour is that the information arrivals, resulting in price changes, arise in bundles rather than being evenly spaced through time [30].

The presence of leverage effects and power effects can be examined through the estimated γ_1 and δ coefficients. All estimated γ_1 parameters are different from zero, and most of them have a low corresponding p-value implying rejection of the null hypothesis. This implies that leverage effects are present. Moreover, the estimates are positive such that past negative shocks have a stronger impact on volatility than past positive shocks. However, an exception is that for Tesla returns the γ_1 coefficients are not significant. This behaviour may be connected to the positive skewness of the returns. It might be the case that there should be a leverage effect in the opposite direction. Furthermore, the δ parameters are estimated to be below two for all return series. This means that autocorrelations in the absolute returns are evident. These findings show how the parameter estimates change when different input data are used and express the models' ability to fit leverage effects and power effects in the data.

Table 8.1: GARCH coefficient estimates and information criteria results for NEL, Tesla, Tomra, Norsk Hydro, General Motors and Aker BP stock returns from 17.11.2010 to 23.03.2017.

Model	a_0	a_1	b_1	γ_1	δ	ν	ξ	BIC	AICc
				Panel A: NEL returns					
G_{st}	0.00132	0.61815	0.38085			2.52605	1.14347	-3.447	**-3.427**
	(0.0114)	(0.0000)	(0.0024)			(0.0000)	(0.0000)		
E_{st}	-1.50042	-0.01456	0.72348	0.64989		2.45813	1.14354	**-3.448**	-3.416
	(0.0002)	(0.8005)	(0.0000)	(0.0000)		(0.0000)	(0.0000)		
A_{st}	0.01137	0.48924	0.50925	0.06560	1.16597	2.44938	1.14274	-3.444	-3.399
	(0.1276)	(0.0000)	(0.0000)	(0.4526)	(0.0000)	(0.0000)	(0.0000)		
				Panel B: Tesla returns					
G_t	0.00001	0.03483	0.95408			3.36551		-4.337	**-4.325**
	(0.0069)	(0.0002)	(0.0000)			(0.0000)			
E_t	-0.13309	-0.02053	0.98064	0.11357		3.37922		**-4.338**	-4.318
	(0.0990)	(0.34552)	(0.0000)	(0.6325)		(0.0000)			
A_t	0.00032	0.06432	0.93660	0.16477	1.20598	3.38427		-4.330	-4.299
	(0.5154)	(0.0018)	(0.0000)	(0.3462)	(0.0022)	(0.0000)			
				Panel C: Tomra returns					
G_t	0.00000	0.02160	0.97740			3.85935		-5.310	**-5.299**
	(0.7880)	(0.0000)	(0.0000)			(0.0004)			
E_t	-2.84005	-0.10552	0.64209	0.38628		3.69455		**-5.316**	-5.296
	(0.0723)	(0.0054)	(0.0012)	(0.0001)		(0.0000)			
A_t	0.00166	0.23563	0.43351	0.24706	1.38395	3.71364		-5.311	-5.280
	(0.5374)	(0.0010)	(0.0907)	(0.0428)	(0.0002)	(0.0000)			
				Panel D: Norsk Hydro returns					
G_t	0.00000	0.02160	0.97740			3.85935		-5.294	**-5.283**
	(0.78800)	(0.0000)	(0.0000)			(0.0004)			
E_t	-2.84005	-0.10552	0.64209	0.38628		3.69455		**-5.303**	**-5.283**
	(0.0723)	(0.0054)	(0.0011)	(0.0001)		(0.0000)			
A_t	0.00166	0.23563	0.43351	0.24706	1.38395	3.71364		-5.298	-5.267
	(0.5374)	(0.0009)	(0.0907)	(0.0427)	(0.0002)	(0.0000)			
				Panel E: General Motors returns					
G_t	0.00000	0.03740	0.95274			4.34846		-5.346	**-5.335**
	(0.54014)	(0.0000)	(0.0000)			(0.0000)			
E_t	-0.07184	-0.05153	0.99109	0.06877		4.36278		**-5.352**	-5.332
	(0.0000)	(0.0002)	(0.0000)	(0.0000)		(0.0000)			
A_t	0.00060	0.03802	0.96201	0.86502	0.76760	4.39385		-5.349	-5.317
	(0.0028)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)			
				Panel F: Aker BP returns					
G_t	0.00001	0.05705	0.94077			3.67901		-4.673	-4.662
	(0.64590)	(0.0637)	(0.0000)			(0.0000)			
E_t	-0.00834	-0.05177	0.99877	0.07952		3.93603		**-4.695**	**-4.675**
	(0.0000)	(0.0000)	(0.0000)	(0.0017)		(0.0000)			
A_t	0.00004	0.04742	0.96405	0.59625	1.12926	4.07154		-4.691	-4.660
	(0.99159)	(0.92119)	(0.0007)	(0.4098)	(0.9410)	(0.0000)			

The fat tails of the return distributions, expressed through the shape parameters ν, are in the range 2.45 to 4.4. For most stocks, however, it seems to be in the range 3.25 to 4.0. NEL appears to have particularly fat tails as the number of degrees of freedom is around 2.5, and General Motors is the least leptokurtic stock with degrees of freedom around 4.4.

Table 8.1 indicates that General Motors returns seem to have a higher significance for the parameter estimates, in addition to having the largest estimated shape parameter. These findings coincide with the theory of stock markets. The U.S. stock market is more liquid than the Norwegian stock market, and it should therefore have less heavy tails and be more stable in modelling. The same observations might be expected for Tesla returns, but this is not entirely true. A reason for this could be the phase of the life cycle the company is in. Tesla is in another phase of the business development, namely a growth phase rather than a steady-state phase. There has been substantial growth in their industry, and the company is valued based on unknown future earnings. This naturally leads to higher volatility. Similar inference can be made for the NEL stock. This may be parts of the reason NEL and Tesla have the lowest estimated ν coefficients.

As the skew student-t distributed innovations were preferred for the NEL stock, the coefficient ξ was also estimated. For all models, the coefficient is significant and estimated to be somewhat above one. This indicates positive skewness for NEL, which was expected from the reported skewness in the data analysis in Section 7.1.

8.2 Out-of-sample Analysis

To further examine the superior GARCH model for each stock return series, the value-at-risk and expected shortfall risk measures are applied. VaR and ES estimates are obtained according to the procedures described in Chapter 5.

8.2.1 Parameter and Implementation Choices

To assess the validity of VaR and ES estimates, they are backtested as described in Chapter 6. We follow the common convention to use the 1% and 5% percentiles, and the confidence level $1-q$ is set to 95%. The VaR backtests are run in R based on functions in the "rugarch" package by Ghalanos [36], while the ESR test is run with the "esback" package by Bayer and Dimitriadis [83]. Using a bootstrap sample of 1 000 observations in the ESR test seems to be sufficient and give reliable results as the deviations of the model statistics remain small. Additionally, a bootstrap sample of 10 000 observations would be computationally challenging and require an unreasonable running time. All models that pass both the unconditional coverage test and the independence test are highlighted by a shaded background for the VaR forecasting results. Moreover, the model with the highest degree of confidence in the conditional coverage test is highlighted with bold numbers. The latter also applies to the ES forecasts that pass the ESR test with the highest degree of confidence.

A rolling window of 1656 daily observations is used to re-estimate each GARCH model, while for historical simulation, the estimation window is set to 1000 observations. Daily risk forecasts are then evaluated for the final 1000 observations. When applying backtesting to obtain VaR and ES estimates, the first step is to predict the one-day-ahead volatility for each day in the testing window. The precise way to do this is to re-estimate the GARCH models for each day. A large testing window could therefore lead to thousands of estimations. However, the estimation set is relatively large, and the parameter estimates will presumably change slowly. A commonly used method to reduce the computational cost of re-estimating the GARCH models for each day is to re-estimate the model parameters only every 50 or 100 days. In this paper, the parameters are re-estimated every 100 days.

8.2.2 NEL

Figure 8.2 views the negative 5% VaR estimates for the period 24.03.2017 to 21.01.2021. The figure demonstrates the inferior performance of the historical simulation approach compared to the standard GARCH model. The reason may be that HS weights the observations equally, and therefore does not adequately adjust to the structural changes in the volatility. Thus, HS is only implemented to be used as a comparison for the other models. Another feature to notice is that the GARCH models fitted with t-distributed innovations seem to adjust more appropriately to the changes in volatility than the normal GARCH model. Considering the in-sample analysis for NEL, this is to be expected.

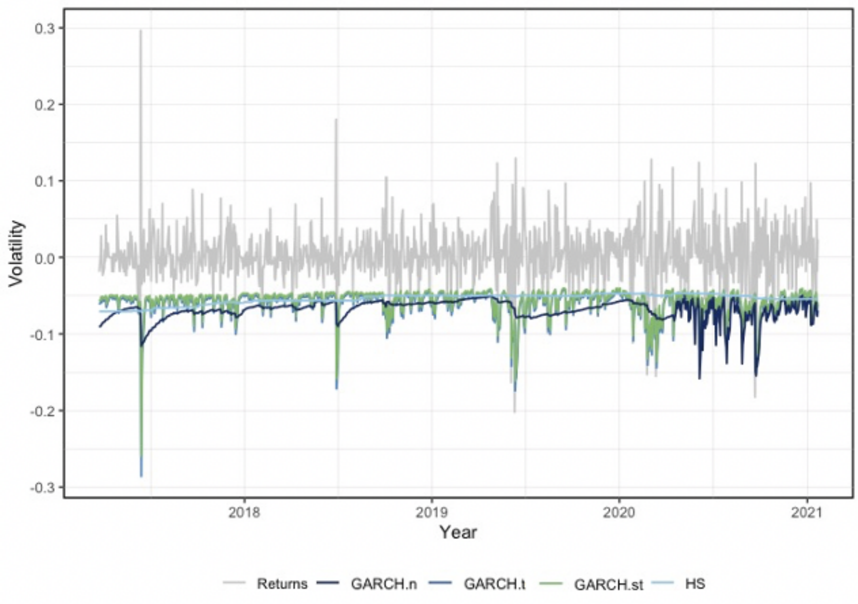


Figure 8.2: Negative 5% VaR estimates from the VaR models based on HS, as well as the parametric approach with the standard GARCH model fitted with normal, student-t and skew student-t distributed innovations plotted together with NEL returns from March 2017 to January 2021.

Value-at-Risk Backtests

Table 8.2 displays the actual number of VaR violations, the violation ratios and the backtesting results of calculating VaR with both the non-parametric and parametric approaches described in Chapter 5. This includes historical simulation as well as the parametric approach of using all three GARCH models with assumptions of normal, student-t and skew student-t distributed innovations. Panels A and B give the results at the 5% and 1% percentile levels, respectively.

Table 8.2: 1% and 5% VaR exceedance for NEL returns from 17.11.2010 to 21.01.2021.

Model	HS	G_n	G_t	G_{st}	E_n	E_t	E_{st}	A_n	A_t	A_{st}
Panel A: Results at percentile level α=0.05, Expected Violations 50										
Violations	84	27	37	45	21	34	41	31	35	42
VR	1.68	0.54	0.74	0.90	0.42	0.68	0.82	0.62	0.70	0.84
UC	0.204	0.000	0.048	0.461	0.000	0.014	0.178	0.003	0.022	0.233
IND	0.189	0.758	0.731	0.984	0.342	0.122	0.061	0.968	0.111	0.515
CC	0.188	0.001	0.134	**0.762**	0.000	0.015	0.070	0.013	0.020	0.397
Panel B: Results at percentile level α=0.01, Expected Violations 10										
Violations	12	10	6	6	6	4	5	12	4	5
VR	1.20	1.00	0.60	0.60	0.60	0.40	0.50	1.20	0.40	0.50
UC	1.000	1.000	0.170	0.170	0.170	0.003	0.079	0.538	0.003	0.079
IND	0.085	0.085	0.788	0.788	0.788	0.858	0.823	0.589	0.858	0.823
CC	0.226	0.226	0.376	0.376	0.376	0.094	0.208	**0.715**	0.094	0.208

According to Danielsson, a good model is indicated by a violation ratio in the interval $[0.8, 1.2]$, whereas an imprecise model gives VR < 0.5 or VR > 1.5 [3]. We will use this terminology in the following model evaluation. By analysing the violation ratios in Panel A of Table 8.2, all VaR models based on GARCH overestimate the 5% VaR as the violation ratios are lower than 1. The skew student-t GARCH models seem to overestimate the least, and the standard GARCH-st model has the VR closest to 1. Only the skew student-t GARCH models have violation ratios that are classified as good models. HS appears to underestimate, and as the violation ratio exceeds 1.5, it is an imprecise model. Panel B shows that the standard GARCH-n model has a violation ratio of 1 at the 1% level. Except for the APARCH-n model and HS that underestimate, the remaining models overestimate the VaR at the 1% level. Moreover, the student-t EGARCH and APARCH models seem to be imprecise VaR models as the violation ratios are below 0.5.

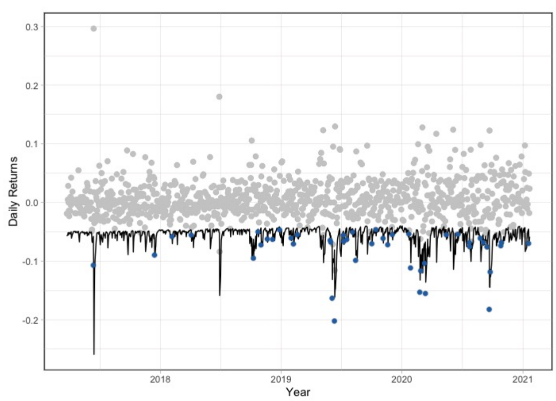
The results of the VaR backtests at the 5% level show that only the skew student-t GARCH models and HS pass the backtesting both for coverage and independence. However, a good model should give the expected number of VaR violations in addition to good results in backtesting. The violation ratio of HS appears too high for this to be a good model. Provided that both the coverage test and the independence test are passed, letting the joint conditional coverage test be the more important test, the standard GARCH-st model is least likely to be rejected. This coincides with the in-sample findings of NEL returns when using AICc as the more important information criterion.

At the 1% percentile level, all models except the EGARCH-t and APARCH-t models pass both the coverage test and the independence test. The normal APARCH model seems to perform best, while there are only minor differences in the p-values of the other models. This contrasts with the in-sample analysis where the t-distributions outperform the normal distribution. However, as we are far out in the tail of the distribution, modelling volatility at the 1% level is more difficult than at the 5% level. More uncertainty could therefore be associated with these results. Moreover, the in-sample and out-of-sample analyses may give different conclusions due to the price dynamics of the NEL stock. GARCH models are known for not adequately modelling the volatility of assets with a rapid price increase.

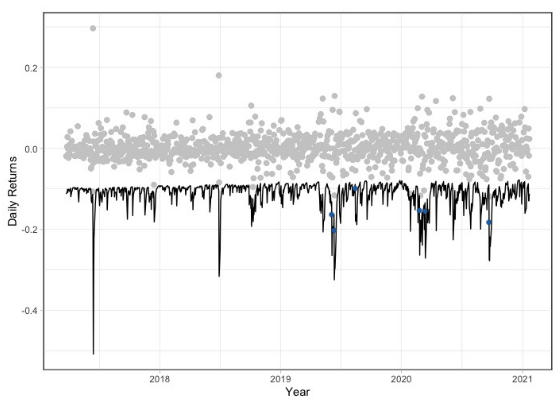
Nearly all VaR models based on GARCH seem to overestimate the risk at both percentile levels. Although some of the models seem to perform better than others, it is difficult to explicitly state whether either of them is significantly better than the other.

For a visual presentation of the VaR results, VaR exceedance plots will be given. For each of the stocks presented, the model that performs best in the conditional coverage test at the 5% level will be used to produce VaR estimates. Additional VaR exceedance plots for the other GARCH models are included in Appendix D.

Figure 8.3 plots the negative 1% and 5% VaR forecasts obtained with the skew student-t GARCH model. A visual examination shows that the VaR violations in blue appear to be evenly spread out in time and not clustered at the 5% level. This is in contrast to the 1% level, where there seems to be some clustering. Moreover, a few larger VaR violations can be observed at the 5% level, whereas no large VaR violations are observed at the 1% level.



(a) VaR(0.05)



(b) VaR(0.01)

Figure 8.3: VaR exceedance plots for the skew student-t GARCH(1,1) model on NEL stock March 2017 to January 2021. The black line indicates the negative VaR estimates, the gray marks the returns and the VaR violations are plotted as blue marks.

Expected Shortfall Backtest

Table 8.3 displays the backtest results of the 5% and 1% ES estimates of NEL returns in Panels A and B, respectively. To understand the ESR test results, it should be remembered that high p-values indicate a good model, while low p-values suggest the opposite. Since a bootstrap test is used, the resulting p-values could have minor differences from time to time, and the size of the deviation between the models should not be interpreted too mechanically.

At the 5% percentile level, all ES models give quite high p-values in the ESR test, and there are only minor differences in model performance. The student-t EGARCH and APARCH models have a p-value of 1 and seem to perform best. However, these models have inferior performance in the 5% VaR analysis as neither the unconditional nor conditional coverage tests are passed. This implies that the models are better at capturing the tail of the return distribution rather than determining a fixed percentile. Ideally, we would like a model that is good at both, and as the skew student-t GARCH models have high p-values in the ESR backtest and good performance in the VaR analysis, these models may be a better choice. Similar conclusions can be made from the 1% percentile level. However, the normal GARCH models have low p-values in the ESR test and seem less competitive with the t-distributed GARCH models.

Table 8.3: 1% and 5% ES exceedance for NEL returns from 17.11.2010 to 21.01.2021.

Model	HS	G_n	G_t	G_{st}	E_n	E_t	E_{st}	A_n	A_t	A_{st}
				Panel A: Results at percentile level α=0.05						
ESR	0.902	0.657	0.999	0.992	0.992	**1.000**	0.998	0.851	**1.000**	0.996
				Panel B: Results at percentile level α=0.01						
ESR	0.717	0.004	0.999	0.984	0.346	**1.000**	0.994	0.058	**1.000**	0.998

8.2.3 Tesla

Value-at-Risk Backtests

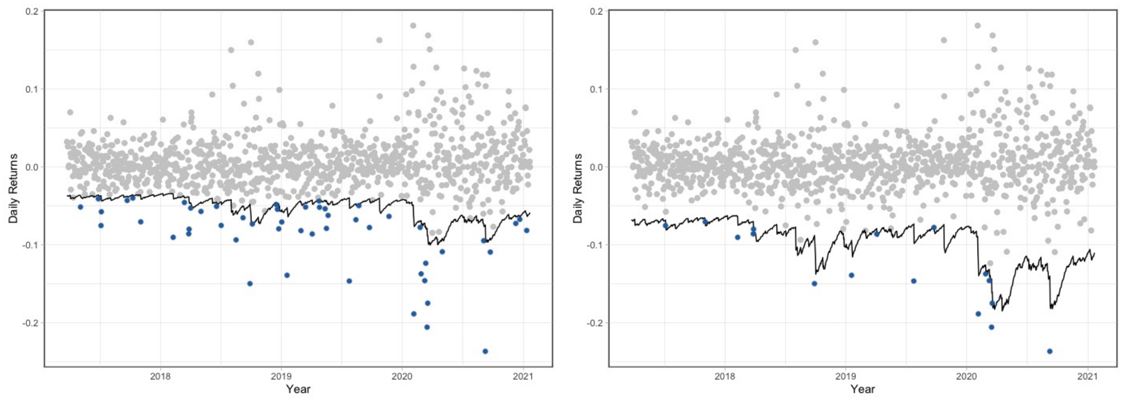
The backtest results of the VaR estimates for Tesla returns are presented in Table 8.4. At the 5% level, all normally distributed GARCH models slightly overestimate the VaR as the violation ratios are below 1. Except for the GARCH-st model, which has a VR equal to 1, the remaining models underestimate the VaR. At the 1% level, all models underestimate the VaR. The normal GARCH models, in addition to the standard GARCH-t model and HS, seem to be imprecise as the violation ratios are above 1.5.

Table 8.4: 1% and 5% VaR exceedance for Tesla returns from 17.11.2010 to 21.01.2021.

Model	HS	G_n	G_t	G_{st}	E_n	E_t	E_{st}	A_n	A_t	A_{st}
				Panel A: Results at percentile level α=0.05, Expected Violations 50						
Violations	89	45	51	50	45	56	55	45	52	51
VR	1.78	0.90	1.02	1.00	0.90	1.12	1.10	0.90	1.04	1.02
UC	0.027	0.461	0.885	1.000	0.461	0.393	0.475	0.461	0.773	0.885
IND	0.034	0.984	0.800	0.747	0.984	0.469	0.508	0.984	0.853	0.682
CC	0.009	0.762	**0.958**	0.949	0.762	0.534	0.622	0.762	0.943	0.910
				Panel B: Results at percentile level α=0.01, Expected Violations 10						
Violations	21	22	16	15	24	14	14	23	14	14
VR	2.10	2.20	1.60	1.50	2.40	1.40	1.40	2.30	1.40	1.40
UC	0.022	0.001	0.079	0.139	0.000	0.231	0.231	0.000	0.231	0.231
IND	0.329	0.505	0.253	0.499	0.603	0.528	0.528	0.109	0.528	0.528
CC	0.046	0.004	0.112	0.266	0.001	**0.399**	**0.399**	0.001	**0.399**	**0.399**

The more formal backtests show that all conditional volatility models pass the tests for coverage and independence at the 5% level. In fact, all GARCH models have relatively high p-values in the conditional coverage test and seem to perform reasonably well. However, the student-t GARCH model has the highest p-value and appears to be the superior model. This coincides with the AICc results from the in-sample analysis. Contrary to the 5% results, only the t-distributed GARCH models pass the tests for coverage and independence at the 1% level. The t-distributed EGARCH and APARCH models seem to have equal and superior performance in the conditional coverage test, and the choice of GARCH model seems to be of little importance compared to the choice of distribution.

Figure 8.4 shows the negative 1% and 5% VaR estimates for the Tesla returns produced by the student-t GARCH model. The VaR exceedances in blue appear to be evenly spread out in time, implying no clustering of VaR violations. However, some violations seem to lie quite far from the VaR(0.05) and VaR(0.01) lines, especially during the outbreak of COVID-19 in the spring of 2020. This means that we observe some larger VaR violations.



(a) VaR(0.05) **(b)** VaR(0.01)

Figure 8.4: VaR exceedance plots for the student-t GARCH(1,1) model on Tesla stock March 2017 to January 2021. The black line indicates the negative VaR estimates, the gray marks the returns and the VaR violations are plotted as blue marks.

Expected Shortfall Backtest

Table 8.5 presents backtest results for the 1% and 5% ES estimates of Tesla returns. The skew student-t GARCH models appear to have similar performance at the 5% level, but the standard GARCH-st model has the highest p-value and seems to perform best. The normally distributed GARCH models and HS seem to have inferior model performance as they do not pass the ESR test at a 5% significance level. Analogous to the 1% VaR analysis, the choice of distribution appears to be of larger importance than the choice of GARCH model. The results at the 1% percentile level give similar conclusions.

Table 8.5: 1% and 5% ES exceedance for Tesla returns from 17.11.2010 to 21.01.2021.

Model	HS	G_n	G_t	G_{st}	E_n	E_t	E_{st}	A_n	A_t	A_{st}
			Panel A: Results at percentile level α=0.05							
ESR	0.000	0.011	0.186	**0.271**	0.008	0.138	0.221	0.004	0.178	0.265
			Panel B: Results at percentile level α=0.01							
ESR	0.001	0.000	0.312	**0.482**	0.000	0.177	0.248	0.000	0.216	0.300

8.2.4 Tomra

Value-at-Risk Backtests

The violation ratios in Panel A of Table 8.6 indicate that all normal GARCH models, in addition to the APARCH-t model, overestimate the 5% VaR for Tomra returns. The APARCH-st model has a violation ratio of 1, whereas the other GARCH models slightly underestimate the VaR. HS appears to have inferior performance at both percentile levels as the violation ratios are larger than 1.5. Except for the APARCH-t model, which still overestimates the VaR, all GARCH models that overestimate at the 5% level seem to underestimate at the 1% level and vice versa.

The 5% VaR backtest results show that the GARCH-n, APARCH-t and APARCH-st models, in addition to the EGARCH models, pass the coverage and independence tests. The APARCH-t model appears to perform best, but there are only minor differences in the performance of the t-distributed EGARCH and APARCH models. At the 1% percentile level, all models pass the tests for coverage and independence. The GARCH-t, GARCH-st and EGARCH-st models seem to perform equally good and better than the other models. However, the differences in model performance are relatively small and it is difficult to choose the superior model.

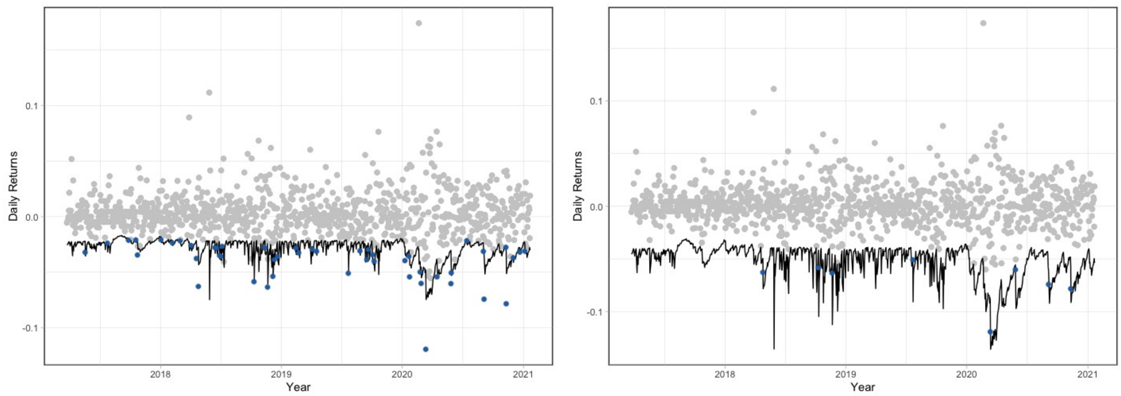
Table 8.6: 1% and 5% VaR exceedance for Tomra returns from 17.11.2010 to 21.01.2021.

Model	HS	G_n	G_t	G_{st}	E_n	E_t	E_{st}	A_n	A_t	A_{st}
Panel A: Results at percentile level α=0.05, Expected Violations 50										
Violations	95	44	51	55	43	52	53	35	49	50
VR	1.90	0.88	1.02	1.10	0.86	1.04	1.06	0.70	0.98	1.00
UC	0.004	0.375	0.885	0.475	0.299	0.773	0.666	0.022	0.884	1.000
IND	0.083	0.170	0.015	0.010	0.415	0.853	0.907	0.156	0.779	0.747
CC	0.004	0.263	0.052	0.028	0.418	0.943	0.905	0.026	**0.951**	0.949
Panel B: Results at percentile level α=0.01, Expected Violations 10										
Violations	20	13	9	9	13	8	9	12	8	8
VR	2.00	1.30	0.90	0.90	1.30	0.80	0.90	1.20	0.80	0.80
UC	0.079	0.362	0.746	0.746	0.362	0.510	0.746	0.538	0.510	0.510
IND	0.470	0.558	0.686	0.686	0.558	0.719	0.686	0.589	0.719	0.719
CC	0.166	0.556	**0.875**	**0.875**	0.556	0.755	**0.875**	0.715	0.755	0.755

The VaR violations appear to be evenly spread out in time from the 1% and 5% VaR estimates shown in Figure 8.5. These VaR estimates are produced by the APARCH model with student-t distributed innovations. By comparing Figures D.7-D.9 in Appendix D.3, the graphs are quite similar for all GARCH models. However, the standard GARCH models seem to capture the characteristics of the returns to a lesser extent. Furthermore, Figure 8.5 indicates some large VaR exceedances at the 5% level, for instance during the outbreak of COVID-19 in the spring of 2020. At the 1% level, all VaR exceedances closely follow the VaR line.

Expected Shortfall Backtest

Table 8.7 displays the backtest results for the 1% and 5% ES forecasts of Tomra returns. The normally distributed GARCH models and HS are clearly outperformed by the t-distributed models at the 5% level. Moreover, the standard GARCH-t model appears to be somewhat better than the other t-distributed models. Similar conclusions can be made from the 1% ES results in Panel B of Table 8.7. However, as the normally distributed GARCH models do not pass the ESR backtest at a 5% significance level, they seem to perform even worse compared to the t-distributed models at the 1% level.



(a) VaR(0.05) **(b)** VaR(0.01)

Figure 8.5: VaR exceedance plots for the student-t APARCH(1,1) model on Tomra stock March 2017 to January 2021. The black line indicates the negative VaR estimates, the gray marks the returns and the VaR violations are plotted as blue marks.

Table 8.7: 1% and 5% ES exceedance for Tomra returns from 17.11.2010 to 21.01.2021.

Model	HS	G_n	G_t	G_{st}	E_n	E_t	E_{st}	A_n	A_t	A_{st}
				Panel A: Results at percentile level α=0.05						
ESR	0.001	0.165	**0.951**	0.916	0.081	0.940	0.896	0.248	0.941	0.902
				Panel B: Results at percentile level α=0.01						
ESR	0.042	0.000	**0.973**	0.951	0.000	0.925	0.908	0.005	0.924	0.885

8.2.5 Norsk Hydro

Value-at-Risk Backtests

Panel A of Table 8.8 shows that all VaR models underestimate at the 5% level for Norsk Hydro returns. Moreover, all conditional volatility models pass the tests for coverage and independence. The EGARCH-n model, followed by the standard GARCH-n model, are least likely to be rejected in the conditional coverage test and seem to perform best. Hence, in contrast to the in-sample analysis, the normal assumption seems to perform well in the out-of-sample analysis. Figure 8.6 also confirms this behaviour as there seems to be little difference in the VaR estimates of the normally and t-distributed GARCH models.

