ERASMUS UNIVERSITY ROTTERDAM
Erasmus School of Economics

Forecasting Conditional Tail Risk Measures by ARMA-GARCH Models

Bachelor Thesis BSc^2 Econometrics/Economics

Elsemieke Gijssel
475841

Supervisor: C. Zhou
Second assessor: B. van Os
July 4, 2021

Abstract

In this paper, we examine the forecasting performance of the tail risk measures Conditional Value-at-Risk (CVaR) and conditional expected shortfall (CES) for variants of the Autoregressive Moving Average (ARMA) model in combination with Generalised Autoregressive Conditional Heteroskedasticity (GARCH) models. For the two simulation studies, we use five simulations fitted to the GARCH, AR-GARCH, GJR-GARCH, and Exponential GARCH (EGARCH) model. For the first simulation study, we compare the GARCH and AR-GARCH models by their bias, Root Mean Squared Error (RMSE), coverage, and interval lengths for three different simulations. For the second, we determine the best estimation model by estimating the five simulations by the GARCH specification with corresponding structure and also by the other four considered estimation models. We assess their accuracy by their RMSE. We find that either the GJR-GARCH or GARCH model is the best estimation model, depending on the thickness of the tail. For the application, we consider a full period and zoom in on two shorter periods characterised by low and high volatility. To conduct the research, we use six different indices, which all represent the returns in a particular country. We conclude that the AR-GARCH model is the best for the full period, while for the low- and high-volatility periods, the best performing model depends on the index. We do, however, find that the more straightforward GARCH model performs better in the high-volatility period than in the low-volatility period and that the conditional coverage is not a reliable measure to assess the models in the high-volatility period.

Contents

1 Introduction **1**

2 Literature **3**

3 Data **5**

4 Methodology **6**
4.1 ARMA-GARCH model 6
4.2 Alternative GARCH models 7
4.3 CVaR and CES estimation 8
4.4 Estimating using EVT 9
4.5 Confidence intervals 10
4.6 Simulation study 11
4.7 Applications 12

5 Simulation **13**
5.1 GARCH versus AR-GARCH 13
5.2 GARCH versus GJR-GARCH and EGARCH 17

6 Application **19**
6.1 Low- versus high-volatility period 21
6.2 High-volatility period 23

7 Conclusion and discussion **28**
7.1 Conclusion 28
7.2 Discussion 29
7.2.1 Simulation 29
7.2.2 Application 29

A Appendix **34**
A.1 Simulation models GARCH 34
A.2 Simulation models GJR-GARCH and EGARCH 34
A.3 Additional simulation figure GARCH versus AR-GARCH 34
A.4 Additional tables GARCH versus GJR-GARCH and EGARCH 35

B Appendix **37**
B.1 Additional tables application 37

C Appendix **40**
C.1 Code explanation 40

1 Introduction

In this research, we study confidence intervals of conditional tail risk measures to assess their accuracy in variations of financial time series models, building upon the research of Hoga (2019). Tail risk indicates unanticipated high or low returns due to an unexpected event, which a model fails to capture. The left tail reflects the most extreme downside performance (losses), while the right tail reflects the extreme upside (gains). Tail risk is often hard to capture in a model, especially in times of crisis. As most investors are more concerned about their potential losses, we focus on modelling the conditional left tail in this research.

In order to capture the tail risk, we consider two well-known risk measures, Conditional Value-at-Risk (CVaR) and conditional expected shortfall (CES). To clarify, CVaR and CES are the conditional counterparts of the risk measures VaR and ES. VaR is the maximum loss that could occur over a given holding period with a pre-specified confidence level. ES is known as tail VaR, which calculates the expected loss in the tail of the distribution beyond the VaR. Financial returns are often dependent on their previous values, indicating the presence of serial dependence. This serial dependence is also present when determining VaR and ES. Thus in order to forecast VaR and ES, we require them to rely on past returns. The CES and CVaR measures take into account the serial relation of the returns, representing the VaR and ES measures conditional on the past returns.

Accurately estimating and forecasting CVaR and CES are fundamental for risk and portfolio managers. It helps them to assess the downside risk of a given stock or portfolio. In practice, stocks with low CVaRs indicate low tail risk, resulting in lower returns, while high-risk stocks often have higher returns due to a risk premium compensating for the higher risk. For investors, especially portfolio managers, this results in a risk-return trade-off when deciding on which stocks to include in their portfolio.

In risk management, the reduction of exposure to potential risks, risk mitigation, is of importance. Investors may use tail risk hedging to reduce their overall portfolio risk, aiming at maximizing their long-term returns while accounting for short-term costs. Essentially, they give up a bit of expected return in exchange for protection against a market crash. Consequently, they do not have to adjust their risk and return expectations after a market crisis. Nowadays, banks already calculate CES for their own risk assessment and in order to set capital reserves. However, estimation remains precarious, tail risk measures are in general not easily captured in a model (Danielsson et al., 2016).

For the estimation and forecasting of the risk measures, we first apply Autoregressive Moving Average (ARMA) models to mimic the univariate residuals and filter out the serial dependence that is present in financial returns. We then combine this with different Generalised Autoregressive Conditional Heteroskedasticity (GARCH) models, resulting in different ARMA-GARCH specifications to model the log-return data. Next, we use the Weissman (1978) estimator and extreme value index estimators to estimate the risk measure of the filtered residuals. Afterwards, we quantify the uncertainty of the estimation by confidence intervals of the CVaR and CES estimators to judge the accuracy of the risk measure estimates. We determine both the theoretical and practical performance of the models by simulations and applications, respectively.

In order to determine the best forecasting method, we formulate the following research

question: *Does the usage of ARMA-GARCH instead of GARCH models lead to an improvement in forecasting performance of the CES and CVaR risk measures?*. We expect that more complex models, like the GJR-GARCH, perform better as they are able to capture the asymmetric reaction to positive and negative shocks. We find that the usage of the AR-GARCH model often leads to an improvement in forecasting accuracy, while the AR-GJR-GARCH and AR-EGARCH do not lead to better forecasts. Furthermore, based on the simulations, we obtain that the GJR-GARCH should, theoretically, be the best model.

After establishing which models work best in general, we broaden the research by distinguishing between model performance in low- and high-risk periods. For this, we investigate the period of the COVID-19 crisis and the period beforehand to determine whether the performance of models differ per sample. We pick the COVID-19 crisis because the fear, uncertainties and related confinement measures that have come with it, have caused an unexpected contraction in economic activities, which led to a fall in stock market indices and an increase in their volatility (Omari et al., 2020). In addition, there is little published research on the COVID-19 crisis in combination with tail risk measures. This paper can thus provide some new insights into risk forecasting. We, therefore, state the following second research question: *Does the forecasting performance of the risk measures differ in periods of low and high volatility?* We expect to find a distinct difference between the forecasting performance in low and high periods of volatility, whereby different models perform the best. The expectations result from the idea that low- and high-volatility influence returns differently. In periods of low volatility, we expect a low number of tail events, resulting in thinner tails on both sides. Therefore, we expect that in low-volatility periods, models that are less persistent in capturing the tail risk might provide forecasts that are just as good as models that focus more on capturing the distribution's tails. Due to the low amount of tail events in the low-volatility period, it easier to capture and forecast them.

On the other hand, in periods of high volatility, we expect more shocks, both positive and negative, because of the uncertainty of the market during these periods. Ghysels et al. (2005) argue that positive shocks have a more considerable overall impact on the conditional mean of returns, take longer to be incorporated into the conditional variance, and are significantly more persistent than negative shocks. The latter accounts for the persistent nature of the conditional variance process. Surprisingly, negative shocks have a larger initial effect on the variance of the returns, though it be transitory. We thus expect that in periods of high volatility, more shocks can potentially cause tail events. This consequently results in fatter tails, for both the left as well as the right tail. This expectation is already partly justified by Agarwalla et al. (2021), which examines the impact of COVID-19 on the behaviour of the tail risk. Additionally, we expect that high volatility results in higher return, compensating for the higher volatility. The forecasting of the tail risk becomes more important because more of the observations are situated in the tails. Therefore, a model that is more focused on the tail is likely to perform better in this period. Additionally, we expect that in the high-volatility period, the sign of the shocks becomes more critical, a positive shock would possibly influence the volatility less than a negative shock. As the COVID-19 period consists of large negative shocks, we expect the models that account for this difference in reaction to certain shocks to outperform.

We find, however, that the results are not in line with our hypothesis. We learn that the

best models for low- and high-volatility periods do not differ much for each index and that the high-volatility period is overall more difficult to capture by all the considered models.

The remainder of this paper is as follows; firstly, we present a literature review in Section 2, followed by Section 3 with a visualization of the data. Subsequently, the methods used are described in Section 4. In order to gain an understanding of the capability of the models under realistic - and controllable - circumstances, a simulation study is performed, which can be found in Section 5. Afterwards, we judge the usefulness of the through applications discussed in Section 6. Finally, Section 7 presents the conclusion and discussion.

2 Literature

This research builds on Hoga (2019). The paper examines the forecasting accuracy of the risk measures CVaR and CES by establishing their confidence intervals. The advantage of CES is that it gives a completer picture of tail risk compared to CVaR, namely it rewards diversification, hereby not encouraging excessive risk-taking (Hoga, 2019). Additionally, according to Taylor (2019), many regulatory frameworks focused on the future put more emphasis on CES than CVaR.

Firstly, in Hoga (2019), the ARMA-GARCH model is used to filter out the serial dependence in the financial returns, after which the paper uses Extreme Value Theory (EVT) on the filtered residuals to estimate CVaR and CES. For the GARCH innovation terms, only the heavy-tailed characteristic is considered without making specific parametric distribution assumptions, following Chan et al. (2007). The advantages of this approach are that it is applicable regardless of the actual distribution of the GARCH innovations and enables one to provide VaR and CES intervals. To take into account the heavy-tailed characteristic, we consider two possible estimation methods; the Laplace and Gaussian Quasi-Maximum Likelihood estimation (QMLE). The first one is used when no fourth moment can be detected in the innovations, suggesting that the innovation terms have thicker tails and follow a skewed-t distribution. It therefore models the fat-tailed feature of a lot of financial market data (Xuan et al., 2017). One performs the Gaussian QMLE when at least four moments are present, indicating asymptotic normality of the innovation terms.

In addition, the Weissman (1978) estimator and an extreme value index for a subsample of the largest residuals are required to provide estimates of the error terms. The two extreme value estimators used in this research are the Hill (1975) estimator and the Moments Ratio (MR) estimator of Danielsson et al. (1996). In order to estimate them, Hoga (2019) determines the beginning of the tail indicated by the top k observations of the data, following the suggestion in Danielsson et al. (2016). Finally, the confidence intervals are constructed in two different ways; one based on a normal approximation and one on self-normalised convergence, as proposed in Shao (2010). Some advantages of the latter are that it is easy to implement as it does not require user-chosen parameters and has good finite-sample properties, in contrast to the various bootstrapping methods like the residual subsample bootstrap proposed in Spierdijk (2016). This approach imposes minimal conditions and allows for heavy-tailed and skewed error distributions, with and without a fourth moment, hereby having the advantage of not having to check the distribution of the error terms. Additionally, this method produces confidence intervals with

accurate coverage, even when other methods fail. However, we have decided not to dig into this method and leave it for further research given its complexity and inefficiency, as noted in Spierdijk (2016).

Hoga (2019) performs a simulation and an application to an empirical data set. From the simulation, Hoga (2019) concludes that the data-driven choice of k leads to good estimates for CVaR and CES in terms of coverage, the spread of their confidence intervals and Root Mean Squared Error (RMSE). Additionally, the confidence intervals based on self-normalisation seem to fit the simulated data better. The application is used to determine the better choice of GARCH model, the AR(1)-GARCH(1,1) or the GARCH(1,1). For this, Hoga (2019) measures the absolute performance of the forecasting models by examining the distribution of CVaR violations and by performing two backtests. Their relative performance is assessed using scoring functions, the quantile score of Giacomini & Komunjer (2005) and the asymmetric Laplace (AL) score function of Taylor (2019). A benefit of the AL score function is that it jointly evaluates CVaR and CES. The paper concludes that the AR(1)-GARCH(1,1) leads to better forecasting performance for both CVaR and CES.

As Hoga (2019) shows that the ARMA-GARCH model performs better than the GARCH model, we build upon this result by introducing alternative ARMA-GARCH models. Possible options for the GARCH component are stated in Horvath & Šopov (2016). Horvath & Šopov (2016) examine how different GARCH models capture tail risk by using EVT to estimate the tail index, which is a characteristic of the tail behaviour for a given distribution. Horvath & Šopov (2016) inspect different GARCH models with a normal distribution and Student's t-distribution for the residuals. The GARCH models evaluated are the simple GARCH, Exponential GARCH (EGARCH) model of Nelson (1991) and GJR-GARCH model of Glosten et al. (1993). The EGARCH model describes different effects for positive and negative innovations on volatility, while the GJR-GARCH model is a simplification of the EGARCH model, which still allows the estimation of the asymmetry effect. Contrary to Hoga (2019), the tail index, which characterises the shape of the tail, is estimated by the modified Hill method introduced by Huisman et al. (2001). The modified Hill uses weighted least squares regression with weight $\frac{1}{\sqrt{k}}$ on all the Hill estimates, resulting in assigning a higher weight to an order statistic higher in the tail. This weighting makes the calculation robust to threshold selection and less dependent on the choice of k (beginning of the tail).

By comparing the simulated distribution of the tail index and the originally estimated ones, Horvath & Šopov (2016) argue that for the normal distribution of the innovations, the fatness of the tails of the actual stock return distribution is underestimated, making the application for practical purposes difficult. In contrast, innovations with the Student's t-distribution better capture the tail shape. The paper concludes that models with a Student's t-distribution for the error term are preferable in modelling tail risk, especially for the EGARCH model. Because of these findings, we are considering the same GARCH specifications. For the estimation of the tail indexes, we are not implementing the modified Hill method because, as already stated, Hoga (2019) introduced a method to find the optimal choice of k, which makes the modified Hill method unnecessary.

Paulauskas & Vaičiulis (2017) explain the generalisation of the Hill and MR estimator.

Paulauskas & Vaičiulis (2017) show that the new estimators have a better asymptotic performance than the classical ones, whereby especially the generalised MR estimator gives improved estimation results. Nevertheless, in this paper we use the original MR estimator instead of the generalised version.

3 Data

The paper of Hoga (2019) uses the log-returns of six global indices, DJIA and NASDAQ for North America, Nikkei 225 and Hang Seng Index (HSI) for Asia, and CAC 40 and DAX 30 for Europe, for the time period of 01/01/1997 to 31/12/2016 ($\approx$ 5000 observations, varying from index to index). The raw closing prices are obtained from the supplemental material in Hoga (2019), which uses the same indices and time period. The sample period contains several events causing high volatility, namely the burst of the internet bubble (2000-2002), 9/11 (2001) and the financial crisis of 2007-2008. The empty observations are removed and the remaining data are transformed into log-returns.

The second research question demands a closer look at the difference in forecasting performance of the risk measures between low- and high-volatility periods of the different models. To maintain comparability, we use the same indices as for the first research question, except for DJIA, as this index can not be extracted from Yahoo Finance. The closing prices of the remaining indices are obtained via Yahoo Finance, and we perform the same data transformations as for the first research question. For the high-volatility period (COVID-19 period), we take the still ongoing COVID-19 crisis from 01/03/2020 to 30/04/2021 ($\approx$ 300 observations), while for the low-volatility period (pre-COVID-19 period), we take the period beforehand of a similar length, from 01/01/2019 to 28/02/2020. In order to assess the forecasting performance of both periods, we apply a rolling window with an estimation window of 1000 observations. Consequently, the entire sample period consists of the period 01/01/2011 till 30/04/2021 ($\approx$ 2000 observations). The descriptive statistics for the considered low- and high-volatility period are given in the following two tables, where the low-volatility period is shown in Table 1 and the high-volatility period in Table 2.

Table 1: Descriptive statistics of the log-returns $\times 10^2$ of stock indices for period January 1, 2019 - February 28, 2020

Stock index	Obs	Mean	St.Dev.	Max	Min	Skewness	Kurtosis	JB
NASDAQ	292	0.088	1.048	4.172	-4.723	-0.883	6.247	166.22
Nikkei	278	0.020	0.956	2.578	-3.738	-0.393	4.584	36.209
HSI	286	0.004	1.039	3.835	-2.941	-0.207	4.087	16.111
CAC 40	297	0.039	0.914	2.688	-4.025	-1.119	6.381	203.450
DAX 30	293	0.041	0.970	3.314	-4.089	-0.840	5.8377	132.798

Note. All Jarque-Bera statistics are significant on a 1% significance level.

