The importance of higher moments in VaR and CVaR estimation.

Faes E.[1]   Mertens de Wilmars S.[2]   Pratesi F.[3]

Antwerp Management School

Prof. dr. Annaert Prof. dr. De Ceuster Prof. dr. Zhang

Master in Finance

June 2021

[1]Enjo.Faes@student.ams.ac.be
[2]Stephane.MertensdeWilmars@student.ams.ac.be
[3]Filippo.Pratesi@student.ams.ac.be

For our families and loved ones

Acknowledgements

First of all, we want to acknowledge the support during the writing of this thesis, given to us by our families and loved ones. Secondly, we are gratefull to our professors Jan Annaert[4], Marc De Ceuster[5] and Hairui Zhang[6] for all the valuable insights during our courses at AMS and dozens of assignments using the R language. We are especially gratefull that they patiently answered all our questions throughout the year.

Secondly, we want to extend our appreciation to Ulrik Lyngs who made it easier to work together in R with a pre-build template for the university of Oxford, without which this thesis could not have been written in this format (Lyngs 2019).

Finally, we thank Alexios Ghalanos for making the implementation of GARCH models integrated in R via his package "rugarch". By doing this, he facilitated the quantitative analysis we did in our methodology.

<div align="right">

Enjo Faes,
Stéphane Mertens de Wilmars,
Filippo Pratesi
Antwerp Management School, Antwerp
17 June 2021

</div>

---

[4] https://www.antwerpmanagementschool.be/nl/faculty/jan-annaert
[5] https://www.antwerpmanagementschool.be/nl/faculty/marc-de-ceuster
[6] https://www.antwerpmanagementschool.be/nl/faculty/hairui-zhang

Abstract

This paper examines the importance of higher-order moments in the estimation of Value at Risk (VaR) and Conditional Value At Risk (CVaR) for the European markets. We use 5 limited cases of the Skewed Generalized T-distribution SGT distributions: the SGED, the GED, the Skewed Student-t (ST), the symmetric Student-t (T) and Normal distribution. We examine the performance of 8 commonly used GARCH models and observe that the skewed distributions SGED and ST are the best performing models in terms of the Akaike Information Criterion (AIC), which is consistent with previous research on US indices. Using a rolling window approach for the out-of-sample (OOS) analysis we find that again this pattern appeared with the ST the best and the SGED a close second best distribution. Next to this, we also explore the Autoregressive Conditional Density (ACD) model, which allows the higher-order moments to be time-varying. The model in-sample provides a better fit according to the AIC, however the complexity of the estimation procedure is in our opinion not the best if we follow the results of our OOS. According to our OOS, we find that the ACD model with a Skewed T-distribution (ST) does not necessarily forecast the VaR and CVaR better than its non time-varying GARCH equivalent. Overall we do found that inclusion of higher-order moments matters in the estimation of VaR and CVaR.@

Contents

Contents

List of Figures

List of Tables

List of Abbreviations

ACD . . . . . . Autoregressive Conditional Density models (Hansen, 1994)

ARCH . . . . . Autoregressive Conditional Heteroscedasticity model (Bollerslev, 1986)

GARCH . . . Generalized Autoregressive Conditional Heteroscedasticity model (Bollerslev, 1986)

IGARCH . . . Integrated GARCH (Bollerslev, 1986)

EGARCH . . Exponential GARCH (Nelson, 1991)

GJRGARCH Glosten-Jagannathan-Runkle GARCH model (Glosten et al. 1993)

NAGARCH . Nonlinear asymmetric GARCH (Engle and Ng, 1993)

TGARCH . . Threshold GARCH (Zakoian, 1994)

TSGARCH . . Also called Absolute Value GARCH or AVGARCH referring to Taylor (1986) and Schwert (1989)

EWMA . . . . Exponentially Weighted Moving Average model

i.i.d, iid . . . . Independent and identically distributed

T . . . . Student's T-distribution

ST . . . . Skewed Student's T-distribution

SGT . . . . . . Skewed Generalized T-distribution

GED . . . . . . Generalized Error Distribution

SGED . . . . . Skewed Generalized Error Distribution

NORM . . . . Normal distribution

VaR . . . . . . Value-at-Risk

CVaR . . . . . Expected shortfall or Conditional Value-at-Risk

List of Symbols

Symbol	Interpretation
α	Significance level, or the percentile chosen for the VaR.
α_0	The constant in the conditional mean equation. In other papers sometimes referred to as μ.
α_1	The slope in the conditional mean equation.
β_0	The constant in the variance equation and (G)ARCH models. In other papers sometimes referred to as ω.
β_1	The parameter estimate of the innovation process in the variance equation and (G)ARCH models. In other papers sometimes referred to as α. This is also known as the ARCH component (p).
β_2	The parameter estimate of the autoregressive component in the variance equation and GARCH models. In other papers sometimes referred to as β. This is also known as the GARCH componenet (q)
γ	The leverage term in the asymmetric GARCH models: GJRGARCH and EGARCH.
ε	The innovation term or "noise" term. The random part or return shock of the previous period in the variance equation in GARCH models.
ζ	The time-varying kurtosis parameter in the ACD specification following a piecewise linear dynamic.
θ	The vector of parameters to estimate by the log likelihood functions in the maximum likelihood estimation of GARCH and distribution parameters.
η	This is the tail-thickness parameter of the unconditional and conditional distribution. Following the notation of Bali, Mo, et al. (2008), the parameter rises in the SGT distribution. For the T distribution this parameter is equal to the $\nu/2$.
κ	The peakedness parameter which is only present in the SGT distribution and equal to 2 for the other distributions used in this paper.
ν	The degree of freedom parameter from the T distribution.
ξ	The skewness parameter in the skewed distributions (SGT, SGED, ST) and in the GARCH models. Sometimes referred to as λ.
ρ	The time-varying skewness parameter in the ACD specification following a piecewise linear dynamic.

Introduction

A general assumption in finance is that stock returns are normally distributed. However, various authors have shown that this assumption does not hold in practice: stock returns are not normally distributed (among which Theodossiou 2000; Subbotin 1923; Theodossiou 2015). For example, Theodossiou (2000) mentions that "empirical distributions of log-returns of several financial assets exhibit strong higher-order moment dependencies which exist mainly in daily and weekly log-returns and prevent monthly, bimonthly and quarterly log-returns from obeying the normality law implied by the central limit theorem. As a consequence, price changes do not follow the geometric Brownian motion." So in reality, stock returns exhibit fat-tails and peakedness (Officer 1972), these are some of the so-called stylized facts of returns. Additionally, a point of interest is the predictability of stock prices (Fama 1965; Fama 1970; Welch and Goyal 2008). This makes it difficult for corporations to manage market risk, i.e. the variability of stock prices.

Risk, in general, can be defined as the volatility of unexpected outcomes (Jorion 2007). Corporations have to manage their risks both for their own sake and due to regulatory requirements. Consequentially, the Value at Risk metric (VaR), developed in response to the financial disaster events of the early 1990s, has been very influential in the financial world. The tool of VaR has become a standard measure of risk for many financial institutions going from banks, that use VaR to calculate the adequacy of their capital structure, to other financial services companies to assess the exposure of their positions and portfolios. The 1% VaR can be defined as the maximum loss of a portfolio, during a time horizon, excluding all the negative events with a combined probability lower than 1% while the Conditional

Introduction

Value at Risk (CVaR) can be defined as the average of the events that are lower than the VaR. Also Bali, Mo, et al. (2008) explains that many implementations of the CVaR have the assumption that asset and portfolio's returns are normally distributed but that this is inconsistent with the evidence. Assuming normality can lead to incorrect VaR and CVaR numbers, an underestimation of the probability of extreme events happening and ultimately a poor understanding of risk exposure.

This paper has *two goals.* The first goal is to *develop a deeper understanding of GARCH models*, their mechanics and the properties of these models in regard to risk management. Second, we aim to *replicate and update the research made by Bali, Mo, et al. (2008) on U.S. indexes, analyzing the dynamics proposed with a European outlook.* Namely, we use GARCH models to predict the volatility of the Euro Stoxx 50 return index. After this estimation, we compute VaR and CVaR and conduct a backtest to see if the GARCH models predict the VaR and CVaR appropriately. This is summarized in our research question:

> **Do higher moments increase accuracy in the estimation of VaR and CVaR?**

The paper is organized as follows. First, chapter 1 discusses the stylized facts, conditional distributions and the GARCH models. Chapter 2 describes the data used and the methodology followed in modeling the volatility with GARCH models. Then a description is given of the control tests used to evaluate the performances of the different GARCH models and underlying distributions. In chapter 3, the findings are presented and discussed. In chapter 4 the results of some diagnostic tests are shown and interpreted. Finally, chapter 5 summarizes the results and answers the research question.

1
Literature review

1.1 Stylized facts of returns

When analyzing returns as a time-series, we look at log returns. The log returns are similar to simple returns, so the stylized facts of returns apply to both. One assumption that is made often in financial applications is that returns are i.i.d., or independently and identically distributed. Another assumption is that they are normally distributed. These assumptions might not be reasonable in reality. Below the stylized facts[1] following Annaert (2021) for returns are given.

- Returns are *small and volatile* (with the standard deviation being larger than the mean on average).

- Returns have very *little serial correlation*, as mentioned by for example Bollerslev (1987).

- Returns exhibit conditional heteroskedasticity, or *volatility clustering*. This effect goes back to Mandelbrot (1963). There is no constant variance (homoskedasticity). Instead it is time-varying. Bollerslev (1987) describes this phenomenon as "rates of return data are characterized by volatile and tranquil periods". Alexander (2008) argues this to have implications for risk models:

---

[1]Stylized facts are the statistical properties that appear to be present in many empirical asset returns (across time and markets)

following a large shock to the market, the volatility changes and the probability of another large shock is increased significantly.

- Returns exhibit *asymmetric volatility*, in the sense that volatility increases more after a negative return shock than after a large positive return shock. This is also called the *leverage effect.* Alexander (2008) mentions that this leverage effect is most pronounced in equity markets: usually there is a strong negative correlation between equity returns and the change in volatility.

- Returns are *not normally distributed*, as found by early work of Fama (1965). Returns have tails fatter than a normal distribution (they are leptokurtotic) and thus carry more risk. Log returns however **can** be assumed to be normally distributed. We will examine this in our empirical analysis. This makes that simple returns follow a log-normal distribution, a skewed density distribution. A good summary is given by Alexander (2008) : "In general, we need to know more about the distribution of returns than its expected return and its volatility. Volatility tells us the *scale* and the mean tells us the *location*, but the dispersion also depends on the **shape** of the distribution. The best dispersion metric would be based on the entire distribution function of returns."

Firms holding a portfolio of various stocks or other investments have a lot of things to consider: expected return of a portfolio, the probability to get a return lower than some threshold, the probability that an asset in the portfolio drops in value when the market crashes. All the previous requires information about the return distribution or so called density function of returns. What we know from the stylized facts of returns that the normal distribution is not appropriate for returns. In appendix part A we summarize some alternative distributions (SGT, SGED, GED, Skewed Student-t, and Student-t) that might better approximate the actual distribution of returns than the normal one.

1.2 Volatility modeling

1.2.1 Rolling volatility

When volatility needs to be estimated on a specific trading day, the method used as a descriptive tool would be to use rolling standard deviations. Engle (2001) explains the calculation of rolling standard deviations, as the standard deviation over a fixed number of the most recent observations[2]. Engle regards this formulation as the first ARCH model.

1.2.2 From ARCH to GARCH models

Autoregressive Conditional Heteroscedasticity (ARCH) models, proposed by Engle (1982), were in the first case not used in financial markets but on inflation. Since then, it has been used as one of the workhorses of volatility modeling.

There are three building blocks of the ARCH model: returns, the innovation process and the variance process (or volatility function), written out for an ARCH(1) in respectively equation (1.1), (1.2) and (1.3). Returns are written as a constant part (α_0) and an unexpected part, called noise or the innovation process ($\alpha_1 \times \varepsilon_t$). The innovation process is the volatility (σ_t) times z_t, which is an independent identically distributed random variable with a mean of 0 (zero-mean) and a variance of 1 (unit-variance). The independent (**i**id), notes the fact that the z-values are not correlated, but completely independent of each other. The distribution is one of the distributions listed in appendix part A. The third component is the variance process or the expression for the volatility. The variance is given by a constant β_0, plus the random part which depends on the return shock of the previous period squared (ε_{t-1}^2). In that sense when the uncertainty or surprise in the last period increases, then the variance becomes larger in the next period. The element σ_t^2

---

[2]For example, for the past month it would then be calculated as the equally weighted average of the squared deviations from the mean from the last 22 observations (the average amount of trading or business days in a month). All these deviations are thus given an equal weight. Also, only a fixed number of past recent observations is examined.

is thus known at time $t-1$, while it is a deterministic function of a random variable observed at time $t-1$ (i.e. ε_{t-1}^2).

$$y_t = \alpha_0 + \alpha_1 \times \varepsilon_t \tag{1.1}$$

$$\varepsilon_t = \sigma_t \times z_t, \ where \ z_t \overset{iid}{\sim} (0,1) \tag{1.2}$$

$$\sigma_t^2 = \beta_0 + \beta_1 \times \varepsilon_{t-1}^2 \tag{1.3}$$

The full description of the ARCH model is given in appendix part A.

An improvement of the ARCH model is the Generalized Autoregressive Conditional Heteroscedasticity (GARCH)[3]. This model and its variants come in to play because of the fact that calculating standard deviations through rolling periods, gives an equal weight to distant and nearby periods, by such not taking into account empirical evidence of volatility clustering, which can be identified as positive autocorrelation in the absolute returns. GARCH models are an extension to ARCH models, as they incorporate both a novel moving average term (not included in ARCH) and the autoregressive component. Furthermore, a second extension is changing the assumption of the underlying distribution. As already explained, the normal distribution is an unrealistic assumption, so other distributions which are described in part A will be used. As Alexander (2008) explains, this does not change the formulae of computing the volatility forecasts but it changes the functional form of the likelihood function[4]. An overview (of a selection) of investigated GARCH models is given in the following table.

---

[3]Stationarity implies that the series on which the ARCH model is used does not have any trend and has a constant expected mean. Only the conditional variance is changing.

[4]which makes the maximum likelihood estimation explained in part 2.2.1 complex with more parameters that have to be estimated.