Table 8.8: 1% and 5% VaR exceedance for Norsk Hydro returns from 17.11.2010 to 21.01.2021.

Model	HS	G_n	G_t	G_{st}	E_n	E_t	E_{st}	A_n	A_t	A_{st}
				Panel A: Results at percentile level α=0.05, Expected Violations 50						
Violations	110	55	59	58	53	61	61	55	63	61
VR	2.20	1.10	1.18	1.16	1.06	1.22	1.22	1.10	1.26	1.22
UC	0.037	0.475	0.204	0.257	0.666	0.122	0.122	0.475	0.069	0.122
IND	0.182	0.571	0.774	0.721	0.479	0.881	0.881	0.268	0.595	0.881
CC	0.047	0.660	0.428	0.494	**0.709**	0.300	0.300	0.419	0.167	0.300
				Panel B: Results at percentile level α=0.01, Expected Violations 10						
Violations	24	19	10	10	19	10	10	21	13	11
VR	2.40	1.90	1.00	1.00	1.90	1.00	1.00	2.10	1.30	1.10
UC	0.362	0.011	1.000	1.000	0.011	1.000	1.000	0.002	0.362	0.754
IND	0.009	0.370	0.085	0.085	0.370	0.085	0.085	0.075	0.558	0.106
CC	0.022	0.026	0.226	0.226	0.026	0.226	0.226	0.002	**0.556**	0.258

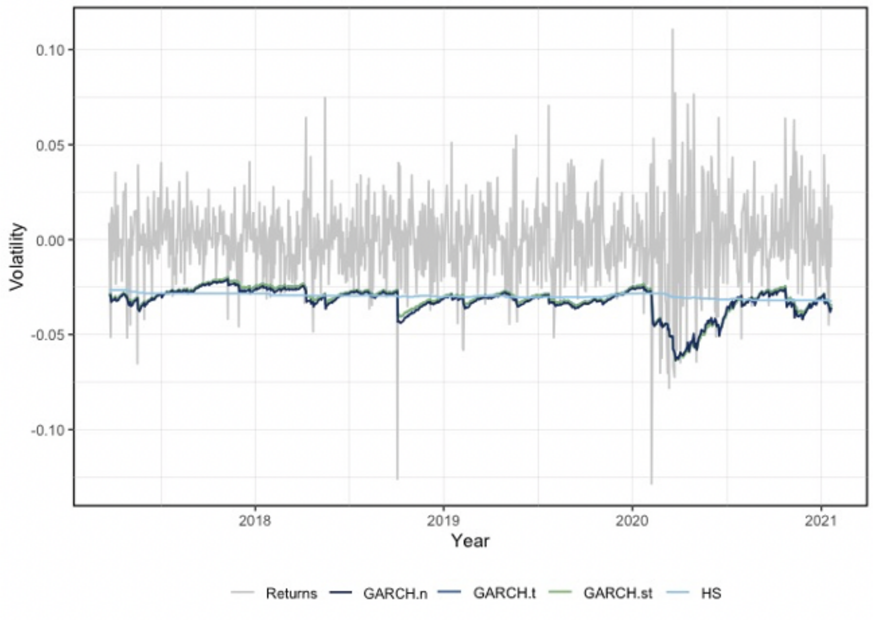
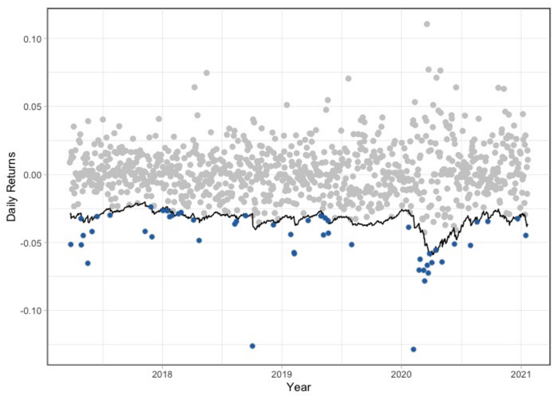


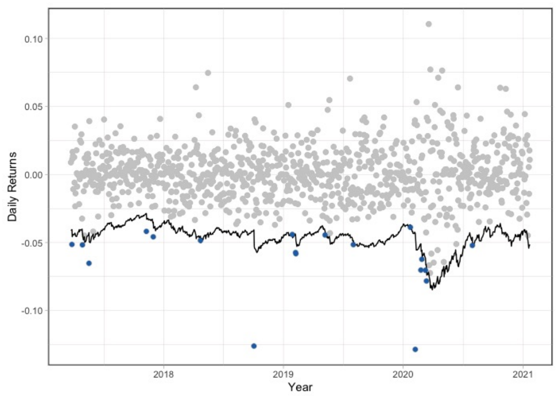
Figure 8.6: Negative 5% VaR estimates from the VaR models based on HS, as well as the parametric approach with the standard GARCH model fitted with normal, student-t and skew student-t distributed innovations plotted together with Norsk Hydro returns from March 2017 to January 2021.

Panel B of Table 8.8 shows that the t-distributed GARCH and EGARCH models have a violation ratio of 1, whereas the other models underestimate the 1% VaR. In particular, the normally distributed GARCH models and HS appear imprecise. This could explain why only the t-distributed EGARCH and APARCH models pass the tests for coverage and independence. Judging by the conditional coverage test, the APARCH-t model seems to perform best. The other models that pass the test have similar and lower performance. Hence, in contrast to the 5% VaR analysis, the t-distributed GARCH models outperform the normal assumption at the 1% level.

The negative 1% and 5% VaR estimates for Norsk Hydro are shown in Figure 8.7. The estimates are produced by the GARCH model with normally distributed errors. No clustering of VaR violations is observed as they appear to be evenly spread out in time. However, at both percentile levels there can be observed some large VaR exceedances.



(a) VaR(0.05)



(b) VaR(0.01)

Figure 8.7: VaR exceedance plots for the normal EGARCH(1,1) model on Norsk Hydro stock March 2017 to January 2021. The black line indicates the negative VaR estimates, the gray marks the returns and the VaR violations are plotted as blue marks.

Expected Shortfall Backtest

The 5% ES backtest results, given in Panel A of Table 8.9, show that all the t-distributed GARCH models pass the ESR test at a 5% significance level. The standard GARCH-st model, followed by the standard GARCH-t model, seem to perform best. This is in contrast to the in-sample analysis, where the ordinary student-t distribution performs slightly better than the skew student-t distribution. Similar conclusions can be made from Panel B, showing the results at the 1% percentile level.

Table 8.9: 1% and 5% ES exceedance for Norsk Hydro returns from 17.11.2010 to 21.01.2021.

Model	HS	G_n	G_t	G_{st}	E_n	E_t	E_{st}	A_n	A_t	A_{st}
Panel A: Results at percentile level α=0.05										
ESR	0.047	0.007	0.318	**0.337**	0.005	0.253	0.266	0.002	0.098	0.214
Panel B: Results at percentile level α=0.01										
ESR	0.101	0.005	0.353	**0.381**	0.001	0.336	0.337	0.000	0.218	0.272

8.2.6 General Motors

Value-at-Risk Backtests

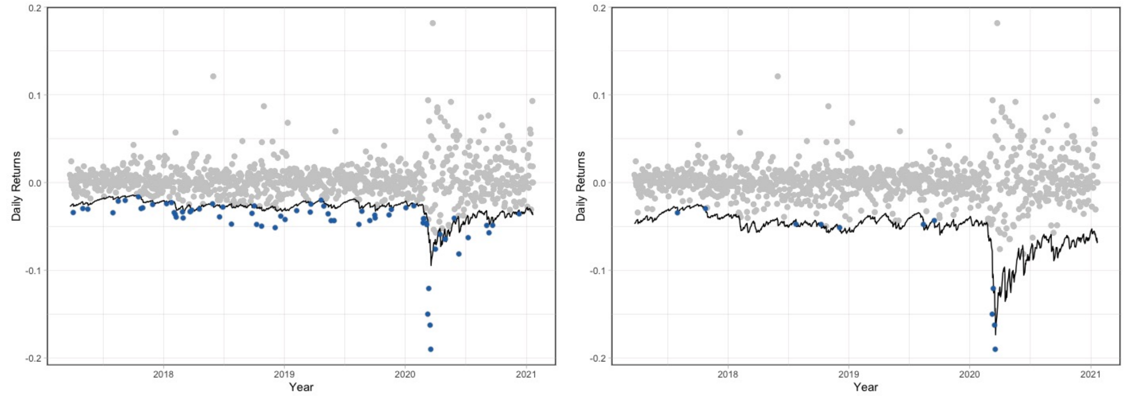
Table 8.10 shows that all conditional volatility models underestimate the 5% VaR for General Motors. The standard GARCH-n and EGARCH-n models seem to perform the best as they have the violation ratio closest to 1. However, neither of the models have a violation ratio that classifies as a good model. This may explain why neither of the VaR models passes the unconditional or conditional coverage tests at the 5% level.

Panel B of Table 8.10 shows that all models underestimate at the 1% level as well and that the EGARCH-st and APARCH-st models have the violation ratios closest to 1. All the other models have too high violation ratios to be classified as good models. The EGARCH-st and APARCH-st models also perform best in the conditional coverage test. Only the models with t-distributed innovations pass the tests for coverage and independence, and these seem to greatly outperform the normally distributed models.

Table 8.10: 1% and 5% VaR exceedance for General Motors returns from 17.11.2010 to 21.01.2021.

Model	HS	G_n	G_t	G_{st}	E_n	E_t	E_{st}	A_n	A_t	A_{st}
Panel A: Results at percentile level α=0.05, Expected Violations 50										
Violations	103	65	68	66	65	71	69	68	71	68
VR	2.06	1.30	1.36	1.32	1.30	1.42	1.38	1.36	1.42	1.36
UC	0.004	0.037	0.013	0.027	0.037	0.004	0.009	0.013	0.004	0.013
IND	0.001	0.182	0.267	0.208	0.182	0.372	0.299	0.122	0.187	0.267
CC	0.000	0.047	0.025	0.039	0.047	0.011	0.019	0.014	0.007	0.025
Panel B: Results at percentile level α=0.01, Expected Violations 10										
Violations	30	27	15	14	27	15	11	31	14	11
VR	3.00	2.70	1.50	1.40	2.70	1.50	1.10	3.10	1.40	1.10
UC	0.000	0.000	0.139	0.231	0.000	0.139	0.754	0.000	0.231	0.754
IND	0.001	0.000	0.499	0.528	0.037	0.499	0.621	0.968	0.528	0.621
CC	0.000	0.000	0.266	0.399	0.000	0.266	**0.842**	0.000	0.399	**0.842**

Figure 8.8 shows the negative 1% and 5% VaR estimates produced by the EGARCH-st model. The VaR exceedances in blue appear to be evenly spread out in time, implying no clustering of VaR violations. Some of the violations during the outbreak of COVID-19 seem to lie quite far from the VaR(0.05) line. This means some larger VaR violations are observed at the 5% percentile level. On the 1% percentile level, there seems to be only one larger VaR violation.



(a) VaR(0.05) **(b)** VaR(0.01)

Figure 8.8: VaR exceedance plots for the skew student-t EGARCH(1,1) model on General Motors stock March 2017 to January 2021. The black line indicates the negative VaR estimates, the gray marks the returns and the VaR violations are plotted as blue marks.

Expected Shortfall Backtest

The 5% VaR analysis shows that all VaR models underestimate and that neither of them passes the coverage tests. Thus, neither of the models seem to perform particularly well at the 5% level. The same behaviour can be seen in Panel A of Table 8.11, showing the 5% ES backtest results. Only the GARCH-st model slightly passes the ESR test. At the 1% level, all the t-distributed GARCH models in addition to HS pass the backtest. The skew student-t distributed models seem to perform better than the student-t distributed GARCH models, and the EGARCH-st model appears to perform best.

Table 8.11: 1% and 5% ES exceedance for General Motors returns from 17.11.2010 to 21.01.2021.

Model	HS	G_n	G_t	G_{st}	E_n	E_t	E_{st}	A_n	A_t	A_{st}
				Panel A: Results at percentile level α=0.05						
ESR	0.000	0.000	0.032	**0.070**	0.000	0.000	0.019	0.000	0.001	0.020
				Panel B: Results at percentile level α=0.01						
ESR	0.256	0.000	0.512	0.687	0.001	0.548	**0.803**	0.000	0.645	0.734

8.2.7 Aker BP

Value-at-Risk

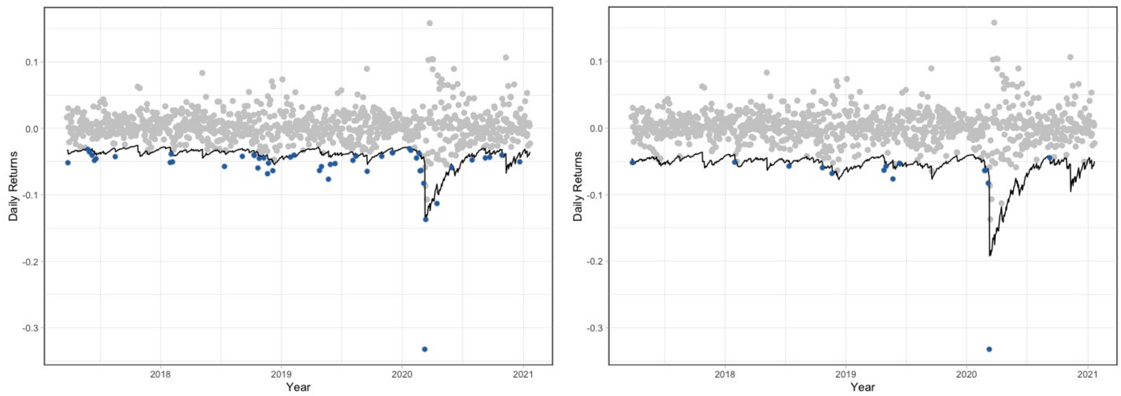
The 1% and 5% VaR results for Aker BP returns are given in Table 8.12. Except for the standard GARCH-n model, which slightly overestimates, all other VaR models underestimate at the 5% level. Moreover, all conditional volatility models pass the coverage and independence tests. The standard GARCH-n model appears to be the superior model as this is least likely to be rejected in the conditional coverage test. However, the performance of the other GARCH models is quite similar, and it is difficult to explicitly state which model is the best.

Panel B of Table 8.12 shows that all models underestimate the 1% VaR and that HS, in addition to the normal EGARCH and APARCH models, are imprecise. Regardless, all conditional volatility models pass the coverage and independence tests and have similar performance in the conditional coverage test. However, the APARCH-st model seems to be somewhat better than the other models.

Table 8.12: 1% and 5% VaR exceedance for Aker BP returns from 17.11.2010 to 21.01.2021.

Model	HS	G_n	G_t	G_{st}	E_n	E_t	E_{st}	A_n	A_t	A_{st}
Panel A: Results at percentile level α=0.05, Expected Violations 50										
Violations	113	49	55	56	53	60	62	53	59	61
VR	2.26	0.98	1.10	1.12	1.06	1.20	1.24	1.06	1.18	1.22
UC	0.257	0.884	0.475	0.393	0.666	0.159	0.093	0.666	0.204	0.122
IND	0.000	0.318	0.268	0.301	0.209	0.457	0.547	0.209	0.414	0.501
CC	0.000	**0.601**	0.419	0.406	0.414	0.281	0.203	0.414	0.319	0.242
Panel B: Results at percentile level α=0.01, Expected Violations 10										
Violations	30	14	11	11	16	11	11	17	11	12
VR	3.00	1.40	1.10	1.10	1.60	1.10	1.10	1.70	1.10	1.20
UC	0.139	0.231	0.754	0.754	0.079	0.754	0.754	0.043	0.754	0.538
IND	0.018	0.186	0.106	0.106	0.253	0.106	0.106	0.290	0.106	0.130
CC	0.020	0.204	0.258	0.258	0.112	0.258	0.258	0.074	0.258	**0.263**

As shown in Figure 8.9, the 1% and 5% VaR exceedances in blue appear evenly spread out in time for Aker BP. These VaR estimates are produced by the standard GARCH-n model. One substantial VaR violation can be observed at both percentile levels during the corona crisis in the spring of 2020.



(a) VaR(0.05) **(b)** VaR(0.01)

Figure 8.9: VaR exceedance plots for the normal GARCH(1,1) model on Aker BP stock March 2017 to January 2021. The black line indicates the negative VaR estimates, the gray marks the returns and the VaR violations are plotted as blue marks.

Expected Shortfall

Contrary to the 5% VaR analysis, the t-distributed models outperform the normal GARCH models in the 5% ES analysis. All the t-distributed GARCH models, in addition to HS, pass the ESR test. Moreover, the GARCH-t model, followed by the GARCH-st model, appear to be the best. The expected shortfall gives the expected size of the loss in the α% worst cases. The risk measure therefore captures the magnitude of the VaR violations and not just the number of violations. Hence, when the exceedances have a large magnitude, as seen from the VaR exceedance plot in Figure 8.9, it is not unreasonable to expect that we need a heavy-tailed distribution to capture the characteristics of the returns.

Similar findings apply to the 1% ES estimates. However, the conditional volatility models seem to be outperformed by HS. In the 1% VaR analysis HS is an imprecise model, and it does not pass the tests for coverage and independence. Depending on the final application, it could be useful to choose a model that performs well in both the VaR and ES analyses.

Table 8.13: 1% and 5% ES exceedance for Aker BP returns from 17.11.2010 to 21.01.2021.

Model	HS	G_n	G_t	G_{st}	E_n	E_t	E_{st}	A_n	A_t	A_{st}
				Panel A: Results at percentile level α=0.05						
ESR	0.111	0.042	**0.314**	0.287	0.008	0.191	0.174	0.005	0.205	0.157
				Panel B: Results at percentile level α=0.01						
ESR	**0.212**	0.040	0.141	0.117	0.011	0.074	0.056	0.011	0.062	0.047

8.3 Comparison Across Market Sectors

The reported results show how different GARCH models adapt to each of the return series and how this affects model performance. Intuitively, this means how well the models adjust to changes in volatility, and thus their ability to model the volatility in different sectors of the stock market. Considering that the stocks represent different market sectors and have different price dynamics, it would be reasonable to assume their volatility to be somewhat different. The choice of the most optimal volatility model could also be affected by the green shift in the stock market. In the following, both the in-sample and out-of-sample analysis will be examined to further discuss which GARCH model appears to be most applicable.

8.3.1 In-sample Analysis

The in-sample analysis shows that models with lag parameters $p = q = 1$ usually perform the best. Another important finding is that t-distributed innovations lead to better average model performance than normally distributed innovations. This finding is consistent with the study of Aparicio and Estrada [11] where they found overall support for the scaled t-distribution, and the study of Ghalanos [13] where he found that most models beat the standard GARCH(1,1) model with normally distributed errors. In our study, the skew student-t distribution seems to give better results for NEL, whereas for the other stocks, the ordinary student-t distribution performs best.

The in-sample analysis also documents that the return characteristics reported in Chapter 7 are present in the time series as most of the GARCH coefficients are significant. The same applies for leverage effects and power effects. As reported in Section 8.1, the fat tails expressed through the shape parameter of the t-distributions is for most stocks in the range 3.25 to 4.0. NEL, however, seems to have particularly fat tails as the shape parameter is around 2.5, and General Motors is the least leptokurtic with degrees of freedom around 4.4. The differences in the estimated shape parameters could result from different market characteristics as well as the phase of the life cycle the companies are in.

Even though both leverage effects and power effects appear to be significant for most of the return series, the standard GARCH model seems to perform best when using AICc as model criterion. This applies to all stocks, except for Aker BP, where the EGARCH model performs best. On the contrary, when using BIC as model criterion, the EGARCH model appears superior for all stocks. This implies that the preferred model captures neither leverage effects nor power effects, or that it captures leverage effects only. However, the model performances are quite similar, and it is difficult to state whether either of the models outperforms the other.

The complexity of the GARCH models is differing. As the EGARCH and APARCH models can capture leverage effects, they have a larger complexity than the standard GARCH model. Additionally, the APARCH model is even more complex as it captures power effects as well. If the model performance is not sufficiently improved by increasing the model complexity, there are at least two main reasons for choosing the simpler model. A lower complexity increases the interpretability, and it also reduces the computational costs. However, having $p = q = 1$, the total number of parameters is quite similar and model fitting does not take an unreasonable amount of time for either of the models.

8.3.2 Value-at-Risk

After the in-sample analysis, the superior models were further analysed using the value-at-risk and expected shortfall risk measures. A good model should in general give the expected number of VaR violations in addition to good results in backtesting. In the following, the most important findings of the out-of-sample analysis will be summarised and discussed.

The 5% VaR analysis shows that the number of actual VaR violations seems to be quite good for all conditional volatility models for the Tesla, Tomra and Aker BP stocks. HS, on the other hand, severely underestimates, and this applies to all stocks. For NEL there is a trend of overestimation, and only the skew student-t GARCH models seem to produce good VaR models. For Norsk Hydro and General Motors there is a trend of underestimation of the conditional volatility models. In fact, neither of the models have a violation ratio within the limit of what we defined as a good model when fitted to General Motors returns. This may explain why neither of them passes the coverage and independence tests at the 5% level.

For the other stocks there are some differences in which model proves to be the most optimal. For NEL the models with skew student-t distributed innovations seem to outperform the other models, and the standard GARCH-st model performs best. This coincides with the in-sample analysis using AICc as model criterion. Moreover, the APARCH-t model seems to be the most optimal for Tomra, but there are only small differences in the t-distributed EGARCH and APARCH models. For Tesla, Norsk Hydro and Aker BP, the normally distributed GARCH models appear to have similar performance as the t-distributed models. Even though the differences are rather small, the standard GARCH-t model appears to perform best for Tesla. For Norsk Hydro, the normal GARCH models seem to perform better than the t-distributed models, and the EGARCH-n model appears most optimal. A similar tendency can be seen for Aker BP, but the standard GARCH-n model appears best. Most of these results are somewhat contradictory to the in-sample results, where the GARCH-t and EGARCH-t models in general appear to perform best.

The reason why normally distributed innovations appear to work better for Norsk Hydro and Aker BP out-of-sample at the 5% level may be due to the differences in return characteristics. All the returns give evidence of non-normality in the data analysis in Chapter 7. However, the Norsk Hydro returns have the smallest reported kurtosis and appear to be closest to the normal distribution. This is also confirmed by the qq-plot of Norsk Hydro in Figure 7.5a. Furthermore, the histogram of Aker BP in Figure 7.5f shows that the non-normality may be caused by the high peakedness at the centre of the distribution rather than fat tails. The t-distributions are mainly to allow for fatter tails than the normal distribution, and as a consequence, the normally distributed GARCH models may perform better for Aker BP. Similar conclusions can be drawn for Tesla, where the normal GARCH models seem to have satisfactory although somewhat inferior performance compared to the t-distributed GARCH models. The qq-plots in Figure 7.3 show that Tesla is the least leptokurtic green stock.

Another reason why normally distributed innovations seem to work better for some of the stocks could be caused by the fact that these stocks are difficult to model. Their behaviour yields large variations over a significant length of time such that even the most advanced volatility models cannot capture the return characteristics in a satisfactory manner. The normally distributed models then perform better because they are more neutral when the model characteristics cannot be precisely defined.

On the 1% percentile level, quite few of the models have a violation ratio within the limit of a good model for NEL. In fact, several of them overestimate and seem to be imprecise. For Tesla, Norsk Hydro, General Motors and Aker BP, all models underestimate the VaR, except for a few models for Norsk Hydro which have a violation ratio equal to 1. Furthermore, HS and the normally distributed GARCH models seem to be imprecise for these stocks. For Tomra, most models have a reasonable violation ratio.

The backtest results at the 1% level show that in addition to the skew student-t GARCH models, the normal GARCH models and the standard GARCH-t model seem to perform well for NEL returns. In fact, the APARCH-n model appears the best, and this also has a reasonable number of actual violations. For Tesla, Norsk Hydro and General Motors, the normally distributed GARCH models are outperformed by the t-distributed models. For Tesla the t-distributed EGARCH and APARCH models have similar performance, and the model choice does not seem to be of importance. For Norsk Hydro the APARCH-t model appears slightly better than the other t-distributed models, whereas for General Motors the skew student-t EGARCH and APARCH models perform best. For Tomra and Aker BP, all conditional volatility models seem to perform well at the 1% level. The GARCH-t, GARCH-st and EGARCH-st models seem equivalent for Tomra, while the APARCH-st model has slightly better performance for Aker BP.

8.3.3 Expected Shortfall

The 5% ES results show that the student-t EGARCH and APARCH models appear better for NEL returns as they have a p-value of 1. However, the VaR analysis shows that these models overestimate and that they do not pass the tests for coverage and independence. The VaR and ES analyses give different conclusions, and the choice of model should therefore depend on the intended application. If the 5% percentile is most interesting, then VaR should be used as the risk measure and the model choice done accordingly. If the size of the loss is most important, ES should be used, which would indicate a different prioritisation of the test statistics. Regardless, a good model should, as mentioned, have an acceptable number of VaR violations in addition to sufficient results in backtesting. This means in most cases we want a model that performs well according to both criteria. All the t-distributed GARCH models seem to perform well for NEL in the ESR backtest, and it could be better to choose a model that performs satisfactorily in both analyses.

For Tesla, Norsk Hydro and General Motors, the standard GARCH-st model appears to perform best. In particular, for General Motors the standard GARCH-st model is the only one that passes the ESR backtest. However, there are only minor deviations in the p-values of the t-distributed GARCH models for all these stocks. Similar trends where the t-distributed models perform well can also be seen for Tomra and Aker BP. For these stocks, the standard GARCH-t model performs slightly better than the other t-distributed models. However, for Tomra the standard GARCH-t model does not pass the tests for coverage and independence. The t-distributed EGARCH and APARCH models, which perform well in both the VaR and ES analyses, might be a better choice.

The 1% ES results give more or less the same conclusions as the 5% ES analysis. One exception can be noticed for Aker BP, where HS now seems better than the conditional volatility models. On the other hand, HS does not have a reasonable number of VaR violations and does not pass the VaR backtests. As all the GARCH models have a more acceptable violation ratio and pass the VaR backtests, these models would presumably be better choices. However, the normal GARCH models do not pass the ESR backtest, and the t-distributed GARCH models should therefore be preferred. Another exception to notice is that the EGARCH-st model now seems to perform best for the General Motors returns.

8.3.4 Comparison Green vs. Non-Green Stocks

Comparing the 1% VaR exceedance plots of the green versus non-green stocks, there seems to be a slight trend of more large VaR violations for the non-green stocks. The only exception is Tesla which has a rather fluctuating VaR(0.01) line and some large violations. The large non-green violations typically happen during the corona crisis in the spring of 2020. Intuitively this indicates that GARCH models adjust more rapidly to rising volatility for the green stocks. In other words, GARCH models seem to be just too slow to adapt to the abruptly changing volatility for the non-green stocks. This leads to worse VaR estimates and thus larger VaR violations. The reason for this could be different volatility structures for different stocks. The green stocks have experienced a rapid increase over the recent years, and the GARCH models are already adapted to data with large volatility. As an example, the data analysis shows that NEL on average has the largest standard deviation, meaning the highest volatility. These observations may indicate that GARCH models react quicker to sudden changes in volatility, such as those caused by COVID-19, for the green stocks.

The reasoning stated above is also supported by the 5% VaR analysis results for the General Motors stock. Figure 7.4c shows that up to the outbreak of COVID-19, General Motors has had a quite stable price dynamic with some minor ups and downs. After this point, however, the company experienced a rapid increase. As the model performance seems inferior for all GARCH models, they seem to have significant problems adjusting to the quickly changing volatility. This is also confirmed by the VaR exceedance plots in Appendix D.5, which show several large VaR violations during the spring of 2020 for all conditional volatility models.

However, it is not necessarily the green image of the company that gives this model feature. Stocks with related price dynamics may experience similar results when modelling their volatility. Structural breaks in the stock data and the volatility structure will have a significant impact on the modelling. There are also other factors that could be more important than the green versus non-green image. Some examples may be the phase of the life cycle the company is in or the market sector it belongs to.

The VaR exceedance plots of Tesla in Figure 8.4 clearly show more large VaR violations for this company compared to the others. This may indicate that the returns of Tesla are more difficult to model, which could be due to the price dynamics of the stock and the phase of the life cycle the company is within. As a result of the advanced car technologies, Tesla is also considered a tech company. Contrary to nearly all other sectors, the technology sector experienced major growth during the corona crisis in 2020. As shown in Figure 7.1c, Tesla experienced significant growth in this time period which clearly has affected the return characteristics of the stock. This growth is also confirmed as Tesla has the highest reported mean in Table 7.1, which indicates a rising stock value. As already pointed out, the GARCH models are known for not being able to model returns with such characteristics well. Furthermore, Tesla is in a growth phase and is valued based on unknown future earnings. This naturally leads to high volatility, which could make the GARCH modelling difficult. A similar tendency can also be seen for NEL. Here the VaR and ES analyses are rather conflicting as the models that clearly have inferior performance in the VaR analysis perform best in the ES analysis. This might indicate that volatility modelling is more difficult for companies in a growth phase rather than a steady-state phase.