Table 2: Descriptive statistics of the log-returns $\times 10^2$ of stock indices for period March 1, 2020 - April 30, 2021

Stock index	Obs	Mean	St.Dev.	Max	Min	Skewness	Kurtosis	JB
NASDAQ	294	0.169	2.156	8.935	-13.149	-0.972	10.722	776.78
Nikkei	286	0.108	1.554	7.731	-6.273	0.147	7.414	233.212
HSI	287	0.040	1.467	4.925	-5.720	-0.457	4.794	48.497
CAC 40	297	0.058	1.897	8.056	-13.098	-1.288	13.626	1479.408
DAX 30	294	0.083	1.925	10.414	-13.055	-0.970	13.915	1505.635

Note. All Jarque-Bera statistics are significant on a 1% significance level.

First of all, we find that the returns in the COVID-19 period, shown in Table 2, are higher than those during the pre-COVID-19 period in Table 1. This can be explained by the higher volatility in the COVID-19 period, as higher volatility results in higher risk and consequently in higher return on average. The observed higher volatility is also characterised by the obtained higher standard deviation. Secondly, we observe a wider spread between the minimum and maximum return during the COVID-19 period. Again, this confirms our hypothesis of higher volatility during this specific period. Thirdly, all the returns have negative skewness for both time periods, except for Nikkei in the COVID-19 period. Both skewness and kurtosis are far away from their normal values of 0 and 3, respectively. The negative skewness indicates that market declines occur more often than market increases, while the positive skewness suggests more large positive returns in comparison to negative returns. Note that the skewness is, on average, a bit more negative for the COVID-19 period than for the pre-COVID-19 period. We thus observe that for most indices the COVID-19 crisis causes more occurrences of negative return than the period beforehand. Lastly, from the kurtosis and JB statistic, we observe that the daily returns in both periods do not follow a normal distribution. The kurtosis and JB values in Table 2 are higher than in Table 1, giving an indication of fatter tails, which is in line with Agarwalla et al. (2021). We therefore replace the normal distribution of the residuals of the GARCH model with the fat-tailed Student's t-distribution.

4 Methodology

The following section describes the econometric techniques which are part of this research. Firstly, Sections 4.1 and 4.2 describe the considered ARMA-GARCH models. Secondly, Section 4.3 specify the theoretical estimation of the CVaR and CES. Followed by the estimation in practice and the construction of confidence intervals in Sections 4.4 and 4.5. Finally, the simulation study and application are described in Sections 4.6 and 4.7.

4.1 ARMA-GARCH model

In order to forecast CVaR and CES, we use an ARMA$(\overline{p}, \overline{q})$ model $\{X_i\}_{i\in\mathbb{Z}}$, where X_i is the log-return data for a specific index, with GARCH(p, q) errors $\{\varepsilon_i\}_{i\in\mathbb{Z}}$. As we investigate the left-tail in this research, we use the right-tail formulation for $X_i^{neg} := -X_i$, which is still an ARMA-GARCH process driven by innovations $U_i^{neg} := -U_i$. For the remainder of this section, X_i^{neg} and U_i^{neg} are denoted by X_i and U_i to make formulation easier.

Following Hoga (2019), the model for $\{X_i\}$ is:

$$X_i = \sum_{j=1}^{\bar{p}} \phi_i^\circ X_{i-j} + \varepsilon_i - \sum_{j=1}^{\bar{q}} \vartheta_j^\circ \varepsilon_{i-j}, \quad \text{with} \tag{1}$$

$$\varepsilon_i = \sigma_i U_i \quad \text{and} \quad \sigma_i^2 = \omega^\circ + \sum_{j=1}^{p} \psi_j^\circ \varepsilon_{i-j}^2 + \sum_{j=1}^{q} \beta_j^\circ \sigma_{i-j}^2, \tag{2}$$

where $\omega^\circ \geq 0$, $\psi_j^\circ \geq 0$ $(j = 1, ..., p)$, $\beta_j^\circ \geq 0$ $(j = 1, ..., q)$ and U_i are independent, identically distributed (i.i.d.), with $\mathrm{E}[U_1] = 0$ and $\mathrm{Var}(U_1) = 1$. The unknown true parameters are captured in the vector:

$$\theta^\circ := (\phi_1^\circ, ..., \phi_{\bar{p}}^\circ, \vartheta_1^\circ, ..., \vartheta_{\bar{q}}^\circ, \omega^\circ, \psi_1^\circ, ..., \psi_p^\circ, \beta_1^\circ, ..., \beta_q^\circ)^\top.$$

We make standard assumptions, explained in detail in Hoga (2019), to ensure stationarity and invertibility of the model, and an unique stationary, causal solution to Equation (2). In order to make sure that the estimators of the GARCH model are consistent, we use Laplace QMLE, explained in Berkes et al. (2004) or standard Gaussian QMLE of Francq et al. (2004), dependent on the number of moments for the innovation terms. We use the latter if the innovations have a finite fourth moment.

For given parameters θ, the feasible estimators for ε_i and σ_i^2 that follow from these assumptions are as follows:

$$\begin{aligned} \hat{\varepsilon}_i(\theta) &= \tilde{X}_i - \sum_{j=1}^{\bar{p}} \phi_i \tilde{X}_{i-j} + \sum_{j=1}^{\bar{q}} \vartheta_j \hat{\varepsilon}_{i-j}(\sigma) \\ \hat{\sigma_i}^2(\theta) &= \omega + \sum_{j=1}^{p} \psi_j \hat{\varepsilon}_{i-j}^2(\theta) + \sum_{j=1}^{q} \beta_j \hat{\sigma}_{i-j}^2(\theta), \end{aligned} \tag{3}$$

where $\tilde{X}_i = X_i$ for $i = 1, ..., n$ and $0 = \tilde{X}_0 = \tilde{X}_{-1}...$ and $0 = \hat{\varepsilon}_0(\theta) = \hat{\varepsilon}_{-1}(\theta) = ...$ and $0 = \hat{\sigma}_0^2 = \hat{\sigma}_{-1}^2 = ...$ as the estimation is based on sample $X_1, ..., X_n$. The parameter θ gives the generic parameter vector for the ARMA-GARCH model, with the estimated parameters $\hat{\theta}$. We can filter out the residuals $\hat{U}_i := \hat{U}_i(\hat{\theta}) := \hat{\varepsilon}_i(\hat{\theta})/\hat{\sigma}_i(\hat{\theta})$, where $\hat{\theta}$ is the estimated parameter vector.

4.2 Alternative GARCH models

We extend the presented ARMA-GARCH model with different GARCH specifications. The different GARCH models considered are mentioned in Horvath & Šopov (2016). The structure of the ARMA-part stays the same as specified in Equation (1), while we alter the structure of Equation (2) for each model. In order to obtain its estimates, we follow the same reasoning as for Equation (3), where the second equation is the estimator of Equation (1).

One of the GARCH models we consider in this extension is the EGARCH model of Nelson (1991), defined as:

$$\ln(\sigma_i^2) = \omega^\circ + \sum_{j=1}^{q} \beta_j^\circ \ln(\sigma_{i-j}^2) + \sum_{j=1}^{p} \psi_j^\circ \left[\frac{|\varepsilon_{i-j}|}{\sigma_{i-j}} - \mathbb{E}\left\{ \frac{|\varepsilon_{i-j}|}{\sigma_{i-j}} \right\} \right] + \sum_{j=1}^{p} \gamma_j^\circ \left(\frac{\varepsilon_{i-j}}{\sigma_{i-j}} \right), \tag{4}$$

where ω° is the true constant parameter, ψ_j° represents a symmetric effect, β_j° measures the persistence in conditional volatility and γ_j° allows for asymmetries, also known as the leverage effect. Additionally $\gamma^\circ < 0$ indicates that negative innovation creates more volatility than positive and the other way around. Due to the logarithmic form of the variance, the coefficients $\omega^\circ, \alpha_j^\circ, \beta_j^\circ$ or γ_j° can attain negative values while σ_i^2 stays positive.

We consider another extension of the GARCH model, namely the GJR-GARCH, which is a simplification of the EGARCH model but still allows the estimation of the asymmetry effect (Glosten et al., 1993). The conditional variance is modelled as follows:

$$\sigma_i^2 = \omega^\circ + \sum_{j=1}^{p} \psi_j^\circ \varepsilon_{i-j}^2 + \sum_{j=1}^{q} \beta_j^\circ \sigma_{i-j}^2 + \sum_{j=1}^{p} \gamma_i^\circ \varepsilon_{i-j}^2 \mathrm{I}_{i-j}, \tag{5}$$

with indicator function:

$$I_{i-j} = \begin{cases} 1 & \varepsilon_{i-j} > 0 \\ 0 & \varepsilon_{i-j} < 0, \end{cases} \tag{6}$$

where the coefficients α_j°, β_j° and ε_i have the same interpretation as in the GARCH model and γ_j° denotes the asymmetry effect. A leverage effect is present when γ_j° is positive, while γ_i° equals zero indicates a symmetric reaction of a change in volatility to returns. The model switches between two different parameters for the error term, ψ_j° and $\psi_j^\circ + \gamma_j^\circ$, depending on the sign of the past shock, represented by Equation (6). There are some additional constraints on the parameters, namely:

$$\omega^\circ \geq 0, \quad \sum_{j=1}^{p} \psi_j^\circ \geq 0, \quad \sum_{j=1}^{q} \beta_j^\circ \geq 0 \quad \text{and} \quad \sum_{j=1}^{p} \psi_j^\circ + \sum_{j=1}^{p} \gamma_i^\circ \geq 0 \tag{7}$$

4.3 CVaR and CES estimation

To estimate both CVaR and CES of observation n, we first need to estimate μ_{n+1}, the mean of observation $n+1$, which is obtained by the invertibility of Equation (2) under the standard assumptions stated in Hoga (2019). The estimate is calculated using the following equation:

$$\hat{\mu}_{n+1}(\theta) = \sum_{j=1}^{r} \tilde{\phi}_j \tilde{X}_{n+1-j} + \sum_{j=1}^{r} \tilde{\phi}_j \sum_{k=1}^{\infty} \sum_{j_1=1}^{r} \dots \sum_{j_k=1}^{r} \vartheta_{j_1} \dots \vartheta_{j_k} \tilde{X}_{n+1-j-j_1-\dots-j_k}, \tag{8}$$

where $r := \max\{\overline{p}, \overline{q}\}$ and $\tilde{\phi}_j := \phi_j - \vartheta_j$, $\vartheta_{j_1}, \dots, \vartheta j_k$ are the parameters estimated in Equation (1) for the observations from 1 till r.

For the ARMA-GARCH model, the estimators of the right-tail CVaR and CES are given by:

$$\begin{aligned} \hat{x}_{a,n} &= \hat{\mu}_{n+1} + \hat{\sigma}_{n+1} \hat{x}_\alpha^U \quad \text{and} \\ \hat{S}_{a,n} &= \hat{\mu}_{n+1} + \hat{\sigma}_{n+1} \hat{S}_\alpha^U \end{aligned} \tag{9}$$

where $\hat{\mu}_{n+1}$ is defined in Equation (8) and $\hat{\sigma}_{n+1}$ specified in Equation (3), both for $\{X_n, \dots, X_1\}$. Furthermore, $\hat{x}_\alpha^U$ denotes the estimator of the $(1-\alpha)$-quantile of U_1 and $\hat{S}_\alpha^U$ is the estimator of $\mathrm{E}[U|U > x_\alpha]$. Both estimators are based on EVT.

4.4 Estimating using EVT

In order to obtain the estimates $\hat{x}_\alpha^U$ and $\hat{S}_\alpha^U$, and to calculate the confidence intervals, we require U_i to satisfy some additional underlying assumptions formulated in Hoga (2019). To illustrate the assumptions for U_i, let $\{U_i\}$ be a sequence of i.i.d. random variables with distribution function F. Define the $(1-1/x)$-quantile $U(x) := F^{\leftarrow}(1-1/x)$, where $F^{\leftarrow}$ denotes the left-continuous inverse of F. Similar to Hoga (2019), we assume $U(\cdot)$ to be varying with *extreme value index* $\gamma > 0$, namely

$$\lim_{x\to\infty} \frac{U(xy)}{U(x)} = y^\gamma \qquad \forall y > 0. \tag{10}$$

As Equation (10) is an asymptotic tail property, we require an integer sequence for the estimation, formulated as follows:

$$k = k_n \to \infty \quad \text{with} \quad 1 \le k < n \quad \text{and} \quad k/n \to 0 \quad \text{as} \quad n \to \infty \tag{11}$$

where k specifies where the tail of the distribution begins and the equation is assumed to hold throughout.

With the tail property and value of k, we can determine $\hat{x}_\alpha^U$. For given sample $\{U_1, ..., U_n\}$ with order statistics $U_{1:n} \le ... \le U_{n:n}$, we use the Weissman (1978) estimator to obtain $\hat{x}_\alpha^U$. Note that $U(1/\alpha)$ is the $(1-\alpha)$-quantile of U_1 we want to estimate. Then define

$$\hat{x}_\alpha^U := U_{n-k:n}\Big(\frac{n\alpha}{k}\Big)^{\hat{\gamma}}, \tag{12}$$

where $U_{n-k:n}$ denotes the order statistic for the sample of random variable $\{U_i\}_{i=k}^n$ and n gives the size of the estimation sample. The estimator of the extreme value index γ, $\hat{\gamma}$, is based on the $(k+1)$ largest order statistics of $\{U_1, ..., U_n\}$. Under the condition of a small α and $\gamma < 1$, ES is a constant multiple of VaR, suggesting the estimator:

$$\hat{S}_\alpha^U := \frac{\hat{x}_\alpha^U}{1-\hat{\gamma}}. \tag{13}$$

where $\hat{\gamma}$ is the estimator of the extreme value index and $\hat{x}_\alpha^U$ is described in Equation (12).

As an estimator of γ, we use the Hill (1975) estimator constructed as

$$\hat{\gamma}_H := \frac{1}{k}\sum_{i=0}^{k} \log\Big(\frac{U_{n-i:n}}{U_{n-k:n}}\Big). \tag{14}$$

Another estimator is the MR estimator of Danielsson et al. (1996),

$$\hat{\gamma}_{MR} := \frac{1}{2}\frac{\frac{1}{k}\sum_{i=0}^{k}\{\log(U_{n-i:n}) - \log(U_{n-k:n})\}^2}{\hat{\gamma}_H}. \tag{15}$$

For both estimators of γ, a choice of k is required. We follow Danielsson et al. (1996), by minimizing the largest distance between a fitted Pareto-type tail and the empirical quantile. We

want to choose k such that the tail approximation, defined as follows:

$$U\big(\frac{n}{j}\big) \approx U_{n-k,n}\big(\frac{j}{l}\big)^{-\hat{\gamma}} =: q(j,k), \tag{16}$$

is good for any value of j. This means that the deviation of $q(j,k)$ from the observed quantile $U_{n-j,n}$ is small for any j and results in the following optimal choice of k:

$$k^* = \underset{k=k_{min},\dots,k_{max}}{\arg\min} \Big[\sup_{j=1,\dots,k_{max}} |U_{n-j,n} - q(j,k)| \Big], \tag{17}$$

where k_{min} and k_{max} are chosen beforehand, and are in this research set to 50 and 200 such that k is small compared to n and $n\alpha$ is small relative to k, which is a requirement according to Theorem 1 in Hoga (2019).

4.5 Confidence intervals

In order to obtain the confidence intervals for the CVaR and CES estimates, we follow Chan et al. (2007) by basing the estimation of γ only on the last residuals, since the first few are affected by the initialization effects of variance Equation (3). We use the subsample $\{\hat{U}_{m_n},...,\hat{U}_n\}$, with chosen starting value $m_n \to \infty$, such that $m_n < n$ and $m_n = o(\sqrt{k})$, as $n \to \infty$. Throughout this research, we use a value of $m_n = 10$.

By this mentioned trimming of the residuals, we are required to formulate a slightly adjusted version of the Hill (1975) and MR estimator of Equations (14) and (15). We use the estimators for the sample $\{\hat{U}_{m_n},...,\hat{U}_{\lfloor nt \rfloor}\}$ with a sample size of $nt - m_n$, with the assumption that the asymptotic tail property in Equation (10) holds for k and where t is a fraction indicating where the considered sample ends.