Table 1.1: GARCH models, the founders

Author(s)	Model
Engle (1982)	ARCH model
Bollerslev (1986)	GARCH model
Bollerslev (1986)	IGARCH model
Nelson (1991)	EGARCH model
Glosten et al. (1993)	GJRGARCH model
Engle and Ng (1993)	NAGARCH model
Zakoian (1994)	TGARCH model
Taylor (1986) and Schwert (1989)	AVGARCH model
Morgan Guaranty Trust Company (1996)	EWMA or RiskMetrics model

1.3   ACD models

An extension to GARCH models was proposed by Hansen (1994), namely the autoregressive conditional density estimation model (referred to as ACD model, sometimes ARCD). It focuses on time variation in higher moments (skewness and kurtosis), because the degree and frequency of extreme events seem to be not expected by traditional models. Some GARCH models are already able to capture the dynamics by relying on a different unconditional distribution than the normal distribution (for example skewed distributions like the SGED or SGT), or a model that allows to incorporate these higher moments. However, Ghalanos (2016) mentions that these models also assume the shape and skewness parameters to be constant (not time-varying). As Ghalanos mentions: "the research on time-varying higher moments has mostly explored different parametrizations in terms of dynamics and distributions with little attention to the performance of the models out-of-sample and ability to outperform a GARCH model with respect to VaR." Also one could question the marginal benefits of the ACD, while the estimation procedure is rather sophisticated (nonlinear bounding specification of higher moment distribution parameters and interaction). So, are skew (skewness parameter) and kurtosis (shape parameter or parameters) time-varying? The literature investigating higher moments has arguments for and against this statement. In part 2.2.2 the specification is given.

1.4   Value at Risk

Value at Risk (VaR) is a risk metric developed simultaneously by Markowitz (1952) and Roy (1952) to calculate how much money an investment, portfolio, department or institution such as a bank could lose in a market downturn, though in this period it remained mostly a theoretical discussion due to lacking processing power and industry demand for risk management measures. Another important document in literature is the *1996 RiskMetrics Technical Document*, composed by RiskMetrics[5] (Morgan Guaranty Trust Company (1996), part of JP Morgan), which gives a good overview of the computation, but also made use of the name "value-at-risk" over equivalents like "dollars-at-risk" (DaR), "capital-at-risk" (CaR), "income-at-risk" (IaR) and "earnings-at-risk" (EaR). According to Holton (2002) VaR gained traction in the last decade of the 20th century when financial institutions started using it to determine their regulatory capital requirements. A VaR_{99} finds the amount that would be the greatest possible loss in 99% of cases. It can be defined as the threshold value F_t. Put differently, in 1% of cases the loss would be greater than this amount. It is specified as in (1.4). Christoffersen et al. (2001) puts forth a general framework for specifying VaR models and comparing between two alternatives models.

$$Pr(y_t \leq F_t | \Omega_{t-1}) \equiv \alpha \tag{1.4}$$

With y_t expected returns in period t, Ω_{t-1} the information set available in the previous period and α the chosen quantile.

---

[5]RiskMetrics Group was the market leader in market and credit risk data and modeling for banks, corporate asset managers and financial intermediaries (Alexander 2008).

1.5   Conditional Value at Risk

One major shortcoming of the VaR is that it does not provide information on the probability distribution of losses beyond the threshold amount. As VaR lacks subadditivity of different percentile outcomes, Artzner et al. (1996) rejects it as a coherent measure of risk. This is problematic, as losses beyond this amount would be more problematic if there is a large probability distribution of extreme losses, than if losses follow say a normal distribution. To solve this issue, they provide a conceptual idea of a Conditional VaR (CVaR) which quantifies the average loss one would expect if the threshold is breached, thereby taking the distribution of the tail into account. Mathematically, a $CVaR_{99}$ is the average of all the VaR with a confidence level equal to or higher than 99. It is commonly referred to as expected shortfall (ES) sometimes and was written out in the form it is used by today by (Bertsimas et al. 2004). It is specified as in (1.5).

To calculate F_t, VaR and CVaR require information on the expected distribution mean, variance and other parameters, to be calculated using the previously discussed GARCH models and distributions.

$$Pr(y_t \leq F_t | \Omega_{t-1}) \equiv \int_{-\infty}^{\alpha} f(y_t | \Omega_{t-1}) \, \mathrm{d}y_t = \alpha \tag{1.5}$$

With the same notations as before, and f the (conditional) probability density function of y_t.

According to the BIS framework, banks need to calculate both VaR_{99} and $VaR_{97.5}$ daily to determine capital requirements for equity, using a minimum of one year of daily observations (Basel Committee on Banking Supervision 2016). Whenever a daily loss is recorded, this has to be registered as an exception or exceedance. Banks can use an internal model to calculate their VaRs, but if they have more than 12 exceptions for their VaR_{99} or 30 exceptions for their $VaR_{97.5}$ they have to follow a standardized approach. Similarly, banks must calculate $CVaR_{97.5}$.

1.6 Past literature on the consequences of higher moments for VaR determination

A small description of the higher moments and VaR determination is given below. A small overview table is given:

Table 1.2: Higher moments and VaR

Author	Higher moments
Hansen (1994)	Skewness and kurtosis extended ARCH-model
Harvey and Siddique (1999)	Skewness, Effect of higher moments on lower moments
Brooks et al. (2005)	Kurtosis, Time varying degrees of freedom

While it is relatively straightforward to include unconditional higher-moments in VaR and CVaR calculations, it is less simple to do so when the higher moments (in addition to the variance) are time-varying. Hansen (1994) extended the ARCH model to include time-varying moments beyond mean and variance. While mean and variance of returns are usually the parameters of most interest, disregarding these higher moments could provide an incomplete description of a conditional distribution. The model proposed by Hansen (1994) allows for skewness and shape parameters to vary in a skewed-t density function through specifying them as functions of their errors in previous periods (in a similar way how variance is estimated). Applications on U.S. Treasuries and exchange rates are discussed by Hansen.

Harvey and Siddique (1999) extended a GARCH(1,1) model to include time-varying skewness by estimating it jointly with time-varying variance using a skewed Student-t distribution. They found a significant impact of skewness on conditional volatility, suggesting that these moments should be jointly estimated for efficiency. Changes in conditional skewness have an impact on the persistence of volatility shocks. They also found that including skewness causes the leverage effects of variance to disappear. They applied their methods on different stock indices (both developed

and emerging) at daily, weekly and monthly frequency.

Brooks et al. (2005) proposed a model based on a Student-t distribution that allows for both the variance and the degrees of freedom to be time-varying, independently from each other. Their model allows for both asymmetric variance and kurtosis through an indicator function (which has a positive effect on these moments only when the shock is in the right tail). They applied their model on different financial assets in the U.S. and U.K. at daily frequency.

2

Data and methodology

2.1 Data

We worked with daily returns on the Euro Stoxx 50 Return Index[1] retrieved from Datastream denoted in EUR from 02 January, 2001 to 19 May, 2021. The choice of daily data is motivated as follows. The primary interest in this paper is (C)VaR models for banks' (or financial institutions') internal trading desks. Their positions are usually short-term, making risk management at the daily level the most appropriate. As such, in reference to the literature review, regulators require VaR forecast for one day in advance. All following analysis could be applied on monthly returns as well. The Euro Stoxx 50, the leading blue-chip index of the Eurozone, was founded in 1999 and covers 50 of the most liquid and largest (in terms of free-float market capitalization) stocks. For its composition and computation we refer to the factsheet (*Calculation guide STOXX ® 2020*). Given that in the 20th century computing return series was time consuming, the Eurostoxx 50 Return index is shorter than the Euro Stoxx 50 Price index (going back to 2001). As a robustness check, we ran all subsequent analysis for the longer price index

---

[1]The same analysis has been performed for the FTSE 100, CAC 40, BEL 20 and the DAX 30 return indices and yielded broadly similar conclusions. The findings of these researches are available upon requests.

2. Data and methodology

as well. This did not yield a qualitative difference in terms of most efficient model(s).

Table 2.1 provides the main statistics describing the return series analyzed. Let daily returns be computed as $R_t = 100\,(\ln P_t - \ln P_{t-1})$, where P_t is the index price at time t and P_{t-1} is the index price at $t-1$.

The arithmetic mean of the series is 0.01% with a standard deviation of 1.44% and a median of 0.02 which translate to an annualized mean of 2.57% and an annualized standard deviation of 22.85%. The skewness statistic is highly significant and negative at -0.22 and the excess kurtosis is also highly significant and positive at 6.54. These 2 statistics give an overview of the distribution of the returns which has fatter tails than the normal distribution with a higher presence of left tail observations. A formal test such as the Jarque-Bera one with its statistic at 739.89 and a high statistical significance, confirms the non-normality.

The right column of table 2.1 exhibits the same descriptive statistics but for the standardizes residuals obtained from a simple GARCH model as mentioned in table 2.1 in Note 2. Again, Skewness statistic at -0.328 with a high statistical significance level and the excess Kurtosis at 1.706 also with a high statistical significance, suggest a non normal distribution of the standardized residuals and the Jarque-Bera statistic at 9511.18, given its high significance, confirms the rejection of the normality assumption.

Table 2.1: Summary statistics of the returns

Statistics	Euro Stoxx 50	Standardized Residuals
Arithmetic Mean	0.0102	-0.0512
Median	0.0152	-0.0314
Maximum	10.4372	5.8561
Minimum	-13.2164	-6.4432
Stdev	1.4391	0.9986
Skewness	-0.2191	-0.3285
	(0***)	(0***)
Excess Kurtosis	6.5382	1.7055
	(0***)	(0***)
Jarque-Bera	9511.1796***	739.8884***

Notes

[1] This table shows the descriptive statistics of the daily percentage returns of Euro Stoxx 50 over the period 2001-01-03 to 2021-05-19 (5316 observations). Including arithmetic mean, median, maximum, minimum, standard deviation. The skewness, excess kurtosis with p-value and signicance and the Jarque-Bera test with significance.

[2] The standardized residual is derived from a maximum likelihood estimation (simple GARCH model) as follows:
$R_t = \alpha_0 + \alpha_1 R_{t-1} + \varepsilon_t$
$\sigma_t^2 = \beta_0 + \beta_1 \varepsilon_{t-1}^2 + \beta_2 \sigma_{t-1}^2,$
Where ε_t is given by $z_t \sigma_t$

[3] *, **, *** represent significance levels at the 5, 1 and <1 procent.

2. Data and methodology

As can be seen in figure 2.1 the Euro area equity and later, since 1999 the Euro Stoxx 50, went up during the tech ("dot com") bubble reaching an ATH of €1734.84. Then, there was a correction to boom again until the burst of the 2008 financial crisis. After which it decreased significantly. With an ATL at 12 March, 2003 of €405.23. There is an improvement, but then the European debt crisis, with its peak in 2010-2012, occurred. From then there was some improvement until the "health crisis", which arrived in Europe, February 2020. This crisis recovered very quickly reaching already values higher then the pre-COVID crisis level.
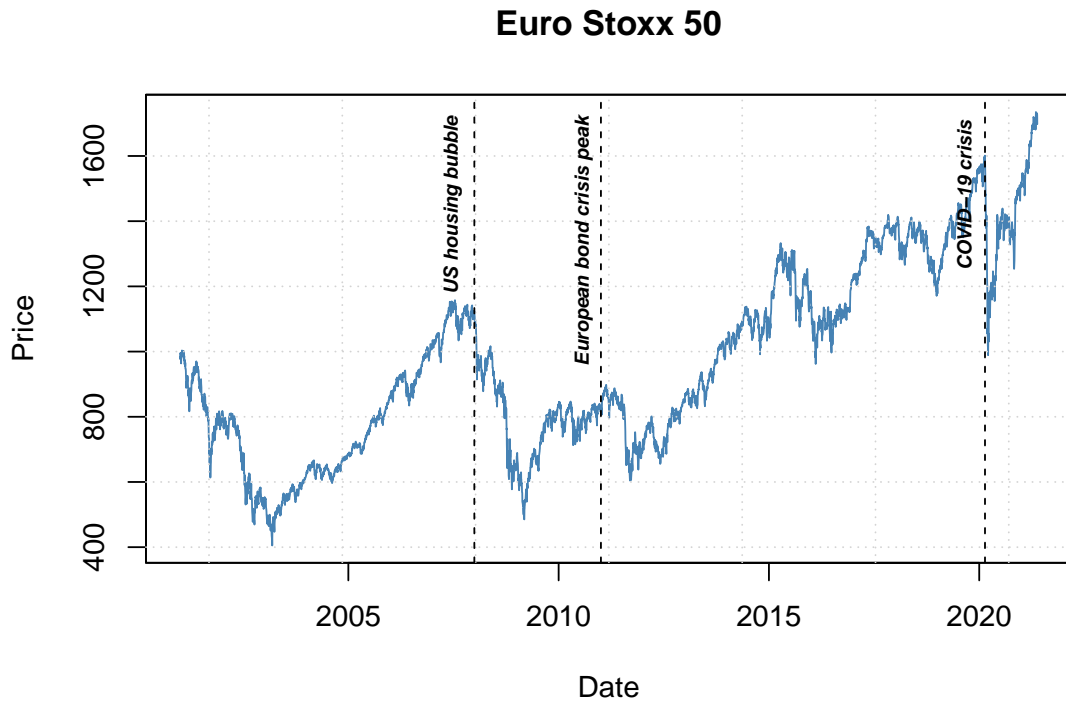
Euro Stoxx 50

Figure 2.1: Euro Stoxx 50 prices

2. Data and methodology

In figure 2.2 the daily log-returns are visualized. A stylized fact that is observable is the volatility clustering. As can be seen: periods of large volatility are mostly followed by large volatility and small volatility by small volatility.
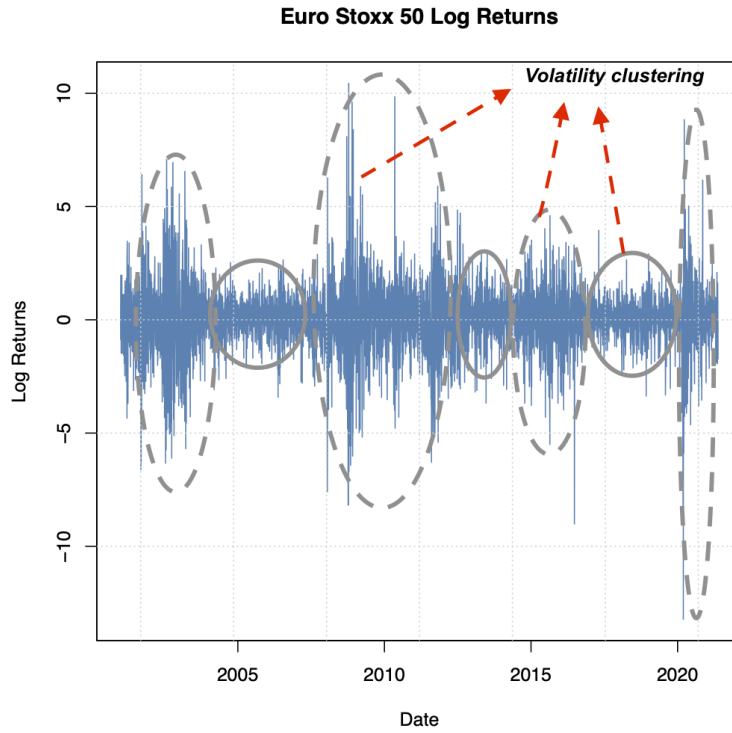
Figure 2.2: Euro Stoxx 50 log returns

In figure 2.3 the density distribution of the log returns are examined. As can be seen, as already mentioned in part 1.1, log returns are not really normally distributed.