8.3.5 Summary of VaR and ES Findings

The results of this study indicate that the model performance is somewhat similar for most models. Neither of the GARCH models appears better for volatility modelling in all scenarios, making it difficult to choose one superior GARCH model. This is true for all three market sectors that we have looked at in this thesis. In addition, there is no clear choice for the green and non-green stocks. This means that depending on the final application, there could be some differences in the choice of model. The model choice should be based on the VaR performance when the percentile is most interesting, and equivalently it should be based on the ES performance when the size of the loss is most important.

The VaR analysis does not clearly show whether either of the GARCH models fit the stock market data better. The 1% and 5% percentile levels give quite different conclusions. For most of the stocks, the ES analysis tends to choose the least complex model, the standard GARCH model with t-distributed innovations. In general, the models with t-distributed innovations seem to perform somewhat better than the models with normally distributed innovations, or at least give more stable results. This applies in both the VaR and ES analysis but is more pronounced in the ES analysis. This result also coincides with the in-sample analysis where the normal GARCH models were outperformed.

When neither of the models clearly stands out as superior, the model complexity should be used as a basis for model selection. If the EGARCH or APARCH models do not significantly outperform the standard GARCH model, the latter is a better choice due to both interpretability and computational complexity.

8.4 Alternative Calibration Method

The previous analyses show that the models have problems adjusting to the rapidly changing volatility during the corona crisis. This could be caused by the fact that the estimation set does not include any particularly stressed periods, and as a consequence, the models are not suited for modelling volatility in crisis periods. To investigate the effect of the estimation sample, we try fitting the GARCH models to the stock return data for the stressed period presented in Section 7.3. In particular, we apply data from 01.01.2006 to 31.12.2010, including the global financial crisis between mid 2007 and early 2009. The models are estimated with 1305 daily observations, and as before, they are backtested on the last 1000 observations covering the period 24.03.2017 to 21.01.2021. We do this for the NEL and Norsk Hydro stocks.

8.4.1 NEL

The VaR and ES results for the NEL stock fitted with the smaller estimation set are provided in Appendix E.1. Both estimation sets give quite similar results at the 5% level for VaR. In addition to the skew student-t distributed models, the GARCH-t and APARCH-t models perform well with the smaller estimation set. The models also seem to overestimate to a lesser extent, meaning that the violation ratios are closer to 1. However, the conclusion remains more or less the same. At the 1% level, the same models pass the tests for coverage and independence with both estimation sets. The GARCH-st model and HS seem to be preferred with the smaller estimation set. The GARCH-st model thus seems to perform best at both percentile levels. The reason HS performs better than some of the conditional volatility models could be due to less structural breaks in the smaller estimation set. Furthermore, at both percentile levels, the ES analysis gives similar conclusions for the two estimation sets. All models appear to perform well at the 5% level, whereas the t-distributed models seem to outperform the normally distributed models at the 1% level.

To further examine the two estimation sets, we compare the 5% VaR exceedance plots in Appendix D.1 and E.1, where the latter shows VaR exceedances when the models are fitted to the smaller estimation set. The VaR violations are shown as blue and green marks when fitted to the larger and smaller estimation set, respectively. There does not seem to be many differences in the VaR exceedance plots of the two estimation sets. The only exception is the GARCH-n model, where the smaller estimation set seems to do a better job capturing the volatility during both the U.S.-China trade war in mid 2019 and during the outbreak of COVID-19 in the spring of 2020.

8.4.2 Norsk Hydro

The VaR and ES results for Norsk Hydro fitted with the smaller estimation set are provided in Appendix E.2. Contrary to the larger estimation set, the normally distributed GARCH models seem to have inferior performance compared to the t-distributed models at the 5% VaR level for the smaller estimation set. The APARCH-t model appears best, followed by the EGARCH-t and GARCH-t models. The t-distributed models also seem to have violation ratios closer to 1. At the 1% level, all conditional volatility models pass the tests for coverage and independence with the smaller estimation set. However, the normally distributed GARCH models still seem to perform poorly compared to some of the t-distributed models. The EGARCH-st, APARCH-t and APARCH-st models seem best. Hence, the APARCH-t model appears to perform well at both percentile levels. The ES analysis shows that HS performs best for Norsk Hydro, followed by the skew student-t distributed models. As mentioned, the reason HS seems to outperform the other models may be that the smaller estimation set contains fewer structural breaks.

Comparing the 5% VaR exceedance plots of Norsk Hydro in Appendix D.4 and E.2, all GARCH models seem to capture the volatility better with the smaller estimation set. This applies for the entire period from March 2017 to January 2021 but particularly during the corona crisis. Hence, there seems to be a trend that the smaller estimation set, which includes data from the financial crisis, captures the volatility better for Norsk Hydro.

8.4.3 Summary Effect of Estimation Sample

Comparing the analyses using the smaller estimation set to the previous analyses, the former appears to give a more definite answer to which GARCH models should be preferred for some of the stocks. The ES analyses have usually given equivalent conclusions at both percentile levels, whereas the conclusions have been somewhat incongruent for the VaR analyses. With the smaller estimation set, the VaR analysis also seems to give a more concrete choice of model. For NEL, the GARCH-st model seems to perform well at both percentile levels in the VaR analysis. Moreover, for Norsk Hydro the APARCH-t model appears to be one of the preferred models at both percentile levels in the VaR analysis. Consequently, the choice of the estimation sample may affect the choice of the preferred GARCH model.

The VaR exceedance plots show that the GARCH models do not seem to capture the volatility of the corona crisis better with an estimation set from the global financial crisis for the NEL stock. On the contrary, this seems to be the case for Norsk Hydro. The data analysis of the smaller estimation set, covered in Section 7.3, shows that Norsk Hydro experienced a very rapid decrease during 2008. As the behaviour of the stock prices during the two crisis periods appears more similar for Norsk Hydro than NEL, it seems reasonable that the GARCH models capture the volatility of the corona crisis better for Norsk Hydro.

The reason that different estimation sets give similar results for NEL can be caused by the fact that the global financial crisis and the corona crisis were different types of crisis periods. By this we mean that the two crises elapsed quite differently. The global financial crisis hit most market sectors at the same time and led to a global recession that lasted almost two years. The corona crisis hit different market sectors at different times, and in general, the economy seems to have stabilised faster. For instance, the technology sector experienced one of the largest declines ever in 2020, while by the end of the year, new highs were reached. As a consequence, the GARCH models are not necessarily better suited for modelling the volatility of the corona crisis when fitted to data that includes the global financial crisis.

The additional analysis on the financial crisis data indicates that comparing two crisis periods gives a more definite model choice than using a rolling data estimating window with the most recent historical data. However, this may not be the case for all stocks, and it will most likely depend on the market sector the stock belongs to.

Chapter 9

Conclusion

This thesis assesses the model performance of the standard GARCH, EGARCH and APARCH models, with three different distributional assumptions, using the return series of the companies NEL, Tesla, Tomra, Norsk Hydro, General Motors and Aker BP from November 2010 to January 2021. For NEL and Norsk Hydro, the data sets are extended to include data from January 2006 to January 2021. As these models can represent time-varying volatility with different levels of complexity, we try to distinguish whether the former three stocks, which are considered green, differ in model characteristics from the three latter stocks, which are considered non-green.

The in-sample analysis for the period November 2010 to March 2017 documents that models with lag parameters $p = q = 1$ usually perform best. Moreover, the analysis shows that models with t-distributed innovations outperform normally distributed GARCH models. Except for NEL, where the skew distribution yields better results, the ordinary student-t distribution appears to give the best performance. In the corresponding out-of-sample analysis, the optimal distributional assumption seems to depend more heavily on the return characteristics of the individual stocks. For NEL, Tesla, Norsk Hydro and Aker BP, some normally distributed models produce good results, but in most cases, their performance is inferior compared to the t-distributed models. In fact, the t-distributed models appear to give reasonable results in most situations. When the normally distributed models perform well, it seems to be due to more normally distributed returns or because even the more advanced GARCH models cannot capture the characteristics of the return series.

The VaR and ES analyses show small differences in the performance of the conditional volatility models. Neither of the GARCH models seem particularly better at VaR and ES estimation or at modelling the volatility. This means that depending on the final application, different models can be preferred. The model choice should be based on the VaR performance when the percentile is most interesting, and equivalently it should be based on the ES performance when the size of the loss is most important. In practice, VaR is typically relevant for daily and regular movements as they impact the profit and loss of portfolios, whereas ES is critical for understanding what kind of capital we need against a portfolio of instruments. Furthermore, if the performance is similar, the least complex model should be preferred.

This result applies to both the green and non-green stocks. However, GARCH models seem to react slower to changing volatility when they are fitted to non-green data. This becomes particularly evident during the market crisis caused by COVID-19. The reason may be that green stocks have experienced larger volatility before the crisis due to the green trend in the market during recent years. Intuitively, this implies that it is not necessarily the company's green image that produces these results, but rather the volatility structure. There could also be other factors such as the phase of the life cycle or the market sectors the companies are within that influence the model performance more heavily.

An example is the additional analysis of NEL and Norsk Hydro returns, where the estimation set includes data from the global financial crisis. For Norsk Hydro the global financial crisis and COVID-19 crisis seem to have similar characteristics, which lead to better VaR exceedance performance when the models are fitted to crisis data compared to a rolling window of the most recent historical returns. For NEL, however, both estimation sets appear to give similar model performance. The additional analysis also documents that the VaR and ES analyses are still somewhat incongruent in the choice of GARCH model. A significant difference is that both the ES and VaR analyses agree on the model choice for the two percentile levels for this alternative calibration. This is in contrast to the previous analyses for which only the ES were in agreement. Thus, careful investigation and consideration of the estimation sample seem to be of importance for the model performance.

9.1 Future Work

As the choice of error distribution seems to be of larger importance than the choice of conditional volatility model, a natural progression of this work is to investigate other distributional assumptions. The ordinary and skew versions of the generalised hyperbolic distribution or the generalised error distribution are common choices [84, 32]. So is the normal inverse Gaussian distribution or Johnson's reparametrised SU distribution [46].

As previously mentioned, the conflicting results in the VaR and ES analyses may indicate that the returns are changing in a way these GARCH models are not able to model. Many other extensions of the GARCH model exist, and the family of autoregressive conditional heteroscedasticity models is still under development. Several models have been published during the last couple of years [85, 86]. Additionally, there exist other models that can be used for volatility modelling. An example is the observation-driven time series models called the generalised autoregressive score (GAS) models proposed by Creal, Koopman and Lucas [87]. These models have the scaled likelihood score as their driving mechanism, and they introduce time-varying parameters in various non-linear models. In fact, the GAS model framework is also inclusive of the GARCH models, among others.

Another feature of this study is that it considers one-day forecasting of volatility only. Future research could be conducted to investigate the impact of both longer and shorter time horizons. For instance, the Basel Committee requires the use of a 10-day time horizon in the calculation of capital charges of market risk [51]. Moreover, intra-day trading has become more prevalent in recent years. This could potentially change the market conditions, and finding the optimal VaR and ES model is therefore a continuous process rather than a one-time event. Additionally, this study addresses volatility modelling of stocks but can be extended to analyse how GARCH models perform on other asset classes such as stock indices, interest rates or currencies. It would also be interesting to investigate how multivariate GARCH models perform on various portfolios.

Bibliography

[1] Fidelity Learning Center. *2020 stock market report.* URL: https://www.fidelity.com/learning-center/trading-investing/markets-sectors/2020-stock-market-report (visited on 1st Mar. 2021).

[2] Alexander J McNeil. *Quantitative risk management : concepts, techniques and tools.* Princeton, N.J, 2005.

[3] Jon Danielsson. *Financial Risk Forecasting: The Theory and Practice of Forecasting Market Risk with Implementation in R and Matlab.* Vol. 590. Wiley finance series. Hoboken: John Wiley & Sons, Incorporated, 2011.

[4] Tim Bollerslev, Robert F. Engle and Daniel B Nelson. ≪ARCH models≫. In: *Handbook of econometrics* 4 (1994), pp. 2959–3038.

[5] Robert F Engle. ≪Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation≫. In: *Econometrica* 50.4 (1982), pp. 987–1007.

[6] Tim Bollerslev. ≪Generalized Autoregressive conditional heteroskedasticity≫. In: *Econometrica* 31 (1986), pp. 307–327.

[7] Paul Wilmott. *The mathematics of financial derivatives : a student introduction.* Cambridge, 1995.

[8] Eugene F. Fama. ≪Mandelbrot and the Stable Paretian Hypothesis≫. In: *The Journal of business (Chicago, Ill.)* 36.4 (1963), pp. 420–429.

[9] Tim Bollerslev. ≪A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return≫. In: *The review of economics and statistics* 69.3 (1987), pp. 542–547.

[10] Carmen Fernández and Mark F. J Steel. ≪On Bayesian Modeling of Fat Tails and Skewness≫. In: *Journal of the American Statistical Association* 93.441 (1998), pp. 359–371.

[11] Felipe M. Aparicio and Javier Estrada. ≪Empirical distributions of stock returns: European securities markets, 1990-95≫. In: *European Journal of Finance* 7 (2001), pp. 1–21.

[12] Peter R Hansen and Asger Lunde. ≪A forecast comparison of volatility models: does anything beat a GARCH(1,1)?≫ In: *Journal of applied econometrics (Chichester, England)* 20.7 (2005), pp. 873–889.

[13] Alexios Ghalanos. *Does anything NOT beat the GARCH(1,1)?* URL: http://www.unstarched.net/2013/01/07/does-anything-not-beat-the-garch11/ (visited on 5th Nov. 2020).

[14] Klaus Neusser. *Time series econometrics.* 2016.

[15] Jonathan D. Cryer. *Time series analysis : with applications in R.* New York, New York, 2008.

[16] Robert H Shumway and David S. Stoffer. *Time Series Analysis and Its Applications : With R Examples.* Cham, 2017.

[17] Jürgen Franke, Wolfgang Karl Härdle and Christian Matthias Hafner. *Statistics of Financial Markets: An Introduction.* Fifth edition. Universitext. Cham: Springer International Publishing AG, 2019.

[18] Christian Francq and Jean-Michel Zakoian. *GARCH Models - Structure, Statistical Inference and Financial Applications.* Chichester, 2019.

[19] Peter J Brockwell and Richard A Davis. *Introduction to Time Series and Forecasting.* Springer texts in statistics. Cham: Springer International Publishing AG, 2016.

[20] David Ruppert and David S. Matteson. *Statistics and data analysis for financial engineering : with R examples*. New York, 2015.

[21] George Casella, Stephen Fienberg and Ingram Okin. *Statistical Models and Methods for Financial Markets*. Springer texts in statistics. New York: Springer New York, 2008.

[22] Carol Alexander. *Market risk analysis : Vol. 2 : Practical financial econometrics*. Chichester, 2008.

[23] Neil Shephard. ≪Statistical aspects of ARCH and stochastic volatility≫. In: *Time Series Models in Econometrics, Finance and Other Fields.* (edited by D.R. Cox, David V. Hinkley and Ole E. Barndorff-Neilsen). Reprinted in the Survey of Applied and Industrial Mathematics, issue on Financial and insurance mathematics, 3, 764-826, Scientific Publisher TVP, Moscow, 1996 (in Russian). London: Chapman & Hall, 1996, pp. 1–67.

[24] Ly Pham. *Time Series Analysis with ARIMA - ARCH/GARCH model in R*. 2013.

[25] Peter F Christoffersen. *Elements of financial risk management*. Waltham, 2012.

[26] Benoit Mandelbrot. ≪The Variation of Certain Speculative Prices≫. In: *The Journal of business (Chicago, Ill.)* 36.4 (1963), pp. 394–419.

[27] Eric Jondeau, Ser-Huang Poon and Michael Rockinger. *Financial Modeling under Non-Gaussian Distributions*. Springer finance. London: Springer London, Limited, 2006.

[28] Carlos M. Jarque and Anil K. Bera. ≪A Test for Normality of Observations and Regression Residuals≫. In: *International statistical review* 55.2 (1987), pp. 163–172.

[29] Ruey S. Tsay. *An introduction to analysis of financial data with R*. Hoboken, New Jersey, 2013.

[30] Chris Brooks. *Introductory econometrics for finance*. Cambridge, 2008.

[31] Rasim Ilker Gokbulut and Mehmet Pekkaya. ≪Estimating and Forecasting Volatility of Financial Markets Using Asymmetric GARCH Models: An Application on Turkish Financial Markets≫. In: *International journal of economics and finance* 6.4 (2014).

[32] Daniel B Nelson. ≪Conditional Heteroskedasticity in Asset Returns: A New Approach≫. In: *Econometrica* 59.2 (1991), pp. 347–370.

[33] Zhuanxin Ding, Clive W.J Granger and Robert F Engle. ≪A long memory property of stock market returns and a new model≫. In: *Journal of empirical finance*. Journal of Empirical Finance 1.1 (1993), pp. 83–106.

[34] Philippe Lambert and Sébastian Laurent. *Modelling financial time series using GARCH-type models with a skewed Student distribution for the innovations*. 2001.

[35] Carol Alexander. *Market risk analysis : Vol. 1 : Quantitative methods in finance*. Chichester, 2008.

[36] Alexios Ghalanos. *rugarch: Univariate GARCH models. R package version 1.4-4*. 2020. URL: https://cran.r-project.org/web/packages/rugarch/rugarch.pdf (visited on 3rd Feb. 2021).

[37] Alexios Ghalanos and Stefan Theussl. *Rsolnp: General Non-Linear Optimization*. 2015. URL: https://cran.r-project.org/web/packages/Rsolnp/Rsolnp.pdf (visited on 2nd Feb. 2021).

[38] Ludwig Fahrmeir et al. *Regression: Models, Methods and Applications*. 2013th ed. Berlin, Heidelberg, 2013.

[39] Kevin Dowd. *Measuring Market Risk*. Wiley finance series. Chichester, UK: John Wiley & Sons, Ltd, 2005.

[40] Paul Wilmott. *Frequently asked questions in quantitative finance : including key models, important formulæ, common contracts, a history of quantitative finance, sundry lists, brainteasers and more*. Chichester, England ; 2007.

[41] Hans-Peter Deutsch and Mark W Beinker. *Derivatives and Internal Models: Modern Risk Management*. Finance and capital markets. Cham: Springer International Publishing AG, 2019.

[42] Banks for International Settlements. *History of the Basel Committee*. URL: https://www.bis.org/bcbs/history.htm (visited on 20th Jan. 2021).

[43] Philippe Artzner et al. «Coherent Measures of Risk». In: *Mathematical finance*. Mathematical Finance 9.3 (1999), pp. 203–228.

[44] J.P. Morgan. *RiskMetrics - Technical document*. New York, 1996.

[45] Leonard Brie, Benoit Genest and Matthieu Arsac. *Back-testing of Expected Shortfall: Main challenges and methodologies*. 2018.

[46] R. A. Rigby and D. M. Stasinopoulos. «Generalized Additive Models for Location, Scale and Shape». In: *Journal of the Royal Statistical Society. Series C (Applied Statistics)*. Journal of the Royal Statistical Society Series C 54.3 (2005), pp. 507–554.

[47] Jón Daníelsson et al. «Fat tails, VaR and subadditivity». In: *Journal of econometrics* 172.2 (2013), pp. 283–291.

[48] Dominique Guégan and Bertrand K Hassani. *Risk Measurement: From Quantitative Measures to Management Decisions*. Cham: Springer International Publishing AG, 2019.

[49] Banks for International Settlements. *Amendment to the capital accord to incorporate market risks*. 1996. URL: https://www.bis.org/publ/bcbs24.pdf (visited on 2nd Mar. 2021).

[50] Banks for International Settlements. *Fundamental review of the trading book: A revised market risk framework*. 2013. URL: https://www.bis.org/publ/bcbs265.pdf (visited on 20th Jan. 2021).

[51] Banks for International Settlements. *Minimum capital requirements for market risk*. 2016. URL: www.bis.org/bcbs/publ/d352.pdf (visited on 17th Mar. 2021).

[52] Carlo Acerbi and Dirk Tasche. «On the coherence of expected shortfall». In: *Journal of banking & finance*. Journal of Banking & Finance 26.7 (2002), pp. 1487–1503.

[53] Paul H Kupiec. «Techniques for verifying the accuracy of risk measurement models». In: *The Journal of Derivatives* (1995).

[54] Peter F. Christoffersen. «Evaluating Interval Forecasts». In: *International economic review (Philadelphia)* 39.4 (1998), pp. 841–862.

[55] Tilmann Gneiting. «Making and Evaluating Point Forecasts». In: *Journal of the American Statistical Association* 106.494 (2011), pp. 746–762.

[56] Alfonso Novales and Laura Garcia-Jorcano. «Backtesting extreme value theory models of expected shortfall». In: *Quantitative finance* 19.5 (2019), pp. 799–825.

[57] L. Carver. «Mooted VaR substitute cannot be backtested, says top quant». In: *Risk* (2013).

[58] Susanne Emmer, Marie Kratz and Dirk Tasche. «What is the best risk measure in practice? A comparison of standard measures». In: *The journal of risk* 18.2 (2015), pp. 31–60.

[59] Tobias Fissler and Johanna F. Ziegel. «HIGHER ORDER ELICITABILITY AND OSBAND'S PRINCIPLE». In: *The Annals of statistics* 44.4 (2016), pp. 1680–1707.

[60] Carlo Acerbi and Balázs Székely. «Backtesting Expected Shortfall». In: *AsiaRisk* (2014), p. 42.

[61] Marie Kratz. «DISCUSSION OF "ELICITABILITY AND BACKTESTING: PERSPECTIVES FOR BANKING REGULATION"». In: *The annals of applied statistics* 11.4 (2017), pp. 1894–1900.

[62] Alexander J McNeil and Rüdiger Frey. «Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach». In: *Journal of empirical finance*. Journal of Empirical Finance 7.3-4 (2000), pp. 271–300.

[63] Bradley Efron. *An Introduction to the bootstrap*. Boca Raton, Fla, 1993.

[64] Paul Embrechts, Roger Kaufmann and Pierre Patie. «Strategic Long-Term Financial Risks: Single Risk Factors». In: *Computational optimization and applications* 32.1 (2005), pp. 61–90.

[65] Natalia Nolde and Johanna F. Ziegel. «ELICITABILITY AND BACKTESTING: PERSPECTIVES FOR BANKING REGULATION». In: *The annals of applied statistics* 11.4 (2017), pp. 1833–1874.

[66] Timo Dimitriadis and Sebastian Bayer. «A joint quantile and expected shortfall regression framework». In: *Electronic journal of statistics* 13.1 (2019), pp. 1823–1871.

[67] Sebastian Bayer and Timo Dimitriadis. «Regression Based Expected Shortfall Backtesting». In: *Journal of Financial Econometrics* (2018).

[68] Andrew J Patton, Johanna F Ziegel and Rui Chen. «Dynamic semiparametric models for expected shortfall (and Value-at-Risk)». In: *Journal of econometrics* 211.2 (2019), pp. 388–413.

[69] Sander Barendse. *Interquantile Expectation Regression*. Tinbergen Institute Discussion Papers 17-034/III. Tinbergen Institute, 2017. URL: https://ideas.repec.org/p/tin/wpaper/20170034.html.

[70] Florentijn Gelling. «Expected Shortfall Backtesting». Master's Thesis. Erasmus University Rotterdam, 2019.

[71] Banks for International Settlements. *Working Paper No. 19 - Messages from the academic literature on risk measurement of the trading book*. 2011. URL: www.bis.org/publ/bcbs_wp19.pdf (visited on 2nd Mar. 2021).

[72] Banks for International Settlements. *Pillar 3 disclosure requirements - consolidated and enhanced framework*. 2017. URL: www.bis.org/bcbs/publ/d400.pdf (visited on 2nd Mar. 2021).

[73] Marcelo Brutti Righi and Paulo Sergio Ceretta. «Individual and flexible expected shortfall backtesting». In: *Journal of risk model validation* 7.3 (2013), pp. 3–20.

[74] Marie Kratz, Yen H Lok and Alexander J McNeil. «Multinomial VaR backtests: A simple implicit approach to backtesting expected shortfall». In: *Journal of banking & finance* 88 (2018), pp. 393–407.

[75] Nick Costanzino and Mike Curran. «Backtesting general spectral risk measures with application to expected shortfall». In: *Journal of risk model validation* 9.1 (2015), pp. 21–31.

[76] Woon K Wong. «Backtesting trading risk of commercial banks using expected shortfall». In: *Journal of banking & finance*. Journal of Banking & Finance 32.7 (2008), pp. 1404–1415.

[77] Jeremy Berkowitz. «Testing Density Forecasts, With Applications to Risk Management». In: *Journal of business & economic statistics* 19.4 (2001), pp. 465–474.

[78] Blogged Finance. *Are You Wondering Is Tesla an ESG Stock?* URL: https://bloggedfinance.com/tesla-esg-stock/ (visited on 22nd Mar. 2021).

[79] James Proberts and Georgi N. Boshnakov. *ARMA-GARCH modelling and white noise tests*. 2014. URL: https://cran.r-project.org/web/packages/sarima/vignettes/garch_tests_example.pdf.

[80] Hasle Stian. *Hydro share to start trading on OTCQX International*. 2009. URL: https://www.hydro.com/no-NO/media/news/2009/hydro-share-to-start-trading-on-otcqx-international/ (visited on 22nd Apr. 2021).

[81] Tim Bollerslev, Ray Y Chou and Kenneth F Kroner. «ARCH modeling in finance: A review of the theory and empirical evidence». In: *Journal of econometrics*. Journal of Econometrics 52.1 (1992), pp. 5–59.

[82] Jon Danielsson. *Univariate Volatility Modelling*. 2020. URL: https://www.financialriskforecasting.com/files/Danielsson-Financial-Risk-Forecasting-Slides-2.pdf (visited on 5th Mar. 2021).

[83] Sebastian Bayer and Timo Dimitriadis. *esback: Expected Shortfall Backtesting. R package version 0.3.0*. 2020. URL: https://cran.r-project.org/web/packages/esback/esback.pdf (visited on 10th Feb. 2021).

[84] Kjersti Aas and Ingrid Hobaek Haff. «The Generalized Hyperbolic Skew Student's t-Distribution». In: *Journal of financial econometrics*. Journal of Financial Econometrics 4.2 (2006), pp. 275–309.

[85] Dong Li et al. «The ZD-GARCH model: A new way to study heteroscedasticity». In: *Journal of econometrics* 202.1 (2018), pp. 1–17.

[86] Philipp Otto, Wolfgang Schmid and Robert Garthoff. «Generalised spatial and spatiotemporal autoregressive conditional heteroscedasticity». In: *Spatial statistics* 26 (2018), pp. 125–145.

[87] Drew Creal, Siem Jan Koopman and André Lucas. *A General Framework for Observation Driven Time-Varying Parameter Models*. Institute of Economic Research, Hitotsubashi University, 2009.

Appendix A

Closed-form Solution of Expected Shortfall

To derive closed-form solutions of the expected shortfall, we first consider the conditional distribution of the return Y_t as $f(y_t|y_t < -VaR_t(\alpha))$. From the well-known definition of conditional probability, we have

$$f(y_t|y_t \leq -\mathrm{VaR}_t(\alpha)) = \frac{f(y_t)}{P(y_t \leq -VaR_t(\alpha))} = \frac{1}{\alpha} f(y_t), \tag{A.1}$$

where we in the last equality have used the definition of VaR in equation (4.1). From equation (6.1) we have

$$\begin{aligned} \mathrm{ES}_t(\alpha) &= -E[Y_t|Y_t \leq -\mathrm{VaR}_t(\alpha)] \\ &= -\int_{-\infty}^{-\mathrm{VaR}_t(\alpha)} y_t f(y_t|y_t < -\mathrm{VaR}_t(\alpha)) dy_t \\ &= -\frac{1}{\alpha} \int_{-\infty}^{-VaR_t(\alpha)} y_t f(y_t) dy_t, \end{aligned} \tag{A.2}$$

where we in the last equality have used the result from equation (A.1).

A.1 Normal Distribution

By assuming standard normally distributed innovations, i.e. $Z_t \sim N(0,1)$, and since $Y_t = \sigma_t Z_t$ we have $Y_t|\mathcal{F}_{t-1} \sim N(0, \sigma_t^2)$. From equation (5.1), the upper limit of the integral in equation (A.2) is -$\mathrm{VaR}_t(\alpha) = \sigma_t \Phi^{-1}(\alpha)$. Moreover, from equation (3.11) we have that $f_n(y_t) = \frac{1}{\sigma_t} f_{n^*}\left(\frac{y_t}{\sigma_t}\right)$, where

$f_{n^*}(\cdot)$ is given in equation (3.10). Inserting this into (A.2) yields

$$\begin{aligned}
\mathrm{ES}_t(\alpha) &= -\frac{1}{\alpha}\int_{-\infty}^{-\mathrm{VaR}_t(\alpha)} y_t f_n(y_t) dy_t \\
&= -\frac{1}{\alpha}\int_{-\infty}^{\sigma_t \Phi^{-1}(\alpha)} y_t \frac{1}{\sigma_t} f_{n^*}\left(\frac{y_t}{\sigma_t}\right) dy_t \\
&= -\frac{1}{\alpha}\int_{-\infty}^{\sigma_t \Phi^{-1}(\alpha)} y_t \frac{1}{\sigma_t}\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\frac{y_t^2}{\sigma_t^2}\right) dy_t \\
&= \frac{1}{\alpha}\frac{1}{\sqrt{2\pi}}\sigma_t \left[\exp\left(-\frac{1}{2}\frac{y_t^2}{\sigma_t^2}\right)\right]_{-\infty}^{\sigma_t \Phi^{-1}(\alpha)} \\
&= \frac{1}{\alpha}\frac{1}{\sqrt{2\pi}}\sigma_t \exp\left(-\frac{1}{2}\left(\Phi^{-1}(\alpha)\right)^2\right) \\
&= \sigma_t \frac{\phi\left(\Phi^{-1}(\alpha)\right)}{\alpha},
\end{aligned}$$

where ϕ is the standard normal density function and Φ is the standard normal cumulative distribution function.