We then estimate the $(1-\alpha)$-quantile and corresponding VaR and ES based on the same subsample $\{\hat{U}_{m_n},...,\hat{U}_{\lfloor nt \rfloor}\}$ by the following adjusted version of Equations (12) and (13):

$$\hat{x}^U_\alpha(t) := \hat{U}_k(t)\Big(\frac{n\alpha}{k}\Big)^{\hat{\gamma}(t)} \quad \text{and} \quad \hat{S}^U_\alpha(t) := \frac{\hat{x}^U_\alpha(t)}{1-\hat{\gamma}(t)}, \tag{18}$$

where $\hat{U}_k(t)$ are the $(\lfloor kt \rfloor + 1)$ largest values of the subsample and k is the optimal level as defined in Equation (17) that gives the optimal value of the beginning of the tail distribution. The tail index estimator, $\hat{\gamma}(t)$, is calculated by applying Equation (14) or (15) to the subsample and replacing k by kt. Then, we determine the estimates of the CVaR and CES, following the motivation of Equation (9):

$$\begin{aligned} \hat{x}_{\alpha,n}(t) &= \hat{\mu}_{n+1} + \hat{\sigma}_{n+1}\hat{x}^U_\alpha(t) \quad \text{and} \\ \hat{S}_{\alpha,n}(t) &= \hat{\mu}_{n+1} + \hat{\sigma}_{n+1}\hat{S}^U_\alpha(t), \end{aligned} \tag{19}$$

where the variables have the same definition as in Equation (9), but now for a specific input value of t.

To construct the confidence intervals for the CVaR and CES estimates, we require Theorem 1, stated in Hoga (2019), to hold. It states that the ARMA-GARCH model and the innovations

need to satisfy four underlying assumptions to obtain asymptotic relations necessary for the confidence intervals. For $\hat{\sigma}^2_{\hat{\gamma},\gamma} = \hat{\gamma}^2_H(1)$ ($\hat{\sigma}^2_{\hat{\gamma},\gamma} = \hat{\gamma}^2_{MR}(1)$) if $\hat{\gamma} = \hat{\gamma}_H$ ($\hat{\gamma} = \hat{\gamma}_{MR}$), the following holds:

$$\frac{1}{\sigma_{\hat{\gamma},\gamma}} \frac{\sqrt{k}}{\log(k/(n\alpha))} \log\Big(\frac{\hat{z}_{\alpha,n}(1)}{z_{\alpha,n}}\Big) \xrightarrow[(n\to\infty)]{\mathcal{D}} \mathcal{N}(0,1) \quad \text{and} \tag{20}$$

$$\frac{\log^2\left(\frac{\hat{z}_{\alpha,n}(1)}{z_{\alpha,n}}\right)}{\int_{t_0}^1 t^2 \log^2\left(\frac{\hat{z}_{\alpha,n}(t)}{\hat{z}_{\alpha,n}(1)}\right)\mathrm{d}t} \xrightarrow[(n\to\infty)]{\mathcal{D}} \frac{W^2(1)}{\int_{t_0}^1 [W(t) - tW(1)]^2 dt} =: V_{t_0}, \tag{21}$$

where $W(\cdot)$ is a standard Brownian motion, k is where the tail of the distribution begins, specified in Equation (17), n indicates the length of the estimation sample and $\alpha = \alpha_n \in (0,1)$ indicates the $(1-\alpha)$-quantile, all for $z \in \{x, S\}$. The estimator of z is thus either the estimator for x or S. Additionally, t_0 is a fraction that indicates from which observation within the sample to start the calculation. Equation (20) depicts a normal approximation, while Equation (21) shows self-normalised convergence.

Under Theorem 1, the following asymptotic $(1-\tau)$-confidence intervals for $z_{\alpha,n}$ are derived:

$$I^{1-\tau}_{na} := \left[\hat{z}_{\alpha,n}(1)\exp\left\{-\Phi\Big(1-\frac{\tau}{2}\Big)\hat{\sigma}_{\hat{\gamma},\gamma}\frac{\log(k/(n\alpha))}{\sqrt{k}}\right\}, \hat{z}_{\alpha,n}(1)\exp\left\{\Phi\Big(1-\frac{\tau}{2}\Big)\hat{\sigma}_{\hat{\gamma},\gamma}\frac{\log(k/(n\alpha))}{\sqrt{k}}\right\}\right],$$

$$I^{1-\tau}_{sn} := \left[\hat{z}_{\alpha,n}(1)\exp\left\{-\sqrt{V_{t_0,1-\tau}\int_{t_0}^1 t^2\log^2\Big(\frac{\hat{z}_{\alpha,n}(t)}{\hat{z}_{\alpha,n}(1)}\Big)\mathrm{d}t}\right\}, \hat{z}_{\alpha,n}(1)\exp\sqrt{V_{t_0,1-\tau}\int_{t_0}^1 t^2\log^2\Big(\frac{\hat{z}_{\alpha,n}(t)}{\hat{z}_{\alpha,n}(1)}\Big)\mathrm{d}t}\right\}\right],$$

where $z_{\alpha,n\in\{x,S\}}$, $V_{t_0,1-\tau}$ denotes the $(1-\tau)$-quantile of V_{t_0}, which is calculated using Equation (21) and $\Phi(\cdot)$ is the cumulative distribution function of standard normal distribution. The first interval is based on Equation (20), while the second is based on Equation (21). Additionally, t_0 is a fraction that represents the observation from which we start the calculation within the sample.

The choice of t_0 is based on a trade-off, whereby choosing t_0 too large can lead to poor approximations as the integral in $I^{1-\tau}_{sn}$ runs over too few log-differences, namely from t_0 up to 1. On the other hand a larger t_0 gives a larger part of the sample being used to calculate $\hat{z}_{\alpha,n}(t_0)$. This would make the estimate more reliable. Thus, the larger the sample size n, the smaller t_0 can be chosen. We take t_0 such that $\hat{z}_{\alpha,n}(t_0)$ is based on at least 200 observations, following Hill (2015).

4.6 Simulation study

We use a simulation study to compare the finite-sample coverage of the 95%-confidence intervals for the left-tail α-CVaR $x_{\alpha,n}$ and α-CES $S_{\alpha,n}$. The two different intervals, $I^{0.95}_{na}$ and $I^{0.95}_{sn}$ are based on a constructed time series with length $n = 1000$. For $I^{0.95}_{sn}$, we use $t_0 = 0.2$ such that we have a minimal sample size of 200 for CVaR and CES estimation. For completeness, we study the robustness of the choice of t_0 by calculating the value of $V_{t_0,\tau}$ for $t_0 = 0.1, 0.2$ and 0.3, for τ in range of 0.50 and 0.995. According to Hoga (2019), a problem in the estimation of CES can arise when t_0 is very small, as then $\hat{\gamma}(t)$ depends on only a few observations for t close to t_0. Consequently, it is possible to obtain an estimate $\hat{\gamma} \geq 1$, making the estimate of

$\hat{S}_\alpha^U(t)$ invalid. To counter this, we use $\hat{x}_\alpha^U(t)/(1-\min\{\hat{\gamma}(t), 0.9\})$ in the simulations. We use 10,000 replications throughout and values $\alpha = 2.5\%$, 1%, 0.5%, where the last is used to assess if our approach is capable of providing accurate estimates further away in the tails. Furthermore, we examine the sensitivity of the coverage of $I_{na}^{0.95}$ and $I_{sn}^{0.95}$ to the choice of k. We use the choice of k from Equation (17), resulting in an optimal k between 5% and 20%. Lastly, the bias and RMSE, calculated as in Gopinathan (1988), of the different estimators for the CVaR and CES are defined to determine whether the Hill (1975) estimator or MR estimator is a better choice and to determine its sensitivity to k. The models we use in the simulations are given in Appendix A.1.

To determine the 'best' performing model of the ones specified in Sections 4.1 and 4.2, we perform a second simulation study. We use six different simulations which are fitted to the GARCH with Laplace estimation, GARCH with Gaussian estimation, ARMA-GARCH, GJR-GARCH and EGARCH models. We expect that the simulations, which are fitted according to the specific models, result in a good performance for that model. Thus, in order to assess a certain model's quality, one has to focus on its estimating performance in the other simulations. We therefore use all the simulations stated in Appendix A.1 and A.2. We hereby consider simulations that take into account asymmetric responses to positive and negative shocks (GJR-GARCH and EGARCH) and those that do not (GARCH and ARMA-GARCH). By doing so, we get a full indication of a model's theoretical performance in different circumstances and with different responses to shocks.

4.7 Applications

For the full period application, we use the historical data, described in Section 3, to investigate whether the ARMA-GARCH models, with a data-dependent choice of k outperform those from simpler GARCH models with a fixed $k = \lfloor 1.5(\log n)^2 \rfloor$, recommended in Chan et al. (2007). Secondly, we use the data to determine whether the usage of the MR estimator is beneficial. First of all, we set a rolling window of n=1000 to capture possible non-stationarities in the long series of returns. We calculate the left tail α-VaR and α-CES estimates, for days $j \in \{n, ..., N-1\}$ and $\alpha = 0.5\%$, by fitting an AR(1)-GARCH(1,1) model to the last n observations, after which we extract the residuals for $m_n = 10$ and calculate the optimal k by Equation (17). We then determine the VaR and ES of the residuals by the Hill and MR estimator. For the GARCH(1,1) model, k is fixed and only the Hill estimator is considered. In total, this results in three different models to estimate and forecast for both the CVaR and CES.

Afterwards, we evaluate the absolute and relative performance of the forecasts. We examine the *absolute* performance of the CVaR by testing the assumption that the sequence of CVaR violations, measured by the indicator $I_{\{j+1 \leq \hat{x}_{\alpha,n}^{j+1}(1)\}}$, is i.i.d. Bernoulli with success probability α, following Hoga (2019). We do this by performing two backtests. Firstly, the test of Kupiec (1995) (UC), which assesses the correct unconditional coverage and is calculated by $\mathrm{E}[I_{\{j+1 \leq \hat{x}_{\alpha,n}^{j+1}(1)\}}] = \alpha$. The inclusion of this test enables us to determine if the number of CVaR violations is close to the desired level, ignoring possible clustering in the violations. Secondly, we apply the Ljung-Box conditional coverage test of Berkowitz et al. (2011) (CC), which tests for correct unconditional coverage and violations being i.i.d..

The p-values obtained from the backtests are not suitable to compare if models are acceptable or not. Hence, we use scoring functions for both CVaR and CES forecasts, described in Fissler et al. (2016), to determine *relative* performance. To evaluate CVaR forecasts of different models, we consider the *quantile score*, defined as:

$$S_{VaR}(\hat{x}_{\alpha,n}, X_{n+1}) = (X_{n+1} - \hat{x}_{\alpha,n})(\alpha - I_{\{X_{n+1} \leq \hat{x}_{\alpha,n}\}}) \tag{22}$$

where $x_{\hat{\alpha},n}$ is the CVaR estimate and X_{n+1} is the loss return value. To evaluate CVaR and CES together, we use the *asymmetric Laplace (AL) log score* of Taylor (2019), given by

$$S_{ES}(\hat{x}_{\alpha,n}, \hat{S}_{\alpha,n}, X_{n+1}) = -\log\left(\frac{\alpha - 1}{\hat{S}_{\alpha,n}}\right) - \frac{(X_{n+1} - \hat{x}_{\alpha,n})(\alpha - I_{\{X_{n+1} \leq \hat{x}_{\alpha,n}\}})}{\alpha \hat{S}_{\alpha,n}} \tag{23}$$

where $\hat{S}_{\alpha,n}$ gives the CES estimate and X_{n+1} the loss return value. Comparing two different forecasts from the different models is then done by comparing the realised scores, where we take the sum for the scoring functions over all the observations in the forecasting sample. The forecast sequence with the lower score is preferred. By taking the sum, we can compare the performance of the CVaR and CES for the three different models for a certain index. Note that the scores for different indices can not be compared. We can assess the significance between the different scores of the GARCH(1,1) and AR(1)-GARCH(1,1) model with the Hill estimator for each index by performing the Diebold Mariano (DM) test of Diebold (2015) for the score functions as described in Equations (22) and (23).

For the second research question, we repeat the application above, however, due to the limited amount of observations and short periods of high and low volatility, we use an in-sample estimation for the ARMA-GARCH parameters. Afterwards, we use a rolling window approach to forecast the CVaR and CES and determine their quality. With the purpose of making a comparison between the two periods, we assess the performance measures for the pre-COVID-19 period and COVID-19 period separately.

5 Simulation

5.1 GARCH versus AR-GARCH

Table 3 shows the quantiles which are required for the calculation of the confidence interval $I_{sn}^{0.95}$. The calculation of $V_{t_0,\tau}$ is according to Equation (21), where τ indicates the considered quantile equal to $(1 - \alpha)$. The results are obtained using a modified version of the R code from Hoga (2019). Appendix C gives the explanation and adaptations of the code.

Table 3: Quantiles $V_{t_0,\tau}$ of V_{t_0}

τ	t_0	0.50	0.60	0.70	0.80	0.90	0.95	0.975	0.99	0.995
$V_{t_0,\tau}$	0.3	4.462	7.495	12.153	20.831	40.769	66.019	95.801	146.903	186.331
	0.2	3.849	6.462	10.544	17.811	34.278	56.254	86.902	137.038	170.692
	0.1	3.496	5.769	9.436	15.749	30.458	48.785	70.499	104.800	135.290

When we compare the obtained results in Table 3 with Hoga (2019), we observe some minor deviations. Especially the values for the larger quantiles, such as for $\tau = 0.995$ with an obtained value of 186.311 and a value in Hoga (2019) of 207.0, differ substantially. The density of the Brownian motion can explain this difference. As we simulate by considering a less dense Brownian motion than the one used in Hoga (2019), we obtain higher quantile values. For a larger value of t_0, the simulations show higher quantile values because a larger value of t_0 results in a larger minimum sample size for the CVaR and CES estimation, giving more estimation values and thus a higher chance for larger outliers. Both observations are in line with Hoga (2019). We find that for the different choices of t_0, the quantile values are still quite close to each other, showing that the results are robust to different choices of t_0.

The simulation results for models (A.1), (A.2) and (A.3) are depicted in Table 4.

Table 4: Average of k, bias, RMSE, coverage probabilities in % (nominal coverage: 95%) and average interval lengths for Hill and MR estimator for the simulation models (A.1), A.2 and (A.3).

							Coverage		Int. length	
Model	Estimator	k^*	z	α	Bias	RMSE	$I_{na}^{0.95}$	$I_{sn}^{0.95}$	$I_{na}^{0.95}$	$I_{sn}^{0.95}$
(A.1)	Hill	68	CVaR	2.5%	-0.034	0.560	49.270	74.180	0.235	0.483
				1.0%	-0.091	0.556	69.150	80.270	0.526	0.793
				0.5%	-0.173	3.621	77.270	84.490	0.822	1.180
			CES	2.5%	-0.074	0.702	39.950	82.710	0.274	0.905
				1.0%	-0.095	0.696	59.580	87.110	0.613	1.431
				0.5%	-0.119	4.241	68.330	88.890	0.958	1.988
	MR	98	CVaR	2.5%	0.059	0.534	60.590	75.350	0.373	0.624
				1.0%	0.002	0.534	81.370	81.730	0.737	0.972
				0.5%	-0.078	3.327	86.350	83.160	1.109	1.369
			CES	2.5%	0.019	0.680	53.420	81.230	0.437	1.089
				1.0%	-0.005	0.697	69.150	83.090	0.863	1.636
				0.5%	-0.030	3.895	73.890	82.830	1.300	2.202
(A.2)	Hill	59	CVaR	2.5%	0.360	2.449	61.090	86.110	2.913	6.460
				1.0%	-0.003	3.425	85.540	90.240	8.456	12.338
				0.5%	-1.192	4.953	91.980	93.610	15.040	20.525
			CES	2.5%	-1.285	4.905	45.320	93.320	4.440	17.534
				1.0%	-3.831	8.164	64.300	94.160	12.896	34.044
				0.5%	-7.308	12.731	68.370	93.350	23.000	52.598
	MR	64	CVaR	2.5%	1.408	2.722	56.410	79.000	3.689	6.940
				1.0%	2.029	3.772	82.330	79.330	10.257	11.814
				0.5%	1.948	4.837	91.890	83.150	17.830	18.146
			CES	2.5%	1.348	4.600	51.880	84.010	5.451	15.305
				1.0%	0.614	6.644	77.920	87.370	15.139	27.625
				0.5%	-0.844	9.609	86.900	89.000	26.368	40.765
(A.3)	Hill	58	CVaR	2.5%	0.225	1.094	56.510	82.560	1.284	2.814
				1.0%	0.108	1.524	84.300	88.880	3.444	4.830
				0.5%	-0.382	2.151	91.220	92.490	5.852	7.777
			CES	2.5%	-0.409	1.786	46.890	91.650	1.755	6.321
				1.0%	-1.282	3.043	66.460	93.310	4.713	11.206
				0.5%	-2.464	4.578	70.250	92.660	8.003	16.844
	MR	66	CVaR	2.5%	0.764	1.307	51.160	73.190	1.688	3.015
				1.0%	1.031	1.795	79.020	75.360	4.238	4.647
				0.5%	0.946	2.155	90.380	80.530	7.011	6.899
			CES	2.5%	0.701	1.756	51.910	81.560	2.255	5.702
				1.0%	0.429	2.589	80.250	86.210	5.663	9.179
				0.5%	-0.115	3.404	88.930	89.140	9.359	13.379

Note. For models (A.2) and (A.3), the values of bias, RMSE and interval lengths are multiplied by 10^3.