In figure 2.4 the prediction errors (in absolute values and squared) are visualized in autocorrelation function plots. It is common practice to check this as in GARCH models the variance is for a large extent driven by the square of the prediction errors. The first component[2] α_0 is set equal to the sample average. As can be seen there is presence of large positive autocorrelation. This reflects, again, the presence of volatility clusters.

---

[2] α_0 is most of the time referred to as the μ in the conditional mean equation. Here we have followed Bali, Mo, et al. (2008).

Returns Histogram Vs. Normal
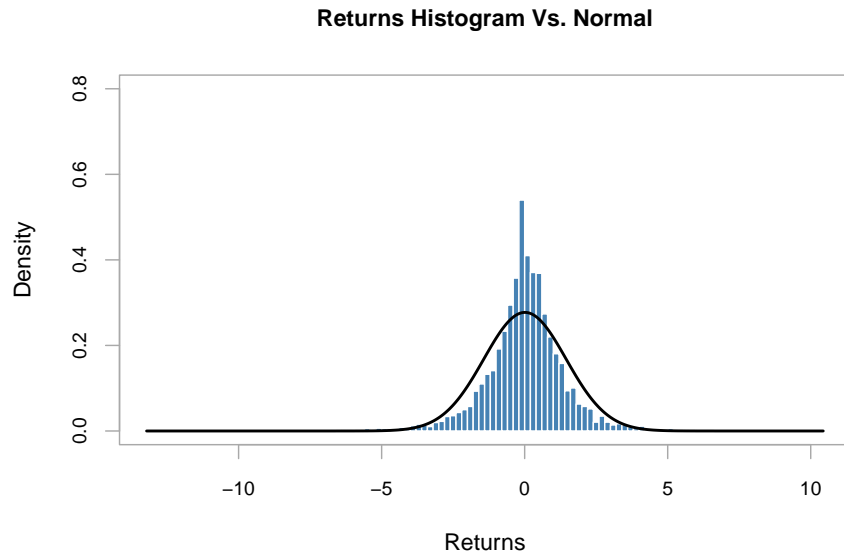
Figure 2.3: Density vs. Normal Euro Stoxx 50 log returns

Absolute Prediction Error

Squared Prediction Error

ACF (absolute prediction errors)

ACF (Squared prediction errors)

Figure 2.4: Absolute prediction errors

2.2 Methodology

2.2.1 Garch models

As already mentioned in part 1.2.2, the following models: SGARCH, EGARCH, IGARCH, GJRGARCH, NGARCH, TGARCH and NAGARCH (or TSGARCH) will be estimated. Additionally the distributions will be examined as well, including the normal, Student-t distribution, skewed Student-t distribution, generalized error distribution, skewed generalized e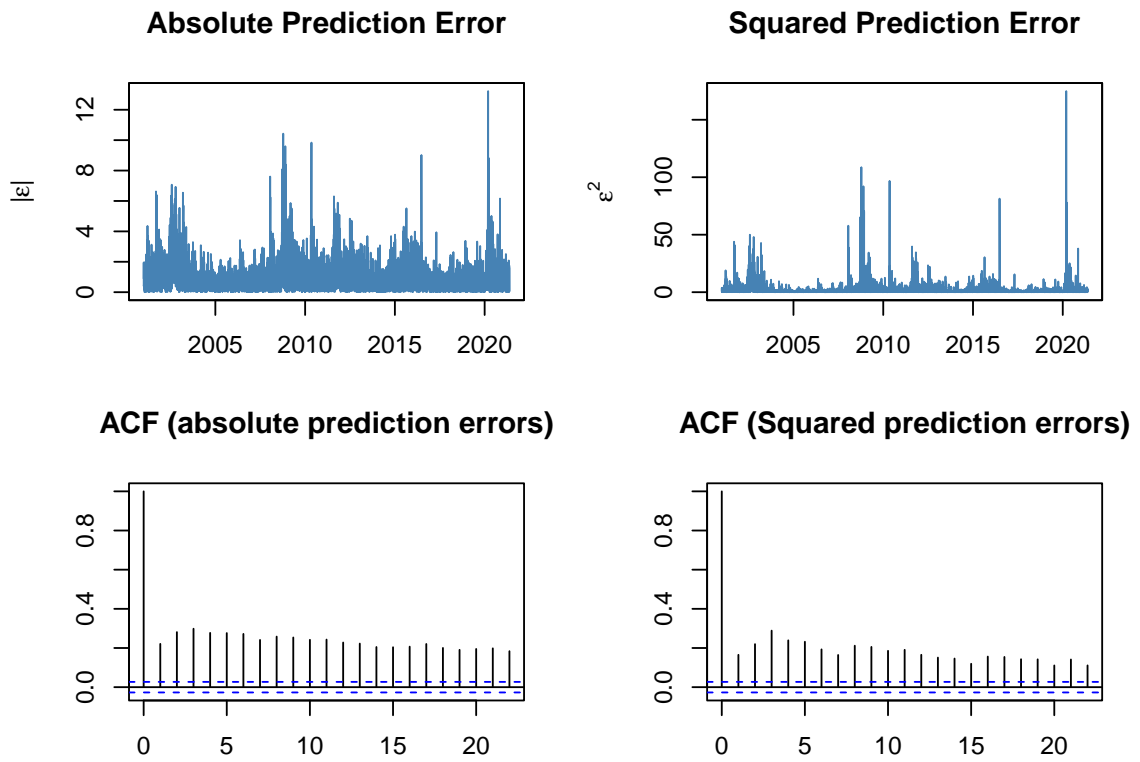rror distribution and the skewed generalized t distribution. Distributions other than the normal are scaled to zero mean and unit variance. They will be estimated using maximum likelihood[3].

Maximum likelihood estimation is a method to find the distribution parameters that best fit the observed data, through maximization of the likelihood function, or the computationally more efficient log-likelihood function (by taking the natural logarithm). It is assumed that the return data is i.i.d. and that there is some underlying parametrized density function f with one or more parameters that generate the data, defined as a vector θ in equation (2.2). These functions are based on the joint probability distribution of the observed data as in equation (2.3). Subsequently, the (log)likelihood function is maximized using an optimization algorithm shown inequation (2.4).

$$y_1, y_2, ..., y_N \sim i.i.d \tag{2.1}$$

$$y_i \sim f(y|\theta) \tag{2.2}$$

$$L(\theta) = \prod_{i=1}^{N} f(y_i|\theta) \tag{2.3}$$

---

[3]As already mentioned, fortunately, Ghalanos (2020b) has made it easy for us to implement this methodology in the R language[^data-meth-4] (R Core Team 2019) with the package "rugarch" v.1.4-4 (*R univariate garch*), which gives us a bit more time to focus on the results and the interpretation.

$$\log(L(\theta)) = \sum_{i=1}^{N} \log f(y_i|\theta)$$

$$\theta^* = arg \max_{\theta}[L] \tag{2.4}$$

$$\theta^* = arg \max_{\theta}[\log(L)] \tag{2.5}$$

After estimation of the GARCH models in-sample, out-sample analysis is done by performing a rolling window approach. With assumptions: a window of 2500 observations and re-estimation every year.

2.2.2 ACD models

Following Ghalanos (2016), arguments of ACD models are specified as in Hansen (1994). The skewness and kurtosis (shape) parameters which are constant in GARCH models (or time-invariant), are here time-varying following a piecewise linear dynamic. In equation (2.6) the parameters of the GARCH-ACD model are specified.

$$
\begin{aligned}
y_t &= \alpha_0 + \alpha_1 \times \varepsilon_t, \\
\varepsilon_t &= \sigma_t \times z_t, \\
z_t &\sim \Delta\left(0, 1, \rho_t, \zeta_t\right), \\
\sigma_t^2 &= \beta_0 + \beta_1 \varepsilon_{t-1}^2 + \beta_2 \sigma_{t-1}^2, \\
\rho_t &= \chi_0 + \chi_1 z_{t-1} I_{z_{t-1}<x} + \chi_2 z_{t-1} I_{z_{t-1}\geqslant x} + \xi_1 \bar{\rho}_{t-1}, \\
\zeta_t &= \kappa_0 + \kappa_1 \left|z_{t-1}\right| I_{z_{t-1}<x} + \kappa_2 \left|z_{t-1}\right| I_{z_{t-1}\geqslant x} + \psi_1 \bar{\zeta}_{t-1},
\end{aligned} \tag{2.6}
$$

where y_t, z_t and σ_t are familiar from GARCH models. ρ_t and ζ_t are respectively the time-varying skewness and shape parameter (with shape parameter meaning here the tail-tickness) and the standardized residuals z_t follows a distribution Δ that has a skewness and shape parameter. ρ_t and $zeta_t$ are following a piecewise linear dynamic with I the indicator variable (taking one if the underlying expression is true, 0 otherwise). x is a treshold value set to 0.

Again Ghalanos (2016) makes it easier to implement the somewhat complex ACD models using the R language with package "racd".

2.2.3   Analysis Tests VaR and CVaR

Unconditional coverage test of Kupiec (1995)

A number of tests are computed to see if the value-at-risk estimations capture the actual losses well. A first one is the unconditional coverage test by Kupiec (1995). The unconditional coverage or proportion of failures method tests if the actual value-at-risk exceedances are consistent with the expected exceedances (a chosen percentile, e.g. 1% percentile) of the VaR model. Following Kupiec (1995) and Ghalanos (2020a), the number of exceedences follow a binomial distribution (with thus probability equal to the significance level or expected proportion) under the null hypothesis of a correct VaR model. The test is conducted as a likelihood ratio test with statistic like in equation (2.7), with p the probability of an exceedence for a confidence level, N the sample size and X the number of exceedences. The null hypothesis states that the test statistic LR^{uc} is χ^2-distributed with one degree of freedom or that the probability of failure $\hat{p}$ is equal to the chosen percentile α.

$$LR^{uc} = -2\ln\left(\frac{(1-p)^{N-X}p^X}{\left(1-\frac{X}{N}\right)^{N-X}\left(\frac{X}{N}\right)^X}\right) \tag{2.7}$$

Conditional coverage test of Christoffersen et al. (2001)

Christoffersen et al. (2001) proposed the conditional coverage test. It is tests for unconditional coverage and serial independence. The serial independence is important while the LR^{uc} can give a false picture while at any point in time it classifies inaccurate VaR estimates as "acceptably accurate" (Bali and Theodossiou 2007). For a certain VaR estimate an indicator variable, $I_t(\alpha)$, is computed as equation (2.8).

$$I_t(\alpha) = \left\{ \begin{array}{ll} 1 & \text{if exceedence occurs} \\ 0 & \text{if no exceedence occurs} \end{array} \right. . \tag{2.8}$$

It involves a likelihood ratio test's null hypothesis is that the statistic is χ^2-distributed with two degrees of freedom or that the probability of violation $\hat{p}$

(unconditional coverage) as well as the conditional coverage (independence) is equal to the chosen percentile α. While it tests both unconditional coverage as independence of violations, only this test has been performed and the unconditional coverage test is not reported.

Dynamic quantile test

Engle and Manganelli (2004) provides an alternative test to specify if a VaR model is appropriately specified by proposing the dynamic quantile test. This test specifies the occurrence of an exceedance (here hit) as in (2.9), with $I(.)$ a function that indicates when there is a hit, based on the actual return being lower than the predicted VaR. θ is the confidence level. They test jointly H_0 that the expected value of hit is zero and that it is uncorrelated with any variables known at the beginning of the period (B), notably the current VaR estimate and hits in previous periods, specified as lagged hits. This is done by regressing hit on these variables as in (2.10). $X\delta$ corresponds to the matrix notation. Under H_0, this regression should have no explanatory power. As a final step, a χ^2-distributed test statistic with m degrees of freedom equal to the parameters to be estimated (constant, number of hits and VaR estimate) is constructed as in (2.11).

$$Hit_t = I\left(R_t < -\mathrm{VaR}_t(1-\alpha)\right) - (1-\alpha), \tag{2.9}$$

$$\begin{aligned} Hit_t = \delta_0 + \delta_1 Hit_{t-1} + \ldots + \delta_p Hit_{t-p} + \delta_{p+1}VaR_t + \\ \delta_{p+2}I_{year1,t} + \ldots + \delta_{p+2+n}I_{yearn,t} + u_t \end{aligned} \tag{2.10}$$

$$Hit_t = X\delta + u_t \quad u_t = \begin{cases} \alpha - 1 & \text{prob } \alpha \\ \alpha & \text{prob } (1-\alpha) \end{cases}$$

$$\frac{\hat{\delta}'_{OLS}X'X\hat{\delta}^a_{OLS}}{(1-\alpha)\alpha} \sim \chi^2(m) \tag{2.11}$$

2. Data and methodology

CVaR Test

The Expected Shortfall (or CVaR) test by McNeil and Frey (2000) tests whether the excess conditional shortfall has a mean of zero. Under the alternative hypothesis, this mean is greater than zero. This test uses a one sided t-statistic and bootstrapped p-values, as the distribution of the excess conditional shortfall is not assumed to be normal. This test is performed as in Ardia et al. (2019).

3
Empirical Findings

3.1 Density of the returns

3.1.1 MLE distribution parameters

In table 3.1 we can see the estimated parameters of the unconditional distribution functions. Note that the Student-t and skewed Student-t distribution are usually noted with degrees of freedom as parameters. For consistency, we have parameterized them using limiting cases of the SGT-distribution. Note that to read the degrees of freedom for the two distributions, it is simply 2η. They are presented for the Skewed Generalized T-distribution (SGT) and limiting cases thereof previously discussed. Additionally, maximum likelihood score and the Akaike Information Criterion (AIC) are reported to compare goodness of fit of the different distributions but also taking into account simplicity of the models. We find that the SGT-distribution has the highest maximum likelihood score of all. All other distributions have relatively similar likelihood scores, though slightly lower and are therefore not the optimal distributions. However, when considering AIC it is a tie between SGT and SGED. This provides some indication that we have a valid case to test the suitability of different SGED-GARCH VaR models as an alternative for the SGT-GARCH VaR models. While sacrificing some goodness of fit, the SGED distribution has the

advantage of requiring one parameter less, which results in easier implementation. For the SGT parameters the standard deviation and skewness are both significant at the 1% level. For the SGED parameters, the standard deviation and the skewness are both significant at respectively the 1% and 5% level. Both distributions are right-skewed. For both distributions the shape parameters are significant at the 1% level, though the η parameter was not estimated as it is by design set to infinity due to the SGED being a limiting case of SGT.[1]

Additionally, for every distribution fitted with MLE, plots are generated to compare the theoretical distribution with the observed returns and reported in the appendix. We find that the normal distribution is not a good approximation. The theoretical SGT, SGED and GED distributions are closer to the actual data, except that they somewhat overestimate peakedness. The theoretical skewed and symmetric Student-t distributions on the other hand underestimate peakedness. Visually, the best match seems to be the SGT distribution, which is to be expected as it has the highest log likelihood score.