A.2 Student-t Distribution

If the innovations are assumed to be scaled t-distributed, i.e. $Z_t \sim t_\nu^*(0,1)$, the value-at-risk in equation (5.2) gives the upper limit -$\mathrm{VaR}_t(\alpha) = \sigma_t t_\nu^{*-1}(\alpha)$, where as before, $t_\nu^*(\alpha)$ denotes the cumulative distribution function of the standardised student-t distribution. Following the same procedure as for the normally distributed innovations, we have for the expected shortfall

$$\begin{aligned}
ES_t(\alpha) &= -\frac{1}{\alpha}\int_{-\infty}^{-\mathrm{VaR}_t(\alpha)} y_t f_t(y_t) dy_t \\
&= -\frac{1}{\alpha}\int_{-\infty}^{\sigma_t t_\nu^{*-1}(\alpha)} y_t \frac{1}{\sigma_t} f_{t^*}\left(\frac{y_t}{\sigma_t}\right) dy_t \\
&= -\frac{1}{\alpha}\int_{-\infty}^{\sigma_t t_\nu^{*-1}(\alpha)} y_t \frac{1}{\sigma_t}\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma(\frac{\nu}{2})\sqrt{\pi\left(\nu-2\right)}}\left(1+\frac{y_t^2}{\sigma_t^2(\nu-2)}\right)^{-(1+\nu)/2} dy_t \\
&= \frac{1}{\alpha}\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma(\frac{\nu}{2})\sqrt{\pi\left(\nu-2\right)}}\frac{\sigma_t(\nu-2)}{(\nu-1)}\left[-\left(1+\frac{y_t^2}{\sigma_t^2(\nu-2)}\right)^{-\frac{1}{2}(\nu-1)}\right]_{\sigma_t t_\nu^{*-1}(\alpha)}^{\infty} \\
&= \frac{1}{\alpha}\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma(\frac{\nu}{2})\sqrt{\pi\left(\nu-2\right)}}\frac{\sigma_t(\nu-2)}{(\nu-1)}\left(1+\frac{t_\nu^{*-1}(\alpha)^2}{\nu-2}\right)^{-\frac{1}{2}(\nu-1)} \\
&= \sigma_t \frac{g_\nu^*\left(t_\nu^{*-1}(\alpha)\right)}{\alpha}\left(\frac{\nu-2+t_\nu^{*-1}(\alpha)^2}{\nu-1}\right),
\end{aligned} \tag{A.3}$$

where $g_\nu^*(\cdot)$ is the density function of the standardised student-t distribution.

From equation (5.2), the standardised student-t distribution function can be expressed as

$$t_\nu^{*-1}(\alpha) = \sqrt{\frac{\nu-2}{\nu}}\, t_\nu^{-1}(\alpha).$$

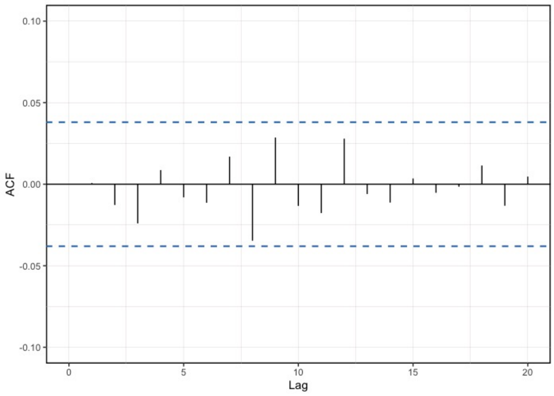
Substituting this into the expression for the expected shortfall in equation (A.3) gives

$$ES_t(\alpha) = \sigma_t \sqrt{\frac{\nu-2}{\nu}}\frac{g_\nu\left(t_\nu^{-1}(\alpha)\right)}{\alpha}\left(\frac{\nu+t_\nu^{-1}(\alpha)^2}{\nu-1}\right),$$

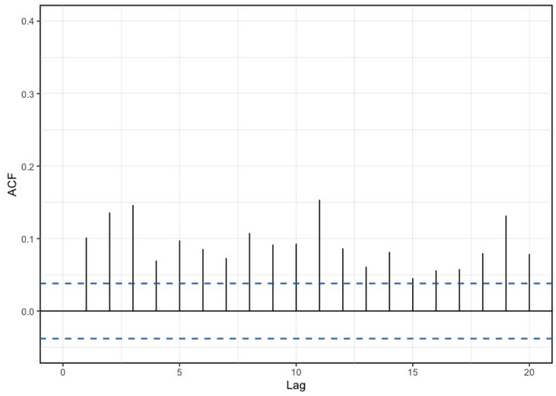
where $g_\nu(\cdot)$ and $t_\nu^{-1}(\cdot)$ denote the density and distribution function of the student-t distribution, respectively.

Appendix B

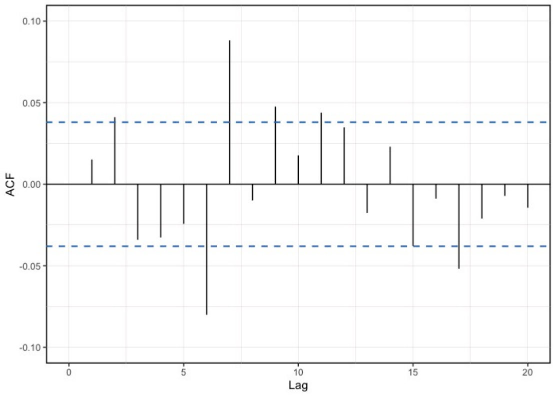
ACF Plots for Non-Green Stocks



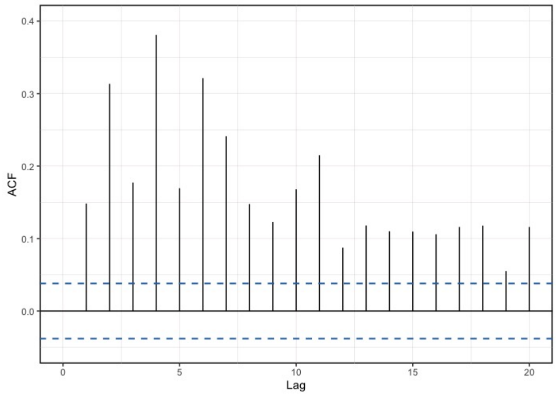
(a) ACF of Norsk Hydro returns.



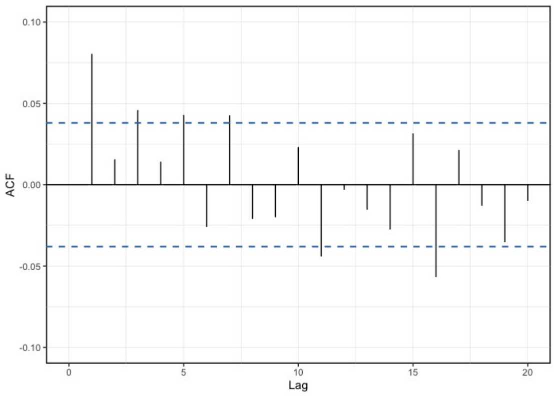
(b) ACF of squared Norsk Hydro returns.



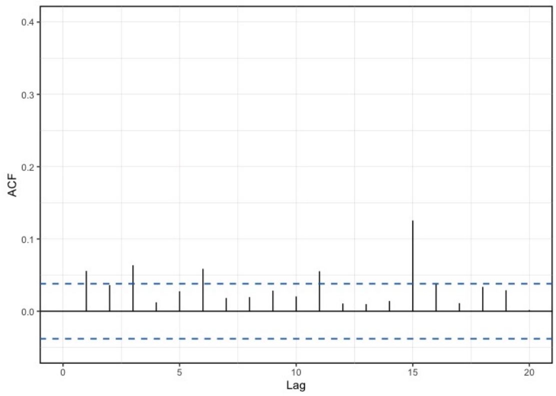
(c) ACF of General Motors returns.



(d) ACF of squared General Motors returns.



(e) ACF of Aker BP returns.



(f) ACF of squared Aker BP returns.

Figure B.1: Sample ACF of returns and squared returns associated to the daily closing values of the Norsk Hydro, General Motors and Aker BP stocks from 17.11.2010 to 21.01.2021. The horizontal dashed lines denote the two standard error limits of the sample ACF.

Appendix C

Individual Stock Model Performance

C.1 NEL

Table C.1: Information criteria for standard GARCH models fitted to NEL returns.

Model	Conditional dist.	AIC	BIC	AICc
GARCH(1,1)	Normal	-2.933	-2.924	-2.919
GARCH(1,1)	Student-t	-3.443	-3.430	-3.419
GARCH(1,1)	Skew Student-t	**-3.463**	**-3.447**	**-3.427**
GARCH(1,2)	Normal	-2.936	-2.923	-2.912
GARCH(1,2)	Student-t	-3.440	-3.424	-3.404
GARCH(1,2)	Skew Student-t	-3.460	-3.440	-3.409
GARCH(2,1)	Normal	-2.933	-2.920	-2.909
GARCH(2,1)	Student-t	-3.440	-3.424	-3.404
GARCH(2,1)	Skew Student-t	-3.460	-3.440	-3.409
GARCH(2,2)	Normal	-2.935	-2.919	-2.899
GARCH(2,2)	Student-t	-3.439	-3.419	-3.388
GARCH(2,2)	Skew Student-t	-3.459	-3.436	-3.391

Table C.2: Information criteria for EGARCH models fitted to NEL returns.

Model	Conditional dist.	AIC	BIC	AICc
EGARCH(1,1)	Normal	-2.970	-2.957	-2.946
EGARCH(1,1)	Student-t	-3.447	-3.431	-3.411
EGARCH(1,1)	Skew Student-t	**-3.467**	**-3.448**	**-3.416**
EGARCH(1,2)	Normal	-2.968	-2.951	-2.931
EGARCH(1,2)	Student-t	-3.445	-3.425	-3.394
EGARCH(1,2)	Skew Student-t	-3.465	-3.442	-3.397
EGARCH(2,1)	Normal	-2.984	-2.965	-2.933
EGARCH(2,1)	Student-t	-3.447	-3.424	-3.379
EGARCH(2,1)	Skew Student-t	**-3.467**	-3.441	-3.379
EGARCH(2,2)	Normal	-2.983	-2.960	-2.915
EGARCH(2,2)	Student-t	-3.446	-3.420	-3.359
EGARCH(2,2)	Skew Student-t	-3.466	-3.437	-3.357

Table C.3: Information criteria for APARCH models fitted to NEL returns.

Model	Conditional dist.	AIC	BIC	AICc
APARCH(1,1)	Normal	-2.938	-2.922	-2.902
APARCH(1,1)	Student-t	-3.447	-3.427	-3.396
APARCH(1,1)	Skew Student-t	**-3.467**	**-3.444**	**-3.399**
APARCH(1,2)	Normal	-2.939	-2.920	-2.888
APARCH(1,2)	Student-t	-3.443	-3.420	-3.375
APARCH(1,2)	Skew Student-t	-3.462	-3.436	-3.375
APARCH(2,1)	Normal	-2.935	-2.912	-2.867
APARCH(2,1)	Student-t	-3.441	-3.415	-3.354
APARCH(2,1)	Skew Student-t	-3.462	-3.432	-3.352
APARCH(2,2)	Normal	-2.976	-2.950	-2.888
APARCH(2,2)	Student-t	-3.441	-3.412	-3.332
APARCH(2,2)	Skew Student-t	-3.461	-3.429	-3.328

C.2 Tesla

Table C.4: Information criteria for standard GARCH models fitted to Tesla returns.

Model	Conditional dist.	AIC	BIC	AICc
GARCH(1,1)	Normal	-4.160	-4.151	-4.146
GARCH(1,1)	Student-t	**-4.350**	**-4.337**	**-4.325**
GARCH(1,1)	Skew Student-t	-4.349	-4.333	-4.313
GARCH(1,2)	Normal	-4.159	-4.146	-4.135
GARCH(1,2)	Student-t	-4.349	-4.332	-4.312
GARCH(1,2)	Skew Student-t	-4.348	-4.329	-4.297
GARCH(2,1)	Normal	-4.159	-4.146	-4.135
GARCH(2,1)	Student-t	-4.349	-4.332	-4.312
GARCH(2,1)	Skew Student-t	-4.348	-4.329	-4.297
GARCH(2,2)	Normal	-4.161	-4.145	-4.125
GARCH(2,2)	Student-t	-4.347	-4.328	-4.296
GARCH(2,2)	Skew Student-t	-4.347	-4.324	-4.279

Table C.5: Information criteria for EGARCH models fitted to Tesla returns.

Model	Conditional dist.	AIC	BIC	AICc
EGARCH(1,1)	Normal	-4.158	-4.145	-4.134
EGARCH(1,1)	Student-t	**-4.354**	**-4.338**	**-4.318**
EGARCH(1,1)	Skew Student-t	-4.353	-4.334	-4.303
EGARCH(1,2)	Normal	-4.157	-4.141	-4.121
EGARCH(1,2)	Student-t	-4.353	-4.333	-4.302
EGARCH(1,2)	Skew Student-t	-4.352	-4.330	-4.284
EGARCH(2,1)	Normal	-4.160	-4.141	-4.109
EGARCH(2,1)	Student-t	-4.352	-4.329	-4.284
EGARCH(2,1)	Skew Student-t	-4.352	-4.326	-4.264
EGARCH(2,2)	Normal	-4.165	-4.142	-4.097
EGARCH(2,2)	Student-t	-4.351	-4.325	-4.264
EGARCH(2,2)	Skew Student-t	-4.351	-4.321	-4.241

Table C.6: Information criteria for APARCH models fitted to Tesla returns.

Model	Conditional dist.	AIC	BIC	AICc
APARCH(1,1)	Normal	-4.160	-4.143	-4.123
APARCH(1,1)	Student-t	**-4.350**	**-4.330**	**-4.299**
APARCH(1,1)	Skew Student-t	-4.349	-4.326	-4.281
APARCH(1,2)	Normal	-4.158	-4.139	-4.107
APARCH(1,2)	Student-t	-4.349	-4.326	-4.281
APARCH(1,2)	Skew Student-t	-4.348	-4.322	-4.261
APARCH(2,1)	Normal	-4.157	-4.134	-4.089
APARCH(2,1)	Student-t	-4.347	-4.321	-4.260
APARCH(2,1)	Skew Student-t	-4.347	-4.318	-4.238
APARCH(2,2)	Normal	-4.158	-4.132	-4.071
APARCH(2,2)	Student-t	-4.346	-4.317	-4.237
APARCH(2,2)	Skew Student-t	-4.346	-4.313	-4.212

C.3 Tomra

Table C.7: Information criteria for standard GARCH models fitted to Tomra returns.

Model	Conditional dist.	AIC	BIC	AICc
GARCH(1,1)	Normal	-5.208	-5.198	-5.193
GARCH(1,1)	Student-t	-5.323	-5.310	**-5.299**
GARCH(1,1)	Skew Student-t	-5.322	-5.306	-5.286
GARCH(1,2)	Normal	-5.208	-5.195	-5.184
GARCH(1,2)	Student-t	**-5.328**	**-5.312**	-5.292
GARCH(1,2)	Skew Student-t	**-5.328**	-5.308	-5.277
GARCH(2,1)	Normal	-5.206	-5.193	-5.182
GARCH(2,1)	Student-t	-5.322	-5.305	-5.285
GARCH(2,1)	Skew Student-t	-5.321	-5.301	-5.270
GARCH(2,2)	Normal	-5.208	-5.192	-5.172
GARCH(2,2)	Student-t	-5.327	-5.308	-5.276
GARCH(2,2)	Skew Student-t	-5.327	-5.304	-5.259

Table C.8: Information criteria for EGARCH models fitted to Tomra returns.

Model	Conditional dist.	AIC	BIC	AICc
EGARCH(1,1)	Normal	-5.220	-5.207	-5.196
EGARCH(1,1)	Student-t	-5.332	-5.316	**-5.296**
EGARCH(1,1)	Skew Student-t	-5.332	-5.312	-5.281
EGARCH(1,2)	Normal	-5.220	-5.204	-5.184
EGARCH(1,2)	Student-t	-5.332	-5.312	-5.281
EGARCH(1,2)	Skew Student-t	-5.331	-5.308	-5.263
EGARCH(2,1)	Normal	-5.236	-5.217	-5.186
EGARCH(2,1)	Student-t	**-5.342**	**-5.320**	-5.274
EGARCH(2,1)	Skew Student-t	**-5.342**	-5.316	-5.254
EGARCH(2,2)	Normal	-5.235	-5.213	-5.167
EGARCH(2,2)	Student-t	-5.341	-5.315	-5.254
EGARCH(2,2)	Skew Student-t	-5.341	-5.311	-5.231

Table C.9: Information criteria for APARCH models fitted to Tomra returns.

Model	Conditional dist.	AIC	BIC	AICc
APARCH(1,1)	Normal	-5.209	-5.192	-5.172
APARCH(1,1)	Student-t	**-5.331**	**-5.311**	**-5.280**
APARCH(1,1)	Skew Student-t	-5.330	-5.307	-5.262
APARCH(1,2)	Normal	-5.211	-5.191	-5.160
APARCH(1,2)	Student-t	-5.330	-5.307	-5.262
APARCH(1,2)	Skew Student-t	-5.330	-5.304	-5.242
APARCH(2,1)	Normal	-5.206	-5.183	-5.138
APARCH(2,1)	Student-t	-5.323	-5.297	-5.236
APARCH(2,1)	Skew Student-t	-5.323	-5.293	-5.213
APARCH(2,2)	Normal	-5.210	-5.184	-5.123
APARCH(2,2)	Student-t	-5.326	-5.296	-5.216
APARCH(2,2)	Skew Student-t	-5.325	-5.292	-5.191

C.4 Norsk Hydro

Table C.10: Information criteria for standard GARCH models fitted to Norsk Hydro returns.

Model	Conditional dist.	AIC	BIC	AICc
GARCH(1,1)	Normal	-5.247	-5.237	-5.232
GARCH(1,1)	Student-t	**-5.307**	**-5.294**	**-5.283**
GARCH(1,1)	Skew Student-t	-5.306	-5.289	-5.269
GARCH(1,2)	Normal	-5.245	-5.232	-5.221
GARCH(1,2)	Student-t	-5.306	-5.289	-5.269
GARCH(1,2)	Skew Student-t	-5.305	-5.285	-5.254
GARCH(2,1)	Normal	-5.247	-5.234	-5.222
GARCH(2,1)	Student-t	-5.306	-5.290	-5.270
GARCH(2,1)	Skew Student-t	-5.305	-5.285	-5.254
GARCH(2,2)	Normal	-5.245	-5.229	-5.209
GARCH(2,2)	Student-t	-5.305	-5.285	-5.254
GARCH(2,2)	Skew Student-t	-5.304	-5.281	-5.236

Table C.11: Information criteria for EGARCH models fitted to Norsk Hydro returns.

Model	Conditional dist.	AIC	BIC	AICc
EGARCH(1,1)	Normal	-5.257	-5.244	-5.233
EGARCH(1,1)	Student-t	**-5.320**	**-5.303**	**-5.283**
EGARCH(1,1)	Skew Student-t	-5.318	-5.299	-5.267
EGARCH(1,2)	Normal	-5.256	-5.240	-5.219
EGARCH(1,2)	Student-t	-5.318	-5.299	-5.267
EGARCH(1,2)	Skew Student-t	-5.317	-5.294	-5.249
EGARCH(2,1)	Normal	-5.256	-5.237	-5.205
EGARCH(2,1)	Student-t	-5.318	-5.295	-5.250
EGARCH(2,1)	Skew Student-t	-5.316	-5.290	-5.229
EGARCH(2,2)	Normal	-5.255	-5.232	-5.187
EGARCH(2,2)	Student-t	-5.317	-5.291	-5.229
EGARCH(2,2)	Skew Student-t	-5.315	-5.286	-5.206

Table C.12: Information criteria for APARCH models fitted to Norsk Hydro returns.

Model	Conditional dist.	AIC	BIC	AICc
APARCH(1,1)	Normal	-5.257	-5.241	-5.221
APARCH(1,1)	Student-t	**-5.317**	**-5.298**	**-5.267**
APARCH(1,1)	Skew Student-t	-5.316	-5.293	-5.248
APARCH(1,2)	Normal	-5.256	-5.237	-5.205
APARCH(1,2)	Student-t	-5.316	-5.293	-5.248
APARCH(1,2)	Skew Student-t	-5.315	-5.289	-5.227
APARCH(2,1)	Normal	-5.256	-5.233	-5.188
APARCH(2,1)	Student-t	-5.315	-5.289	-5.228
APARCH(2,1)	Skew Student-t	-5.314	-5.285	-5.205
APARCH(2,2)	Normal	-5.255	-5.229	-5.167
APARCH(2,2)	Student-t	-5.314	-5.285	-5.205
APARCH(2,2)	Skew Student-t	-5.313	-5.280	-5.179

C.5 General Motors

Table C.13: Information criteria for standard GARCH models fitted to General Motors returns.

Model	Conditional dist.	AIC	BIC	AICc
GARCH(1,1)	Normal	-5.279	-5.269	-5.264
GARCH(1,1)	Student-t	**-5.359**	**-5.346**	**-5.335**
GARCH(1,1)	Skew Student-t	-5.358	-5.342	-5.322
GARCH(1,2)	Normal	-5.279	-5.266	-5.255
GARCH(1,2)	Student-t	**-5.359**	-5.343	-5.323
GARCH(1,2)	Skew Student-t	-5.358	-5.338	-5.307
GARCH(2,1)	Normal	-5.278	-5.265	-5.254
GARCH(2,1)	Student-t	**-5.359**	-5.342	-5.322
GARCH(2,1)	Skew Student-t	-5.358	-5.338	-5.307
GARCH(2,2)	Normal	-5.278	-5.261	-5.241
GARCH(2,2)	Student-t	-5.358	-5.338	-5.307
GARCH(2,2)	Skew Student-t	-5.357	-5.334	-5.289

Table C.14: Information criteria for EGARCH models fitted to General Motors returns.

Model	Conditional dist.	AIC	BIC	AICc
EGARCH(1,1)	Normal	-5.287	-5.274	-5.263
EGARCH(1,1)	Student-t	**-5.368**	**-5.352**	**-5.332**
EGARCH(1,1)	Skew Student-t	-5.367	-5.348	-5.316
EGARCH(1,2)	Normal	-5.286	-5.270	-5.250
EGARCH(1,2)	Student-t	**-5.368**	-5.348	-5.317
EGARCH(1,2)	Skew Student-t	-5.367	-5.344	-5.299
EGARCH(2,1)	Normal	-5.289	-5.269	-5.238
EGARCH(2,1)	Student-t	-5.367	-5.345	-5.299
EGARCH(2,1)	Skew Student-t	-5.366	-5.340	-5.279
EGARCH(2,2)	Normal	-5.288	-5.265	-5.220
EGARCH(2,2)	Student-t	-5.366	-5.340	-5.279
EGARCH(2,2)	Skew Student-t	-5.365	-5.336	-5.256

Table C.15: Information criteria for APARCH models fitted to General Motors returns.

Model	Conditional dist.	AIC	BIC	AICc
APARCH(1,1)	Normal	-5.286	-5.270	-5.250
APARCH(1,1)	Student-t	**-5.368**	**-5.349**	**-5.317**
APARCH(1,1)	Skew Student-t	-5.367	-5.344	-5.299
APARCH(1,2)	Normal	-5.285	-5.266	-5.234
APARCH(1,2)	Student-t	-5.367	-5.344	-5.299
APARCH(1,2)	Skew Student-t	-5.366	-5.340	-5.279
APARCH(2,1)	Normal	-5.283	-5.260	-5.215
APARCH(2,1)	Student-t	-5.362	-5.336	-5.275
APARCH(2,1)	Skew Student-t	-5.364	-5.334	-5.254
APARCH(2,2)	Normal	-5.283	-5.256	-5.195
APARCH(2,2)	Student-t	-5.363	-5.333	-5.253
APARCH(2,2)	Skew Student-t	-5.361	-5.329	-5.228

C.6 Aker BP

Table C.16: Information criteria for standard GARCH models fitted to Aker BP returns.

Model	Conditional dist.	AIC	BIC	AICc
GARCH(1,1)	Normal	-4.471	-4.461	-4.457
GARCH(1,1)	Student-t	-4.686	**-4.673**	**-4.662**
GARCH(1,1)	Skew Student-t	-4.686	-4.670	-4.650
GARCH(1,2)	Normal	-4.472	-4.459	-4.448
GARCH(1,2)	Student-t	**-4.688**	-4.671	-4.651
GARCH(1,2)	Skew Student-t	**-4.688**	-4.668	-4.637
GARCH(2,1)	Normal	-4.470	-4.457	-4.446
GARCH(2,1)	Student-t	-4.685	-4.668	-4.648
GARCH(2,1)	Skew Student-t	-4.685	-4.665	-4.634
GARCH(2,2)	Normal	-4.471	-4.454	-4.434
GARCH(2,2)	Student-t	-4.686	-4.667	-4.635
GARCH(2,2)	Skew Student-t	-4.686	-4.663	-4.618

Table C.17: Information criteria for EGARCH models fitted to Aker BP returns.

Model	Conditional dist.	AIC	BIC	AICc
EGARCH(1,1)	Normal	-4.461	-4.448	-4.437
EGARCH(1,1)	Student-t	-4.711	**-4.695**	**-4.675**
EGARCH(1,1)	Skew Student-t	-4.711	-4.692	-4.661
EGARCH(1,2)	Normal	-4.461	-4.445	-4.425
EGARCH(1,2)	Student-t	-4.712	-4.692	-4.661
EGARCH(1,2)	Skew Student-t	-4.712	-4.689	-4.644
EGARCH(2,1)	Normal	-4.459	-4.439	-4.408
EGARCH(2,1)	Student-t	**-4.716**	-4.693	-4.648
EGARCH(2,1)	Skew Student-t	**-4.716**	-4.689	-4.628
EGARCH(2,2)	Normal	-4.459	-4.436	-4.391
EGARCH(2,2)	Student-t	-4.714	-4.688	-4.627
EGARCH(2,2)	Skew Student-t	-4.714	-4.685	-4.605

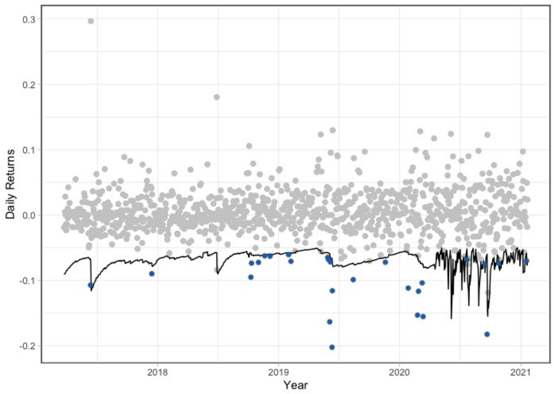
Table C.18: Information criteria for APARCH models fitted to Aker BP returns.