From Table 4, we notice an increasing bias for the Hill estimator for CES when α decreases,

while for the MR estimator the bias does not show a clear pattern. Furthermore, the RMSE increases as well for higher α. For each level of α, we observe lower RMSE values for CVaR than CES. All results mentioned above are in line with Hoga (2019), however, the obtained values differ in absolute terms from those in Hoga (2019). Note that the RMSE values of model (A.1), for $\alpha = 0.5\%$ are quite large compared to alternative values of α and that the RMSE differs significantly from the RMSE value stated in Hoga (2019). Section 7.2 discusses an possible explanation.

Moreover, the comparable RMSE values for both Hill and MR estimators might be surprising, as according to Theorem 1 of Hoga (2019), the theoretical asymptotic variances of CVaR and CES based on Hill estimator are smaller by a factor two in comparison with the MR estimator. However, the obtained results are consistent with Wagner & Marsh (2004). An exception to the comparable values for the estimators is for model (A.2), for CES and an α of 0.5% as the RMSE for the Hill and MR estimator are 12.731 and 9.609, respectively.

Regarding the coverage, we obtain coverage that is, on average, relatively low for I_{na}, except for the CVaR in combination with the MR estimator. We observe an improvement when using the self-normalised confidence intervals I_{sn}, especially for $\alpha = 2.5\%$. Additionally, we find, consistent with Hoga (2019), higher coverage improvement for CES than CVaR. This can be explained by Equation (20), where σ is unchanged and consequently the variance is the same for both CVaR and CES. For the self-normalised data, we use calculations specific for a particular risk measure, thereby being able to more adequately capture the larger variability in the CES estimates. Consequently leading to more accurate coverage. Overall, the coverage obtained using the self-normalised confidence intervals, $I_{sn}^{0.95}$, appear to be sufficient according to Chan et al. (2007). We observe that higher coverage leads to a longer interval length. Furthermore, the confidence interval lengths between CVaR and CES do not differ much when based on the normal approximation $I_{na}^{0.95}$, while they do differ much more for $I_{sn}^{0.95}$ based on a self-normalised approximation. A possible explanation for this finding is the larger improvement of CES in comparison with CVaR.

The quality of the confidence intervals depends on the quality of the approximations of Equations (20) and (21). To demonstrate, Figure 1 shows the probability-probability (PP) plots of the random variables on the left and right sides of Equations (20) and (21). The PP plot is of simulation model (A.2) and left-tail α-CES with $\alpha = 1\%$.

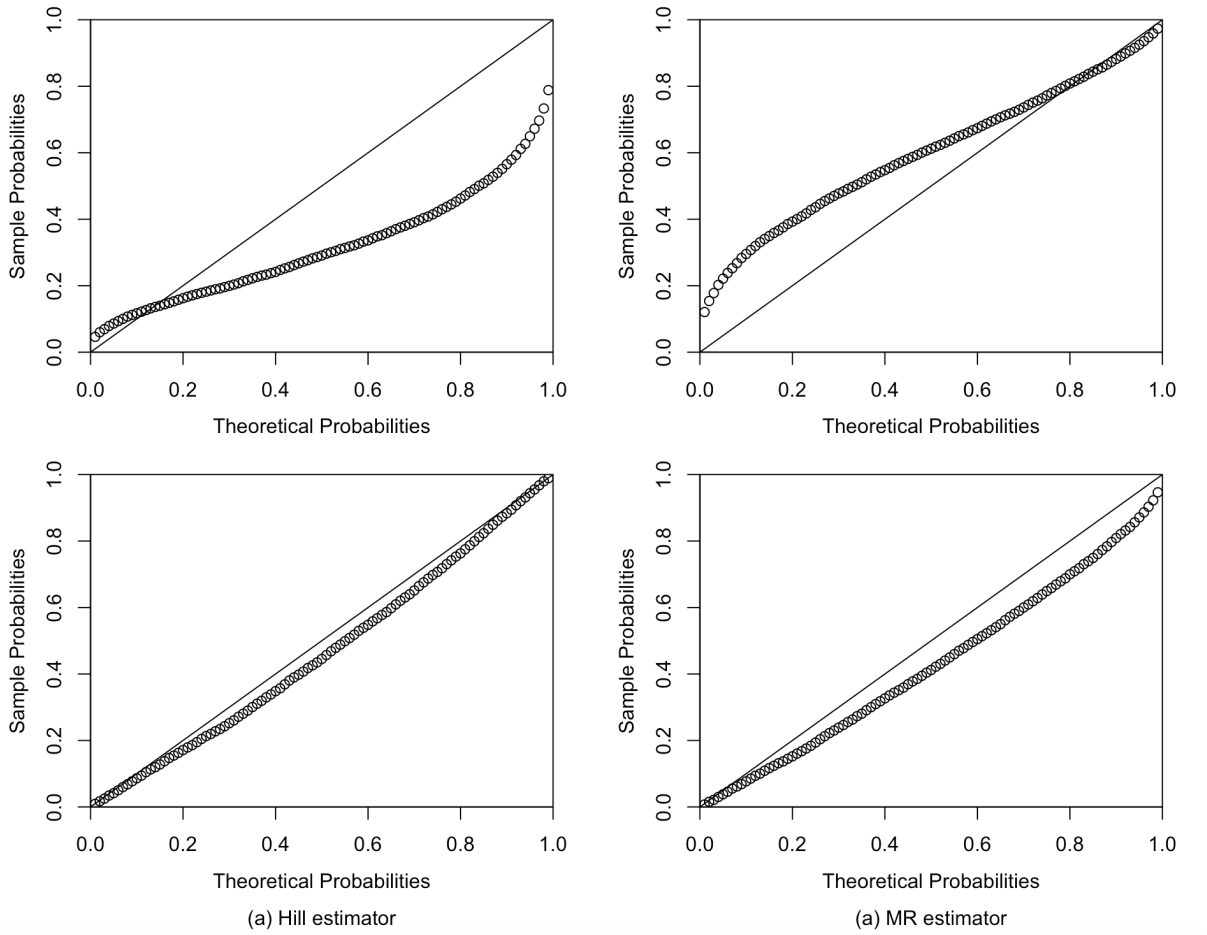


Figure 1: PP plots for random variables of Equation (20) in top subfigures and (21) in bottom subfigures for left-tail 1%-CES of simulation model (A.2)

Similar to Hoga (2019), we find that the self-normalised approximation gives a better fit than the normal approximation for both the Hill and MR estimator.

In order to determine the sensitivity of the results in Table 4 to the choice of k, we consider additional figures which show the coverage probability, interval length, bias and RMSE in relation with $k \in [50, 200]$. We consider the left-tail 1% CES for both $I_{na}^{0.95}$ and $I_{sn}^{0.95}$. The results using the Hill estimator are depicted left, while those using the MR estimator are shown right in Figures 2 and A1. Figure 2 shows the results of simulation model (A.2).

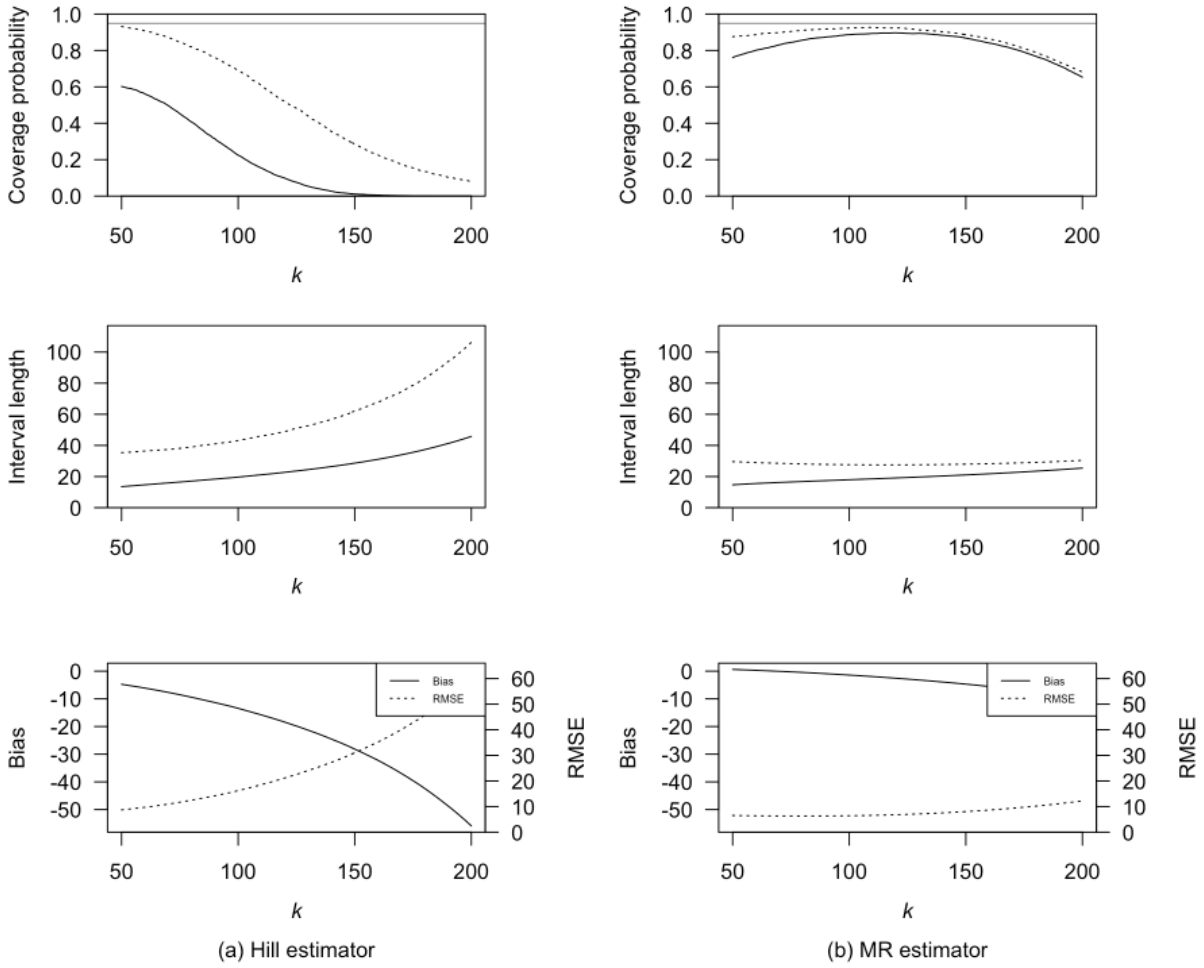


Figure 2: Characteristics as a function of k for left-tail 1% CES for $I_{na}^{0.95}$ (solid) and $I_{sn}^{0,95}$ (dotted) for model (A.2). Interval lengths, bias and RMSE multiplied by 10^3. The horizontal line in the top plots indicates nominal coverage of 95%.

The obtained subfigures of Figure 2 follow the same pattern as the ones in Hoga (2019). We observe that especially the coverage, interval length, bias and RMSE for the Hill estimator are sensitive to the choice of k, whereas for the MR estimator, the choice of k has less effect on the considered variables. Additionally, we find that the average choice of k^*, presented in Table 4, gives the highest coverage probabilities (closest to the nominal level of 95%), relatively short interval lengths and low bias and RMSE values. All results are in accordance with the simulation study results in Hoga (2019). For model (A.1), we can construct Appendix Figure A1 with the same structure as Figure 2. We observe results close to the ones for model (A.2) and those reported in Hoga (2019).

5.2 GARCH versus GJR-GARCH and EGARCH

In order to assess the performance of GARCH, GJR-GARCH and EGARCH specifications and determine which one performs best, we perform the second simulation study described in Section 4.6. The performance of the models is quantified by their RMSE, whereby the best model attains the lowest RMSE value. For the GARCH specification, we use two GARCH(1,1) models, one with a Laplace QMLE, denoted by GARCH-t and a Gaussian QMLE, denoted by GARCH.

As we have found comparable values for the Hill and MR estimator in the previous simulation study, we only consider the results of the Hill estimator to maintain readability and keep this research concise.

The outcomes of this study are shown in Table 5 for $\alpha = 1.0\%$. The top bar of the table shows the estimation model and the simulation model that is being evaluated is depicted on the left hand side. The simulations have a specific structure related to a particular model, specified in Appendices A.1 and A.2. To sum up, model (A.1) is simulated using the structure of the GARCH model, model (A.2) the structure of the GARCH-t, model (A.3) of the AR-GARCH model, model (A.4) is fitted according to the GJR-GARCH structure and model (A.5) uses the EGARCH model as basis of the simulation. In bold are the values that indicate the best model (with the lowest RMSE) for that specific simulation and the underlined values the second best. Moreover, the results for $\alpha = 0.5\%$ and 2.5% are stated in Appendix A.4. Note that we take $\alpha = 0.5\%$ mainly to determine how far in the tails the models are capable of providing good forecasts. Thus to determine which model performs best and apply in real-life applications, we take these results less into account and focus more on the performance for the other two values of α.

Table 5: Performance matrix for different simulations and models for $\alpha = 1.0\%$ and Hill estimator

		GARCH-t	GARCH	AR-GARCH	GJR-GARCH	EGARCH
(A.1)	CVaR	<u>0.556</u>	1.037	1.010	**0.518**	8.531
	CES	<u>0.696</u>	2.467	1.311	**0.649**	5.053
(A.2)	CVaR	**2.919**	3.425	<u>3.273</u>	3.227	11.754
	CES	7.951	8.164	**7.670**	<u>7.786</u>	15.345
(A.3)	CVaR	2.878	2.957	<u>1.524</u>	**1.515**	10.474
	CES	4.269	4.428	<u>3.043</u>	**2.914**	13.571
(A.4)	CVaR	<u>0.176</u>	0.180	0.179	**0.134**	0.616
	CES	0.346	<u>0.345</u>	0.350	**0.230**	0.837
(A.5)	CVaR	<u>0.164</u>	0.659	0.490	0.166	**0.145**
	CES	<u>0.253</u>	1.145	0.880	0.255	**0.230**

Note. For models (A.2) and (A.3), the values are multiplied by 10^3.

Overall, the results in Table 5 show that the best estimates for $\alpha = 1.0\%$ are obtained by the GJR-GARCH model. For all simulations and for both CVaR and CES, this model performs the best or second best, except for model (A.5) and CVaR estimation for model (A.2). This gives an indication to use the GJR-GARCH model specification when there is uncertainty about the structure of the data.

Interestingly, we observe that the best performing estimation and forecasting model is not always the one that is initially fitted to construct the simulation and is not necessarily the same for both risk measures. For example, for model A.2 $\alpha = 1.0\%$, we find that the GARCH-t model is the best for CVaR, while for CES the AR-GARCH gives the most accurate estimates. Additionally, for models A.1 and A.3 observe that the GJR-GARCH and EGARCH model give the best estimates for both estimators and not the GARCH-t and AR-GARCH, as initially expected.

In Appendix A.4, the results for $\alpha = 0.5\%$ are depicted in Table A1. Again, these suggest that

the GJR-GARCH model performs best. Similarly, as for $\alpha = 1.0\%$, both the AR-GARCH and GJR-GARCH perform well. Note that for simulation model A.2, the GARCH-t, the GARCH specification with Laplace QMLE instead of Gaussian QMLE performs best, which is somewhat surprising because it is not the originally fitted model.

Table A2 shows the results of the Hill estimator for $\alpha = 2.5\%$. We observe that in this case the GJR-GARCH performs worse than for the lower α values, while the simpler GARCH-t specification performs the best by attaining the lowest RMSE for four of the five models. The second-best model is the AR-GARCH model, which attains the lowest RMSE for its corresponding simulation and is the second-best option in two of the four remaining simulations.