---

[1]To check whether the relative ranking of distributions still holds in different periods, we have calculated the maximum likelihood score and AIC for three smaller periods: The period up to the dotcom collapse (1987-2001), up to the Global Financial Crisis (2002-2009) and up to the present Covid-crash (2009-2021). There is no qualitative difference in relative ranking with these subsamples.

Table 3.1: Maximum likelihood estimates of unconditional distribution functions

dist	α	β	ξ	κ	η	LLH	AIC
SGT	0.014	1.441	-0.02	1.233	4.959	-8850.419	17710.84
	(0.025)	(0.031)***	(0.016)	(0.085)***	(1.368)***		
SGED	0.015	1.42	0.008	0.898	Inf	-8859.217	17710.84
	(0.003)***	(0.015)***	(0.002)***	(0.018)***			
GED	0	1.418	0	0.899	Inf	-8859.537	17725.07
	(0.002)	(0.023)***		(0.022)***			
ST	0.012	1.635	-0.043	2	2.817	-8873.843	17755.69
	(0.02)	(0.076)***	(0.016)***		(0.132)***		
T	0.045	1.64	0	2	1.403	-8877.165	17760.33
	(0.015)***	(0.078)***			(0.133)***		
Normal	0.01	1.439	0	2	Inf	-9477.660	18959.32
	(0.02)	(0.014)***					

Notes:

Table contains parameter estimates for SGT-distribution and some of its limiting cases. The underlying data is the daily return series of the Euro Stoxx 50 for the period between December 31. 1986 and April 27. 2021. Standard errors are reported between brackets. *LLH* is the maximum log-likelihood value. *, ** and *** point out significance at 10, 5 and 1 percent level.

3.2 Constant higher moments

To compare the goodness-of-fit of different combinations of GARCH models and distributions, we have portrayed the AICs in table 3.2. The smaller the criterium the better. As you can see in table 3.2 the AIC for the skewed Student-t distribution (ST) is the best for almost all the models. As also shown in appendix part B. Only for the SGARCH and IGARCH the SGED distribution has a lower AIC. The best goodness-of-fit over all distributions seems to be the NAGARCH model.

Due to 3.2 being the result of an in-sample estimation, we will further examine the five best performing GARCH models according to the AIC (EGARCH, GJRGARCH, NAGARCH, AVGARCH and TGARCH).

Table 3.2: Model selection according to AIC

	SGARCH	IGARCH	EWMA	EGARCH	GJRGARCH	NAGARCH	TGARCH	AVGARCH
N	3.174	3.176	3.198	3.114	3.124	3.107	3.111	3.107
T	3.130	3.130	3.140	3.079	3.089	3.074	3.077	3.074
ST	3.127	3.127	3.135	3.072	3.083	3.067	3.071	3.067
GED	3.128	3.128	3.139	3.080	3.089	3.075	3.079	3.076
SGED	3.125	3.126	3.136	3.075	3.084	3.069	3.073	3.069

Notes
This table shows the AIC value for the respective model.
With on the rows the distributions.

3. Empirical Findings

Table 3.3 presents the maximum likelihood estimates for 8 GARCH models based on the ST distribution with constant skewness and kurtosis parameters (robust white errors are presented in parenthesis). The parameter α_0 is only statistically significant for the SGARCH, IGARCH and EWMA model with a value close to 0. The AR(1) coefficient, α_1, has parameters going from -0.049 to -0.032 with p values ranging from 0.011 to 0.014 suggesting significance, but indicating very small negative autocorrelation. The GARCH parameters in the conditional variance equations (β_0) are generally statistically significant except for the EGARCH model. The results of β_1 and β_2 show the presence of significant time-variation in the conditional volatility of the Euro Stoxx 50, in fact, the sum of β_1 and β_2 for the GARCH parameters is close to one (from 0.827 to 1), suggesting the presence of persistence in the volatility of the returns. The parameter ξ is highly significant for all the 8 models tested with values ranging from 0.885 to 0.918 confirming the presence of skewness in the returns. The shape parameter η, which, in our case, measures the number of degrees of freedom divided by two, determining the tail behavior, is significant for all the models and ranges between 3.153 and 4.0635. The parameter γ, which is present only for EGARCH and GJRGARCH is significant and with values around 0.14. The absolute value function in family GARCH models (NAGARCH, TGARCH and AVGARCH) is subject to the *shift* and the *rot* parameters whose values are always positive and statistically significant. According to the log likelihood values (LLH), displayed in table 3.3, the model with the highest value is AVGARCH, while excluding the non-standard (or family) GARCH models from the analysis, the model that performs best is EGARCH.

Table 3.4 shows a very similar picture for the GARCH-SGED models. Compared to the GARCH-ST models

Table 3.3: Maximum likelihood estimates of the GARCH-ST models with constant skewness and kurtosis parameters

	SGARCH	IGARCH	EGARCH	GJRGARCH	EWMA	NAGARCH	TGARCH	AVGARCH
α_0	0.052	0.052	0.014	0.019	0.055	0.004	0.013	0.003
	(0.012)***	(0.013)***	(0.011)	(0.012)	(0.014)***	(0.012)	(0.013)	(0.013)
α_1	-0.048	-0.048	-0.033	-0.043	-0.049	-0.034	-0.038	-0.032
	(0.013)***	(0.013)***	(0.012)***	(0.012)***	(0.014)***	(0.012)***	(0.011)***	(0.012)**
β_0	0.017	0.014	0.004	0.021	0	0.024	0.023	0.024
	(0.004)***	(0.004)***	(0.003)	(0.004)***		(0.003)***	(0.004)***	(0.002)***
β_1	0.095	0.1	-0.155	0	0.072	0.06	0.078	0.068
	(0.011)***	(0.011)***	(0.011)***	(0.01)	(0.008)***	(0.013)***	(0.008)***	(0.004)***
β_2	0.899	0.9	0.982	0.902	0.928	0.787	0.922	0.899
	(0.011)***		(0)***	(0.013)***		(0.022)***	(0.009)***	(0)***
ξ	0.918	0.918	0.89	0.895	0.915	0.885	0.891	0.885
	(0.017)***	(0.017)***	(0.017)***	(0.016)***	(0.016)***	(0.017)***	(0.017)***	(0.017)***
η	3.301	3.153	3.95	3.868	3.624	4.062	4.0335	4.0635
	(0.325)***	(0.273)***	(0.4355)***	(0.422)***	(0.285)***	(0.46)***	(0.451)***	(0.457)***
γ			0.107	0.177				
			(0.011)***	(0.02)***				
$shift$						1.567		0.393
						(0.332)***		(0.014)***
rot							1	1
							(0.069)***	(0.117)***
LLH	-8303.694	-8304.437	-8158.11	-8186.06	-8328.667	-8143.563	-8154.785	-8143.141

Notes:

This table shows the maximum likelihood estimates of various GARCH-ST models. The daily returns used on the Euro Stoxx 50 cover the period from 03 January, 2001 to 19 May, 2021 (5316 observations).

The mean process is modeled as follows: $R_t = \alpha_0 + \alpha_1 \times R_{t-1} + \varepsilon_t$ Where, in the 8 GARCH models estimated, γ is the asymmetry in volatility, ξ, κ and η are constant and robust standard errors based on the method of White (1982)) are displayed in parenthesis. LLH is the maximized log likelihood value.

Table 3.4: Maximum likelihood estimates of the GARCH-SGED models with constant skewness and kurtosis parameters

	SGARCH	IGARCH	EGARCH	GJRGARCH	EWMA	NAGARCH	TGARCH	AVGARCH
α_0	0.046 (0.011)***	0.045 (0.012)***	0.008 (0.008)	0.014 (0.012)	0.046 (0.015)***	0 (0.008)	0.007 (0.012)	-0.002 (0.012)
α_1	-0.055 (0.013)***	-0.055 (0.014)***	-0.041 (0.012)***	-0.05 (0.013)***	-0.056 (0.016)***	-0.043 (0.012)***	-0.047 (0.011)***	-0.041 (0.012)***
β_0	0.02 (0.005)***	0.015 (0.004)***	0.005 (0.002)**	0.023 (0.004)***	0	0.025 (0.003)***	0.024 (0.005)***	0.025 (0.002)***
β_1	0.096 (0.012)***	0.104 (0.012)***	-0.151 (0.01)***	0 (0.01)	0.07 (0.008)***	0.059 (0.007)***	0.077 (0.01)***	0.067 (0.003)***
β_2	0.895 (0.012)***	0.896	0.981 (0)***	0.901 (0.014)***	0.93	0.793 (0.009)***	0.922 (0.01)***	0.897 (0)***
ξ	0.935 (0.021)***	0.936 (0.023)***	0.901 (0.019)***	0.906 (0.02)***	0.93 (0.02)***	0.896 (0.017)***	0.902 (0.019)***	0.896 (0.018)***
η								
γ			0.107 (0.01)***	0.175 (0.021)***				
$shift$						1.547 (0.147)***		0.435 (0.012)***
rot							1 (0.076)***	0.959 (0.101)***
LLH	-8300.515	-8302.804	-8164.576	-8189.218	-8330.811	-8149.596	-8160.239	-8149.518

Notes:
This table shows the maximum likelihood estimates of various GARCH-SGED models. The daily returns used on the Euro Stoxx 50 Price index cover the period from 03 January, 2001 to 19 May, 2021 (5316 observations). The mean process is modeled as follows: $R_t = \alpha_0 + \alpha_1 \times R_{t-1} + \varepsilon_t$ Where, in the 8 GARCH models estimated, γ is the asymmetry in volatility, ξ, κ and η are constant and robust standard errors based on the method of White (1982)) are displayed in parenthesis. LLH is the maximized log likelihood value.

3.2.1 Value-at-risk

As already mentioned 2 candidate models seem to be most appropriate: EGARCH and NAGARCH. To check if they perform well out-of-sample we conduct a backtest by using a rolling forecasting technique. A simple graph is shown in figure 3.1 for the EGARCH-ST model. It seems that the VaR model for $\alpha = 0.05$ underestimates the downside events, while the VaR model for $\alpha = 0.01$ captures more of the downside events.
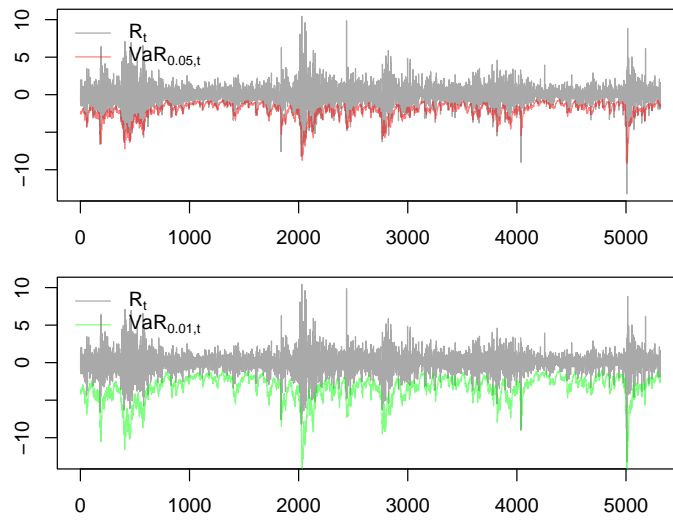

Figure 3.1: Value-at-Risk (in-sample) for the EGARCH-ST model

Let us examine this further using a rolling window approach whilst forecasting 1-day ahead results (using a sample data of 2500 observations or approximately 10 years) with re-estimating parameters every year.

Figure 3.2 shows that choosing an appropriate forecast period is important (with here the Eurobond crisis, the Brexit and Covid-crisis), so in order to avoid a look-ahead bias this rolling window approach was used instead of a static forecast method.

Euro Stoxx 50

Figure 3.2: Selected period to start forecast from

If we look at the results of the rolling window, we can for example compare as in figure 3.3 the EGARCH-ST (with skewed Student-t distribution) with the EGARCH-N (with normal distribution). The EGARCH-N seems to capture the extreme events a bit less compared to EGARCH-ST. Next, we will apply some diagnostic tests to investigate this formally.
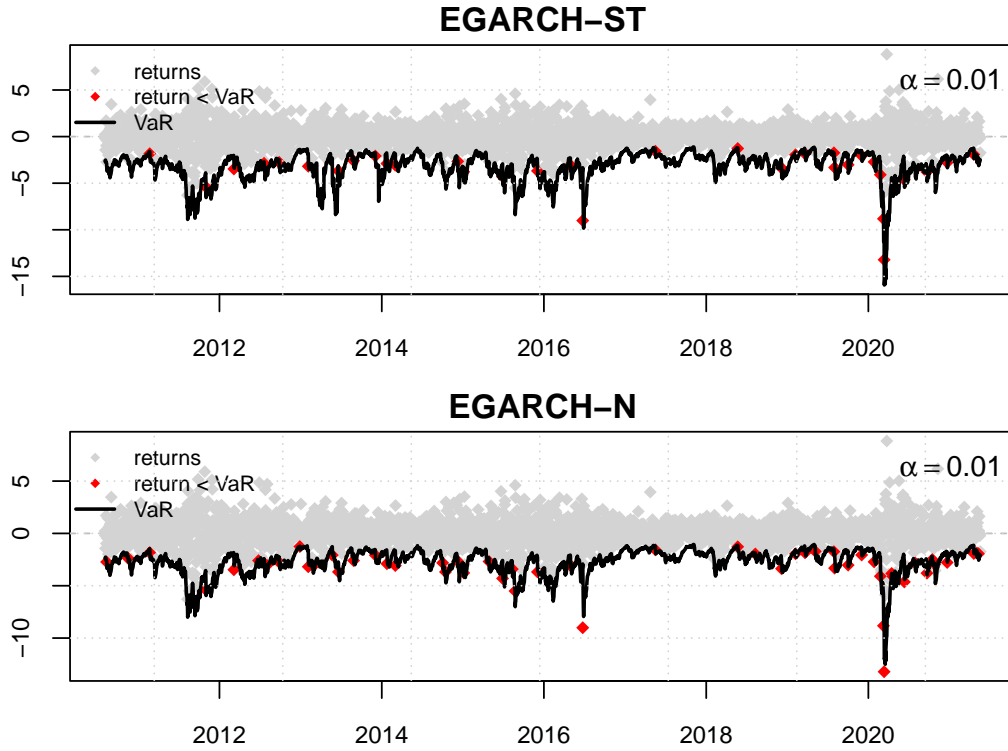
Figure 3.3: Comparison between VaR-EGARCH-ST and VaR-EGARCH-N

In table 3.5 the results of these tests are given. With the SGED as underlying distribution, the ratio of actual to expected VaR exceedances is highest for the GJRGARCH model, with an excess exceedance rate of 24%. The other GARCH models still have a positive but slightly lower excess exceedance rate of 21%. The ratio of actual to expected CVaR exceedances gives a similar picture, with the highest excess exceedance rate being the GJRGARCH with 25% and for the other models 21%. These results are significant at the 5% level, except for the NAGARCH, which is significant only at the 10% level. This points that the difference between actual and expected exceedances is significantly different from 0. For all models under the SGED, the unconditional and conditional cover test and the dynamic quantile test statistics are insignificant. This suggests that the SGED is a good distribution for calculating VaR and CVaR.