Model	Conditional dist.	AIC	BIC	AICc
APARCH(1,1)	Normal	-4.478	-4.461	-4.441
APARCH(1,1)	Student-t	-4.711	**-4.691**	**-4.660**
APARCH(1,1)	Skew Student-t	-4.711	-4.688	-4.643
APARCH(1,2)	Normal	-4.478	-4.458	-4.427
APARCH(1,2)	Student-t	**-4.712**	-4.689	-4.644
APARCH(1,2)	Skew Student-t	**-4.712**	-4.686	-4.624
APARCH(2,1)	Normal	-4.475	-4.452	-4.407
APARCH(2,1)	Student-t	-4.708	-4.682	-4.621
APARCH(2,1)	Skew Student-t	-4.708	-4.679	-4.599
APARCH(2,2)	Normal	-4.475	-4.449	-4.388
APARCH(2,2)	Student-t	-4.709	-4.680	-4.600
APARCH(2,2)	Skew Student-t	-4.709	-4.677	-4.576

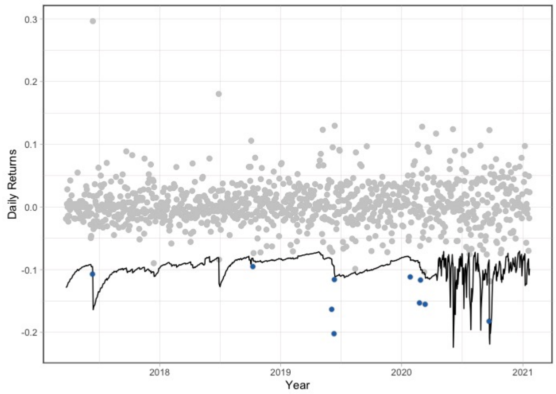
Appendix D

Individual Stock VaR Exceedance

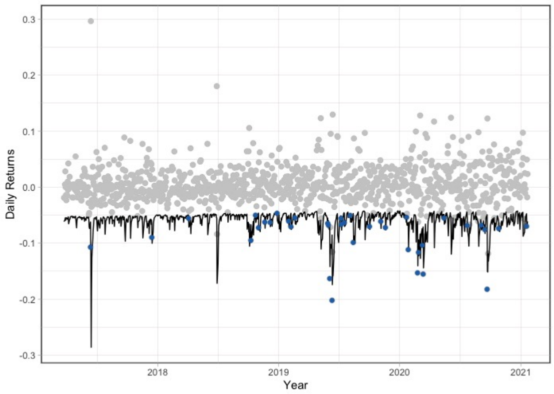
D.1 NEL



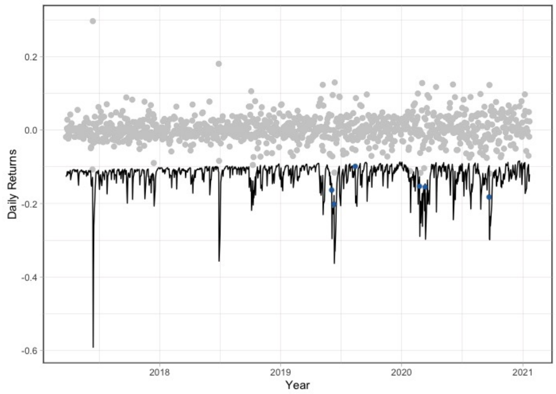
(a) Normal VaR(0.05)



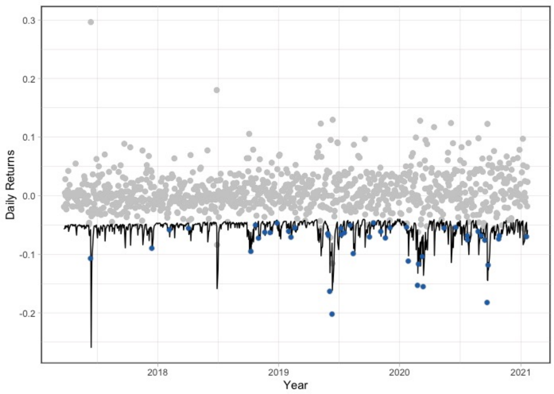
(b) Normal VaR(0.01)



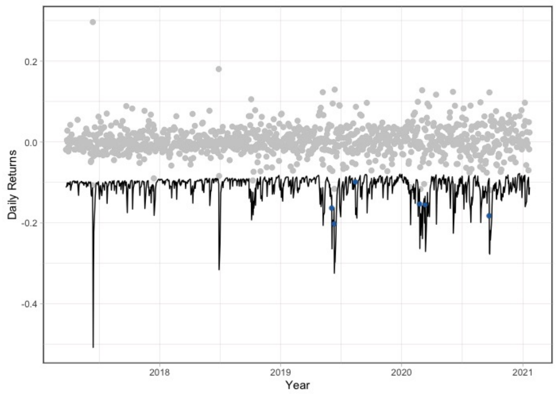
(c) Student-t VaR(0.05)



(d) Student-t VaR(0.01)

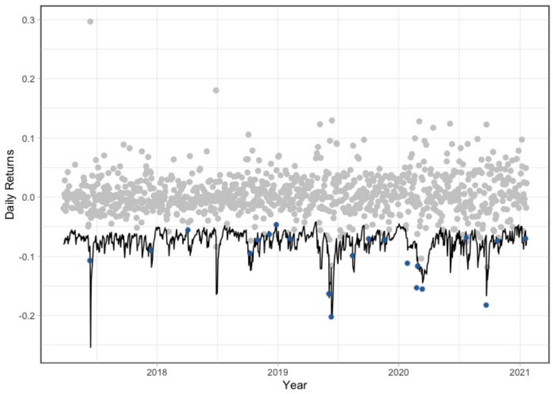


(e) Skew Student-t VaR(0.05)

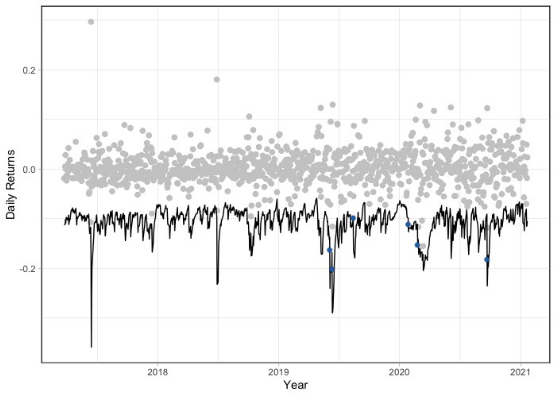


(f) Skew Student-t VaR(0.01)

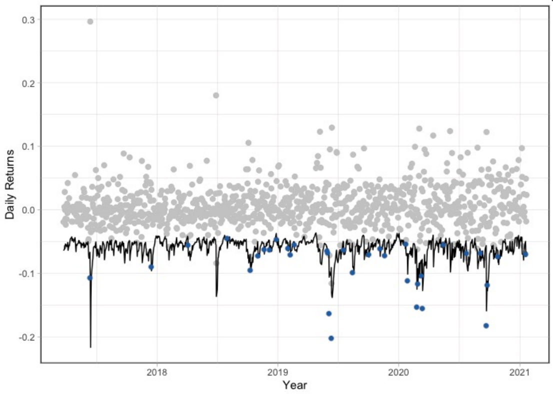
Figure D.1: VaR exceedance plots for GARCH(1,1) model on NEL stock March 2017 to January 2021. The black line indicates the negative VaR estimates, the gray marks the returns and the VaR violations are plotted as blue marks.



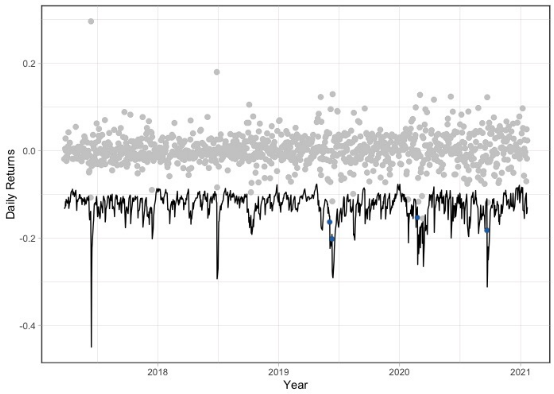
(a) Normal VaR(0.05)



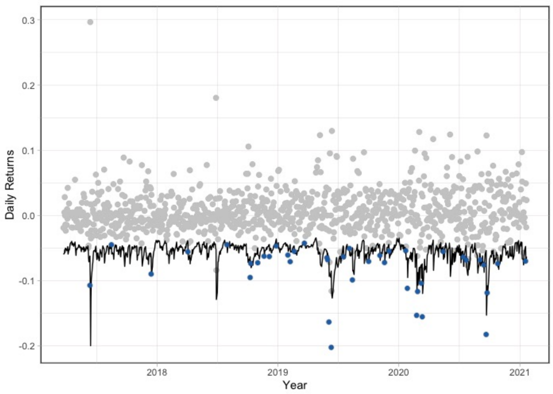
(b) Normal VaR(0.01)



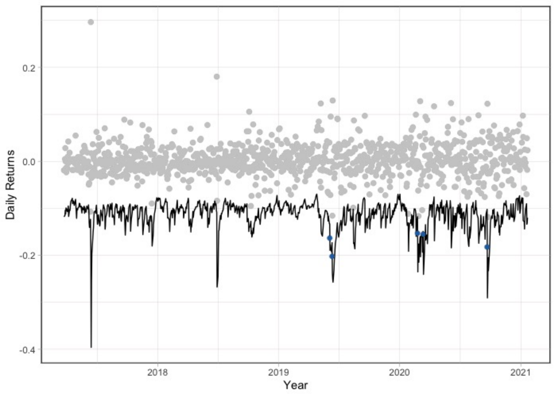
(c) Student-t VaR(0.05)



(d) Student-t VaR(0.01)

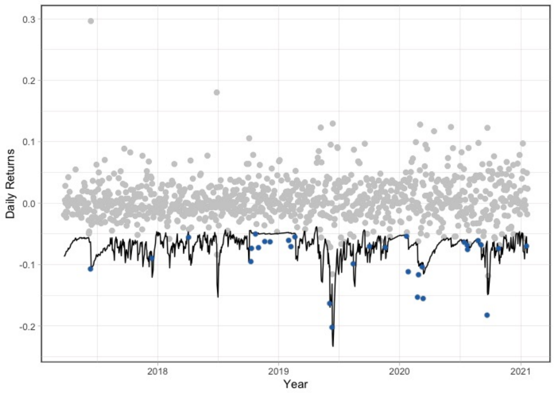


(e) Skew Student-t VaR(0.05)

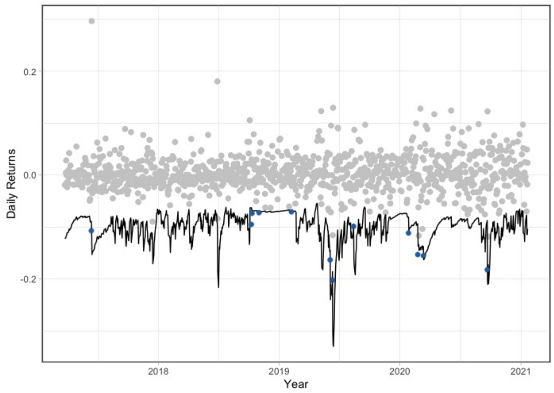


(f) Skew Student-t VaR(0.01)

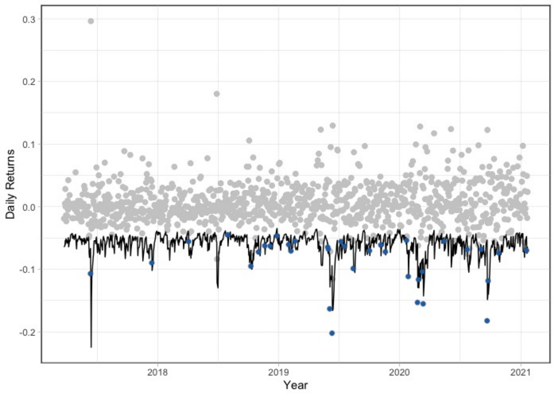
Figure D.2: VaR exceedance plots for EGARCH(1,1) model on NEL stock March 2017 to January 2021. The black line indicates the negative VaR estimates, the gray marks the returns and the VaR violations are plotted as blue marks.



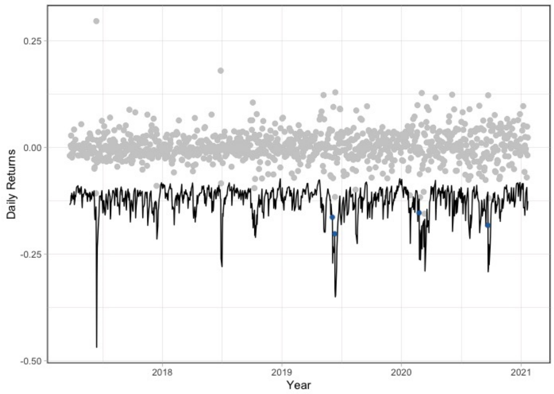
(a) Normal VaR(0.05)



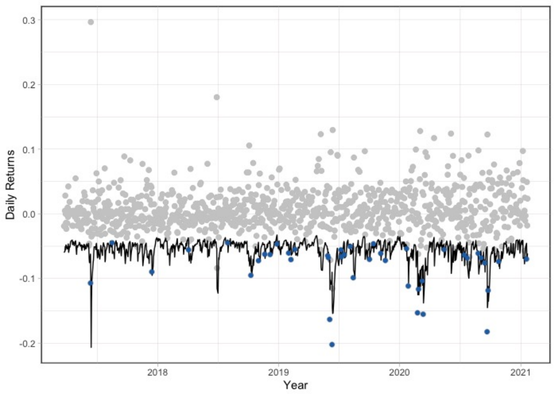
(b) Normal VaR(0.01)



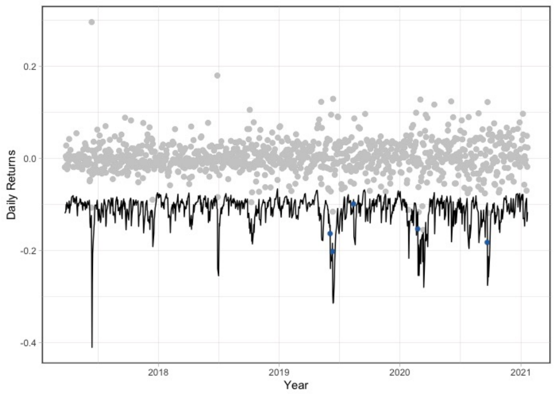
(c) Student-t VaR(0.05)



(d) Student-t VaR(0.01)



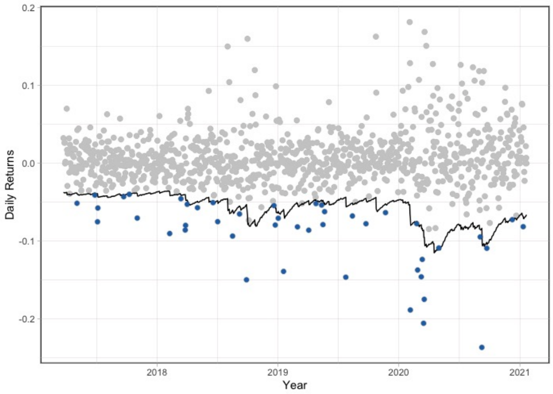
(e) Skew Student-t VaR(0.05)



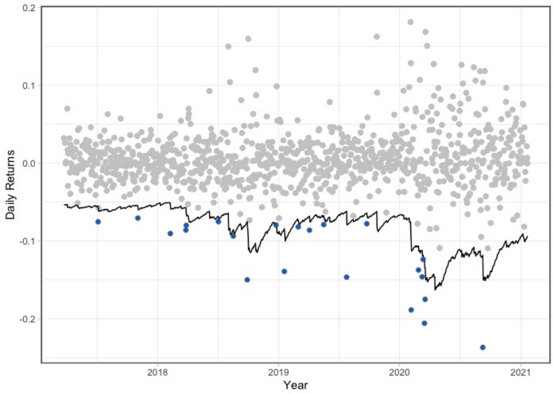
(f) Skew Student-t VaR(0.01)

Figure D.3: VaR exceedance plots for APARCH(1,1) model on NEL stock March 2017 to January 2021. The black line indicates the negative VaR estimates, the gray marks the returns and the VaR violations are plotted as blue marks.

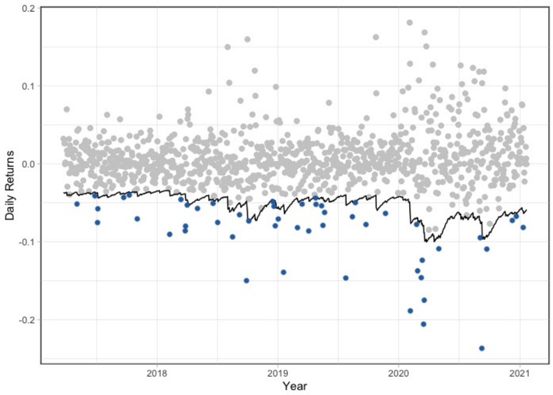
D.2 Tesla



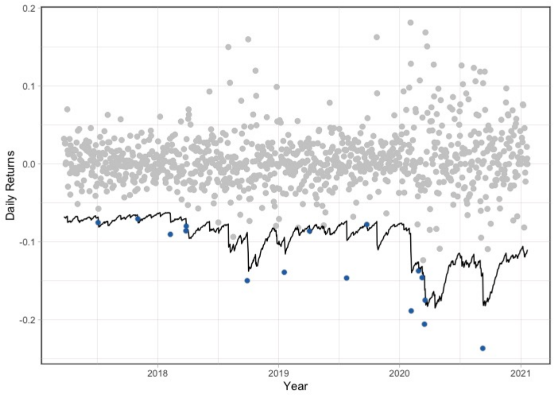
(a) Normal VaR(0.05)



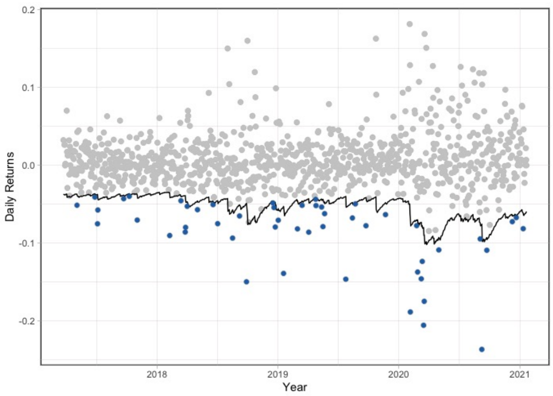
(b) Normal VaR(0.01)



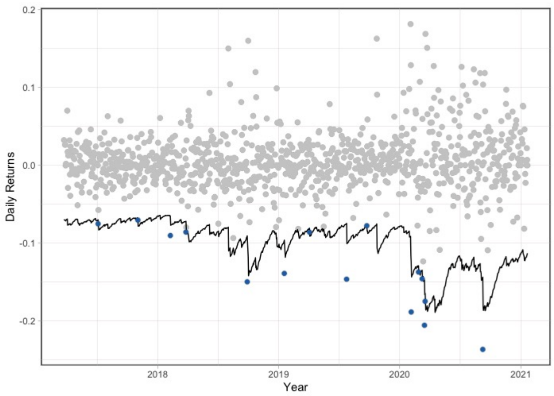
(c) Student-t VaR(0.05)



(d) Student-t VaR(0.01)

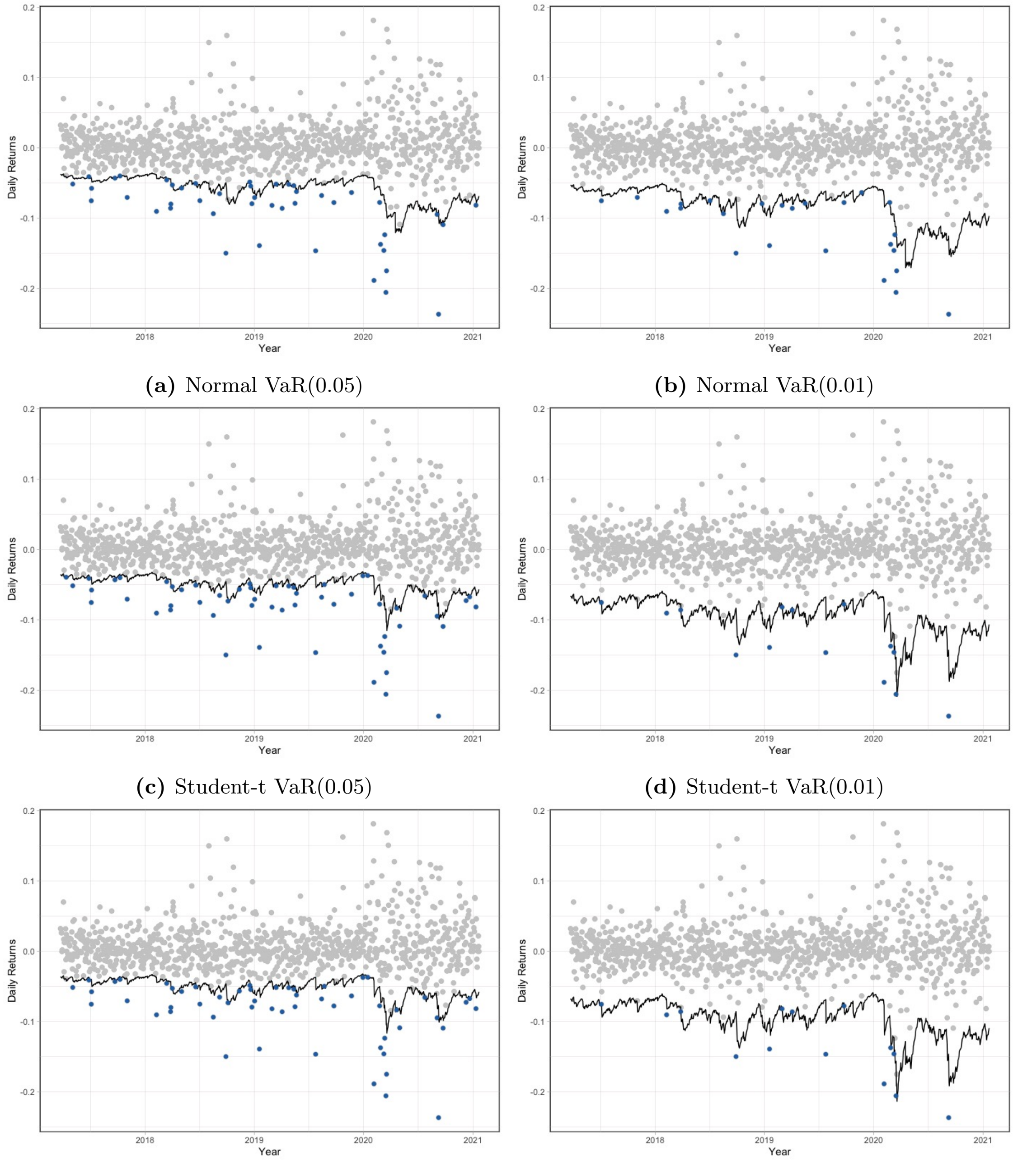


(e) Skew Student-t VaR(0.05)



(f) Skew Student-t VaR(0.01)

Figure D.4: VaR exceedance plots for GARCH(1,1) model on Tesla stock March 2017 to January 2021. The black line indicates the negative VaR estimates, the gray marks the returns and the VaR violations are plotted as blue marks.



(a) Normal VaR(0.05)

(b) Normal VaR(0.01)

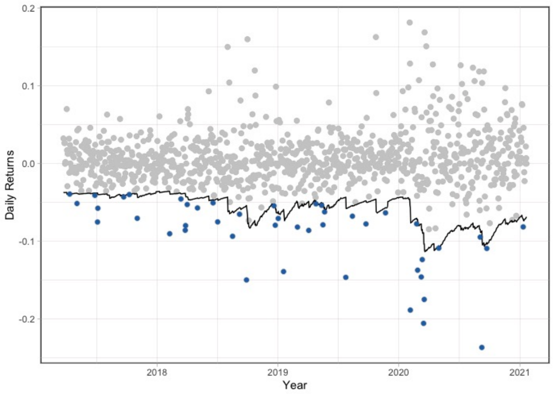
(c) Student-t VaR(0.05)

(d) Student-t VaR(0.01)

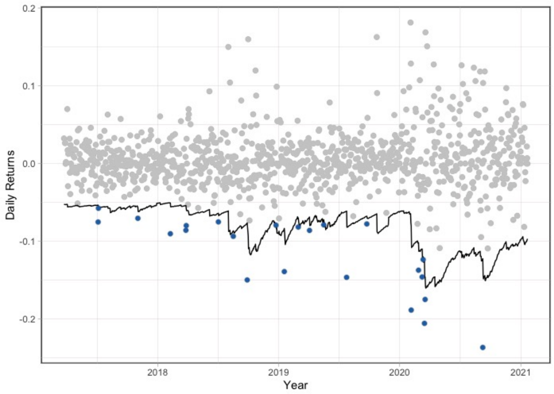
(e) Skew Student-t VaR(0.05)

(f) Skew Student-t VaR(0.01)

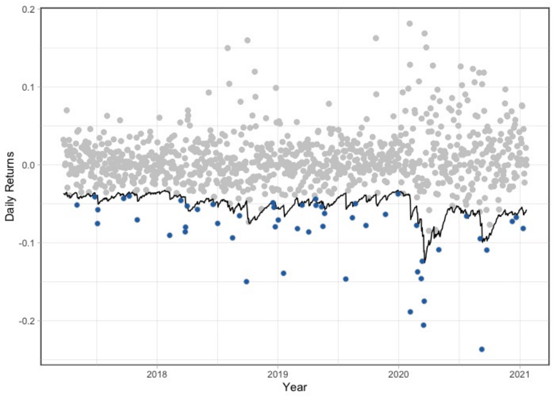
Figure D.5: VaR exceedance plots for EGARCH(1,1) model on Tesla stock March 2017 to January 2021. The black line indicates the negative VaR estimates, the gray marks the returns and the VaR violations are plotted as blue marks.



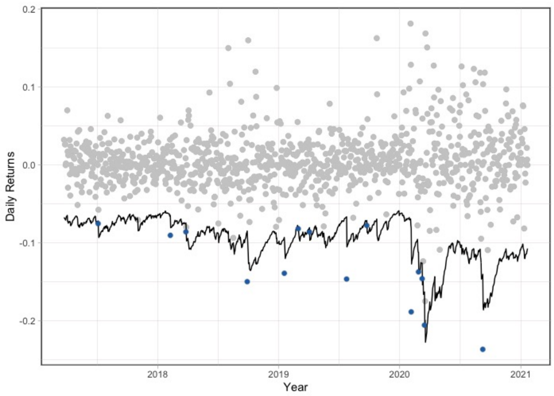
(a) Normal VaR(0.05)



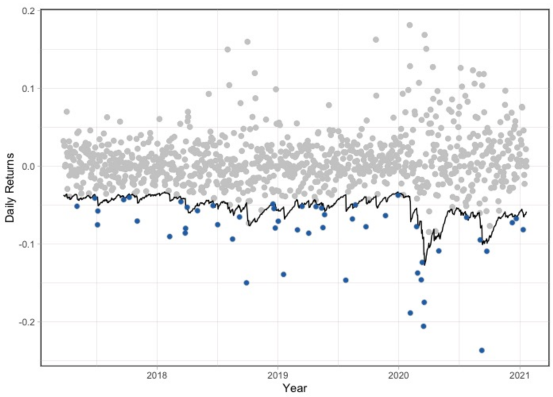
(b) Normal VaR(0.01)



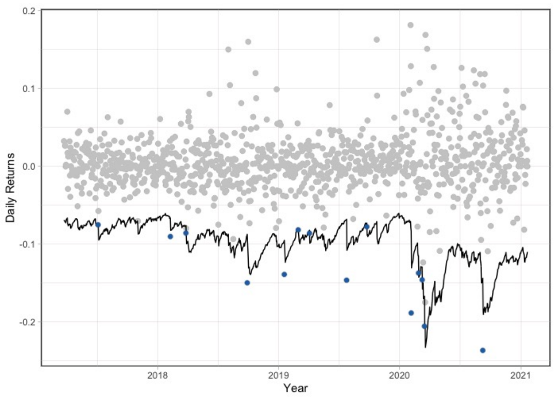
(c) Student-t VaR(0.05)



(d) Student-t VaR(0.01)



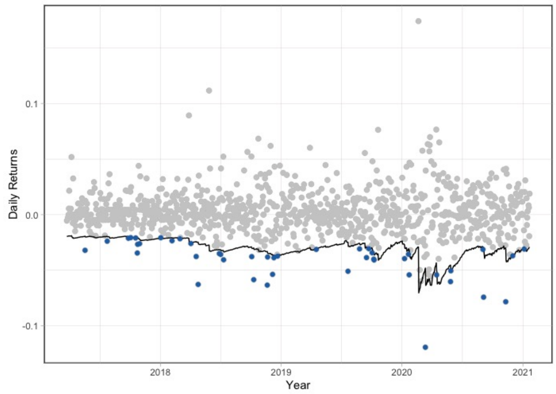
(e) Skew Student-t VaR(0.05)



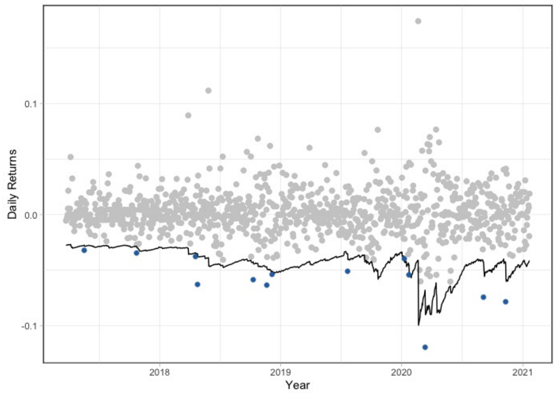
(f) Skew Student-t VaR(0.01)

Figure D.6: VaR exceedance plots for APARCH(1,1) model on Tesla stock March 2017 to January 2021. The black line indicates the negative VaR estimates, the gray marks the returns and the VaR violations are plotted as blue marks.