Some further remarks, we observe higher RMSE values for CES than CVaR and a decrease in RMSE for an increase in α. This aligns with what we have seen in the previous simulation study and Hoga (2019). Besides, we observe that the EGARCH estimation model does not perform well for all simulation models and values of α, except for model A.5, which is the one based on the EGARCH model its structure. A possible explanation for this is that the GJR-GARCH is a simplification of the EGARCH model which still accounts for an asymmetric effect to positive and negative shocks, but may produce less noise in the estimation due to being a simpler model. Moreover, it is worth mentioning that where the GARCH-t model, with QMLE based on Laplace residuals performs well, the GARCH model using estimation based on the Gaussian residuals is never the best or second best model. This observation indicates that the Laplace distribution describes the error terms more accurately than the Gaussian one. A potential explanation is given in Xuan et al. (2017), stating that the Laplace distribution can accurately describe the fat-tail feature of financial market data. Additionally, Xuan et al. (2017) explain the good performance of the GARCH-t model for $\alpha = 2.5\%$ by stating that the larger value of α we use, the fatter tail we take for CVaR and CES estimation. The Laplace distribution focuses on capturing fat tails, thus taking a large value for α leads to better performance of the GARCH-t model.

6 Application

The results of the first application, described in Section 4.7 for the given indices for the period 01/01/1997 to 31/12/2016, are displayed below. For further reference, we call this application the full period application. The method G-k-H represents the GARCH(1,1) model with the Hill estimator, the method AG-k^*-H indicates the AR(1)-GARCH(1,1) model with the Hill estimator and AG-k^*-MR the AR(1)-GARCH(1,1) model with MR estimator. Additionally, the methods AGJR-k^*-H and AGJR-k^*-MR model the AR(1)-GJR-GARCH(1,1) model with Hill and MR estimator, while AEG-k^*-H and AEG-k^*-MR indicate the AR(1)-EGARCH(1,1) model for the same estimators. The k^* illustrates that the optimal choice of k is determined by Equation 17. For the unconditional coverage (UC), the test of Kupiec (1995) is performed while for the conditional coverage (CC), we use the test of Berkowitz et al. (2011). The quantile score indicates the performance of the CVaR forecasts, while the AL log score quantifies the forecasting accuracy of the CES. The results of the model evaluation are displayed in Table 6 for $\alpha = 0.5\%$.

Table 6: p-values of backtests (UC and CC) and quantile scores and AL log scores for different stock indices

Index	Method	UC	CC	Quantile score	AL log score
DJIA	G-k-H	0.352	0.108	0.788	-9141.403
	AG-k^*-H	0.971	0.129	0.770†	**-9269.830**†
	AG-k^*-MR	0.004***	0.002***	**0.767**	-9259.075
	AGJR-k^*-H	0.002***	0.001***	1.156	-7750.613
	AGJR-k^*-MR	0.013**	0.001***	1.079	-7823.585
	AEG-k^*-H	0.041**	1.000	0.844	-8943.505
	AEG-k^*-MR	0.041**	1.000	0.844	-8943.505
NASDAQ	G-k-H	0.110	0.091*	0.953	-8489.051
	AG-k^*-H	0.480	0.114	**0.937**†	**-8605.109**†
	AG-k^*-MR	0.193	0.001***	0.943	-8515.700
	AGJR-k^*-H	0.795	0.000***	1.092	-7851.398
	AGJR-k^*-MR	0.280	0.000***	1.092	-7688.839
	AEG-k^*-H	0.008**	0.000***	1.028	-7793.307
	AEG-k^*-MR	0.000***	0.000***	1.034	-7605.064
Nikkei	G-k-H	0.567	0.104	1.134	-7365.226
	AG-k^*-H	0.730	0.109	**1.128**	**-7402.499**
	AG-k^*-MR	0.214	0.000***	1.138	-7313.831
	AGJR-k^*-H	0.913	0.000***	1.301	-6787.419
	AGJR-k^*-MR	0.009**	0.000***	1.322	-6627.037
	AEG-k^*-H	0.913	0.000***	1.205	-6895.978
	AEG-k^*-MR	0.054*	0.000***	1.213	-6867.086
HSI	G-k-H	0.132	0.084*	0.980	-8049.843
	AG-k^*-H	0.408	0.101	**0.966**	**-8167.439**†
	AG-k^*-MR	0.152	0.000***	0.968	-8140.278
	AGJR-k^*-H	0.201	0.090*	1.048	-7774.578
	AGJR-k^*-MR	0.059*	0.000***	1.054	-7749.629
	AEG-k^*-H	0.227	0.000***	1.115	-7220.375
	AEG-k^*-MR	0.000***	0.000***	1.197	-6760.409
CAC 40	G-k-H	0.225	1.000	1.024	8324.582
	AG-k^*-H	0.917	0.134	**1.016**	**-8376.782**
	AG-k^*-MR	0.315	0.010***	1.017	-8304.801
	AGJR-k^*-H	0.180	0.000***	0.596	-2638.854
	AGJR-k^*-MR	0.000***	0.000***	0.620	-2340.806
	AEG-k^*-H	0.002***	0.000***	0.484	-2708.829
	AEG-k*-MR	0.000***	0.000***	0.511	-2454.803
DAX 30	G-k-H	0.159	0.099*	1.010	-8253.591
	AG-k^*-H	0.603	0.122	0.984†	**-8383.700**†
	AG-k^*-MR	0.018**	0.017***	**0.982**	-8327.681
	AGJR-k^*-H	0.940	0.000***	1.238	-7213.556
	AGJR-k^*-MR	0.416	0.000***	1.230	-7139.880
	AGJR-k^*-H	0.940	0.000**	1.238	-7213.556
	AGJR-k^*-MR	0.416	0.000**	1.230	-7139.880
	AEG-k^*-H	0.032**	0.016**	1.097	-7697.795
	AEG-k^*-MR	0.000***	0.000***	1.134	-7476.433

Notes. Significance of 10%, 5%, 1% marked with *, **, ***, The lowest quantile and AL log score are printed in bold and the dagger marks significant difference by DM statistic for 5%.

We find that AG-k^*-H is never rejected at the 10% level, making it the best model in general, in line with the results discussed in Hoga (2019). The performance of AG-k^*-H is satisfactory, while the AG-k^*-MR and AGJR-k^*-MR methods are not because of their poor conditional coverage. Additionally, both AR-EGARCH models are also almost always rejected at 1% level, suggesting poor performance as well. For the AGJR-k^*-H model, we find significant unconditional coverage, while the conditional coverage is rarely rejected at the 10% level. In

contrast to the simulation study, where the MR estimator performed best in terms of RMSE, we observe that the Hill estimator performs better in this application. A possible explanation is given by the additional simulations performed in Hoga (2019). By evaluating these, Hoga (2019) concludes that, if the tail of innovations follows the true Pareto distribution more accurately, then the Hill estimator performs better relatively. The innovations of the indices thus seem to closely follow the shape of the Pareto distribution in this full period application. Regarding the score functions, we obtain that AG-k^*-H gives the best CVaR and CES forecasts overall. Moreover, note that the score functions for the AR-GJR-GARCH and AR-EGARCH models are never the best and often one of the worst, especially for CAC 40. This is in contrast with findings in Nugroho et al. (2019), which concludes that the GJR-GARCH(1,1) model outperforms the GARCH(1,1) model. A potential reason for this inconsistency is the usage of the AR lag which makes our model too complex, consequently adding noise and no additional explanatory power. Another possible explanation is that the volatility of the returns of the application does not follow have an asymmetric reaction to shocks, making the models that do take into account such reactions, like the GJR-GARCH and EGARCH, perform worse.

6.1 Low- versus high-volatility period

For the second application, the low/high period application, we compare the difference in model performance in a low- and high-volatility period. We do so by performing an in-sample estimation of the parameter estimates for the same models as the full period application. Both the low- and high-volatility period and the total estimation period are summarised in Section 3.

The in-sample estimates for the entire estimation period can be found in Appendix B1. First of all, note that no AR(1) lag is significant, except for the AR-GARCH model for the DAX 30. The insignificance indicates that the AR(1) lag adds little to no explanatory power and only causes noise, similar to what we have found for the full period application. Additionally, from the parameter estimates we can identify the presence of a leverage effect for the AR-GJR-GARCH model, as the γ estimate is positive and significant for all indices, suggesting that negative shocks more heavily influence the volatility than positive shocks. The definition of γ and ψ is switched in comparison with our discussed specification in Section 4 since the R code considers a different specification of the EGARCH model. A more detailed description of the R code is given in Appendix C. For the AR-EGARCH model, we find a positive γ for all indices, indicating the presence of a significantly positive symmetry effect and a negative value for ψ, which again shows that negative shocks create more volatility than positive ones for all indices.

The period before the COVID-19 crisis is used as the low-volatility period. For the high-volatility period, we take the COVID-19 period. Tables 7a 7b show the outcomes of these two periods respectively for $\alpha = 0.5\%$. Significance for coverage is indicated by *, ** and *** for 10%, 5% and 1%. The lowest value for the quantile score and AL log score, referring to the best model is made bold. The dagger represents a significant difference between the two lowest Hill estimator quantile scores and AL log scores by the DM statistic for 5%.

Table 7: p-values of backtests (UC and CC) and quantile scores and AL log scores for different stock indices in a low- and high-volatility period

Index	Method	UC	CC	Quantile score	AL log score
NASDAQ	G-k-H	0.226	1.000	0.060	-619.755
	AG-k^*-H	0.226	1.000	**0.059**	**-632.604**
	AG-k^*-MR	0.201	0.997	0.063	-631.198
	AGJR-k^*-H	0.201	0.000***	0.080	-544.489
	AGJR-k^*-MR	0.035**	0.000***	0.089	-503.634
	AEG-k^*-H	0.654	1.000	0.063	-611.900
	AEG-k^*-MR	0.003***	0.000***	0.069	-567.325
Nikkei	G-k-H	0.237	1.000	0.060	-567.225
	AG-k^*-H	0.237	1.000	0.057	-593.121
	AG-k^*-MR	0.604	1.000	**0.051**	**-635.044**
	AGJR-k^*-H	0.237	1.000	0.064	-556.252
	AGJR-k^*-MR	0.604	0.000***	0.056	-607.516
	AEG-k^*-H	0.237	1.000	0.056	-608.806
	AEG-k^*-MR	0.237	1.000	**0.051**	-625.604
HSI	G-k-H	0.231	1.000	0.054	-630.085
	AG-k^*-H	0.231	1.000	0.052	-645.295
	AG-k^*-MR	0.188	0.996	0.051	-658.870
	AGJR-k^*-H	0.231	1.000	0.055	-626.083
	AGJR-k^*-MR	0.231	1.000	**0.049**	-664.372
	AEG-k^*-H	0.718	1.000	0.050	**-667.945**
	AEG-k^*-MR	0.633	1.000	0.051	-657.990
CAC 40	G-k-H	0.210	0.998	**0.067**	**-618.730**
	AG-k^*-H	0.004***	0.947	0.077	-555.435
	AG-k^*-MR	0.004***	0.947	0.084	-498.978
	AGJR-k^*-H	0.038***	0.000***	0.078	-556.875
	AGJR-k^*-MR	0.000***	0.000***	0.092	-453.894
	AEG-k^*-H	0.004***	0.947	0.072	-577.595
	AEG-k*-MR	0.000***	0.873	0.083	-509.734
DAX 30	G-k-H	0.003***	0.940	**0.064**	**-623.605**
	AG-k^*-H	0.003***	0.940	0.067	-601.145
	AG-k^*-MR	0.000***	0.860	0.073	-552.672
	AGJR-k^*-H	0.003***	0.940	0.082	-528.028
	AGJR-k^*-MR	0.000***	0.000***	0.093	-435.041
	AEG-k^*-H	0.003***	0.940	0.067	-579.677
	AEG-k^*-MR	0.003***	0.940	0.071	-552.072

(a) Pre-COVID-19 period: January 1, 2019 - February 28, 2020

Index	Method	UC	CC	Quantile score	AL log score
NASDAQ	G-k-H	0.661	1.000	0.128	-421.713
	AG-k^*-H	0.661	1.000	0.126	-418.917
	AG-k^*-MR	0.206	0.996	0.125	-394.996
	AGJR-k^*-H	0.224	1.000	0.108†	-471.504†
	AGJR-k^*-MR	0.206	0.996	**0.106**	**-485.357**
	AEG-k^*-H	0.004***	0.935	0.200	730.231
	AEG-k^*-MR	0.000***	0.000***	0.260	919.371
Nikkei	G-k-H	0.231	1.000	0.082	-520.588
	AG-k^*-H	0.231	1.000	0.076	-552.030
	AG-k^*-MR	0.718	1.000	**0.068**	**-595.988**
	AGJRk^*-H	0.633	0.000***	0.095	-477.240
	AGJR-k^*-MR	0.633	0.000****	0.096	-497.950
	AEG-k^*-H	0.031**	0.000***	0.092	-379.278
	AEG-k^*-MR	0.003***	0.000***	0.099	-330.965
HSI	G-k-H	0.716	1.000	0.077	-545.060
	AG-k^*-H	0.636	1.000	**0.076**	**-559.741**
	AG-k^*-MR	0.636	1.000	0.079	-551.201
	AGJR-k^*-H	0.032***	0.000***	0.085	-512.977
	AGJR-k^*-MR	0.003***	0.000***	0.106	-423.754
	AEG-k^*-H	0.000***	0.000***	0.090	-470.579
	AEG-k^*-MR	0.000***	0.000***	0.114	-354.566
CAC 40	G-k-H	0.675	1.000	**0.095**	**-528.351**
	AG-k^*-H	0.215	0.000***	0.115	-523.109
	AG-k^*-MR	0.039***	0.000***	0.127	-503.670
	AGJR-k^*-H	0.004***	0.000***	0.270	38.829
	AGJR-k^*-MR	0.004***	0.000***	0.284	208.695
	AEG-k^*-H	0.000***	0.000***	0.146	-317.449
	AEG-k^*-MR	0.000***	0.000***	0.170	-228.960
DAX 30	G-k-H	0.206	0.000***	**0.150**	**-421.091**
	AG-k^*-H	0.036	0.000***	0.153	-413.311
	AG-k^*-MR	0.004	0.000***	0.165	-375.880
	AGJR-k^*-H	0.036***	0.000***	0.279	210.931
	AGJR-k^*-MR	0.004***	0.000***	0.284	322.689
	AEG-k^*-H	0.000***	0.000***	0.194	-83.738
	AEG-k^*-MR	0.000***	0.000***	0.210	-21.010

(b) COVID-19 period: March 1, 2020 - April 30, 2021

When we zoom in on the performance of the models in the pre-COVID-19 period in Table 7a, we find that the GARCH model with both Hill and MR estimator are never rejected at the 10% level for the conditional coverage, while only rejecting the unconditional coverage for the DAX 30, making it the best model overall. This is different from what we have observed in the full period application, where the AG-k^*-H model performs the strongest, while in this case it is the second best model. From the score functions, we derive that there is no clear best model for both the CVaR and CES estimator. The best model differs for each index; nevertheless, we find that the AG-k^*-MR and AG-k^*-H models are often the best or second best. Consequently, this suggests that the models that account for an asymmetric response to positive and negative shocks do not significantly improve the forecasting performance.

Additionally, we observe that for the NASDAQ, HSI and DAX 30 the coverage values for G-k-H and AG-k^*-H are pretty close. This can be attributed to similar estimation model coefficients (Appendix Table B1). We find that the AR(1) coefficient is very small, indicating that the AR(1)-GARCH model captures the data almost the same as way as the GARCH model, therefore giving similar outcomes. Following this reasoning, we can explain the better performance of the G-k-H in comparison to AG-k^*-H for the DAX 30. Because from the parameter estimates, we can deduct that the AR(1) coefficient is very low, implying that the addition of the AR-lag does not add much explanatory power to the model. Instead, it only adds noise when forecasting, resulting in an inferior performance of the AG-k^*-H compared

to the standard GARCH specification. Furthermore, note that AR-GJR-GARCH and AR-EGARCH models almost always behave poorer than the 'simpler' AR-GARCH and GARCH models, pointing out that modelling an asymmetric reaction to positive and negative shocks does not lead to forecasting improvements. The poor performance of the asymmetric models is in line with what we have found in the full period application.

Recall that for the full period application, we find that the Hill estimator performs better than the MR estimator. However, for the low-volatility period we do not have a clear winner. Regarding conditional and unconditional coverage, the Hill estimator gives better estimates. However, judging by the quantile and AL log score, we obtain that its dependent on the index whether the Hill and MR estimator give the lowest scores. Overall, the Hill estimator seems to perform slightly better, as it more often attains the highest coverage and lowest score function value. This result is consistent with those of the full period application, shown in Table 4. Another striking observation is that for the two European indices the GARCH model performs the best, suggesting a difference in return structure compared to the NASDAQ and Nikkei, for which the AR-GARCH specifications perform better. The HSI index requires an asymmetric reaction to shocks to be captured as the AR-GJR-GARCH and AR-EGARCH are the most accurate.