With the ST as underlying distribution, the ratio of actual to expected VaR exceedances is highest for the TGARCH model, with an excess exceedance rate of

28%. The NAGARCH model has the lowest VaR excess exceedance rate of 17%. For the CVaR excess exceedance rate, there is a similar picture, with the TGARCH having an excess exceedance rate of 29% and the NAGARCH of 18%. These results are insignificant, thus it cannot be concluded that the excess exceedance for VaR and CVaR are significantly different from 0. Both the unconditional and conditional cover tests are insignificant. However, the dynamic quantile test is significant at the 10% level for the EGARCH and TGARCH. This suggests that the ST distribution is a good distribution for calculating VaR and CVaR, though not as good as the SGED.

The GED, T and Normal distribution have significant unconditional and conditional and dynamic quantile test statistics for most models, which suggest they are less appropriate for modeling VaR and CVaR.

Table 3.5: VaR and CVaR test statistics

	EGARCH	GJRGARCH	TGARCH	NAGARCH	AVGARCH
Panel A: SGED					
AE VaR	1.207	1.243	1.207	1.207	1.207
AE CVaR	1.214**	1.25**	1.214**	1.214*	1.214**
UC	1.147	1.558	1.147	1.147	1.147
CC	1.979	2.439	1.979	1.979	1.979
DQ	7.757	13.286	12.578	5.223	5.846
Panel B: GED					
AE VaR	1.456	1.456	1.491	1.456	1.385
AE CVaR	1.464***	1.464**	1.5***	1.464**	1.393***
UC	5.184**	5.184**	5.969**	5.184**	3.764*
CC	6.396**	6.396**	7.242**	6.396**	4.859*
DQ	20.235***	16.564**	19.584**	11.688	10.508
Panel C: ST					
AE CVaR	1.207	1.243	1.278	1.172	1.243
AE CVaR	1.214	1.25	1.286	1.179	1.25
UC	1.147	1.558	2.026	0.796	1.558
CC	1.979	2.439	2.959	1.579	2.439
DQ	15.071*	13.002	13.935*	4.552	6.841
Panel D: T					
AE VaR	1.527	1.562	1.562	1.491	1.598
AE CVaR	1.536	1.571	1.571*	1.5	1.607
UC	6.803***	7.683***	7.683***	5.969**	8.61***
CC	8.137**	9.081**	9.081**	7.242**	10.072***
DQ	0.012	18.986**	22.759***	17.376**	25.561***
Panel E: N					
AE VaR	1.953	1.882	1.953	1.74	1.776
AE CVaR	1.964***	1.893***	1.964***	1.75***	1.786***
UC	20.217***	17.575***	20.217***	12.76***	13.904***
CC	22.409***	19.609***	22.409***	14.496***	15.712***
DQ	36.498***	32.87***	39.472***	27.102***	37.614***

Notes:

Table contains the ratio of actual to expected exceedances for VaR and Conditional VaR, the unconditional and conditional coverage test statistic and the dynamic quantile test statistic for VaR. Significance levels for the VaR ratio not reported. *, ** and *** point out significance at 10, 5 and 1 percent level.

3.3 Time-varying higher moments

As we already pointed out in part 1.3, it might also be interesting to look at time-varying moments and check if there is an improvement for estimating the VaR and CVaR as Bali, Mo, et al. (2008) did for example.

Table 3.6: Maximum likelihood estimates of the ST-ACD model with constant skewness and kurtosis parameters

	ACD GARCH
α_0	0.054 (0.111)
α_1	-0.044 (0.216)
β_0	0.02
β_1	0.098 (0.002)***
β_2	0.892 (0.009)***
χ_0	-1.548 (13.208)
χ_1	0.046 (0.401)
χ_2	0.032 (0.488)
ξ_1	0.563 (3.778)
κ_0	0.127 (0.196)
κ_1	0 (0.175)
κ_2	1 (0.249)***
ψ_1	0.788 (0.101)***

Notes:
This table shows the maximum likelihood estimates of various GARCH-SGED models. The daily returns used on the Euro Stoxx 50 Price index cover the period from 03 January, 2001 to 19 May, 2021 (5316 observations).
The mean process is modeled as follows: $R_t = \alpha_0 + \alpha_1 \times R_{t-1} + \varepsilon_t$ Where, in the model estimated, γ is the asymmetry in volatility, (calculated using robust standard errors based on the method of White (1982)) are displayed in parenthesis.

The following figure 3.4 plots the conditional mean, the conditional volatility. The implied conditional time varying skewness and excess kurtosis for the Euro Stoxx 50 series, which are similar to illustrations given by Galanos[2].

---

[2]`http://www.unstarched.net/2013/04/22/time-varying-higher-moments-with-the-racd-package/`
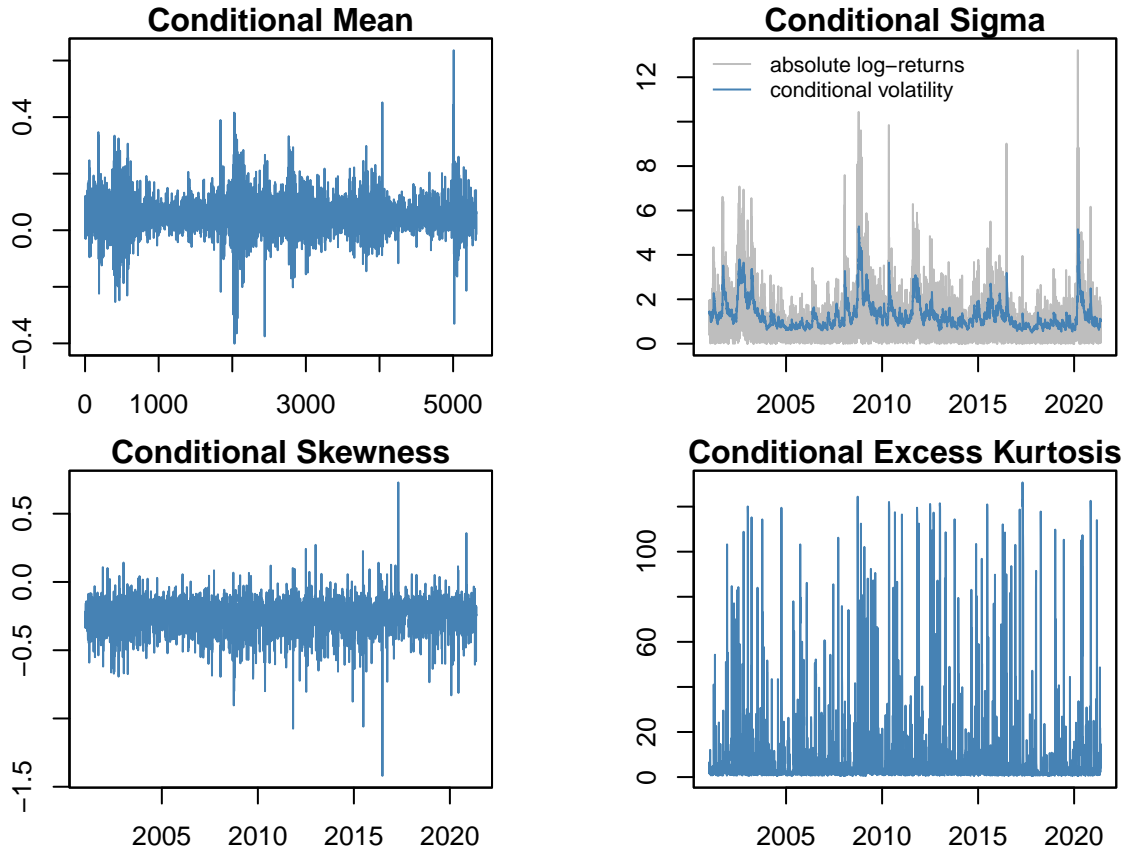
Figure 3.4: Dynamics of the ACD model

After performing a backtest[3], we observe that the ACD model is slightly worse than the GARCH equivalent in estimating the VaR. Although there are much less models tested here because of the difficulty of the procedure of ACD models, we find less convincing evidence of the merit of using conditional higher moments in VaR and CVaR estimation. So contrary to Bali, Mo, et al. (2008), we find that the ACD models seem to underperform relative to the GARCH model. With a rejection of unconditional coverage test statistic at 10% significance level. It is not surprising that the GARCH model performs bad as well looking at the rejection of the dynamic quantile test and the CVaR test.

---

[3]This backtest contains the following features: recursive, window of 2500, refitted every 250 trading days (approximately one year).

Table 3.7: VaR and ES test statistics (ACD-ST vs GARCH-ST)

	ACD	GARCH
AE VaR	1.357	1.286
AE CVaR	1.357**	1.286**
UC	3.131*	2.026
CC	4.171	2.959
DQ	16.118**	14.587*

Notes:

Table contains the ratio of actual to expected exceedances for VaR and Expected Shortfall, the unconditional and conditional coverage test statistic and the dynamic quantile test statistic for VaR. Significance levels for the VaR ratio not reported. *, ** and *** point out significance at 10, 5 and 1 percent level.

4

Robustness checks

In order to check if the models are specified correctly, some diagnostic checks have to be performed. The specification checks have to be done on the standardized residuals of the estimated GARCH model given by the following equation:

$$\hat{Z}_t = \frac{\hat{\varepsilon}_t}{\hat{\sigma}_t} = \frac{R_t - \hat{\mu}}{\hat{\sigma}_t}$$

.

Table 4.1 contains the results of two tests. Panel A displays the Ljung-box test on the squared standardized residuals of the GARCH models. Panel B displays the ARCH LM test on the squared standardized residuals. The ARCH LM test checks if the ARCH process is adequately fitted. As can be seen for both tests, the EGARCH, AVGARCH, NAGARCH and TGARCH do not suffer from any correlation in the squared residuals. For EWMA there is highly significant correlation of the residuals for all underlying distributions, suggesting that the model is inappropriate for VaR and CVaR estimation.

Table 4.1: Diagnostic Tests for Heteroscedasticity

	SGARCH	EGARCH	AVGARCH	NAGARCH	GJRGARCH	TGARCH	IGARCH	EWMA
Panel A: Ljung Box Test on the standardized squared values of the residuals								
Norm	31.157*	25.321	22.263	23.03	32.186*	26.208	32.727*	44.066***
T	33.907**	24.81	21.321	21.862	34.34**	26.658	34.183**	41.765***
ST	33.961**	25.024	21.412	22.051	34.607**	26.811	34.187**	40.605***
GED	32.493*	24.826	21.63	22.191	33.365*	26.142	33.361*	42.627***
SGED	32.569*	25.065	21.57	22.341	33.747*	26.351	33.342*	41.333***
Panel B: ARCH LM Test on the standardized squared values of the residuals								
Norm	32.322*	26.461	23.081	23.474	34.475**	26.991	33.857*	42.773***
T	34.687**	25.958	22.063	22.218	37.875**	27.634	34.912**	40.719***
ST	34.605**	26.138	22.129	22.36	38.174**	27.718	34.756**	39.559**
GED	33.431*	25.973	22.433	22.598	36.379**	27.023	34.228**	41.433***
SGED	33.393*	26.173	22.319	22.698	36.859**	27.167	34.071**	40.155***

Notes

Table displays the Ljung box statistics and the ARCH LM Test for the standardized squared residuals of the models analyzed. The underlying data is the daily return series of the Euro Stoxx 50 for the period between 2001-01-03 and 2021-05-19.

*, ** and *** point out respectively significance at 10, 5 and 1 percent level.

The null hypothesis of the test in both panels are described as follows:

H_0: $Corr(Z_t^2, Z_{t-1}^2) = Corr(Z_t^2, Z_{t-2}^2) = ... = Corr(Z_t^2, Z_{t-22}^2) = 0$

4. Robustness checks

The Generalized Method of Moments (GMM) test by Hansen (1982) is also a test that checks if the model is correctly specified or not, by examining the moments of the standardized residuals. Here only the results of the t-tests for the four individual moments are examined from the GMM test[^findings-3]. The mean (first moment) should be equal to zero ($E\left[z_t\right] = 0$). The variance (second moment) should be equal to one ($E\left[z_t^2 - 1\right] = 0$). The skewness should be equal to zero and the excess kurtosis (third and fourth moment) should be equal to zero (respectively $E\left[z_t^3\right] = 0$ and $E\left[z_t^4 - 3\right] = 0$). Table 4.2 portrays the results of the GMM test. Only the AVGARCH and NAGARCH models fail to reject the null-hypothesis that the moments are significantly different from the normal distribution for all underlying distributions. TGARCH fails to reject the null-hypothesis for all distributions, except for the Student-t. Similarly, EGARCH fails to reject it for the GED and Student-t. The SGARCH, IGARCH and EWMA reject the null-hypothesis with high significance for all distributions and are thus not suited for VaR and CVaR estimation.