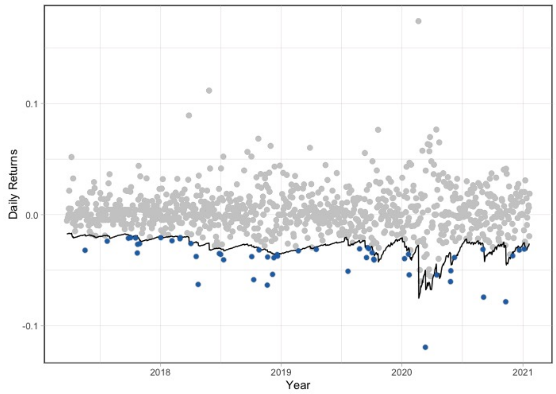
D.3 Tomra



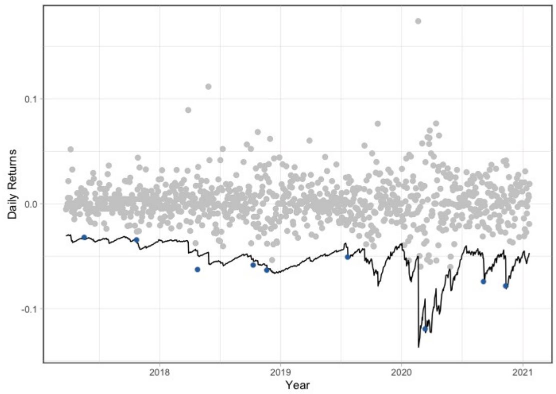
(a) Normal VaR(0.05)



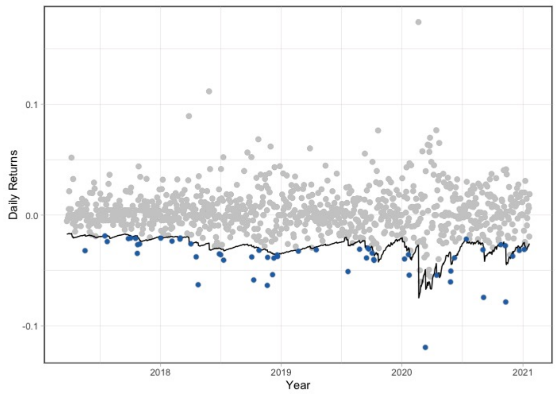
(b) Normal VaR(0.01)



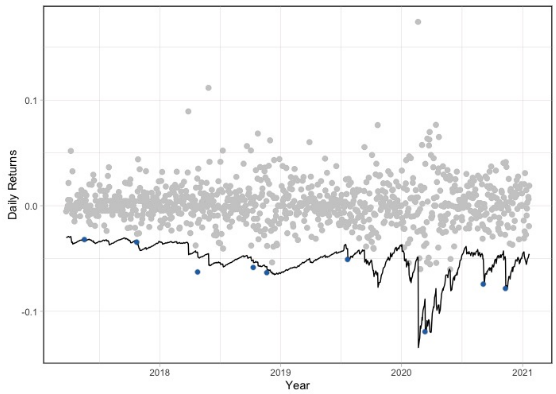
(c) Student-t VaR(0.05)



(d) Student-t VaR(0.01)

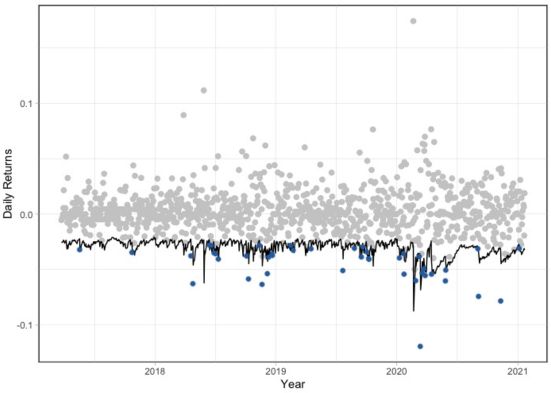


(e) Skew Student-t VaR(0.05)

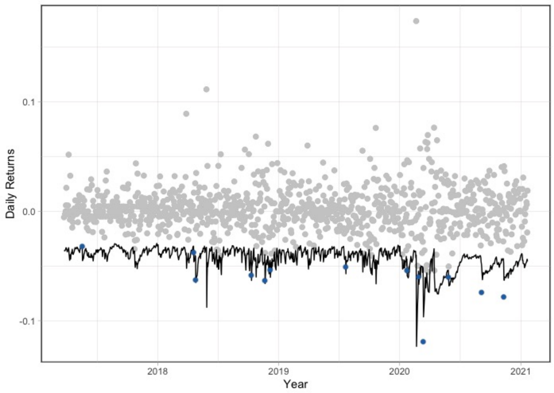


(f) Skew Student-t VaR(0.01)

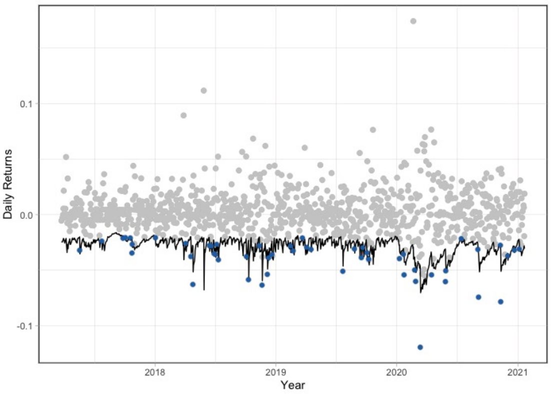
Figure D.7: VaR exceedance plots for GARCH(1,1) model on Tomra stock March 2017 to January 2021. The black line indicates the negative VaR estimates, the gray marks the returns and the VaR violations are plotted as blue marks.



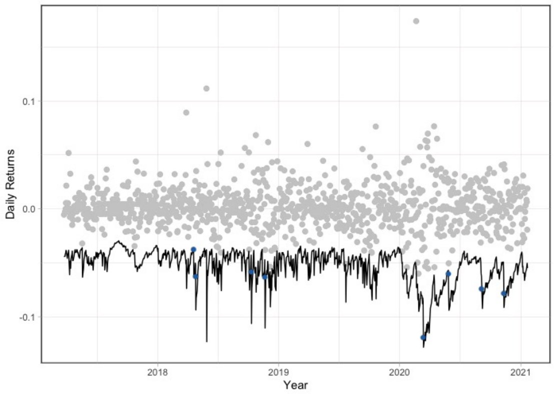
(a) Normal VaR(0.05)



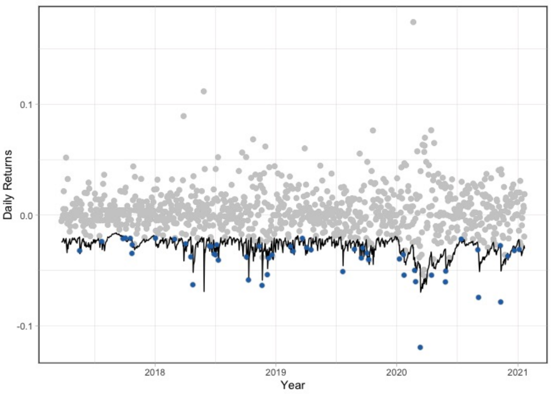
(b) Normal VaR(0.01)



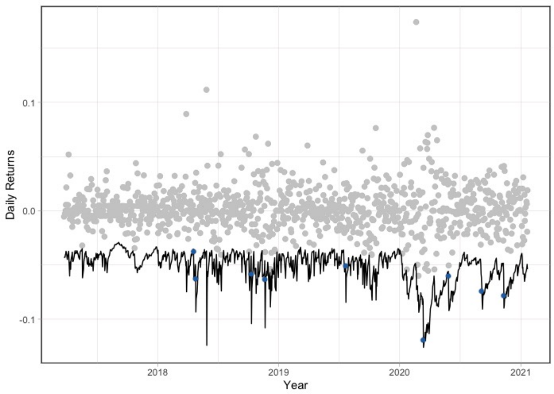
(c) Student-t VaR(0.05)



(d) Student-t VaR(0.01)

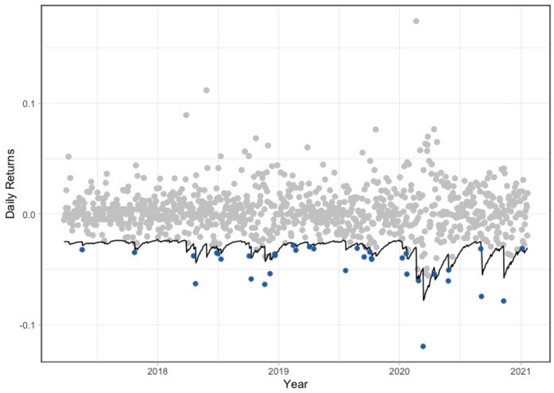


(e) Skew Student-t VaR(0.05)

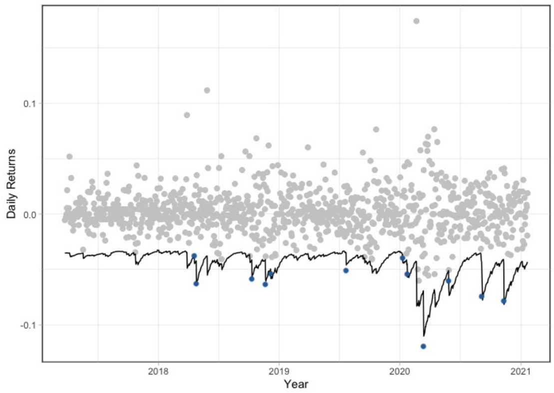


(f) Skew Student-t VaR(0.01)

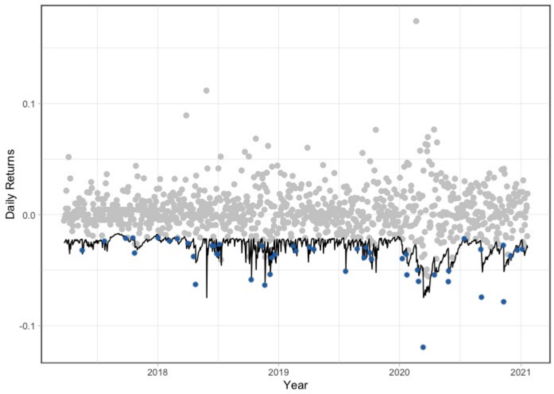
Figure D.8: VaR exceedance plots for EGARCH(1,1) model on Tomra stock March 2017 to January 2021. The black line indicates the negative VaR estimates, the gray marks the returns and the VaR violations are plotted as blue marks.



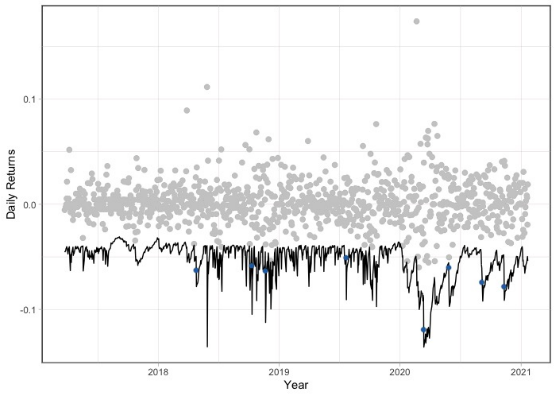
(a) Normal VaR(0.05)



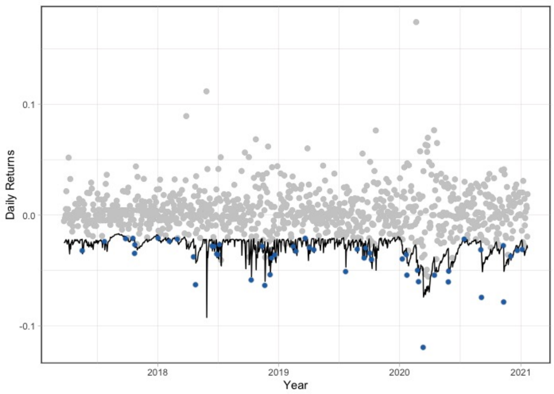
(b) Normal VaR(0.01)



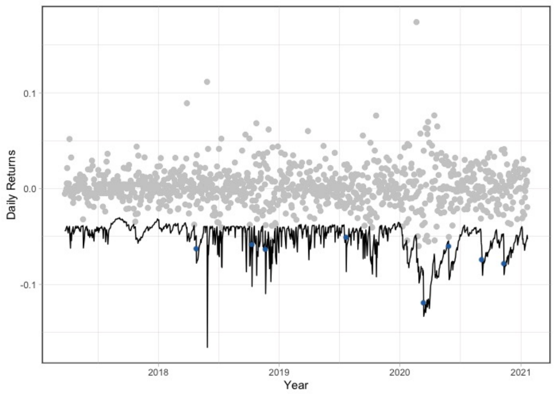
(c) Student-t VaR(0.05)



(d) Student-t VaR(0.01)



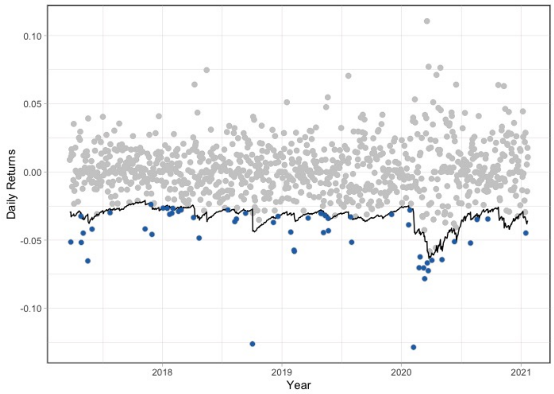
(e) Skew Student-t VaR(0.05)



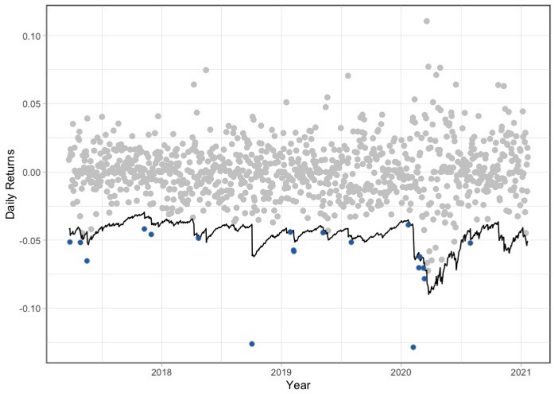
(f) Skew Student-t VaR(0.01)

Figure D.9: VaR exceedance plots for APARCH(1,1) model on Tomra stock March 2017 to January 2021. The black line indicates the negative VaR estimates, the gray marks the returns and the VaR violations are plotted as blue marks.

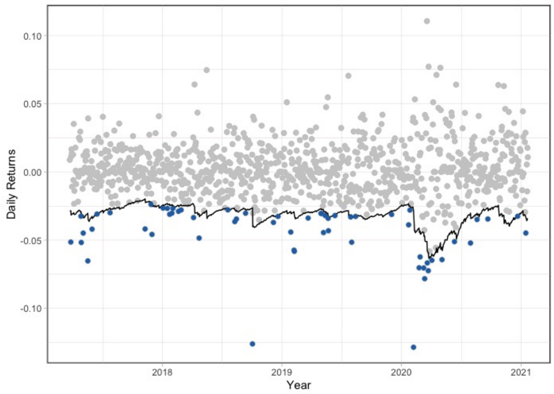
D.4 Norsk Hydro



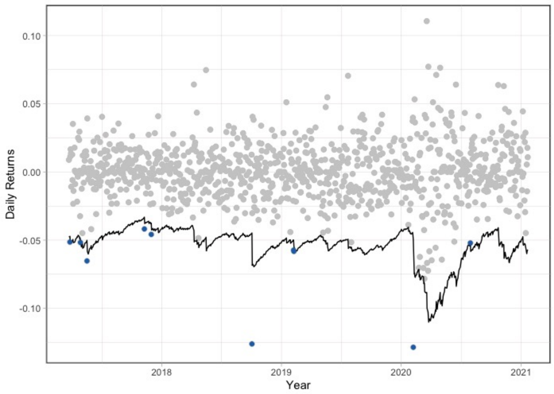
(a) Normal VaR(0.05)



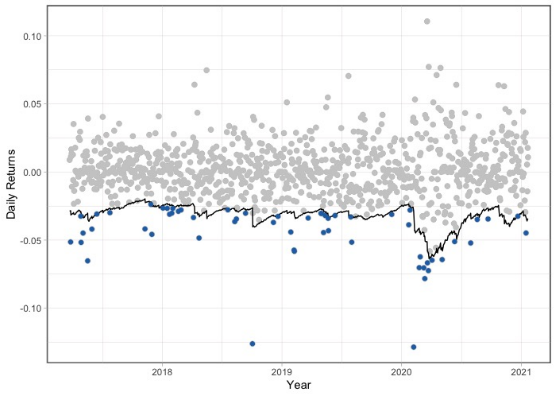
(b) Normal VaR(0.01)



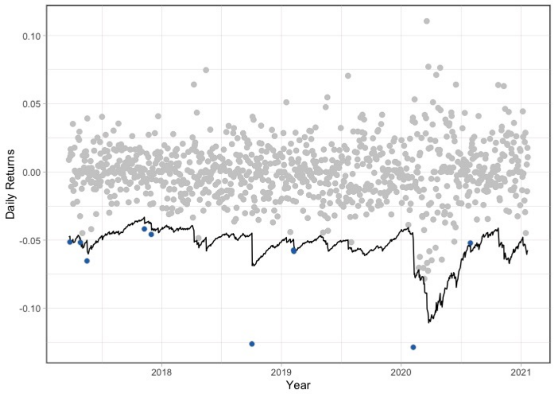
(c) Student-t VaR(0.05)



(d) Student-t VaR(0.01)

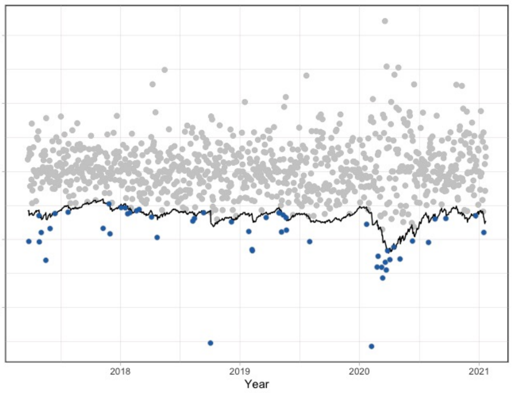


(e) Skew Student-t VaR(0.05)

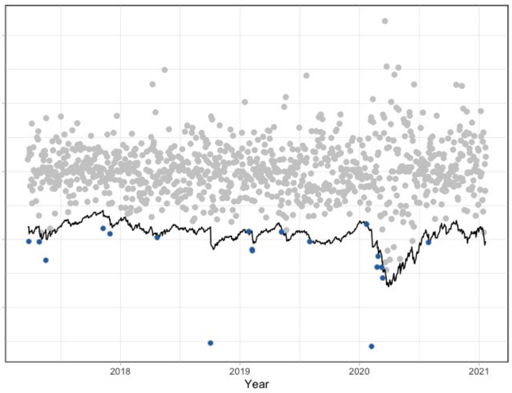


(f) Skew Student-t VaR(0.01)

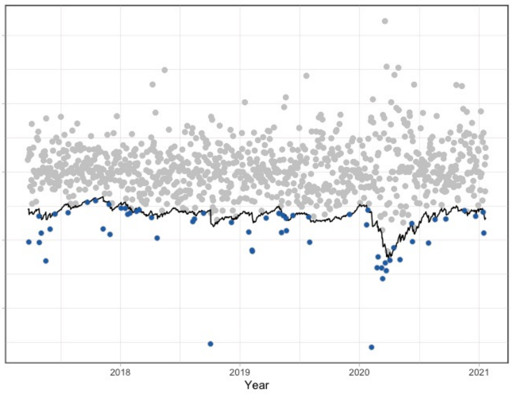
Figure D.10: VaR exceedance plots for GARCH(1,1) model on Norsk Hydro stock March 2017 to January 2021. The black line indicates the negative VaR estimates, the gray marks the returns and the VaR violations are plotted as blue marks.



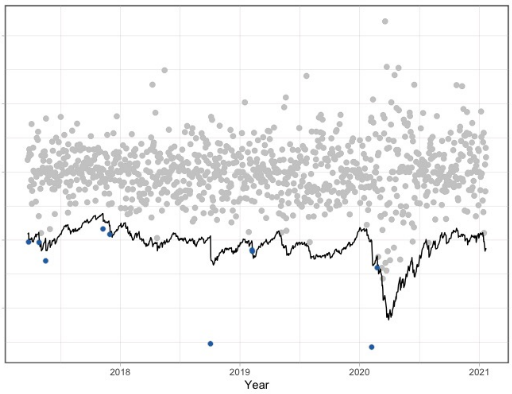
(a) Normal VaR(0.05)



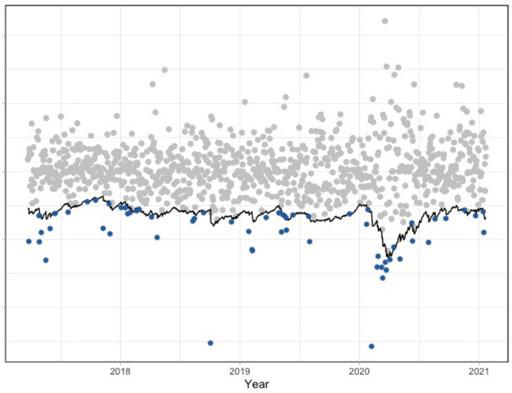
(b) Normal VaR(0.01)



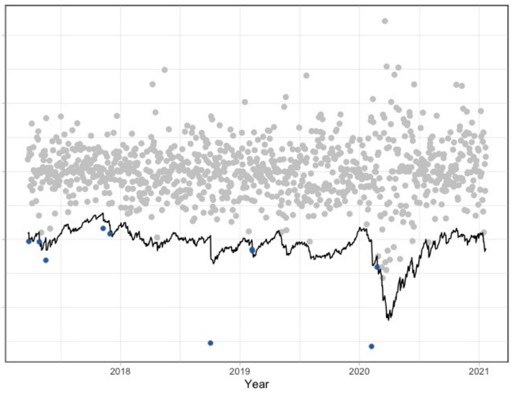
(c) Student-t VaR(0.05)



(d) Student-t VaR(0.01)

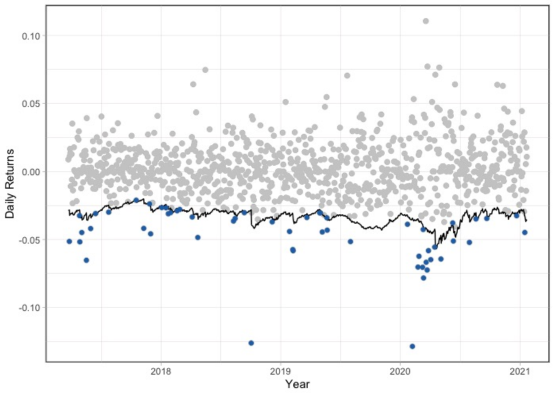


(e) Skew Student-t VaR(0.05)

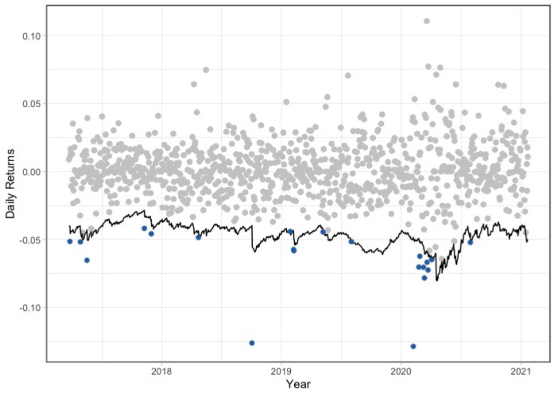


(f) Skew Student-t VaR(0.01)

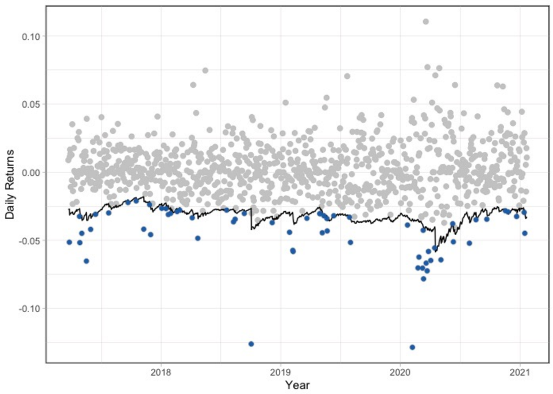
Figure D.11: VaR exceedance plots for EGARCH(1,1) model on Norsk Hydro stock March 2017 to January 2021. The black line indicates the negative VaR estimates, the gray marks the returns and the VaR violations are plotted as blue marks.



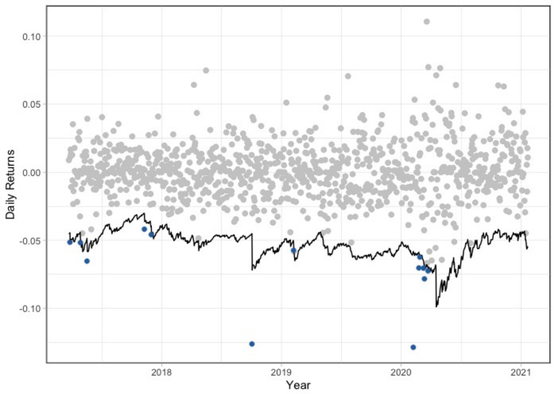
(a) Normal VaR(0.05)



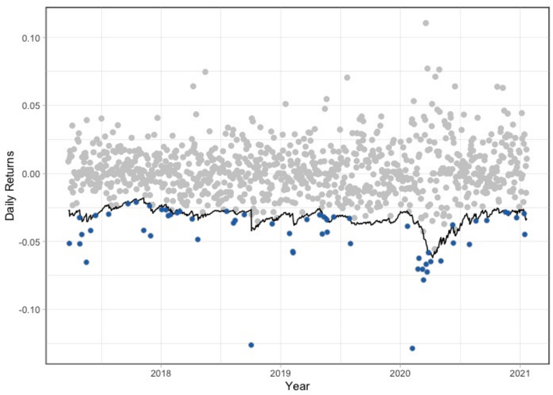
(b) Normal VaR(0.01)



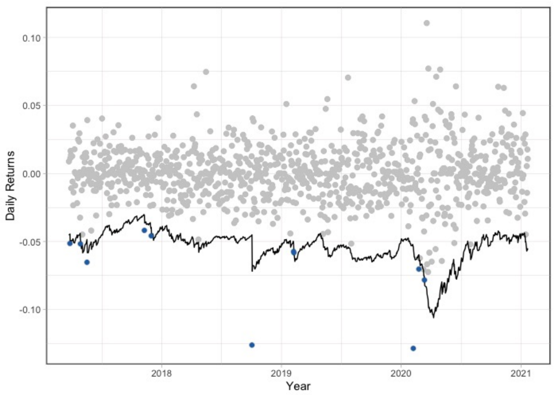
(c) Student-t VaR(0.05)



(d) Student-t VaR(0.01)



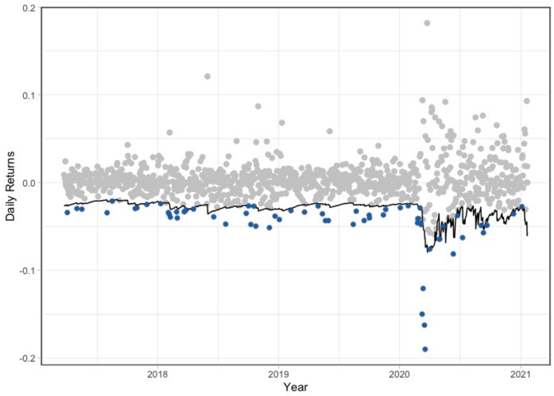
(e) Skew Student-t VaR(0.05)



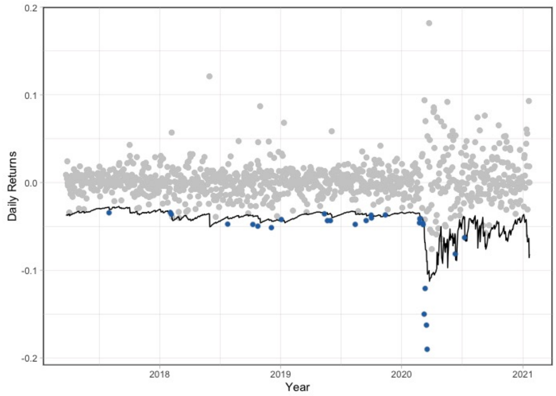
(f) Skew Student-t VaR(0.01)

Figure D.12: VaR exceedance plots for APARCH(1,1) model on Norsk Hydro stock March 2017 to January 2021. The black line indicates the negative VaR estimates, the gray marks the returns and the VaR violations are plotted as blue marks.