Table 7b provides the results for the COVID-19 period. In summary, we find that the GARCH and AR-GARCH models perform better than the AR-GJR-GARCH and AR-EGARCH, indicating that modelling for an asymmetric response does not improve forecasting ability. A simpler model thus seems to work best in high-volatility periods, as is the case in low-volatility periods. Compared to the low-volatility period, the high-volatility period makes it more challenging for models to produce accurate forecasts, illustrated by a lower conditional coverage and higher score values on average.

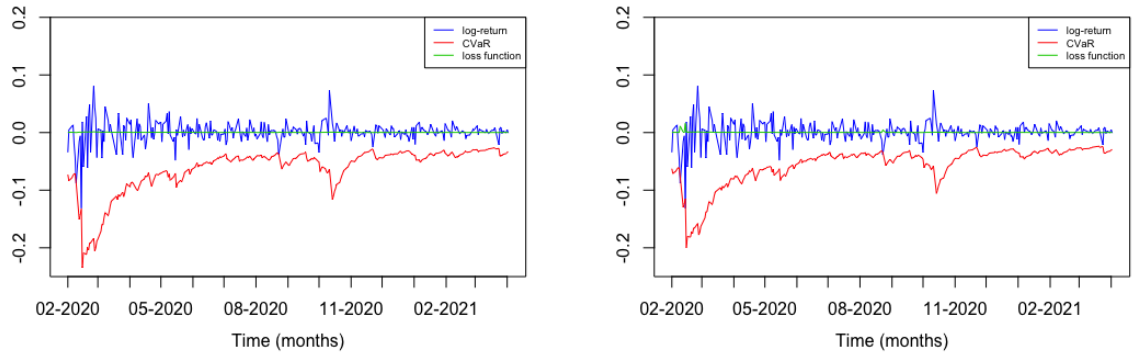
For CAC 40 and DAX 30, the two European indices, the conditional and unconditional coverage are almost always rejected at 1% level, making all of the models inappropriate to implement as a forecasting model. The significance of the unconditional coverage suggests that there are too few VaR violations, indicating an overly conservative VaR (Campbell, 2005). Additionally, note that for the NASDAQ, the AR-EGARCH shows high positive score values, indicating very inaccurate CES forecasts. While for the CAC 40 and DAX 30 we observe similar behaviour, however now for the AR-GJRGARCH model. Again, this can possibly be explained by the fact that the models are more complex, potentially producing noisier forecasts than a relatively easy model such as the GARCH model.

6.2 High-volatility period

Now, we zoom in on the COVID-19 situation in Europe to clarify the poor model performance, especially the significant conditional coverage, for the European indices CAC 40 (France) and DAX 30 (Germany), and compare it with the North American and Asian market. First of all, conditional coverage is a combination of unconditional coverage and the independence property. The independence property places strong restrictions on ways in which violations of the VaR may occur. An important one, which might not hold in our case, is the independence of a current violation with the previous ones. If previous violations influence a future VaR violation, which is

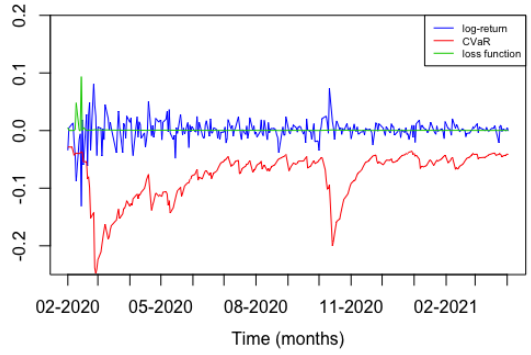
the case for a clustering of violations, then the reported VaR does not accurately reflect the loss that can be expected. As a consequence, market risk capital requirements can be underfunded for an extended period in times of increased volatility, such as the COVID-19 crisis. Moreover, as we evaluate for $\alpha = 0.5\%$ and a sample of approximately 300 observations, we expect to find 0 to 3 violations. The low number of violations causes difficulty in determining the conditional coverage and makes the outcomes unreliable. This potentially explains the rather surprising values for the conditional coverage for the low-volatility period in general.

To determine if this is actually the case, we plot the log-returns in combination with the estimates of the CVaR and the loss function in Equation (22), which evaluates the difference between the CVaR forecasts and the actual values. Figure 3 shows the resulting figures for the four different models for the Hill estimator for the CAC 40. We only consider the performance of the Hill estimator because by the simulation study and first application, we obtain that the Hill estimator often performs the best, making its poor performance in the COVID-19 period more surprising and more essential to understand. Figure 3a shows the results of the G-k-H model, Figure 3b the outcomes of the AG-k^*-H model, Figure 3c the ones of model AGJR-k^*-H and Figure 3d displays the results of model AEG-k^*-H, all for $\alpha = 0.5\%$.

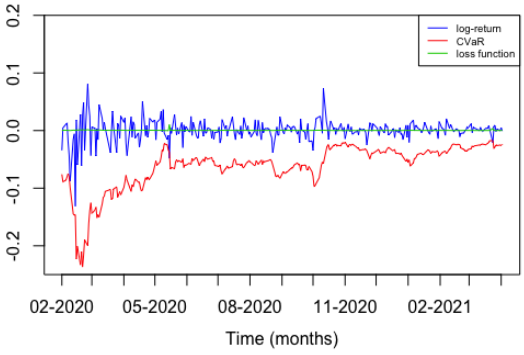


(a) GARCH model with Hill estimator

(b) AR-GARCH model with Hill estimator



(c) AR-GJRGARCH model with Hill estimator



(d) AR-EGARCH model with Hill estimator

Figure 3: Log-returns, CVaR estimates and loss function for CAC 40 for COVID-19 period March 1, 2020 - April 30, 2021

Using Figures 3, 4, 5, and Appendix Figures B1 and B2, we find evidence that indeed the AR-GJR-GARCH and AR-EGARCH model often perform inaccurately, with the loss function as an indicator for inaccurate CVaR forecasts. A potential cause for this are the clusters of high and low log-returns. Moreover, we find evidence that the conditional coverage does not perform well due to violations on consecutive days, which suggests clustering that complicate the determination of the conditional coverage. Especially, the beginning of the COVID-19 period is not correctly captured in the CVaR estimates because of the series of unexpected news items about the severeness of COVID-19.

To begin with, note that from Figure 3, we can identify the start of the COVID-19 period in February 2020 with the occurrence of spikes indicating high and low log-returns. Additionally, we observe a second spike of high CVaR log-returns in October 2020, which characterises the moment that most European countries went into lockdown. The lockdown was again an unexpected shock that caused the volatility to rise, however all models capture this shock accurately in their forecasts this time. For the GARCH model, we find only one violation, namely around the beginning of February 2020. We can attribute this to the rise of the COVID-19 virus outside of China and the restrictions on travellers leaving China (Smith & Goldberg, 2020), which is received as an unexpected shock in the CAC 40 index and is consequently difficult to capture. Because there is only one violation, we can not test independence. We hence obtain the conditional coverage value of 1.000. The conditional coverage is thus transformed into an unconditional coverage which gives a value of 1.000 when there is only one violation.

When performing the same analysis for the AR-GARCH model, we observe two violations. One again at the beginning of February 2020 and represents the same event as the one for the GARCH model. The other one is a few days later and is caused by the news of the first COVID-19 death in France and Europe, stated in the Financial Times (Mallet & Peel, 2020). This news caused an unanticipated shock for France and possibly whole of Europe. Both violations are very close to each other, verifying the clustering of the violations and making it impossible for them to be independent, resulting in a conditional coverage of 0.000. The conditional coverage can in this situation again be interpreted as more of an unconditional coverage measure.

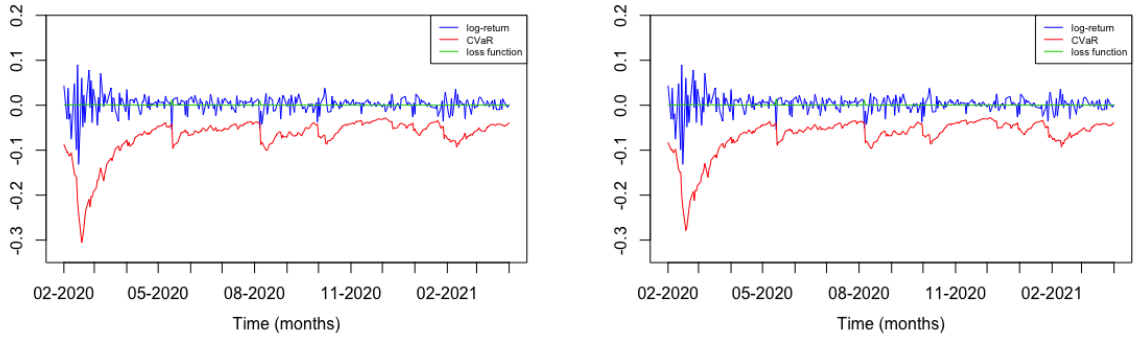
For the AR-GJR-GARCH model, Figure 3c shows many consecutive violations and a high peak in the loss function for almost the whole month of February. This suggests that the model performs poorly, corresponding with the quantile score that we obtain in Table 7. Again, we observe violations at similar moments as for the GARCH and AR-GARCH models and thus have the same news items as reason for the violations. We can identify an additional significant violation for the AR-GJR-GARCH model, which is near the end of February. This one can be attributed to the second death in France, stated in the Financial Times news section (Smith & Goldberg, 2020).

Lastly, for the AR-EGARCH model, we again find two violations, similar to the AR-GARCH model, caused by the same events. Furthermore, the violation in May 2020 can also be identified for the American and Asian indices. We discuss the cause of this violation when evaluating the Asian market, as especially the volatility spike for these indices is large.

From Figure B1, we find similar results for DAX 30 as for CAC 40. We observe the same consecutive series of violations, indicating that the same shocks have caused them. Hence the

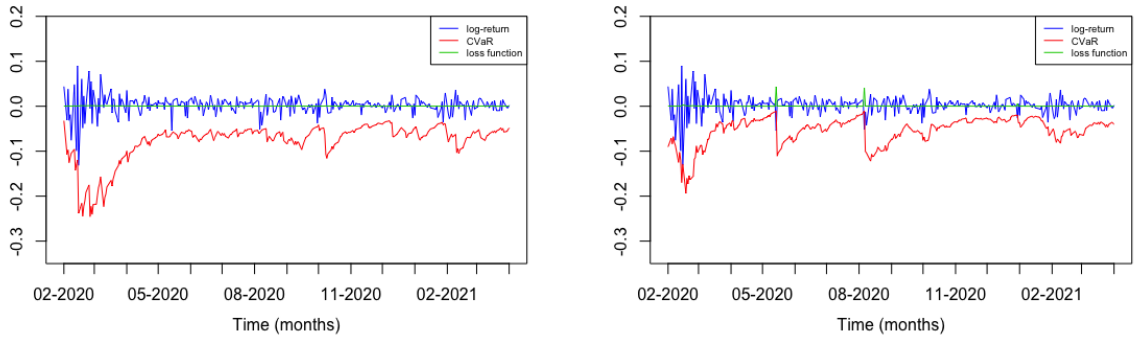
news items surrounding COVID-19 are unexpected and challenging to capture in any model in both France and Germany, and by continuing this reasoning probably in the whole of Europe or the world. The series of violations cause the conditional coverage to attain a value of 0.000, similarly to the CAC 40, suggesting that the conditional coverage does not seem to be a reliable measure to base conclusions on when evaluating the performance of the models.

In comparison with the results for the North American index (NASDAQ), stated in Figure 4, we do not identify consecutive violations like we have found for both European indices. However, we observe a higher number of violations, causing the conditional coverage to act accordingly and give more reliable results.



(a) GARCH model with Hill estimator

(b) AR-GARCH model with Hill estimator



(c) AR-GJRGARCH model with Hill estimator

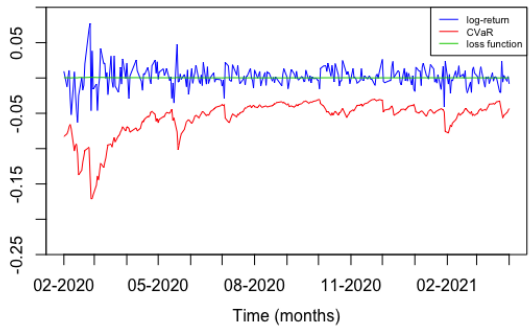
(d) AR-EGARCH model with Hill estimator

Figure 4: Log-returns, CVaR estimates and loss function for NASDAQ for COVID-19 period March 1, 2020 - April 30, 2021

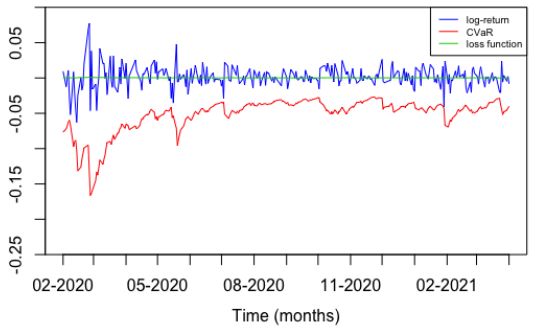
The results suggest that the effect of the news article of AJMC Staff (2021) in February has a less extreme but still noticeable effect in comparison with the European indices. On the other hand, an event in August 2020 is especially important to the technology industry in the US because we only locate the violation for the NASDAQ. This event is described in Bloomberg as 'more upside panic', causing a rise in demand for call options for tech stocks, which feeds into gains in the stocks and consequently also in the NASDAQ Peterseil, Yakob and Greifeld, Katherine and Barnet, Jan-Patrick (2020). Moreover, we observe quite some violations for the AR-EGARCH model which is confirmed by the high spikes in the loss function, and explains

the model its poor performance.

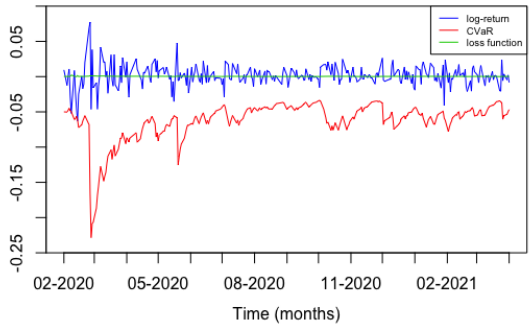
Lastly, we examine the Asian market in a high-volatility period by the Nikkei (Japan) and HSI (Hong Kong) index, shown in Figures 5 and B2.



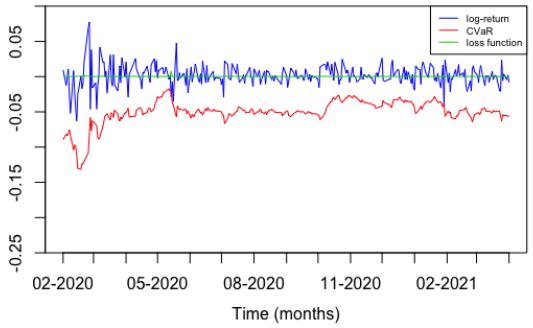
(a) GARCH model with Hill estimator



(b) AR-GARCH model with Hill estimator



(c) AR-GJRGARCH model with Hill estimator



(d) AR-EGARCH model with Hill estimator

Figure 5: Log-returns, CVaR estimates and loss function for the Nikkei for COVID-19 period March 1, 2020 - April 30, 2021

For the Nikkei, we can not identify any violations for the GARCH model, while we identify only one for the AR-GARCH model. This observation explains the conditional coverage of 1.000 for both, since 0 or 1 violation is always independent. For the AR-GJR-GARCH and AR-EGARCH model, the violations are again consecutive, giving a conditional coverage of 0.000. The violations of the AR-GJR-GARCH model in February 2020 can again be attributed to the beginning of the COVID-19 crisis. Those of the AR-EGARCH are in May 2020. Note the presence of an additional violation in February 2021 for the AR-EGARCH model. This violation seems to be specific for the Asian markets, as also the HSI shows violations around this time. A potential cause for the unexpected returns is the military coup in Myanmar (Smith et al., 2021).

For the HSI, we find a similar violation pattern as for the Nikkei for the AR-GJR-GARCH and AR-EGARCH model. Note that the violations in May 2020 cause an extremely low log-return for the HSI and an violation for all models, indicating that the event that caused it was of great importance to Hong Kong. We thus expect that the violations in May 2020 for the HSI, Nikkei and NASDAQ were caused by protests in Hong Kong (Reuters staff, 2020). These were

particularly important for the HSI and the Nikkei, however also influence the NASDAQ, as the US and especially its technology industry is strongly connected to Asia and with that also the protests and its outcome.