Table 4.2: GMM Tests

	SGARCH	EGARCH	AVGARCH	NAGARCH	GJRGARCH	TGARCH	IGARCH	EWMA
Panel A: SGED								
Mean	-0.036**	-0.002	0.007	0.006	-0.006	-0.001	-0.036**	-0.039***
Variance	0.015**	0.009	0.007	0.007	0.01	0.008	-0.016**	0.121***
Skewness	-0.46**	-0.392	-0.374	-0.381	-0.406	-0.374	-0.443**	-0.566***
Excess Kurtosis	1.974**	1.782	1.809	1.803	1.693	1.696	1.755**	3.821***
Panel B: GED								
Mean	-0.051***	-0.024*	-0.015	-0.017	-0.026*	-0.022	-0.05***	-0.051***
Variance	0.004***	0.004*	0.004	0.004	0.003*	0.003	-0.024***	0.12***
Skewness	-0.49***	-0.452*	-0.438	-0.444	-0.459*	-0.435	-0.476***	-0.602***
Excess Kurtosis	1.858***	1.763*	1.815	1.809	1.662*	1.689	1.661***	3.824***
Panel C: ST								
Mean	-0.042***	-0.009	0.002	0.001	-0.012	-0.007	-0.042***	-0.049***
Variance	0.006***	0.005	0.003	0.003	0.004	0.005	-0.016***	0.124***
Skewness	-0.473***	-0.414	-0.388	-0.396	-0.421	-0.393	-0.46***	-0.596***
Excess Kurtosis	1.944***	1.79	1.809	1.808	1.671	1.702	1.777***	3.855***
Panel D: T								
Mean	-0.059***	-0.03**	-0.019	-0.02	-0.032**	-0.027*	-0.059***	-0.062***
Variance	-0.007***	0.001**	0.003	0.002	-0.001**	0.002*	-0.024***	0.124***
Skewness	-0.511***	-0.473**	-0.452	-0.458	-0.476**	-0.451*	-0.501***	-0.638***
Excess Kurtosis	1.812***	1.783**	1.848	1.84	1.658**	1.707*	1.674***	3.856***
Panel E: N								
Mean	-0.051***	-0.007	0.004	0.002	-0.01	-0.004	-0.051***	-0.047***
Variance	0.001***	0.001	0	0	0.001	0	-0.035***	0.118***
Skewness	-0.482***	-0.391	-0.371	-0.378	-0.405	-0.372	-0.465***	-0.583***
Excess Kurtosis	1.775***	1.684	1.736	1.721	1.605	1.616	1.516***	3.807***

Notes   Table displays the GMM test statistics for the standardized residuals. The underlying data is the daily return series of the Euro Stoxx 50 for the period between 2001-01-03 and 2021-05-19.   The null hypothesis of the test for each variable are described as follows: $H_0 : E[z_t] = 0$ for the mean, $H_0 : E[z_t^2 - 1] = 0$ for the variance. $H_0 : E[z_t^3] = 0$ for the skewness and $H_0 : E[z_t^4 - 3] = 0$ for the excess kurtosis.

Figure 4.1 shows again that the normal distribution is not a good choice as an assumption for the innovation process. The best models seem to be the ST and the SGED. While this is just eye-balling, it should be noted that our results point in the same direction: skewed distributions seem to model the data better.
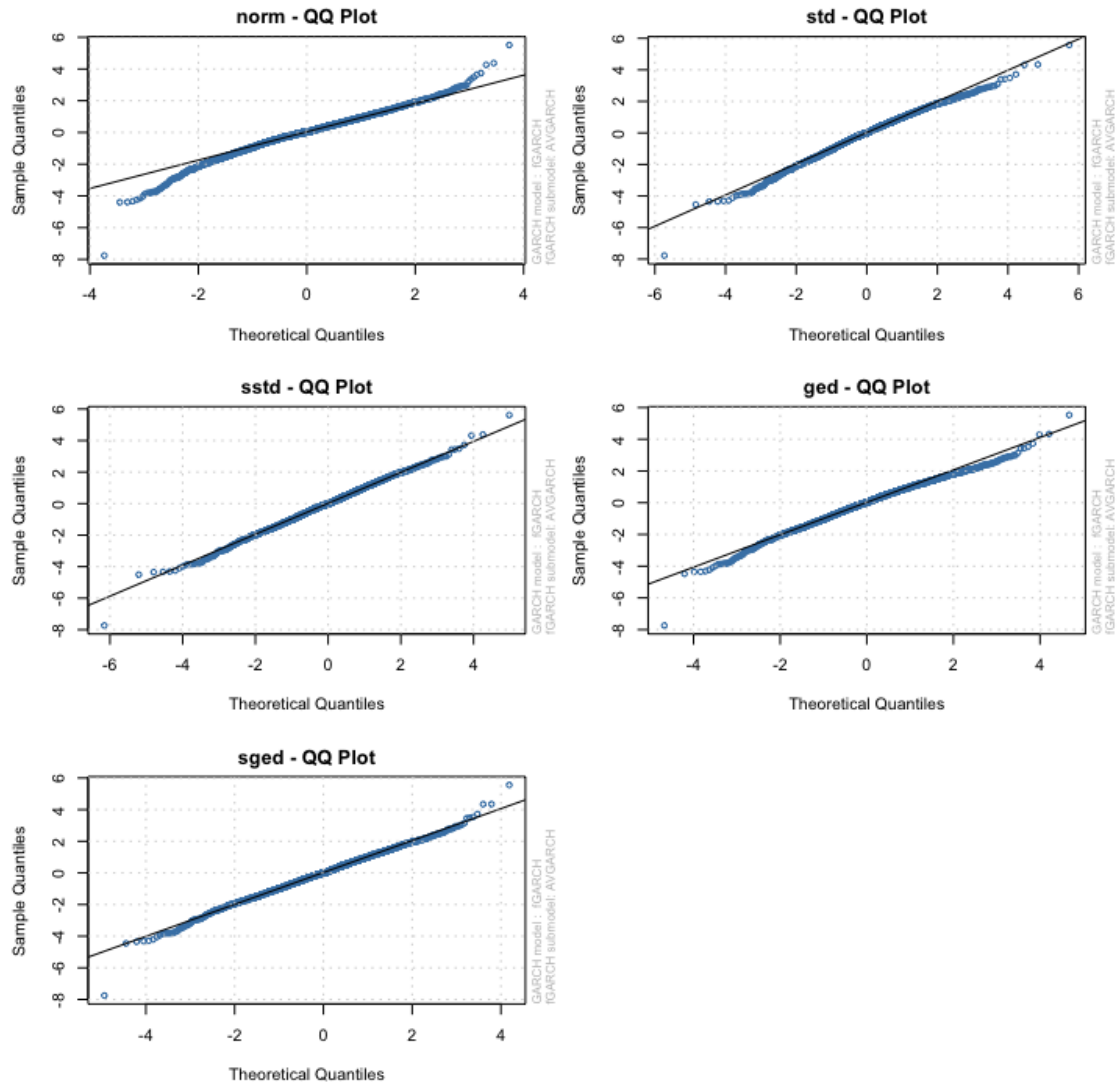

Figure 4.1: QQ plots of AVGARCH residuals using various distributions

5
Conclusion

This paper tested whether including higher-order moments in VaR and CVaR would make for more accurate risk measurement. The starting point for this was that there is a strong body of literature that suggests market returns are leptokurtic and right-skewed. We chose to focus on daily data and forecast one period ahead, as this is the most relevant for short-term trading desks and the requirements of regulators. Inspired by Bali, we selected 5 different distributions and 8 different GARCH models, for a total of 40 combinations, to estimate VaR and CVaR (different variations of ARMA order shown in the appendix B suggest that this only slighlthy affects results in comparison with the choice of the underlying distribution or GARCH model choice). We used daily data from the Euro Stoxx 50 return index between 2001 and 2021, a period which includes both major market crises from the 21th century: the Global Financial Crisis and the Covid-19 pandemic.

Our studied distributions are all limiting cases of the SGT-distribution studied by Bali, namely the SGED, GED, ST, T and normal distribution. To provide some intuition on which distributions most closely match the observed log returns, we performed maximum likelihood estimation and compared the Aikaike Information Criterions to penalize complexity. We found that the SGED distribution is as good

5. Conclusion

as the SGT (due to having one less parameter to estimate, even though it fits the data slightly worse). The normal distribution fitted the observed returns the worst, with both the maximum likelihood score and the AIC underperforming the other distributions. This affirmed previous literature that the normal distribution is inadequate to model returns.

As a second step we estimated the 8 GARCH models (SGARCH, IGARCH, EGARCH, EWMA, GJRGARCH, NAGARCH, TGARCH and AVGARCH) with the 5 underlying distributions in-sample and again reported the AIC. We found that the best model to forecast 1 day returns was the AVGARCH-ST. The ST distribution is generally the most performing over all GARCH models, with SGED being slightly better only for the SGARCH and IGARCH. We reported on the parameters of the different GARCH models for these two distributions, and find that the skewness and shape parameters are highly significant, suggesting that indeed higher order moments are relevant for estimating VaR and CVaR.

To avoid look-ahead bias as introduced by in-sample estimation, we applied a rolling window technique to 25 combinations (selecting the 5 best GARCH models on the basis of AIC). Here we refitted every 22 trading days and forecast 1 trading day. Subsequently we computed the VaR and cVaR and compared results for 4 performance tests: the ESTest McNeil and Frey (2000), Unconditional Coverage test, Conditional Coverage test and Dynamic Quantile test. We found that the smallest difference between predicted and actual exceedances over the VaR_{99} was for the NAGARCH-ST. For most other models the SGED distribution has the smallest difference. For the $CVaR_{99}$ and associated ESTest, the results are that the SGED distribution gives the smallest difference between predicted and actual exceedances for the TGARCH and the AVGARCH and the same as the ST distribution for EGARCH and GJRGARCH. For both distributions, the uc, cc and dq are insignificant. The other distributions have significant values for the cc and uc test and for the most part for the dq test and are thus not as suitable for (C)VaR

5. Conclusion

estimation.

To our knowledge, there has been little research on including time-varying moments beyond the variance in (C)VaR estimation. We allow for skewness and shape parameters to vary using an ACD model as proposed by Ghalanos (2016) We compared the results of the SGARCH-ST model with the ACD-ST model and concluded that allowing higher moments to vary over time gives better goodness-of-fit in terms of maximum log-likelihood and AIC and more parameters (especially the shape parameters) significant. However it did not improve the (C)VaR estimation for any of the 4 performance tests. Future research might study ACD models for different GARCH models, than the standard one, better (C)VaR tests.

We performed 4 different robustness checks on the in-sample GARCH estimations to test if the residuals are normally distributed. First, we performed the Ljung-Box Test on the residuals and reject the null-hypothesis that there is no serial correlation of variance. Second, we applied the ARCH LM test on absolute, and squared residuals of our GARCH models and came to the same conclusion as with the Ljung-Box test (these are similar tests, except that the former tests the ARMA process and the latter the ARCH process). Third, we performed the GMM test to see if the residuals have moments equal to those of of the normal distribution. We find that this hypothesis can be rejected with high significance. Last, we performed the Jarque-Bera test to test if the kurtosis and skewness matches the normal distribution. We reject this hypothesis as well. There is thus ample evidence that the residuals are non-normal.

To conclude by answering the research question "Do higher moments increase accuracy in the estimation of VaR and CVaR?", we can answer in the positive, though caution that more model complexity might not always mean better VaR and cVaR estimation.

Appendices

A
Appendix to literature review

Alternatives to the normal distribution

SGT (Skewed Generalized t-distribution)   The SGT distribution is introduced by Theodossiou (1998) and applied by Bali and Theodossiou (2007) and Bali, Mo, et al. (2008). According to Bali, Mo, et al. (2008) the proposed solutions (use of historical simulation, student's t-distribution, generalized error distribution or a mixture of two normal distributions) to the non-normality of standardized financial returns only partially solved the issues of accounting for skewness and leptokurtosis.The Pdf of the SGT distribution is given by eqution (A.1). B is the beta function (also called Euler integral).

$$f_{SGT}(x;\mu,\sigma,\xi,\kappa,\eta) = \frac{\kappa}{2v\sigma\eta^{1/\kappa}B\left(\frac{1}{\kappa},\eta\right)\left(\frac{|x-\mu+m|^{\kappa}}{\eta(v\sigma)^{\kappa}(\xi\,\mathrm{sign}(x-\mu+m)+1)^{\kappa}}+1\right)^{\frac{1}{\kappa}+\eta}}$$

$$m = \frac{2v\sigma\xi\eta^{\frac{1}{\kappa}}B\left(\frac{2}{\kappa},\eta-\frac{1}{\kappa}\right)}{B\left(\frac{1}{\kappa},\eta\right)}$$

$$\text{(A.1)}$$

$$v = \frac{\eta^{-\frac{1}{\kappa}}}{\sqrt{(3\xi^2+1)\frac{B\left(\frac{3}{\kappa},\eta-\frac{2}{\kappa}\right)}{B\left(\frac{1}{\kappa},\eta\right)}-4\xi^2\frac{B\left(\frac{2}{\kappa},\eta-\frac{1}{\kappa}\right)^2}{B\left(\frac{1}{\kappa},\eta\right)^2}}}$$
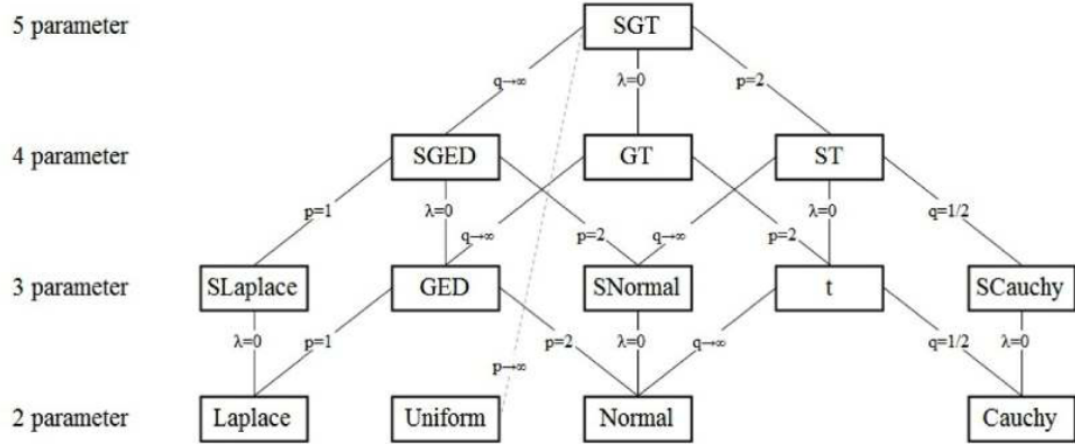
A. Appendix to literature review

Figure A.1: SGT distribution and limiting cases

Following Theodossiou (1998) however, there are two parameters, κ[1]and η[2]) for the shape in the SGT distribution. κ is the peakedness parameter. η is the tail-thickness parameter. It is equal to the degrees of freedom ν divided by 2 if $\xi = 0$ and $\kappa = 2$. As shown in the following figure[3] A.1 by Carter Davis, from the SGT the other distributions in the figure are limiting cases of the SGT.