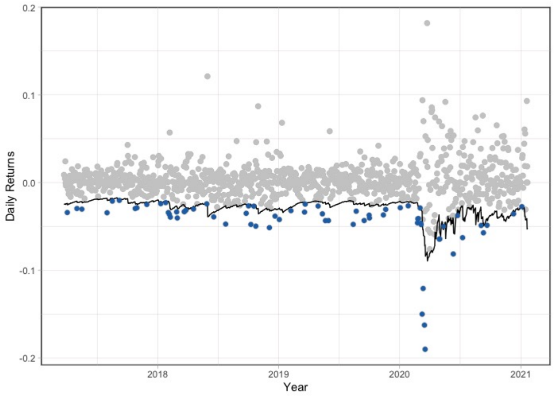
D.5 General Motors



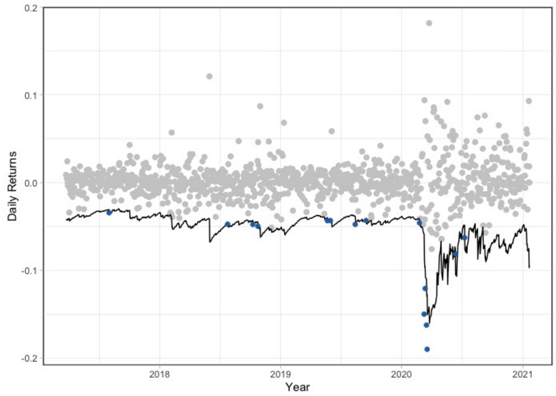
(a) Normal VaR(0.05)



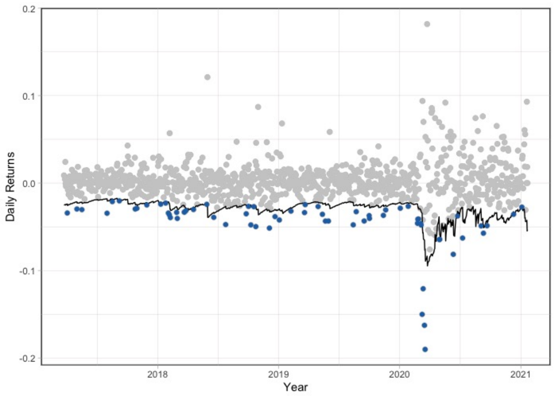
(b) Normal VaR(0.01)



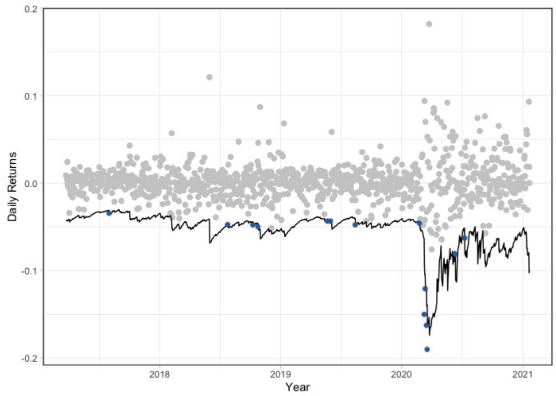
(c) Student-t VaR(0.05)



(d) Student-t VaR(0.01)

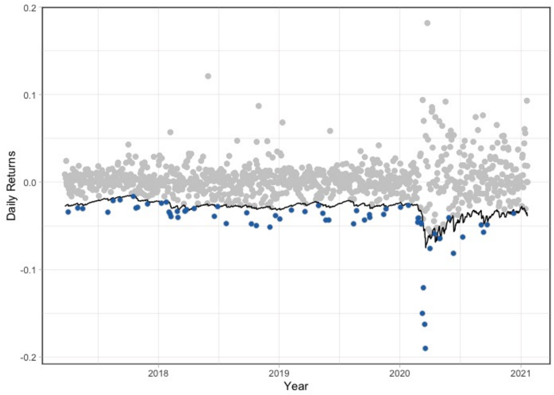


(e) Skew Student-t VaR(0.05)

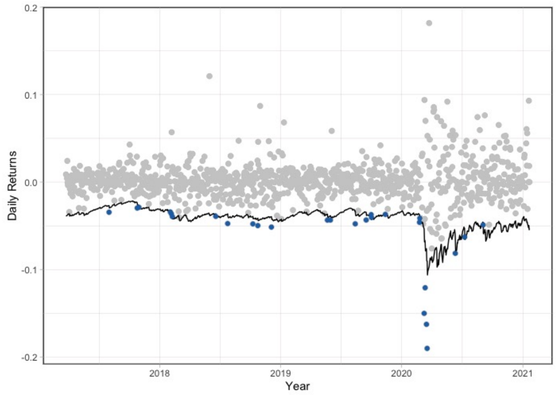


(f) Skew Student-t VaR(0.01)

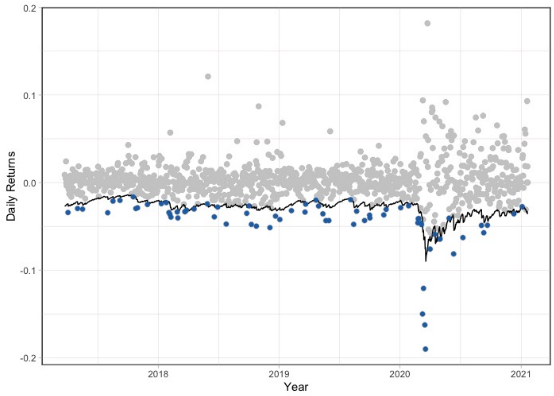
Figure D.13: VaR exceedance plots for GARCH(1,1) model on General Motors stock March 2017 to January 2021. The black line indicates the negative VaR estimates, the gray marks the returns and the VaR violations are plotted as blue marks.



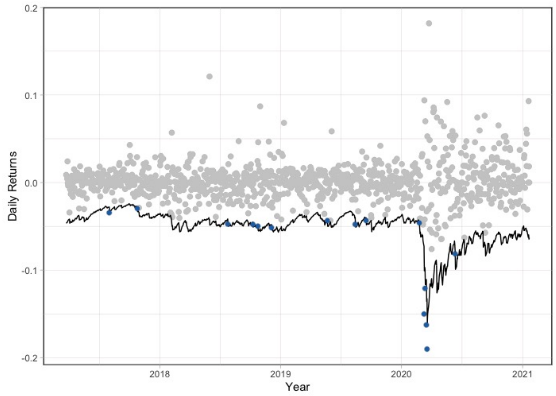
(a) Normal VaR(0.05)



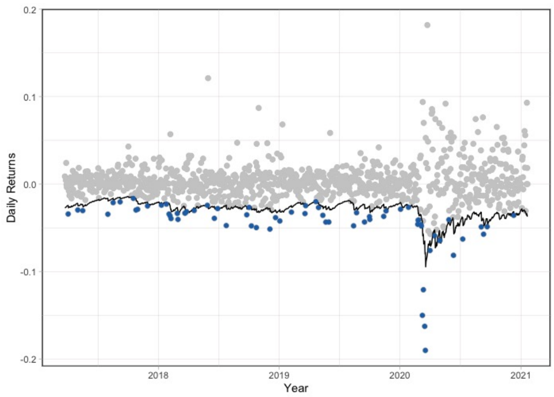
(b) Normal VaR(0.01)



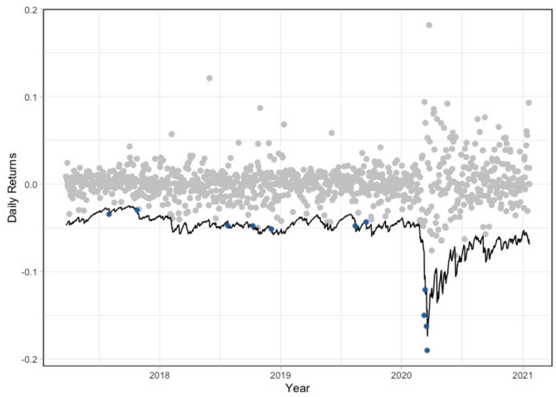
(c) Student-t VaR(0.05)



(d) Student-t VaR(0.01)

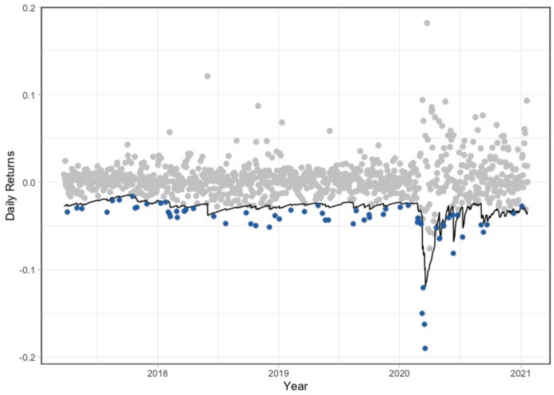


(e) Skew Student-t VaR(0.05)

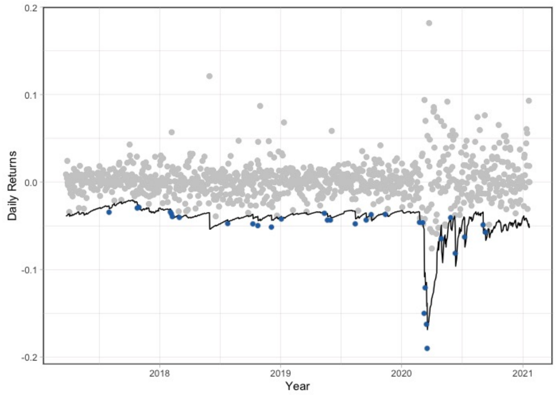


(f) Skew Student-t VaR(0.01)

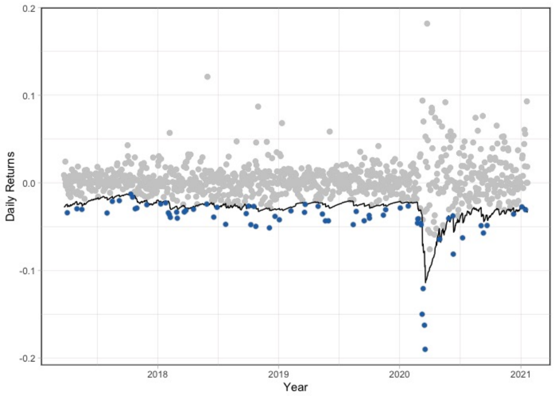
Figure D.14: VaR exceedance plots for EGARCH(1,1) model on General Motors stock March 2017 to January 2021. The black line indicates the negative VaR estimates, the gray marks the returns and the VaR violations are plotted as blue marks.



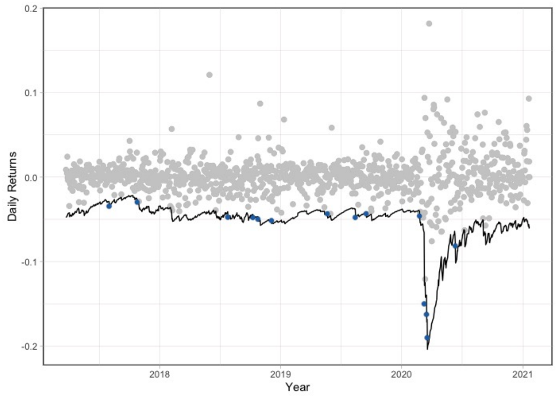
(a) Normal VaR(0.05)



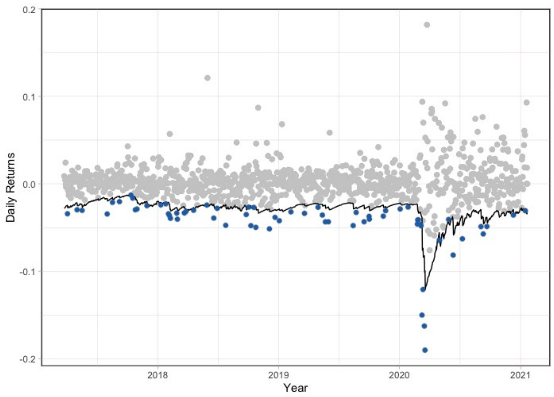
(b) Normal VaR(0.01)



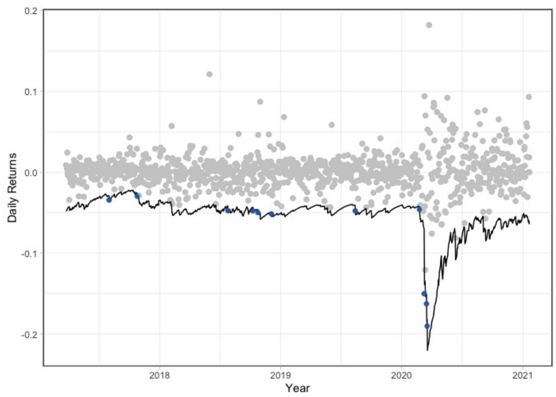
(c) Student-t VaR(0.05)



(d) Student-t VaR(0.01)



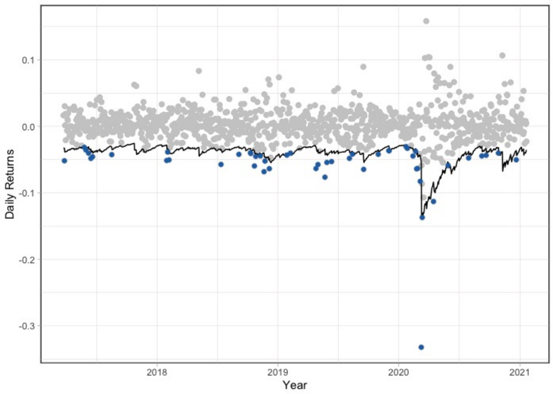
(e) Skew Student-t VaR(0.05)



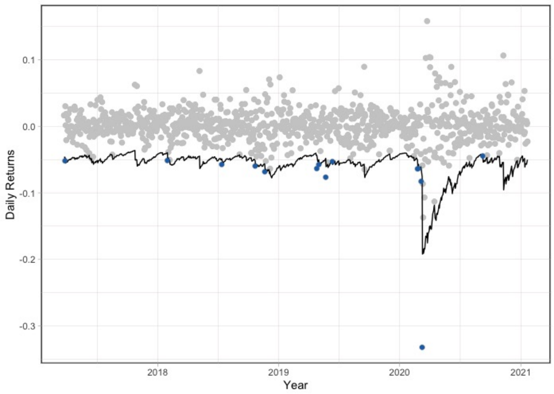
(f) Skew Student-t VaR(0.01)

Figure D.15: VaR exceedance plots for APARCH(1,1) model on General Motors stock March 2017 to January 2021. The black line indicates the negative VaR estimates, the gray marks the returns and the VaR violations are plotted as blue marks.

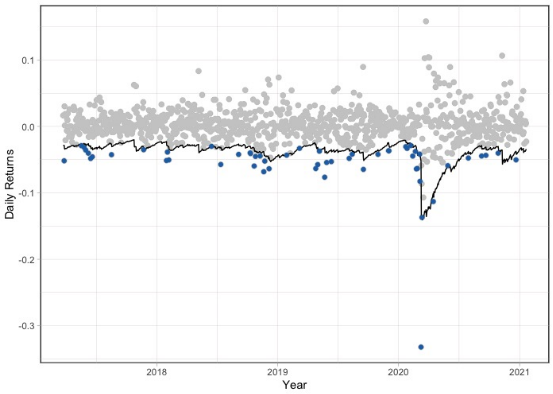
D.6 Aker BP



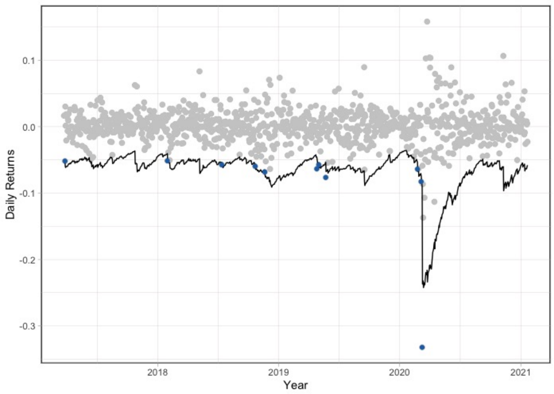
(a) Normal VaR(0.05)



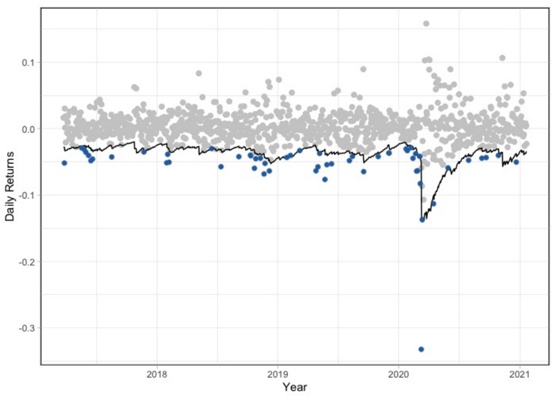
(b) Normal VaR(0.01)



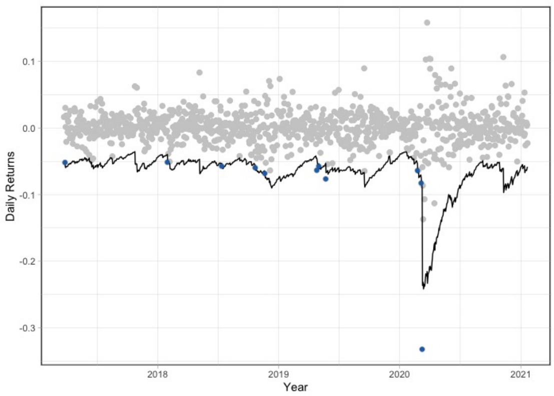
(c) Student-t VaR(0.05)



(d) Student-t VaR(0.01)

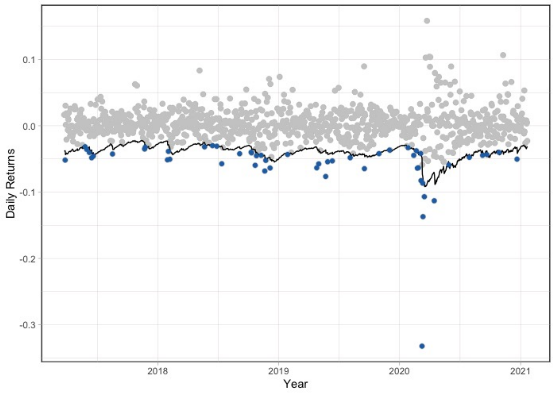


(e) Skew Student-t VaR(0.05)

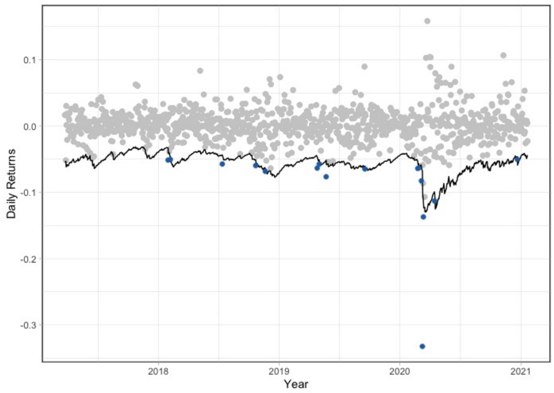


(f) Skew Student-t VaR(0.01)

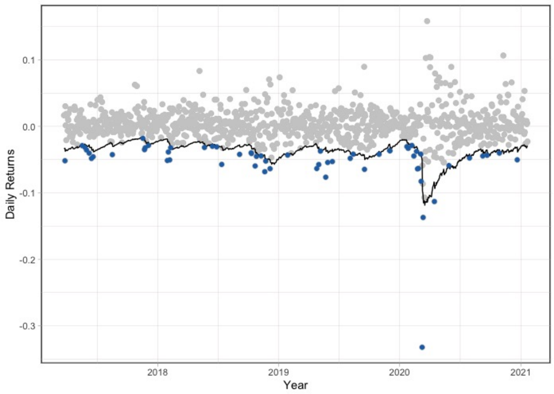
Figure D.16: VaR exceedance plots for GARCH(1,1) model on Aker BP stock March 2017 to January 2021. The black line indicates the negative VaR estimates, the gray marks the returns and the VaR violations are plotted as blue marks.



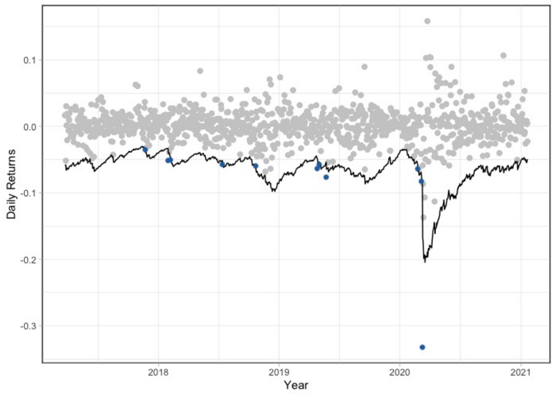
(a) Normal VaR(0.05)



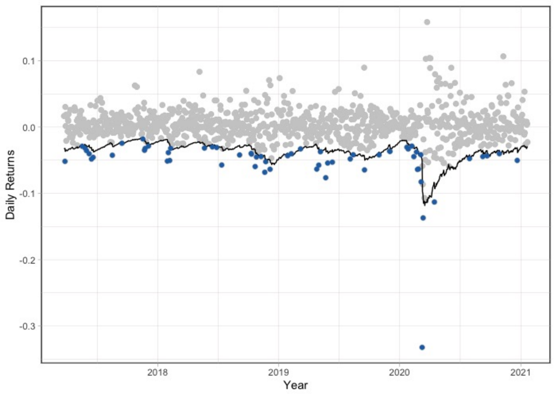
(b) Normal VaR(0.01)



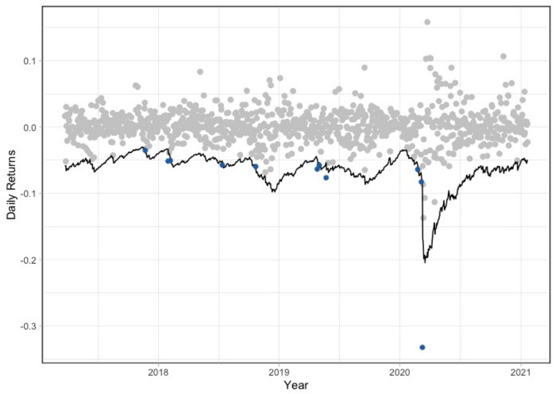
(c) Student-t VaR(0.05)



(d) Student-t VaR(0.01)

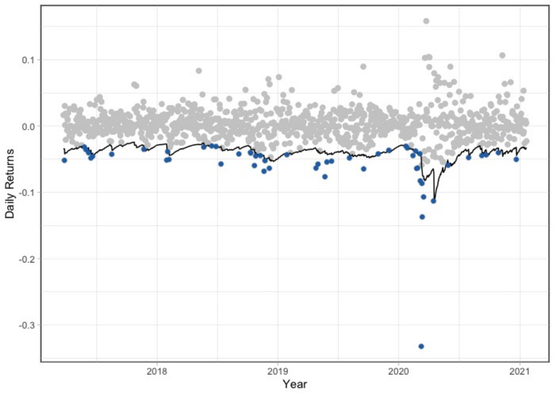


(e) Skew Student-t VaR(0.05)

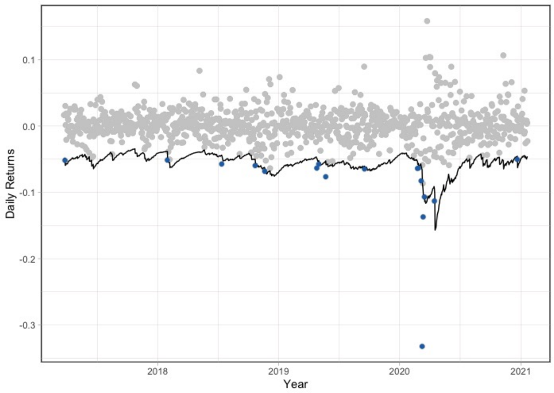


(f) Skew Student-t VaR(0.01)

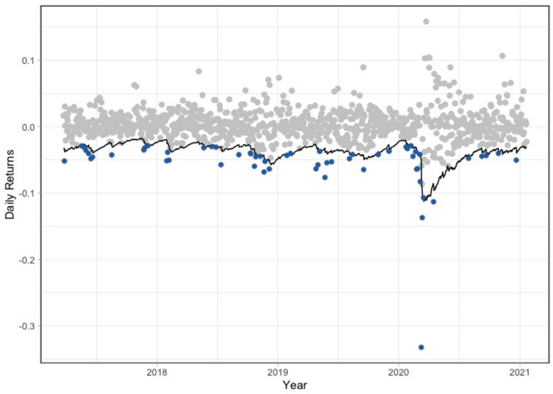
Figure D.17: VaR exceedance plots for EGARCH(1,1) model on Aker BP stock March 2017 to January 2021. The black line indicates the negative VaR estimates, the gray marks the returns and the VaR violations are plotted as blue marks.



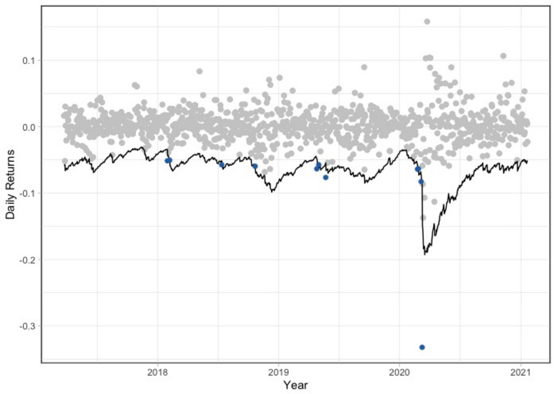
(a) Normal VaR(0.05)



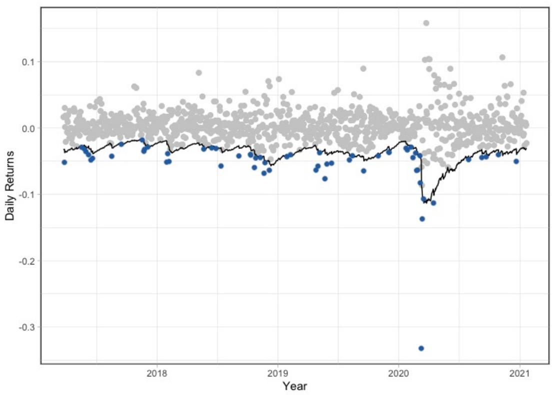
(b) Normal VaR(0.01)



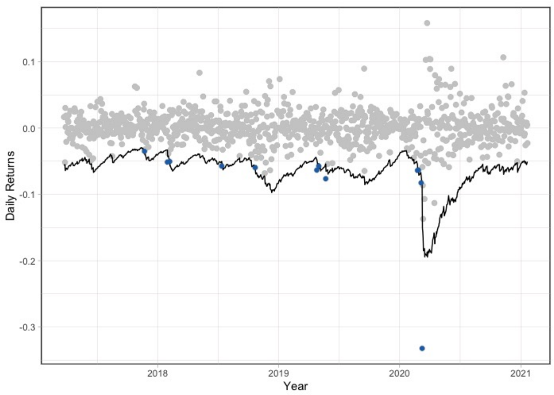
(c) Student-t VaR(0.05)



(d) Student-t VaR(0.01)



(e) Skew Student-t VaR(0.05)



(f) Skew Student-t VaR(0.01)

Figure D.18: VaR exceedance plots for APARCH(1,1) model on Aker BP stock March 2017 to January 2021. The black line indicates the negative VaR estimates, the gray marks the returns and the VaR violations are plotted as blue marks.

Appendix E

Additional Calibration Method Results

E.1 NEL

Value-at-Risk Backtests

Table E.1: 1% and 5% VaR exceedance for NEL returns. The models are estimated with data from 01.01.2006 to 31.12.2010 and backtested with data from 23.04.2017 to 21.01.2021.

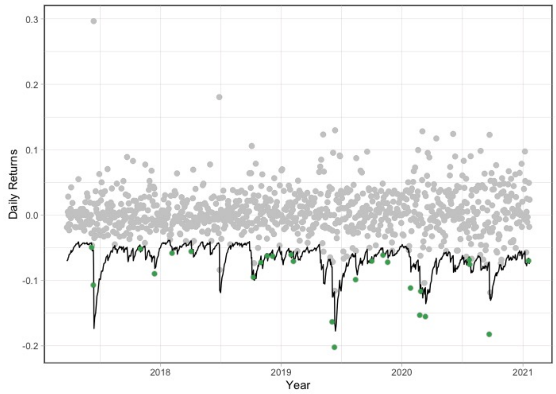
Model	HS	G_n	G_t	G_{st}	E_n	E_t	E_{st}	A_n	A_t	A_{st}
Panel A: Results at percentile level α=0.05, Expected Violations 50										
Violations	67	26	48	50	34	45	51	38	50	54
VR	1.34	0.52	0.96	1.00	0.68	0.90	1.02	0.76	1.00	1.08
UC	0.133	0.000	0.770	1.000	0.014	0.461	0.885	0.070	1.000	0.566
IND	0.300	0.705	0.313	0.730	0.122	0.039	0.394	0.083	0.730	0.524
CC	0.189	0.001	0.576	**0.942**	0.015	0.091	0.688	0.043	**0.942**	0.208
Panel B: Results at percentile level α=0.01, Expected Violations 10										
Violations	13	7	5	8	12	4	5	14	3	5
VR	1.30	0.70	0.50	0.80	1.20	0.40	0.50	1.40	0.30	0.50
UC	0.510	0.314	0.079	0.510	0.538	0.030	0.079	0.231	0.009	0.079
IND	0.719	0.753	0.823	0.719	0.589	0.858	0.823	0.528	0.893	0.823
CC	**0.755**	0.573	0.208	**0.755**	0.715	0.094	0.208	0.399	0.033	0.208

Expected Shortfall Backtest

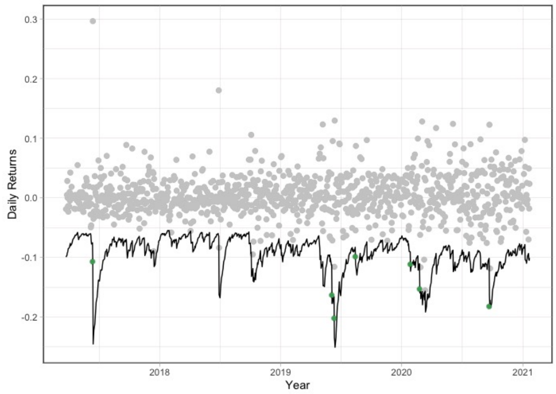
Table E.2: 1% and 5% ES exceedance for NEL returns. The models are estimated with data from 01.01.2006 to 31.12.2010 and backtested with data from 23.04.2017 to 21.01.2021.