7 Conclusion and discussion

7.1 Conclusion

For the first research question: *Does the usage of ARMA-GARCH instead of GARCH models lead to an improvement in forecasting performance of the CES and CVaR risk measures?*, we conclude that the use of the ARMA-GARCH specification gives better forecasts in practice, especially when using a data-dependent choice of k. Mind that we should still consider the GJR-GARCH and GARCH specifications as, according to the simulations, they both still provide accurate estimates and forecasts.

The conclusion is based on the results of the simulation studies and the full period application. The first simulation study compares the performance of three different simulation models for CVaR and CES estimates for both the Hill and MR estimator. By this study, we conclude that CVaR estimates are more accurately estimated than those of CES because the RMSE values are lower for CVaR than CES for all models. Additionally, we find that the confidence intervals based on the self-normalised approximation are more accurate than those based on the normal approximation due to their higher coverage.

Moreover, from the second simulation study, we derive that the best model is almost always the same for the CVaR and CES forecasts. We thus do not have to take the choice of risk measure into account when deciding on the model. Additionally, for $\alpha = 0.5\%$ and 1.0%, we conclude that the GJR-GARCH model performs best, while for $\alpha = 2.5\%$, the simpler GARCH-t model outperforms the other models. This shows that the best model depends on the size of the tails. When forecasting for a fatter tail, we are better of with a simpler model, while for a thinner tail, a more complex model which incorporates an asymmetric response to volatility shocks is the better option.

The results of the first application show that both GARCH and AR-GARCH can produce extreme CVaR forecasts for the log-returns of the six indices with correct conditional and unconditional coverage. Furthermore, we conclude that the AR-GARCH model with a data-dependent choice of k in comparison with the simpler GARCH model with a fixed choice of k, improves the forecasts of CVaR and in particular of CES.

The second research question, *Does the forecasting performance of the risk measures differ in periods of low and high volatility?* is answered by the second application in Section 6. We conclude that the best forecasting model stays the same for three of the five cases, giving no clear indication that a different model would give a better forecasting performance. This goes against our hypothesis of better performance by more complex and asymmetrical models. Additionally, we can conclude that for the European indices, the GARCH model gives the most accurate forecasts. Furthermore, we notice that in a high-volatility period, the conditional coverage behaves more like an unconditional coverage, giving an indication for 0 or 1 close violations when the conditional coverage is 1.000 and suggesting a clustered series of violations when the

unconditional coverage is 0.000.

7.2 Discussion

In this subsection, we discuss the causes of some striking simulation and application results and outcomes which deviate from previous literature.

7.2.1 Simulation

Some of the differences in Table 4 compared to Hoga (2019), can be attributed to Hoga (2019) using different simulations in the supplementary R code than described in the paper. In this research, we follow the approach of the paper and not of the code. As the research does not specify whether the results are based on the code or the specification in the paper, this can be a potential explanation of the difference in simulation results. Another aspect that causes the difference between Table 4 and Hoga (2019) is the random simulation of the data. For this, we use a random t-distribution to model the innovation terms, which are consequently different for each simulation and for each observation. The simulated innovation terms ultimately determine the model parameters and thus give slightly different results in absolute terms for each separate simulation.

For the second simulation, our results are somewhat different from what we expected. In contrast to what we predicted, we find that the model that corresponds to a certain simulation is not always the best one. For example, for $\alpha = 1.0\%$, the GJR-GARCH works best instead of the GARCH-t. Furthermore, it is striking that the GJR-GARCH model performs so well, while the EGARCH estimation model is rarely the best. We expected them to have a similar performance, because the GJR-GARCH model is a simplified version of the EGARCH model since both models capture the asymmetric response to shocks. A possible clarification for what we obtain is that the EGARCH is a more complex model than the GJR-GARCH and might be too complex, giving the additional parameters no explanatory power. To investigate the overall strong performance of the GJR-GARCH model in simulations, we would need to perform some additional simulations to get a completer understanding of the model. This is left for further research.

7.2.2 Application

For the full period application, we find results in line with Hoga (2019). We now find that the 'simpler' GARCH and AR-GARCH still perform the best and not the AR-GJR-GARCH or EGARCH. A potential reason for this has already been discussed in the results, namely that the AR-lag does not have any explanatory power but only adds noise. An additional potential reason for the poor performance of the EGARCH model, which possesses some theoretical advantages over the simpler GARCH models, is that in practice the likelihood maximiser is highly dependent on the choice of starting values (Hartz et al., 2006).

When considering the low/high period application, we find results that are not in line with our initial hypothesis of 'easier' models performing better in low-volatility periods and more extensive models are needed to capture the different shocks in high volatility periods. We can

partly explain the unexpected results in Table 7 by the backtesting procedure that we apply for the CVaR and CES, which is in accordance with Hoga (2019). This procedure assumes no structural breaks in the data, yet financial markets are continually evolving with new technologies, assets, markets and institutions that affect the statistical properties of market prices and consequently the properties of returns as well (Danielsson, 2011). As we are investigating such structural breaks in this case, this can potentially influence the return property and consequently the accuracy of the estimates of the GARCH models. For further research, we could model this structural break. Another improvement to our research can be made by diving deeper into the conditional coverage in the high-volatility period. To solve the unreliable conditional coverage, we should take a longer sample, use a higher α or uncluster the violations. All of these options could be more thoroughly investigated in further research.

Lastly, note that for the applications we have used indices that are focused on particular countries, however, the economical situation in these countries can also be influenced by political decisions and reforms, hereby changing what the index represents, making the application results for the low/high period not completely representative for the corresponding continent. Take for example the HSI, due to the unrest in Hong Kong and its political changes in the last year, it might be that the index does not capture the same changes in volatility and return as the Asian continent anymore. For the CAC 40 and DAX 30, we only look at the economic situation in France and Germany. One could argue whether these two countries are depicting a complete picture of Europe. Moreover, the NASDAQ mostly consists of technology companies, which is only a small part of the North American economy and could potentially give a distorted view of the US' economy and its reaction to certain shocks. To tackle this, we could conduct further research with other indices that are more representative for a whole continent. For example, by taking the S&P500 to examine the North American market and STOXX Europe 600 for Europe.

References

Agarwalla, S. K., Varma, J. R., & Virmani, V. (2021). The impact of COVID-19 on tail risk: Evidence from Nifty index options. *Economics Letters*, 109878.

AJMC Staff. (2021). *A Timeline of COVID-19 Developments in 2020.* Retrieved 28-06-2021, from `https://www.ajmc.com/view/a-timeline-of-covid19-developments-in-2020`

Azzalini, A., & Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *65*(2), 367–389.

Berkes, I., Horváth, L., et al. (2004). The efficiency of the estimators of the parameters in GARCH processes. *Annals of statistics*, *32*(2), 633–655.

Berkowitz, J., Christoffersen, P., & Pelletier, D. (2011). Evaluating Value-at-Risk models with desk-level data. *Management Science*, *57*(12), 2213–2227.

Campbell, S. D. (2005). A review of backtesting and backtesting procedures. *The Journal of Risk*, *9*(2), 1.

Chan, N. H., Deng, S.-J., Peng, L., & Xia, Z. (2007). Interval estimation of Value-at-Risk based on GARCH models with heavy-tailed innovations. *Journal of Econometrics*, *137*(2), 556–576.

Danielsson, J. (2011). *Financial risk forecasting: the theory and practice of forecasting market risk with implementation in R and Matlab* (Vol. 588). John Wiley & Sons.

Danielsson, J., Ergun, L. M., de Haan, L., & de Vries, C. G. (2016). Tail index estimation: Quantile driven threshold selection. *Available at SSRN 2717478*.

Danielsson, J., Jansen, D. W., & De Vries, C. G. (1996). The method of moments ratio estimator for the tail shape parameter. *Communications in Statistics-Theory and Methods*, *25*(4), 711–720.

Diebold, F. X. (2015). Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of Diebold–Mariano tests. *Journal of Business & Economic Statistics*, *33*(1), 1–1.

Fissler, T., Ziegel, J. F., & Gneiting, T. (2016). Expected shortfall is jointly elicitable with Value at Risk-Implications for backtesting. *Risk Magazine*.

Francq, C., Zakoian, J.-M., et al. (2004). Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli*, *10*(4), 605–637.

Ghysels, E., Santa-Clara, P., & Valkanov, R. (2005). There is a risk-return trade-off after all. *Journal of Financial Economics*, *76*(3), 509–548.

Giacomini, R., & Komunjer, I. (2005). Evaluation and Combination of Conditional Quantile Forecasts. *Journal of Business & Economic Statistics*, *23*(4), 416–431.

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

Gopinathan, K. (1988). A general formula for computing the coefficients of the correlation connecting global solar radiation to sunshine duration. *Solar energy*, *41*(6), 499–502.

Hafner, C. M., & Linton, O. (2017). An almost closed form estimator for the EGARCH model. *Econometric Theory*, *33*(4), 1013–1038.

Hartz, C., Mittnik, S., & Paolella, M. S. (2006). *Accurate Value-at-Risk forecast with the (good old) normal-GARCH model* (Tech. Rep.). Center for Financial Studies (CFS).

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

Hill, J. B. (2015). Expected shortfall estimation and Gaussian inference for infinite variance time series. *Journal of Financial Econometrics*, *13*(1), 1–44.

Hoga, Y. (2019). Confidence intervals for conditional tail risk measures in ARMA–GARCH models. *Journal of Business & Economic Statistics*, *37*(4), 613–624.

Horvath, R., & Šopov, B. (2016). GARCH models, tail indexes and error distributions: An empirical investigation. *North American Journal of Economics and Finance*, *37*, 1–15.

Huisman, R., Koedijk, K. G., Kool, C. J. M., & Palm, F. (2001). Tail-index estimates in small samples. *Journal of Business & Economic Statistics*, *19*(2), 208–216.

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

Mallet, V., & Peel, M. (2020). *France reports first coronavirus death outside Asia.* Retrieved 25-06-2021, from `https://www.ft.com/content/e771a77a-4fe9-11ea-8841-482eed0038b1`

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the Econometric Society*, *59*, 347–370.

Nugroho, D., Kurniawati, D., Panjaitan, L., Kholil, Z., Susanto, B., & Sasongko, L. (2019). Empirical performance of GARCH, GARCH-M, GJR-GARCH and log-GARCH models for returns volatility. In *Journal of physics: Conference series* (Vol. 1307, p. 012003).

Omari, C., Mundia, S., Ngina, I., et al. (2020). Forecasting Value-at-Risk of Financial Markets under the Global Pandemic of COVID-19 Using Conditional Extreme Value Theory. *Journal of Mathematical Finance*, *10*(04), 569.

Paulauskas, V., & Vaičiulis, M. (2017). A class of new tail index estimators. *Annals of the Institute of Statistical Mathematics*, *69*(2), 461–487.

Peterseil, Yakob and Greifeld, Katherine and Barnet, Jan-Patrick. (2020). *Tech Traders Say Options Hedging Is Firing Up Rally in Nasdaq.* Retrieved 04-07-2021, from `https://www.bloomberg.com/news/articles/2020-09-01/nasdaq-volatility-twist-prompts-theories-on-storm-in-tech-stocks`

Reuters staff. (2020). *Timeline: key dates in Hong Kong's anti-government protests.* Retrieved 28-06-2021, from `https://www.reuters.com/article/us-hongkong-protests-timeline/timeline-key-dates-in-hong-kongs-anti-government-protests-idUSKBN236080`

Shao, X. (2010). A self-normalized approach to confidence interval construction in time series. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *72*(3), 343–366.

Smith, G., Creery, J., & Goldberg, E. (2021). *FirstFT: Today's top stories.* Retrieved 28-06-2021, from `https://www.ft.com/content/8cb6372e-c70b-4b25-8f11-8db55e03653d`

Smith, G., & Goldberg, E. (2020). *FirstFT: Today's top stories.* Retrieved 25-06-2021, from `https://www.ft.com/content/d570e17e-57d9-11ea-a528-dd0f971febbc`

Spierdijk, L. (2016). Confidence intervals for ARMA–GARCH Value-t-Risk: The case of heavy tails and skewness. *Computational Statistics & Data Analysis*, *100*(C), 545–559.

Taylor, J. W. (2019). Forecasting value at risk and expected shortfall using a semiparametric approach based on the asymmetric Laplace distribution. *Journal of Business & Economic Statistics*, *37*(1), 121–133.

Wagner, N., & Marsh, T. A. (2004). Tail index estimation in small samples Simulation results for independent and ARCH-type financial return models. *Statistical Papers*, *45*(4), 545–561.

Weissman, I. (1978). Estimation of parameters and large quantiles based on the k largest observations. *Journal of the American Statistical Association*, *73*(364), 812–815.

Xuan, H., Song, L., Amin, M., & Shi, Y. (2017). Quasi-Maximum Likelihood estimator of Laplace (1, 1) for GARCH models. *Open Mathematics*, *15*(1), 1539–1548.

A Appendix

A.1 Simulation models GARCH

We use the following three models for our simulations, which we run in R version 4.0.5.

$$
\begin{aligned}
&X_i = \varepsilon_i \quad \text{where} \quad \varepsilon_i = \sigma_i U_i, \quad U_i \sim st_3(5), \\
&\sigma_i^2 = 0.95 \cdot \frac{20^2}{252} + 0.15 \cdot \varepsilon_{i-1}^2 + 0.8 \cdot \sigma_{i-1}^2, \qquad \text{(A.1)} \\
&X_i = \varepsilon_i \quad \text{where} \quad \varepsilon_i = \sigma_i U_i, \quad U_i \sim st_{4.2}(0), \\
&\sigma_i^2 = 3.2 \cdot 10^{-6} + 0.0349 \cdot \varepsilon_{i-1}^2 + 0.9373 \cdot \sigma_{i-1}^2 \qquad \text{(A.2)} \\
&X_i = 0.2714 X_{i-1} + \varepsilon_i, \quad \text{where} \quad \varepsilon_i = \sigma_i U_i, \\
&\sigma_i^2 = 3.4 \cdot 10^{-6} + 0.1407 \cdot \varepsilon_{i-1}^2 + 0.7914 \cdot \sigma_{i-1}^2, \\
&U_i \sim st_{5.3}(0.8531), \qquad \text{(A.3)}
\end{aligned}
$$

where $st_\nu(\xi)$ is the skewed Student's t distribution with ν degrees of freedom and ξ is the skewness parameter (Azzalini & Capitanio, 2003). If $\xi = 0$, it reduces to the standard t distribution. Furthermore, Equations (A.1) and (A.2) both show an GARCH(1,1) model, while Equation (A.3) represents an AR(1)-GARCH(1,1) model. We standardise the skewed-t values by dividing them by their standard deviation to obtain the required Var(U) = 1.

A.2 Simulation models GJR-GARCH and EGARCH

In order to assess the quality of the GJR-GARCH and EGARCH, we add the following models to simulation models already given in Section A.1.

$$
\begin{aligned}
&X_i = \varepsilon_i \quad \text{where} \quad \varepsilon_i = \sigma_i U_i, \quad U_i \sim st_4(0), \\
&\sigma_i^2 = 0.06 + 0.38\varepsilon_{i-j}^2 + 0.4\sigma_{i-j}^2 - 0.31\varepsilon_{i-j}^2 \mathrm{I}_{i-j}, \qquad \text{(A.4)} \\
&X_i = \varepsilon_i \quad \text{where} \quad \varepsilon_i = \sigma_i U_i, \quad U_i \sim st_{3.5}(2), \\
&\ln(\sigma_i^2) = -0.03 + 0.9 \ln(\sigma_{i-j}^2) + 0.5 \left[\frac{|\varepsilon_{i-j}|}{\sigma_{i-j}} - \mathbb{E}\left\{ \frac{|\varepsilon_{i-j}|}{\sigma_{i-j}} \right\} \right] - 0.1 \left(\frac{\varepsilon_{i-j}}{\sigma_{i-j}} \right) \qquad \text{(A.5)}
\end{aligned}
$$

where $st_\nu(\xi)$ is the skewed Student's t distribution with ν degrees of freedom, ξ is the skewness parameter (Azzalini & Capitanio, 2003) and I_{i-j} is the indicator function as defined in Equation (6). Equation (A.4) is a GJR-GARCH(1,1) model and is performed following Nugroho et al. (2019) while Equation (A.5) uses the simulation values proposed in Hafner & Linton (2017), making it a simulation of an EGARCH(1,1) model. Again, we standardise the skewed-t values similarly as in the simulation in Appendix A.1.