Student's t-distribution A common alternative for the normal distribution is the Student t distribution. Similarly to the normal distribution, it is also symmetric (skewness is equal to zero if $\nu > 3$). The probability density function (pdf), consistent with Ghalanos (2020a), is given by equation (A.2). As will be seen in 1.2, GARCH models are used for volatility modeling in practice. Bollerslev (1987) examined the use of the GARCH-Student or GARCH-t model as an alternative to the standard Normal distribution, which relaxes the assumption of conditional normality by assuming the standardized innovation to follow a standardized Student t-distribution (Bollerslev 2008).

---

[1]Referred to as κ by Theodossiou (1998) and Bali, Mo, et al. (2008), but p by Carter Davis in the "sgt" package.

[2]Also referred to as n by Theodossiou (1998) and η by Bali, Mo, et al. (2008), but q by Carter Davis in the "sgt" packages.

[3]Source: https://cran.r-project.org/web/packages/sgt

$$f(x; \mu, \sigma, \nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\sigma\pi\nu}} \left(1 + \frac{(x-\mu)^2}{\sigma\nu}\right)^{-(\nu+1)/2} \tag{A.2}$$

where μ, σ and ν are respectively the mean, scale and shape (tail-thickness) parameters. $\nu/2$ is equal to the η^4 parameter of the SGT distribution with other restrictions (see part A). The symbol Γ is the Gamma function.

Unlike the normal distribution, which depends on two parameters only, the student t distribution allows for fatter tails. This kurtosis coefficient is given by equation (A.3) if $\nu > 4$. This is useful while the standardized residuals of stock returns appear to have fatter tails than the normal distribution following Bollerslev (2008).

$$kurt = 3 + \frac{6}{\nu - 4} \tag{A.3}$$

Generalized Error Distribution   The GED distribution is nested in the generalized t distribution by McDonald and Newey (1988) and is used in the GED-GARCH model by Nelson (1991) to model stock market returns. This model replaced the assumption of conditional normally distributed error terms by standardized innovations that following a generalized error distribution. It is a symmetric, unimodal distribution (location parameter is the mode, median and mean). This is also sometimes called the exponential power distribution (Bollerslev 2008). The conditional density (pdf) is given by equation (A.4) following Ghalanos (2020a).

$$f(x; \mu, \sigma, \kappa) = \frac{\kappa e^{-\frac{1}{2}\left|\frac{x-\mu}{\sigma}\right|^\kappa}}{2^{1+1/\kappa}\sigma\Gamma(1/\kappa)} \tag{A.4}$$

where μ, σ and κ are respectively the mean, scale and shape parameters.

---

[4]Also referred to as n by Theodossiou (1998) or q by Carter Davis in the "sgt" package.

Skewed t-distribution The density function can be derived following Fernández and Steel (1998) who showed how to introduce skewness into uni-modal standardized distributions (Trottier and Ardia 2015). The first equation from Trottier and Ardia (2015), here equation (A.5) presents the skewed t-distribution.

$$f_\xi(z) \equiv \frac{2\sigma_\xi}{\xi + \xi^{-1}} f_1(z_\xi), \quad z_\xi \equiv \begin{cases} \xi^{-1}(\sigma_\xi z + \mu_\xi) & \text{if } z \geq -\mu_\xi/\sigma_\xi \\ \xi(\sigma_\xi z + \mu_\xi) & \text{if } z < -\mu_\xi/\sigma_\xi \end{cases} \tag{A.5}$$

where $\mu_\xi \equiv M_1(\xi - \xi^{-1})$, $\sigma_\xi^2 \equiv (1 - M_1^2)(\xi^2 + \xi^{-2}) + 2M_1^2 - 1$, $M_1 \equiv 2\int_0^\infty uf_1(u)du$ and ξ between 0 and ∞. $f_1(\cdot)$ is in this case equation (A.2), the pdf of the student t distribution coming to equation (A.6), which has the parameterization following the SGT parameters.

$$f_{ST}(x; \alpha, \beta, \xi, \eta) = \frac{\Gamma\left(\frac{1}{2} + \eta\right)}{\sqrt{\beta\pi\eta}\Gamma(\eta)\left(\frac{|x - \alpha + m|^2}{\eta\beta(\xi\,\text{sign}(x-\alpha+m)+1)^2} + 1\right)^{\frac{1}{2} + \eta}}$$

(A.6)

$$m = \frac{2\xi\sqrt{\beta\eta}\Gamma\left(\eta - \frac{1}{2}\right)}{\sqrt{\pi}\Gamma\left(\eta + \frac{1}{2}\right)}$$

According to Giot and Laurent (2003) as well as Giot and Laurent (2004), the skewed t-distribution outperforms the symmetric density distributions.

Skewed Generalized Error Distribution A further distribution to analyse is the SGED distribution of Theodossiou (2000). It is applied in GARCH models by Lee et al. (2008). The SGED distribution extends the Generalized Error Distribution (GED) to allow for skewness and leptokurtosis. The density function can be derived following Fernández and Steel (1998) who showed how to introduce skewness into uni-modal standardized distributions (Trottier and Ardia 2015). It can also be found in Theodossiou (2000). The pdf is then given by the same equation (A.5) as the skewed t-distribution but with $f_1(\cdot)$ equal to equation (A.4). To then get equation (A.7).

A. Appendix to literature review

$$f_{SGED}(x; \mu, \sigma, \xi, \kappa) = \frac{\kappa e^{-\left(\frac{|x-\mu+m|}{v\sigma(1+\xi \, \text{sig}(x-\mu+m))}\right)^{\kappa}}}{2v\sigma\Gamma(1/\kappa)}$$

$$m = \frac{2^{\frac{2}{\kappa}} v\sigma\xi\Gamma\left(\frac{1}{2}+\frac{1}{\kappa}\right)}{\sqrt{\pi}} \tag{A.7}$$

$$v = \sqrt{\frac{\pi\Gamma\left(\frac{1}{\kappa}\right)}{\pi(1+3\xi^2)\Gamma\left(\frac{3}{\kappa}\right)-16^{\frac{1}{\kappa}}\xi^2\Gamma\left(\frac{1}{2}+\frac{1}{\kappa}\right)^2\Gamma\left(\frac{1}{\kappa}\right)}}$$

ARCH models

From the components of the ARCH model described in part 1.2.2 we could look at the conditional moments (or expected returns and variance). We can plug in the component σ_t into the conditional mean innovation ε_t and use the conditional mean innovation to examine the conditional mean return. In equation (A.8) and (A.9) they are derived. Because the random variable z_t is distributed with a zero-mean, the conditional expectation is 0. As a consequence, the conditional mean return in equation (A.9) is equal to the unconditional mean in the most simple case. But variations are possible using ARMA (eg. AR(1)) processes.

$$\mathbb{E}_{t-1}(\varepsilon_t) = \mathbb{E}_{t-1}(\sqrt{\beta_0 + \beta_1 \times \varepsilon_{t-1}^2} \times z_t) = \sigma_t \mathbb{E}_{t-1}(z_t) = 0 \tag{A.8}$$

$$\mathbb{E}_{t-1}(y_t) = \alpha_0 + \mathbb{E}_{t-1}(\varepsilon_t) = \alpha_0 \tag{A.9}$$

For the conditional variance, knowing everything that happened until and including period $t-1$ the conditional innovation variance is given by equation (A.10). This is equal to σ_t^2, while the variance of z_t is equal to 1. Then it is easy to derive the conditional variance of returns in equation (A.11), that is why equation (1.3) is called the variance equation.

$$var_{t-1}(\varepsilon_t) = \mathbb{E}_{t-1}(\varepsilon_t^2) = \mathbb{E}_{t-1}(\sigma_t^2 \times z_t^2) = \sigma_t^2 \mathbb{E}_{t-1}(z_t^2) = \sigma_t^2 \tag{A.10}$$

$$var_{t-1}(y_t) = var_{t-1}(\varepsilon_t) = \sigma_t^2 \tag{A.11}$$

The unconditional variance is also interesting to derive, while this is the long-run variance, which will be derived in equation (A.15). After deriving this using the law of iterated expectations and assuming stationarity for the variance process, one would get equation (A.12) for the unconditional variance, equal to the constant β_0 and divided by $1 - \beta_1$, the slope of the variance equation.

$$\sigma^2 = \frac{\beta_0}{1 - \beta_1} \tag{A.12}$$

This leads to the properties of ARCH models: Stationarity[5] condition for variance: $\beta_0 > 0$ and $0 \leq \beta_1 < 1$. But also, zero-mean innovations and uncorrelated innovations. Thus a weak white noise process ε_t. The unconditional 4th moment, kurtosis $\mathbb{E}(\varepsilon_t^4)/\sigma^4$ of an ARCH model is given by equation (A.13). This term is larger than 3, which implicates fat-tails.

$$3\frac{1 - \beta_1^2}{1 - 3\beta_1^2} \tag{A.13}$$

Another property of ARCH models is that it takes into account volatility clustering. Because we know that $var(\varepsilon_t) = \mathbb{E}(\varepsilon_t^2) = \sigma^2 = \beta_0/(1 - \beta_1)$, we can plug in β_0 for the conditional variance $var_t(\varepsilon_{t+1}) = \mathbb{E}(\varepsilon_{t+1}^2) = \sigma_{t+1}^2 = \beta_0 + \beta_1 \times \varepsilon_t^2$. Thus it follows that equation (A.14) displays volatility clustering. If we examine the RHS, as $\beta_1 > 0$ (condition for stationarity), when shock ε_t^2 is larger than what you expect it to be on average σ^2 the LHS will also be positive. Then the conditional variance will be larger than the unconditional variance. Briefly, large shocks will be followed by more large shocks.

$$\sigma_{t+1}^2 - \sigma^2 = \beta_1 \times (\varepsilon_t^2 - \sigma^2) \tag{A.14}$$

---

[5]Stationarity implies that the series on which the ARCH model is used does not have any trend and has a constant expected mean. Only the conditional variance is changing.

A. Appendix to literature review

Excess kurtosis can be modeled, even when the conditional distribution is assumed to be normally distributed. The third moment, skewness, can be introduced using a skewed conditional distribution as we saw in part A. The serial correlation for squared innovations is positive if fourth moment exists (equation (A.13), this is volatility clustering once again.

How will then the variance be forecasted? Well, the conditional variance for the k-periods ahead , denoted as period $T + k$, is given by equation (A.15). This can already be simplified, while we know that $\sigma^2_{T+1} = \beta_0 + \beta_1 \times \varepsilon^2_T$ from equation (1.3).

$$
\begin{aligned}
\mathbb{E}_T(\varepsilon^2_{T+k}) &= \beta_0 \times (1 + \beta_1 + ... + \beta_1^{k-2}) + \beta_1^{k-1} \times \sigma^2_{T+1} \\
&= \beta_0 \times (1 + \beta_1 + ... + \beta_1^{k-1}) + \beta_1^{k} \times \sigma^2_T
\end{aligned}
\tag{A.15}
$$

It can be shown that then the conditional variance in period $T+k$ is equal to equation (A.16). The LHS is the predicted conditional variance k-periods ahead above its unconditional variance, σ^2. The RHS is the difference current last-observed return residual ε^2_T above the unconditional average multiplied by β_1^k, a decreasing function of k (given that $0 \leq \beta_1 < 1$). The further ahead predicting the variance, the closer β_1^k comes to zero, the closer to the unconditional variance, i.e. the long-run variance.

$$
\mathbb{E}_T(\varepsilon^2_{T+k}) - \sigma^2 = \beta_1^k \times (\varepsilon^2_T - \sigma^2)
\tag{A.16}
$$

GARCH models

All the GARCH models are estimated using the package "rugarch" by Ghalanos (2020b). We use specifications similar to Ghalanos (2020a). Parameters have to be restricted so that the variance output always is positive, except for the EGARCH model, as this model mathematically ensures the output is positive.

Symmetric (normal) GARCH model The standard GARCH model (Bollerslev 1986) is written consistent with Ghalanos (2020a) as in equation (A.17) without external regressors.

$$\sigma_t^2 = \beta_0 + \beta_1 \varepsilon_{t-1}^2 + \beta_2 \sigma_{t-1}^2 \tag{A.17}$$

where σ_t^2 denotes the conditional variance, β_0 the intercept and ε_t^2 the residuals from the used mean process. The GARCH order is defined by (q, p) (ARCH, GARCH), which is here $(1, 1)$. As Ghalanos (2020a) describes: "one of the key features of the observed behavior of financial data which GARCH models capture is volatility clustering which may be quantified in the persistence parameter $\hat{P}$ specified as in equation (A.18) for a GARCH model of order $(1, 1)$.

$$\hat{P} = \beta_1 + \beta_2. \tag{A.18}$$

The unconditional variance of the standard GARCH model of Bollerslev is very similar to the ARCH model, but with the Garch parameter (β_2) included as in equation (A.19).

$$\begin{aligned} \hat{\sigma}^2 &= \frac{\beta_0}{1 - \hat{P}} \\ &= \frac{\beta_0}{1 - \beta_1 - \beta_2} \end{aligned} \tag{A.19}$$

IGARCH model Following Ghalanos (2020a), the integrated GARCH model (Bollerslev 1986) can also be estimated. This model assumes the persistence $\hat{P} = 1$. This is done by Ghalanos, by setting the sum of the ARCH and GARCH parameters to 1. Because of this unit-persistence, the unconditional variance cannot be calculated.

GJRGARCH model The GJRGARCH model (Glosten et al. 1993), which is an alternative for the asymmetric GARCH (AGARCH) by Engle (1990) and Engle and Ng (1993), models both positive as negative shocks on the conditional variance asymmetrically by using an indicator variable $I_t - 1$, it is specified as in equation (A.20).

A. Appendix to literature review

$$\sigma_t^2 = \beta_0 + \beta_1 \varepsilon_{t-1}^2 + \gamma_j I_{t-1} \varepsilon_{t-1}^2 + \beta_2 \sigma_{t-1}^2 \tag{A.20}$$

where γ_j represents the *leverage* term. The indicator function I takes on value 1 for $\varepsilon \leq 0$, 0 otherwise. Because of the indicator function, persistence of the model now crucially depends on the asymmetry of the conditional distribution used according to Ghalanos (2020a).

EGARCH model   The EGARCH model or exponential GARCH model (Nelson 1991) is defined as in equation (A.21). The advantage of the EGARCH model is that there are no parameter restrictions, since the output is log variance (which cannot be negative mathematically), instead of variance.

$$\log_e(\sigma_t^2) = \beta_0 + \beta_1 z_{t-1} + \gamma_1(|z_{t-1}| - E|z_{t-1}|) + \beta_2 \log_e(\sigma_{t-1}^2) \tag{A.21}$$

where β_1 captures the sign effect and γ_j the size effect.