Model	HS	G_n	G_t	G_{st}	E_n	E_t	E_{st}	A_n	A_t	A_{st}
Panel A: Results at percentile level α=0.05										
ESR	0.996	0.821	0.997	0.962	0.572	**1.000**	0.987	0.394	0.999	0.976
Panel B: Results at percentile level α=0.01										
ESR	0.980	0.122	0.994	0.972	0.040	0.997	0.993	0.016	**1.000**	**1.000**

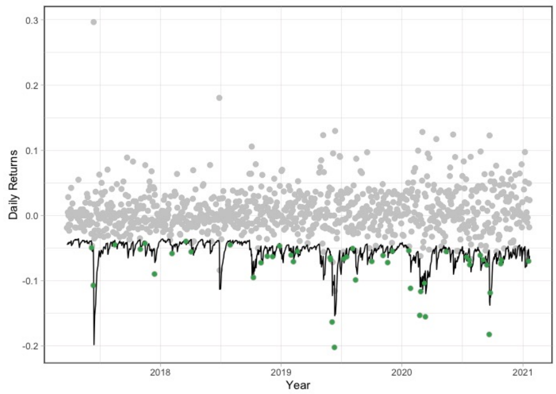
E.1.1 VaR Exceedance



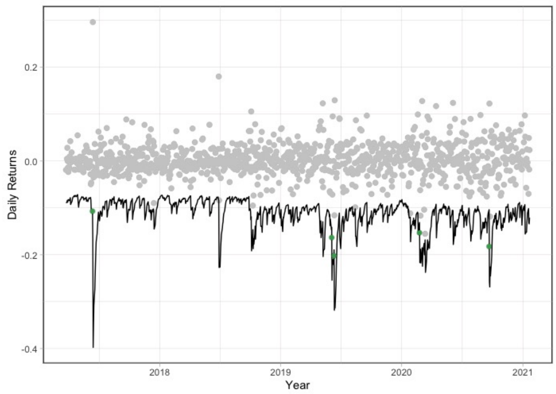
(a) Normal VaR(0.05)



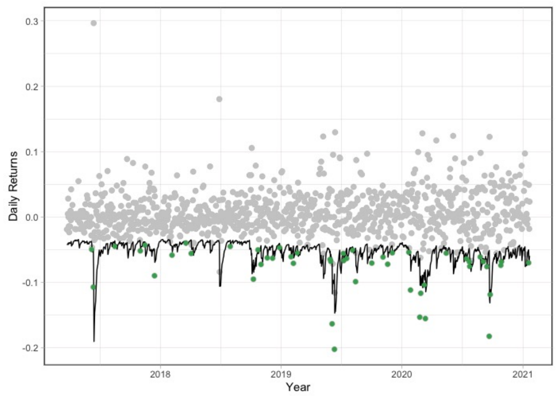
(b) Normal VaR(0.01)



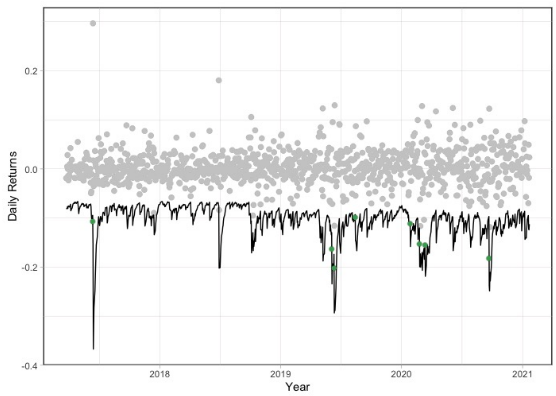
(c) Student-t VaR(0.05)



(d) Student-t VaR(0.01)

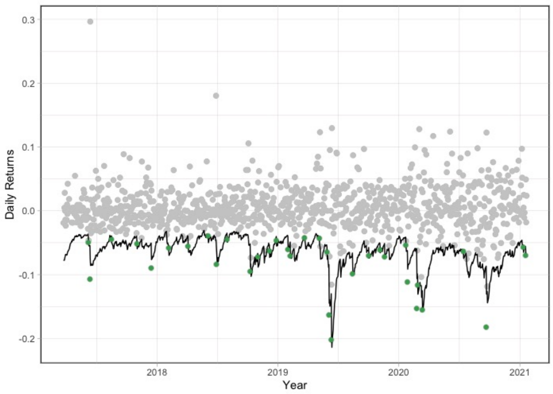


(e) Skew Student-t VaR(0.05)

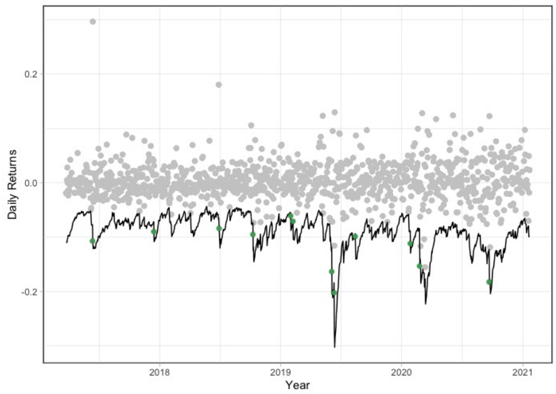


(f) Skew Student-t VaR(0.01)

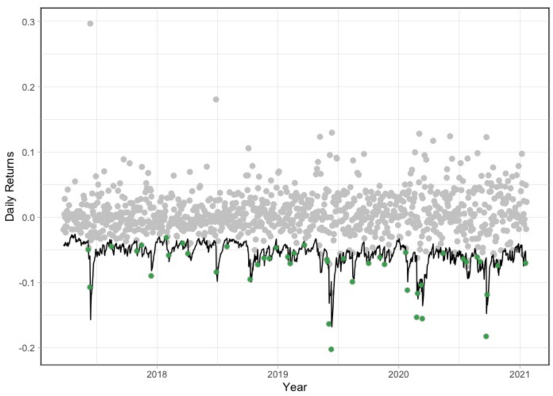
Figure E.1: VaR exceedance plots for GARCH(1,1) model on NEL stock March 2017 to January 2021. The models are estimated on the data from 01.01.2006 to 31.12.2010. The black line indicates the negative VaR estimates, the gray marks the returns and the VaR violations are plotted as green marks.



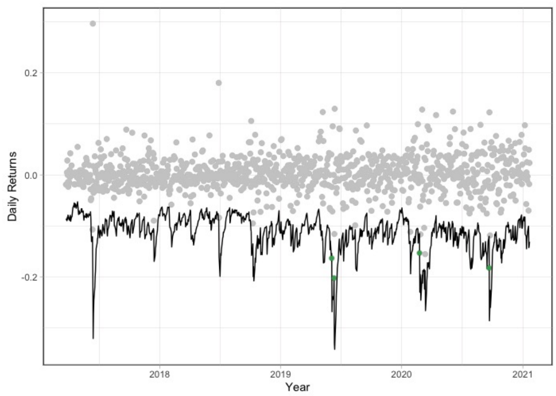
(a) Normal VaR(0.05)



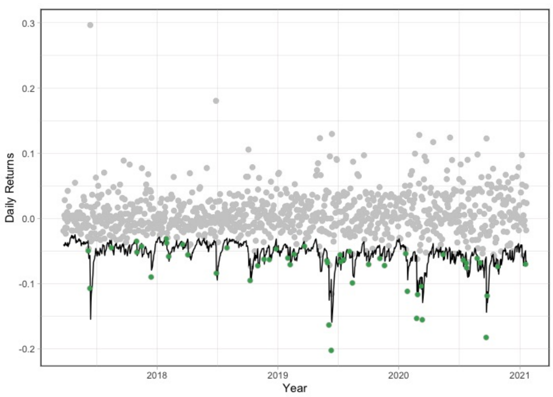
(b) Normal VaR(0.01)



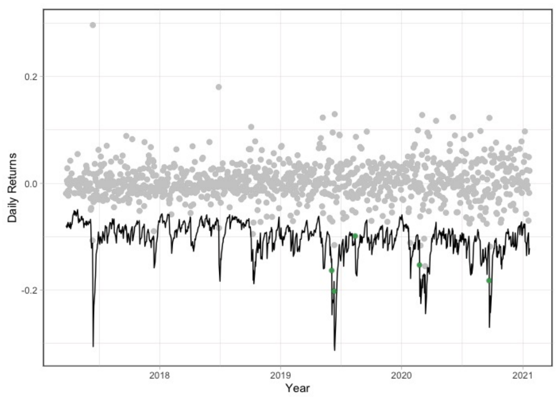
(c) Student-t VaR(0.05)



(d) Student-t VaR(0.01)

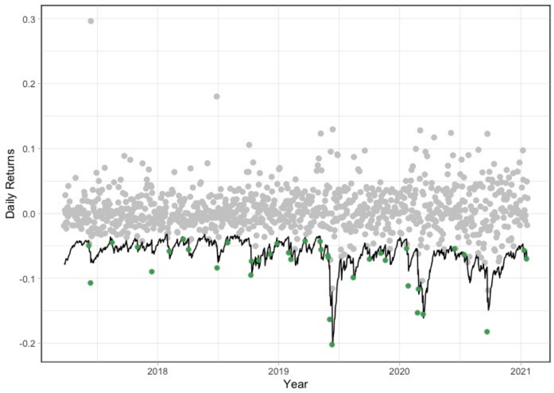


(e) Skew Student-t VaR(0.05)

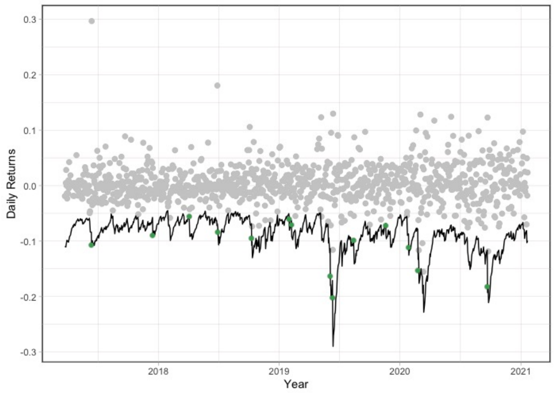


(f) Skew Student-t VaR(0.01)

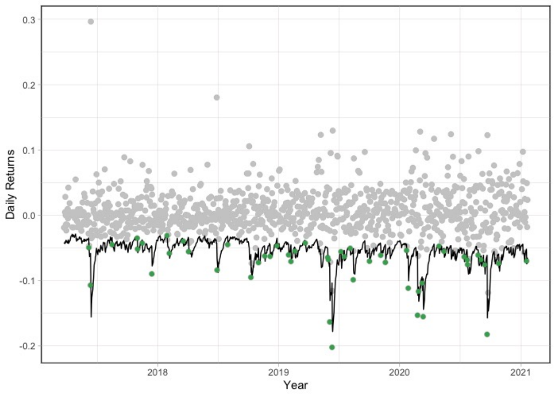
Figure E.2: VaR exceedance plots for EGARCH(1,1) model on NEL stock March 2017 to January 2021. The models are estimated on the data from 01.01.2006 to 31.12.2010. The black line indicates the negative VaR estimates, the gray marks the returns and the VaR violations are plotted as green marks.



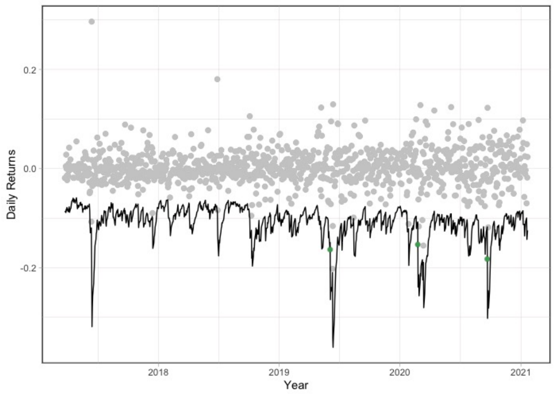
(a) Normal VaR(0.05)



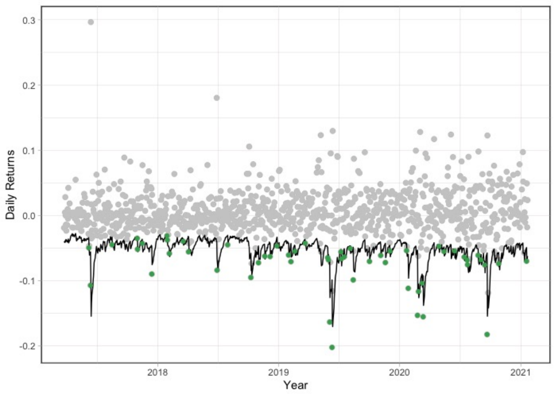
(b) Normal VaR(0.01)



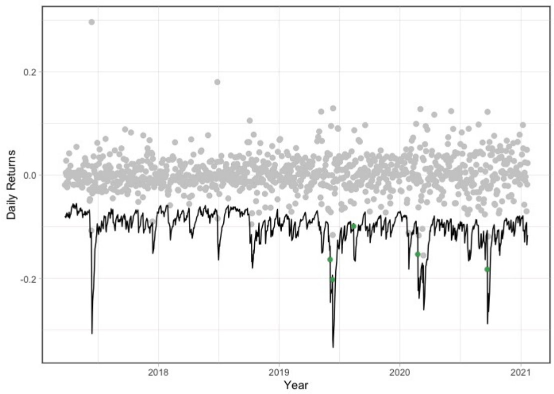
(c) Student-t VaR(0.05)



(d) Student-t VaR(0.01)



(e) Skew Student-t VaR(0.05)



(f) Skew Student-t VaR(0.01)

Figure E.3: VaR exceedance plots for APARCH(1,1) model on NEL stock March 2017 to January 2021. The models are estimated on the data from 01.01.2006 to 31.12.2010. The black line indicates the negative VaR estimates, the gray marks the returns and the VaR violations are plotted as green marks.

E.2 Norsk Hydro

Value-at-Risk Backtests

Table E.3: 1% and 5% VaR exceedance for Norsk Hydro returns. The models are estimated with data from 01.01.2006 to 31.12.2010 and backtested with data from 23.04.2017 to 21.01.2021.

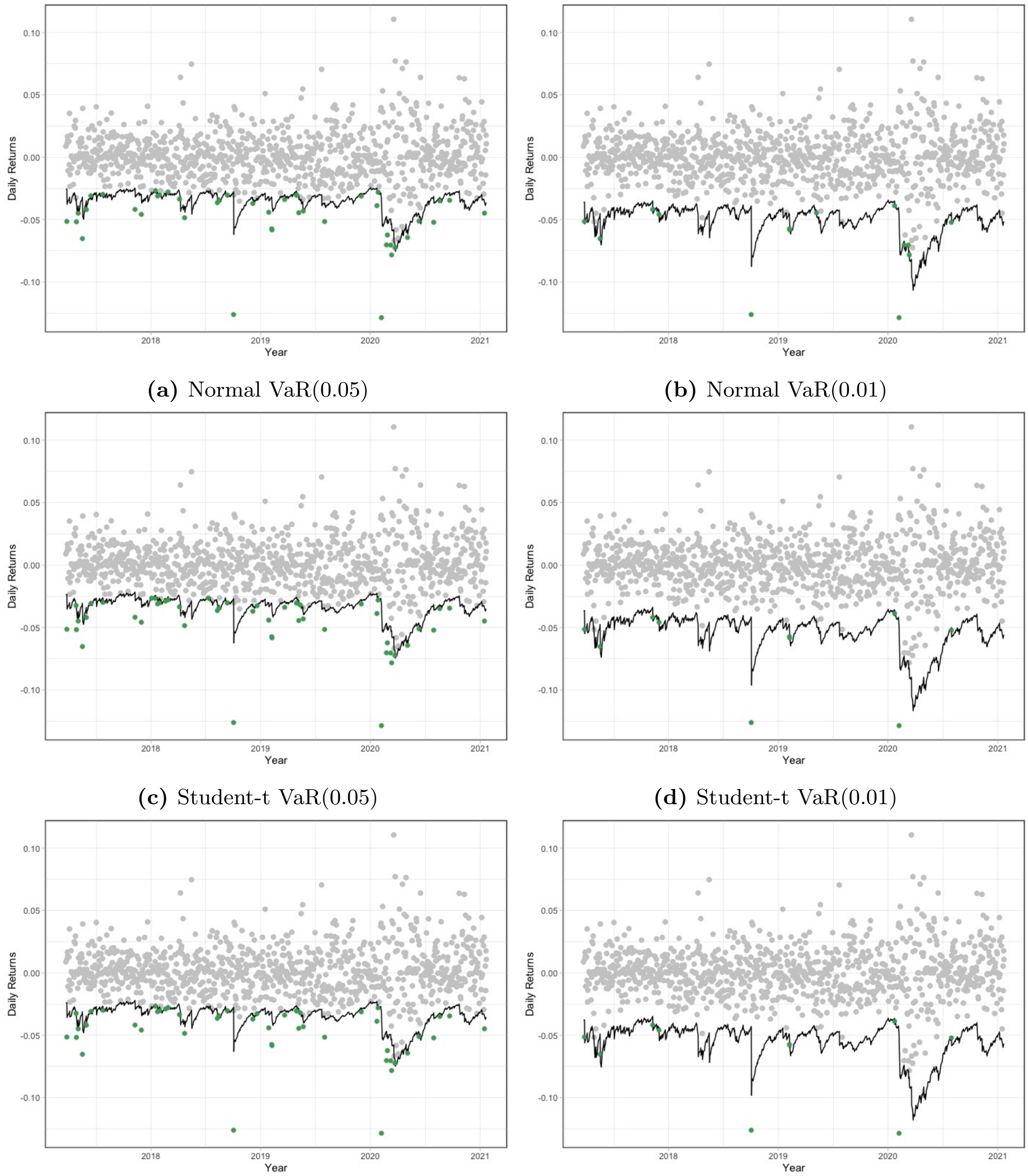
Model	HS	G_n	G_t	G_{st}	E_n	E_t	E_{st}	A_n	A_t	A_{st}
Panel A: Results at percentile level α=0.05, Expected Violations 50										
Violations	41	44	50	48	43	50	46	42	51	47
VR	0.82	0.88	1.00	0.96	0.86	1.00	0.92	0.84	1.02	0.94
UC	0.048	0.375	1.000	0.770	0.299	1.000	0.557	0.233	0.885	0.660
IND	0.011	0.440	0.730	0.829	0.415	0.747	0.548	0.374	0.800	0.880
CC	0.006	0.501	0.942	0.936	0.418	0.949	0.702	0.331	**0.958**	0.898
Panel B: Results at percentile level α=0.01, Expected Violations 10										
Violations	13	14	10	9	15	11	10	14	10	10
VR	1.30	1.40	1.00	0.90	1.50	1.10	1.00	1.40	1.00	1.00
UC	0.362	0.231	1.000	0.746	0.139	0.754	1.000	0.231	1.000	1.000
IND	0.009	0.186	0.085	0.686	0.218	0.621	0.653	0.186	0.653	0.653
CC	0.022	0.204	0.226	0.875	0.157	0.842	**0.904**	0.204	**0.904**	**0.904**

Expected Shortfall Backtest

Table E.4: 1% and 5% ES exceedance for Norsk Hydro returns. The models are estimated with data from 01.01.2006 to 31.12.2010 and backtested with data from 23.04.2017 to 21.01.2021.

Model	HS	G_n	G_t	G_{st}	E_n	E_t	E_{st}	A_n	A_t	A_{st}
Panel A: Results at percentile level α=0.05										
ESR	**0.998**	0.154	0.344	0.462	0.082	0.307	0.485	0.165	0.354	0.491
Panel B: Results at percentile level α=0.01										
ESR	**0.964**	0.018	0.139	0.217	0.012	0.145	0.236	0.021	0.184	0.259

E.2.1 VaR Exceedance



(a) Normal VaR(0.05)

(b) Normal VaR(0.01)

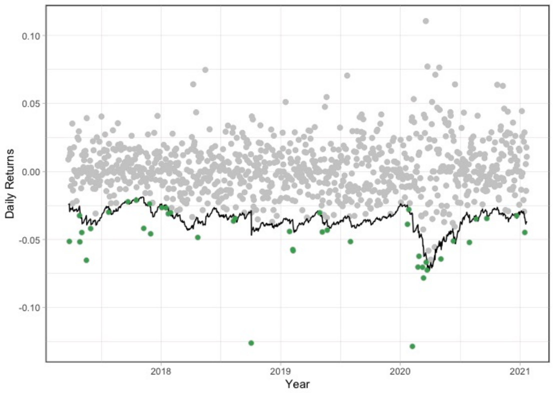
(c) Student-t VaR(0.05)

(d) Student-t VaR(0.01)

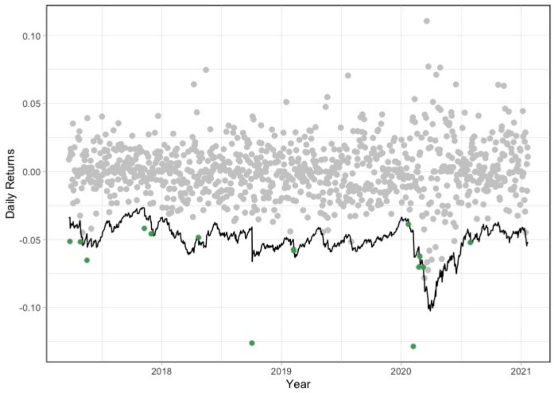
(e) Skew Student-t VaR(0.05)

(f) Skew Student-t VaR(0.01)

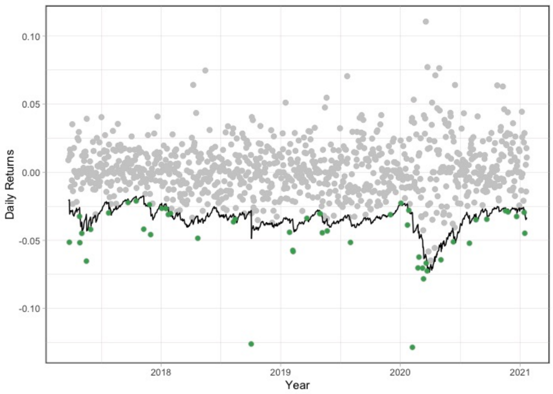
Figure E.4: VaR exceedance plots for GARCH(1,1) model on Norsk Hydro stock March 2017 to January 2021. The models are estimated on the data from 01.01.2006 to 31.12.2010. The black line indicates the negative VaR estimates, the gray marks the returns and the VaR violations are plotted as green marks.



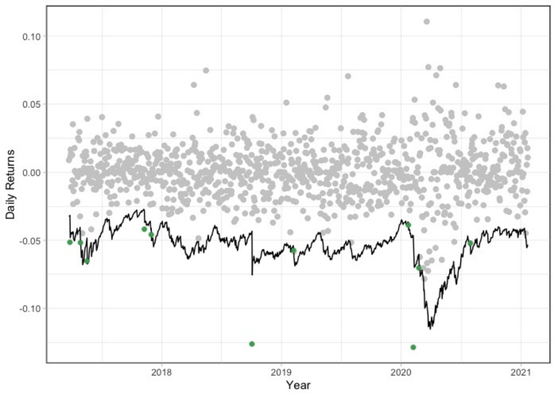
(a) Normal VaR(0.05)



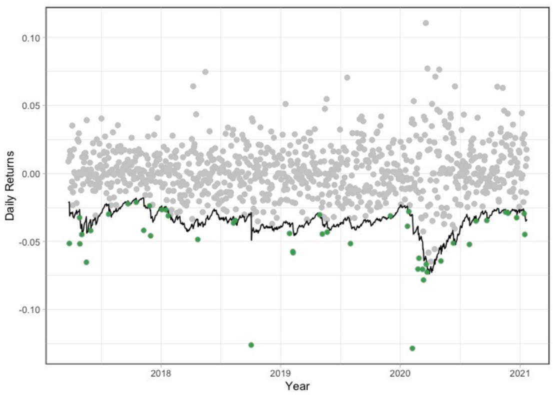
(b) Normal VaR(0.01)



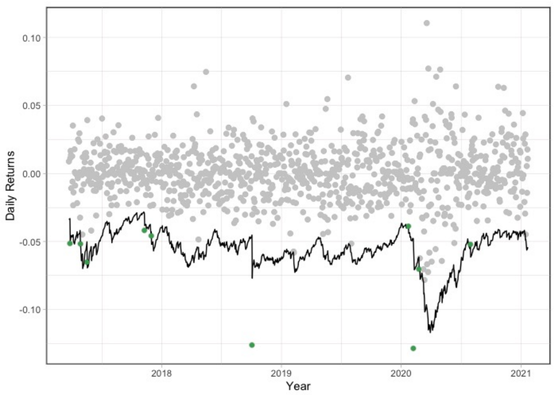
(c) Student-t VaR(0.05)



(d) Student-t VaR(0.01)

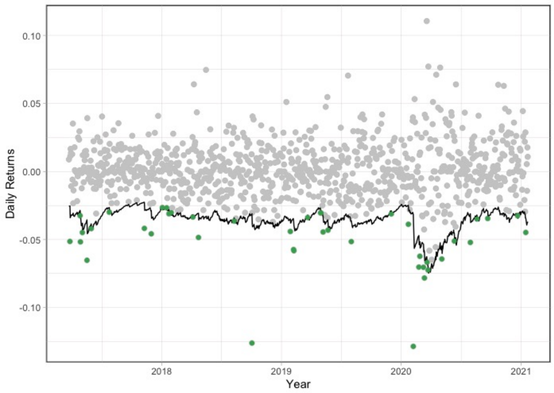


(e) Skew Student-t VaR(0.05)

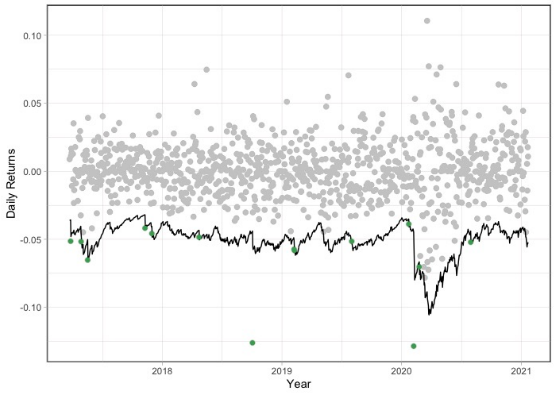


(f) Skew Student-t VaR(0.01)

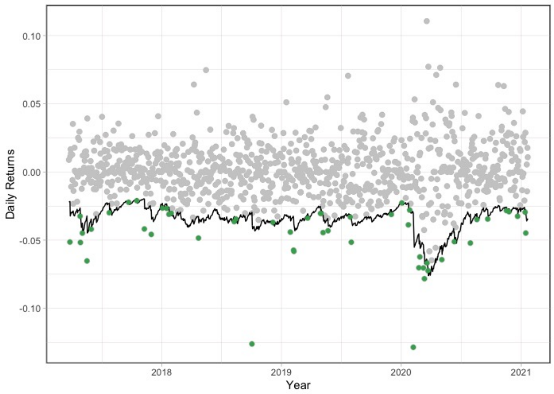
Figure E.5: VaR exceedance plots for EGARCH(1,1) model on Norsk Hydro stock March 2017 to January 2021. The models are estimated on the data from 01.01.2006 to 31.12.2010. The black line indicates the negative VaR estimates, the gray marks the returns and the VaR violations are plotted as green marks.



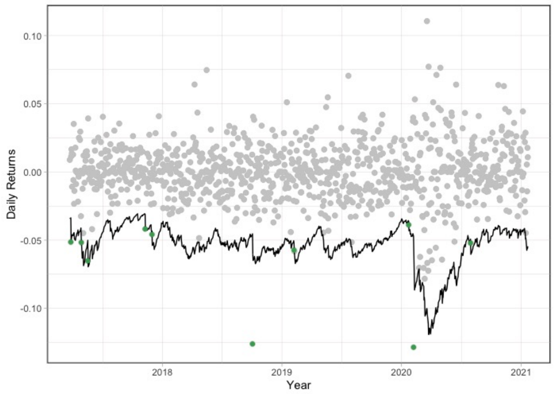
(a) Normal VaR(0.05)



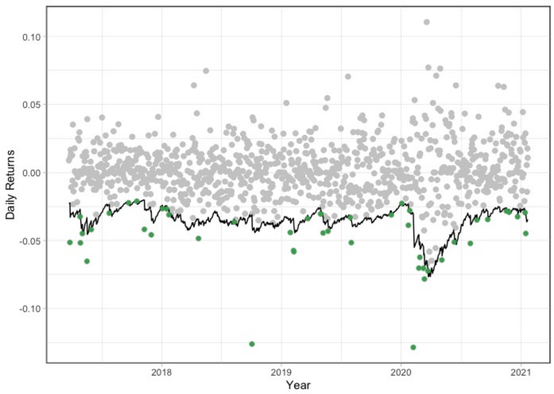
(b) Normal VaR(0.01)



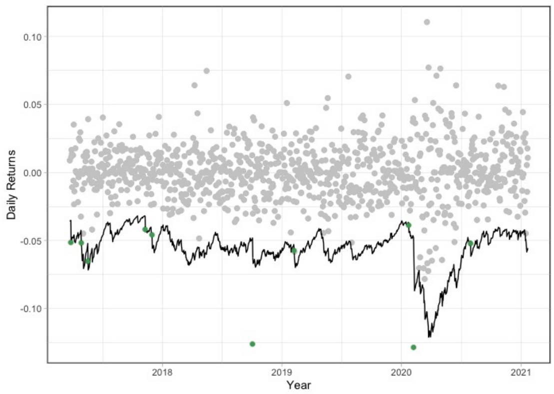
(c) Student-t VaR(0.05)



(d) Student-t VaR(0.01)



(e) Skew Student-t VaR(0.05)



(f) Skew Student-t VaR(0.01)

Figure E.6: VaR exceedance plots for APARCH(1,1) model on Norsk Hydro stock March 2017 to January 2021. The models are estimated on the data from 01.01.2006 to 31.12.2010. The black line indicates the negative VaR estimates, the gray marks the returns and the VaR violations are plotted as green marks.

Catharina Bratterud Lilleengen

NTNU
Norwegian University of
Science and Technology