A.3 Additional simulation figure GARCH versus AR-GARCH

The figure shows the coverage probability, interval length, bias and RMSE in relation with $k \in \{50, 200\}$. We consider the left-tail 1% CES for both $I_{na}^{0.95}$ and $I_{sn}^{0.95}$. The results using the Hill estimator are depicted left, while those using the MR estimator are shown right in the

figure. Figure 2 shows the results for simulation model (A.2).

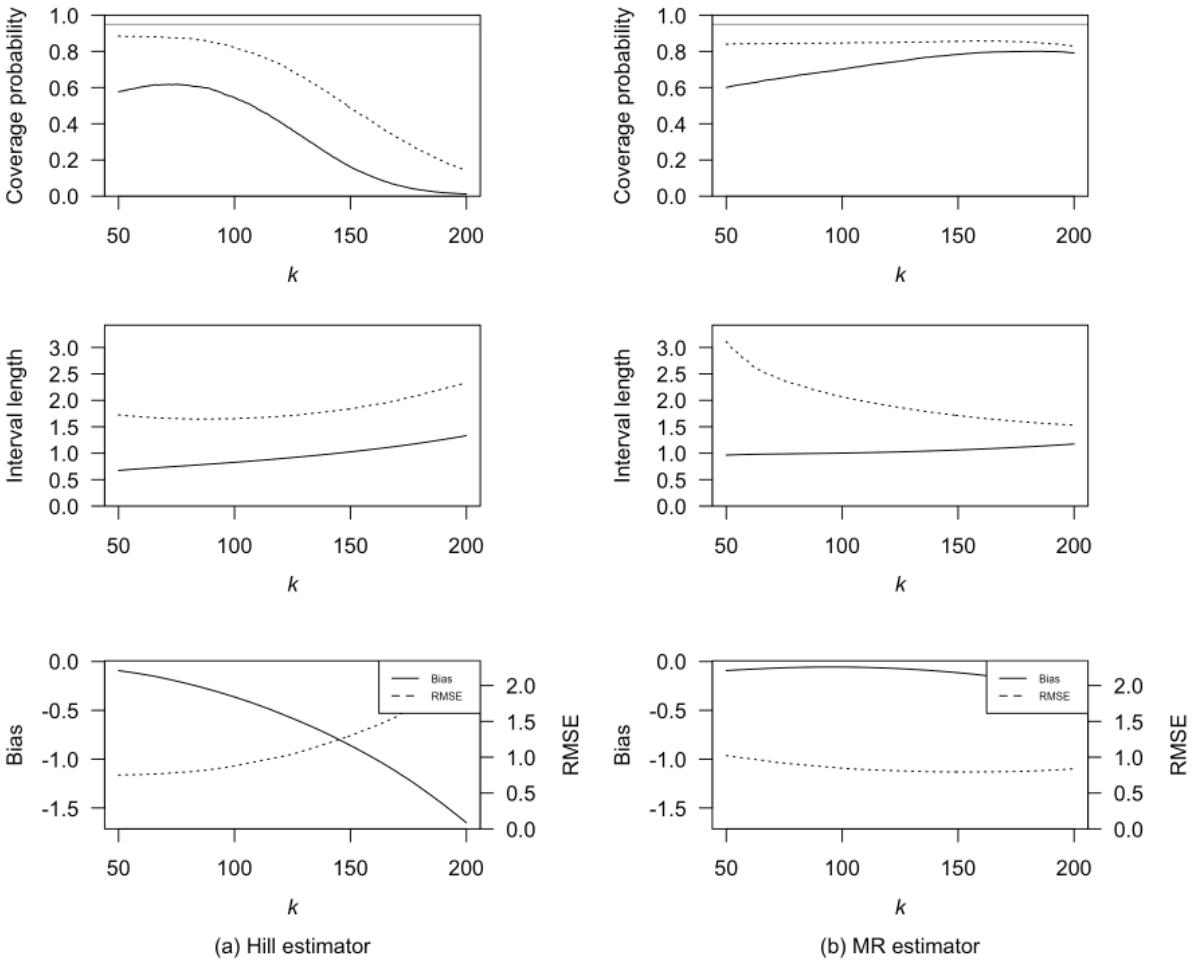


Figure A1: Characteristics as a function of k for left-tail 1% CES for $I_{na}^{0.95}$ (solid) and $I_{sn}^{0,95}$ (dotted) for model (A.1). The horizontal line in the top plots indicates nominal coverage of 95%.

A.4 Additional tables GARCH versus GJR-GARCH and EGARCH

In this section, we give the additional performance matrices corresponding to the second simulation study discussed in Section 5.2 in which we assess the forecasting performance of the GARCH, AR-GARCH, GJR-GARCH and EGARCH model by their RMSE. The top bar of the tables state the estimation models, while the left column depicts the simulation model evaluated. In bold are the values that attain the lowest RMSE, indicating the best forecasting model, and the underlined values point out the second best forecasting model for the specific simulation.

Table A1: Performance matrix for different simulations and models for $\alpha = 0.5\%$ and Hill estimator

		GARCH-t	GARCH	ARGARCH	GJRGARCH	EGARCH
(A.1)	CVaR	3.621	1.185	1.319	**0.671**	10.549
	CES	4.241	2.241	1.864	**0.830**	9.687
(A.2)	CVaR	**4.489**	4.953	5.249	5.074	12.927
	CES	**12.283**	12.731	12.681	12.786	17.452
(A.3)	CVaR	3.611	3.277	2.151	**2.097**	11.758
	CES	5.883	5.464	4.578	**4.533**	14.702
(A.4)	CVaR	**0.239**	0.241	0.243	0.246	0.599
	CES	0.523	0.522	0.526	**0.521**	0.779
(A.5)	CVaR	0.211	0.749	0.665	0.200	**0.194**
	CES	0.343	1.418	1.244	0.324	**0.320**

Table A2: Performance matrix for different simulations and models for $\alpha = 2.5\%$ and Hill estimator

		GARCH-t	GARCH	ARGARCH	GJRGARCH	EGARCH
A.1	CVaR	**0.560**	0.699	0.860	0.635	.320
	CES	**0.702**	0.910	1.122	0.775	7.647
A.2	CVaR	**1.717**	2.449	2.094	2.137	8.625
	CES	**3.888**	4.905	4.225	4.400	11.685
A.3	CVaR	2.451	2.510	**1.094**	1.117	8.267
	CES	3.086	3.165	**1.786**	1.797	10.213
A.4	CVaR	**0.126**	0.131	0.129	0.134	0.483
	CES	**0.216**	0.226	0.219	0.230	0.603
A.5	CVaR	0.128	0.374	0.289	0.555	**0.115**
	CES	**0.180**	0.684	0.549	0.690	0.184

B Appendix

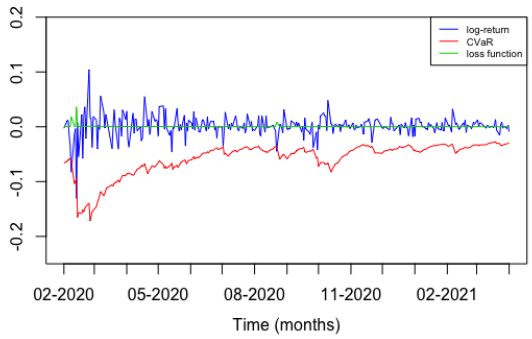
B.1 Additional tables application

Table B1: Parameter estimates of in-sample fit for the full period

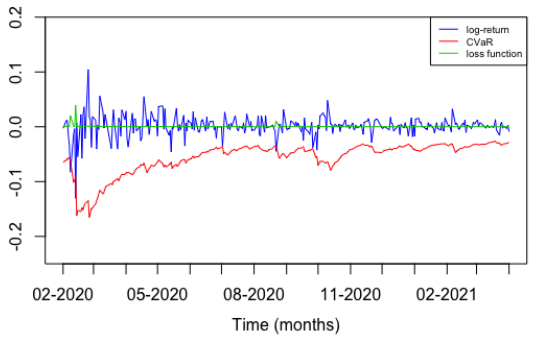
	ϕ	ω	ψ	β	γ
NASDAQ					
GARCH	-	0.000***	0.148***	0.807***	-
		(0.000)	(0.022)	(0.026)	
AR-GARCH	-0.027	0.000***	0.147***	0.808***	-
	(0.022)	(0.000)	(0.022)	(0.026)	
AR-GJRGARCH	-0.013	0.000***	0.000	0.834***	0.276***
	(0.020)	(0.000)	(0.005)	(0.012)	(0.031)
AR-EGARCH	-0.012	-0.423***	-0.201***	0.953***	0.165***
	(0.021)	(0.018)	(0.018)	(0.002)	(0.021)
Nikkei 225					
GARCH	-	0.000***	0.139***	0.820***	-
		(0.000)	(0.025)	(0.032)	
AR-GARCH	-0.013	0.000***	0.139***	0.820***	-
	(0.023)	(0.000)	(0.026)	(0.033)	
AR-GJRGARCH	0.002	0.000***	0.027***	0.809***	0.246***
	(0.021)	(0.000)	(0.006)	(0.014)	(0.035)
AR-EGARCH	0.001	-0.545***	-0.163***	0.938***	0.222***
	(0.020)	(0.048)	(0.019)	(0.938)	(0.025)
HSI					
GARCH	-	0.000**	0.051***	0.932***	-
		(0.000)	(0.010)	(0.014)	
AR-GARCH	0.025	0.000**	0.052***	0.931***	-
	(0.020)	(0.000)	(0.010)	(0.015)	
AR-GJRGARCH	0.008	0.000	0.004	0.937***	0.080***
	(0.019)	(0.000)	(0.009)	(0.011)	(0.020)
AR-EGARCH	0.008	-0.209***	-0.076***	0.977***	0.097***
	(0.019)	(0.001)	(0.011)	(0.000)	(0.0067)
CAC 40					
GARCH	-	0.000***	0.125***	0.854***	-
		(0.000)	(0.029)	(0.032)	
AR-GARCH	-0.011	0.000***	0.124***	0.854***	-
	(0.022)	(0.000)	(0.029)	(0.032)	
AR-GJRGARCH	-0.005	0.000	0.000	0.875***	0.233***
	(0.019)	(0.000)	(0.015)	(0.033)	(0.067)
AR-EGARCH	0.004	-0.214***	-0.178***	0.976***	0.119***
	(0.018)	(0.002)	(0.012)	(0.000)	(0.010)
DAX 30					
GARCH	-	0.000***	0.091***	0.890***	-
		(0.000)	(0.019)	(0.021)	
AR-GARCH	0.006***	0.000***	0.092***	0.889***	-
	(0.021)	(0.000)	(0.019)	(0.021)	
AR-GJRGARCH	0.010	0.000	0.000	0.895***	0.188
	(0.019)	(0.000)	(0.058)	(0.040)	(0.142)
AR-EGARCH	0.013	-0.239	-0.159***	0.973***	0.136***
	(0.017)	(0.013)	(0.016)	(0.001)	(0.011)

Notes. t-statistics are in parentheses; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

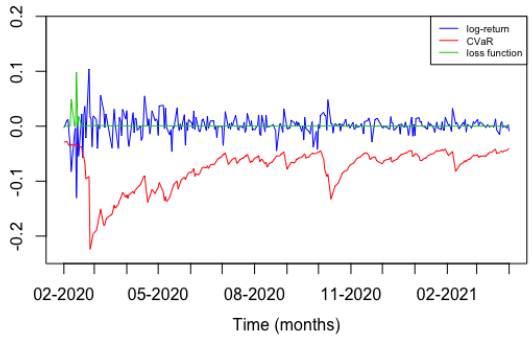
Figure B1a shows the results of the G-k-H model, Figure B1b the outcomes of the AG-k^*-H model, Figure B1c the ones of model AGJR-k^*-H and Figure B1d displays the results of model AEG-k^*-H, all for the DAX 30.



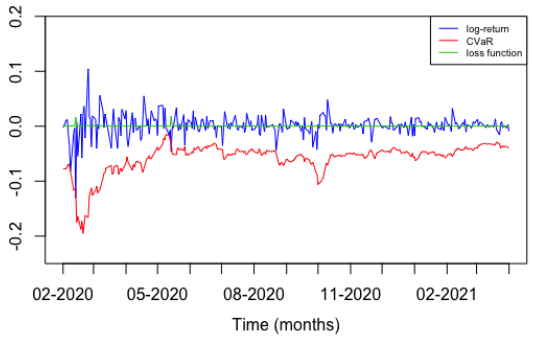
(a) GARCH model with Hill estimator



(b) AR-GARCH model with Hill estimator



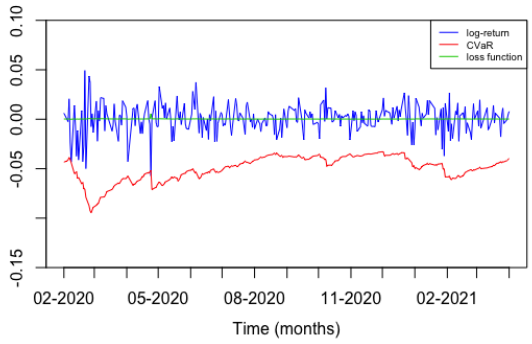
(c) AR-GJRGARCH model with Hill estimator



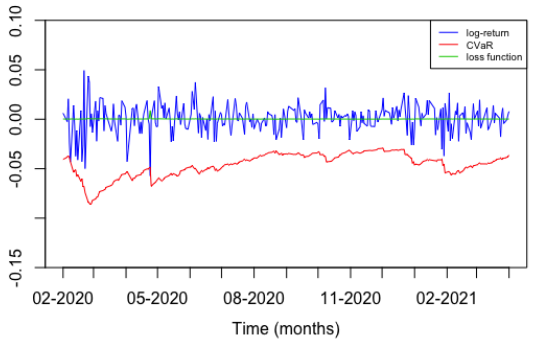
(d) AR-EGARCH model with Hill estimator

Figure B1: Log-returns, CVaR estimates and loss function for DAX 30 for COVID-19 period March 1, 2020 - April 30, 2021

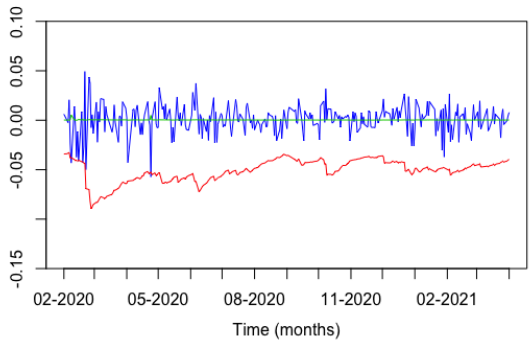
Figure B2a shows the results of the G-k-H model, Figure B2b the outcomes of the AG-k^*-H model, Figure B2c the ones of model AGJR-k^*-H and Figure B2d displays the results of model AEG-k^*-H all for the HSI.



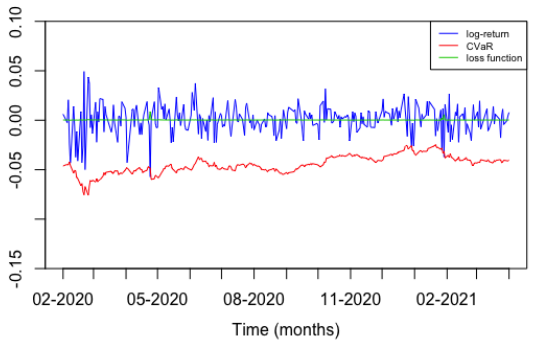
(a) GARCH model with Hill estimator



(b) AR-GARCH model with Hill estimator



(c) AR-GJRGARCH model with Hill estimator



(d) AR-EGARCH model with Hill estimator

Figure B2: Log-returns, CVaR estimates and loss function for the HSI for COVID-19 period March 1, 2020 - April 30, 2021

C Appendix

C.1 Code explanation

For the coding, we use as basis the R code which is given as supplementary material in Hoga (2019). We have made some small adaptations to the codes. For 'Simulation Replication', we have altered some values in the already existing simulations and have removed some of the variables we are not using, like the CV1 variable. For the 'Results Replication', we have not adjusted any of the code and used the exact same as Hoga (2019). For the two extensions, we have created new files, namely 'Simulation Extension' and 'Results Extension'. For both, we used again a large part of the already existing code, however we have now added two models to both of them, namely the GJR-GARCH and EGARCH model. Note that for the EGARCH model, the interpretation of the γ and ψ are switched in comparison with our Methodology in Section 4. Furthermore, for the 'Simulation Extension' we also incorporated two additional simulation models. Lastly, we wrote the 'Plots Additional Extension' which gives the code to reproduce the plots which are used in to further analyse the COVID-19 period in Section 4.7, in which we use some of the functions of our earlier 'Extension results' file.