NAGARCH model   The NAGARCH or nonlinear asymmetric model (Engle and Ng 1993). It is specified as in equation (A.22). The model is *asymmetric* as it allows for positive and negative shocks to differently affect conditional variance and *nonlinear* because a large shock is not a linear transformation of a small shock.

$$\sigma_t^2 = \beta_0 + \beta_1(\varepsilon_{t-1} + \gamma_1\sqrt{\sigma_{t-1}})^2 + \beta_2 \sigma_{t-1}^2 \tag{A.22}$$

As before, γ_1 represents the *leverage* term.

TGARCH model   The TGARCH or threshold model (Zakoian 1994) also models asymmetries in volatility depending on the sign of the shock, but contrary to the GJRGARCH model it uses the conditional standard deviation instead of conditional variance. It is specified as in (A.23).

$$\sigma_t = \beta_0 + \beta_1^+ \varepsilon_{t-1}^+ - \beta_1^- \varepsilon_{t-1}^- + \beta_2 \sigma_{t-1} \tag{A.23}$$

A. Appendix to literature review

where ε_{t-1}^{+} is equal to ε_{t-1} if the term is negative and equal to 0 if the term is positive. The reverse applies to ε_{t-1}^{-}. They cite Davidian and Carroll (1987) who find that using volatility instead of variance as scaling input variable gives better variance estimates. This is due to absolute residuals (contrary to squared residuals with variance) more closely predicting variance for non-normal distributions.

TSGARCH model   The absolute value Garch model or TS-Garch model, as named after Taylor (1986) and Schwert (1989), models the conditional standard deviation and is intuitively specified like a normal GARCH model, but with the absolute value of the shock term. It is specified as in (A.24).

$$\sigma_t = \beta_0 + \beta_1 \left| \varepsilon_{t-1} \right| + \beta_2 \sigma_{t-1} \tag{A.24}$$

EWMA   An alternative to the series of GARCH models is the exponentially weighted moving average or EWMA model. This model calculates conditional variance based on the shocks from previous periods. The idea is that by including a smoothing parameter λ more weight is assigned to recent periods than distant periods. The λ must be less than 1. It is specified as in (A.25).

$$\sigma_t^2 = (1 - \lambda) \sum_{j=1}^{\infty} (\lambda^j \varepsilon_{t-j}^2) \tag{A.25}$$

In practice a λ of 0.94 is often used, such as by the financial risk management company RiskMetrics^{TM} model of J.P. Morgan (Morgan Guaranty Trust Company 1996).

B

Appendix to findings

MLE fitted distributions

Below in figure B.1 are the plots of the distributions fitted to the Euro Stoxx 50 data using Maximum Likelihood Estimation.
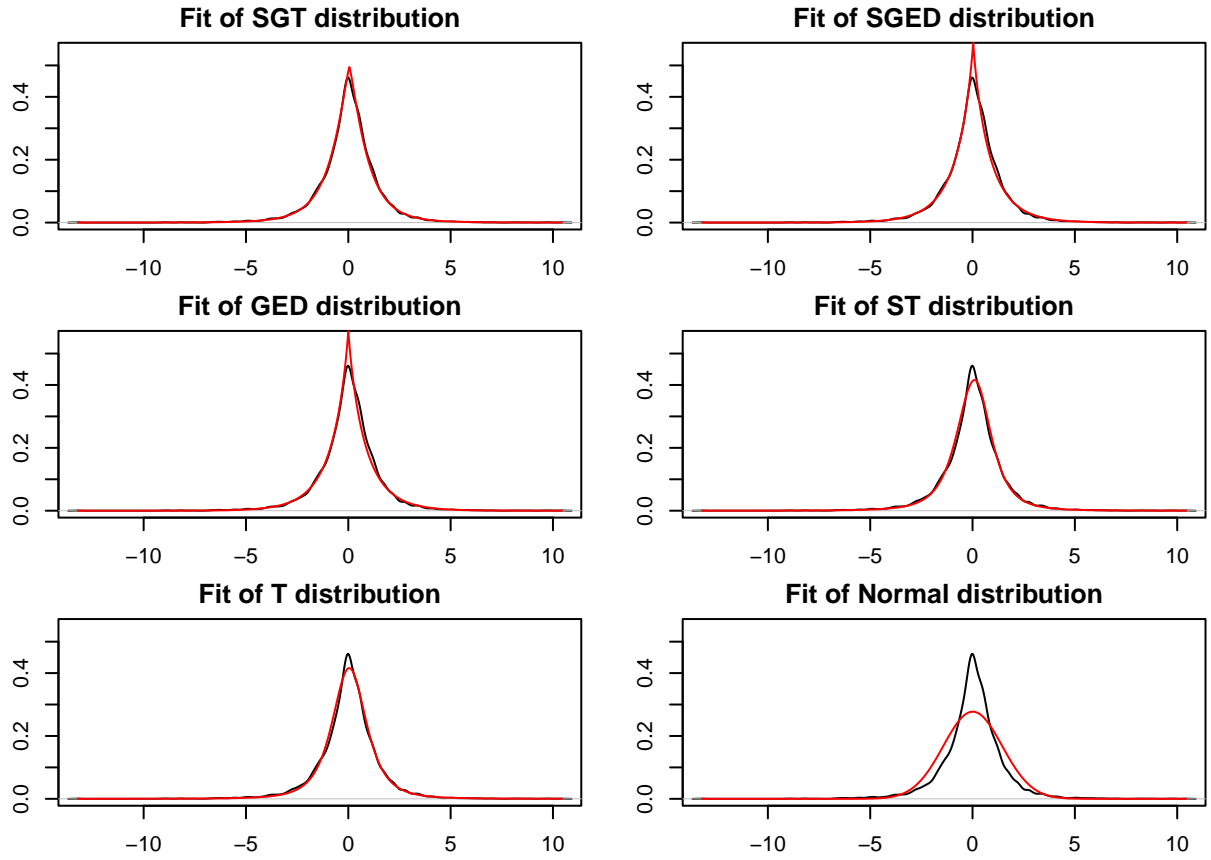
Figure B.1: Theoretical (red) vs. Actual (black) distributions of Euro Stoxx 50

Goodness of fit

As already mentioned, next to testing the models in part 3, we also tested other models using the different distributions. This we did in order to check if distributions that capture the higher moment effects are really better in terms of goodness of fit. We did a small data mining experiment with 140 models that were estimated. This can lead to overfitting because of the fit in-sample. However, we can decide if there is a trend using the different distributions for the several GARCH models. Thus, in this experiment, our rule of thumb was to examine general trends. Six cases were examined.

B. Appendix to findings

First, in figure B.2, symmetric GARCH with symmetric distributions are looked at. As you can see the general error distribution (GED) seems to perform slightly better than the the T distribution. While the student's t distribution (T) performs better than the normal distribution (NORM) according to the AIC. Which is consistent with the literature that found that the assumption of the normal distribution is a rather poor assumption.
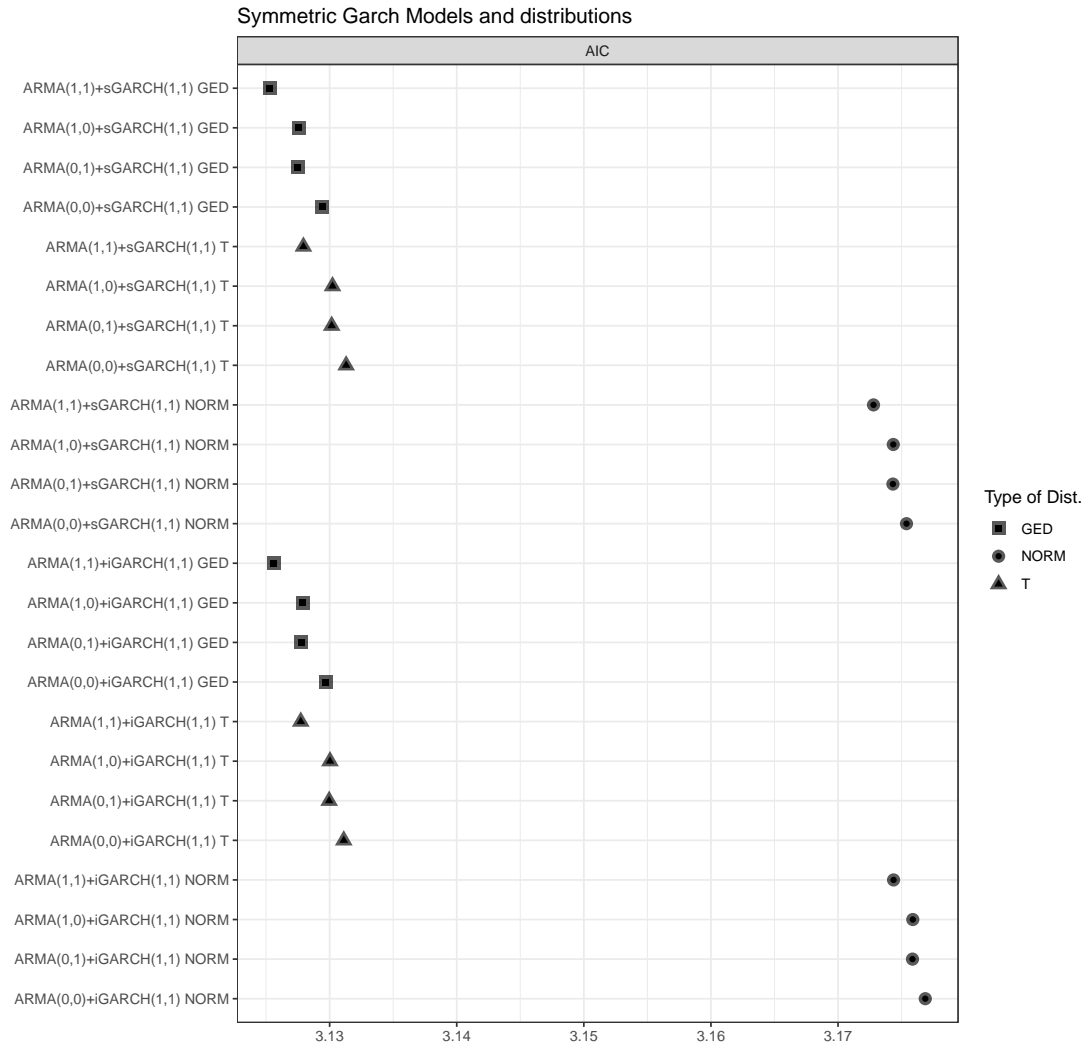

Figure B.2: Goodness of fit symmetric GARCH and distributions

B. Appendix to findings

Second, in figure B.3, symmetric GARCH models with the best symmetric distribution (T) and other distributions (SGED, ST) are looked at. As you can see the SGED seems to perform better than the skewed student's t distribution (ST) and the latter is better than the scaled T distribution. This is consistent with Giot and Laurent (2003) that the skewed student's t (ST) distribution outperforms the symmetric distributions.
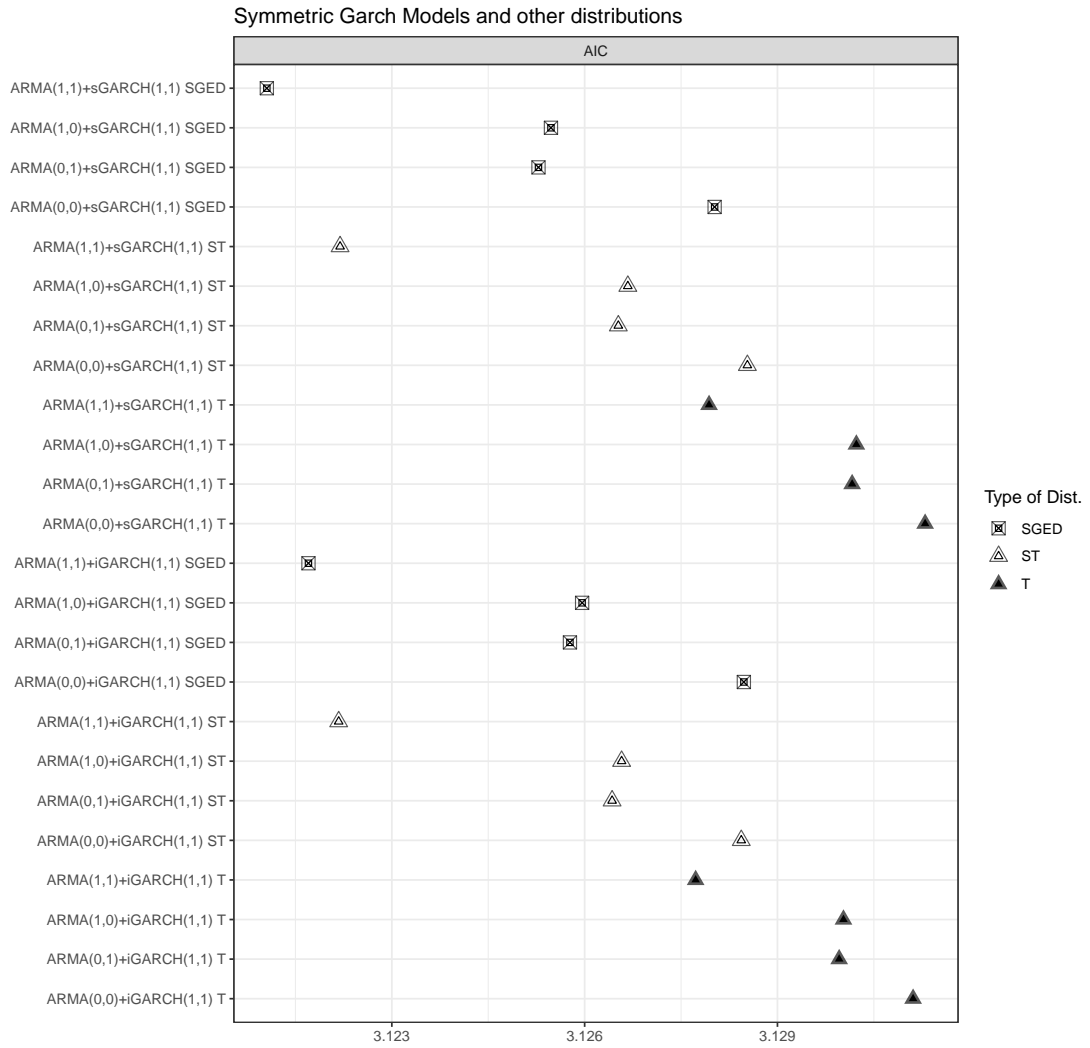

Figure B.3: Goodness of fit symmetric GARCH and other distributions

In figure B.4 you can see the same patter as in figure B.2, the student's t distribution performs best among the symmetric distributions.
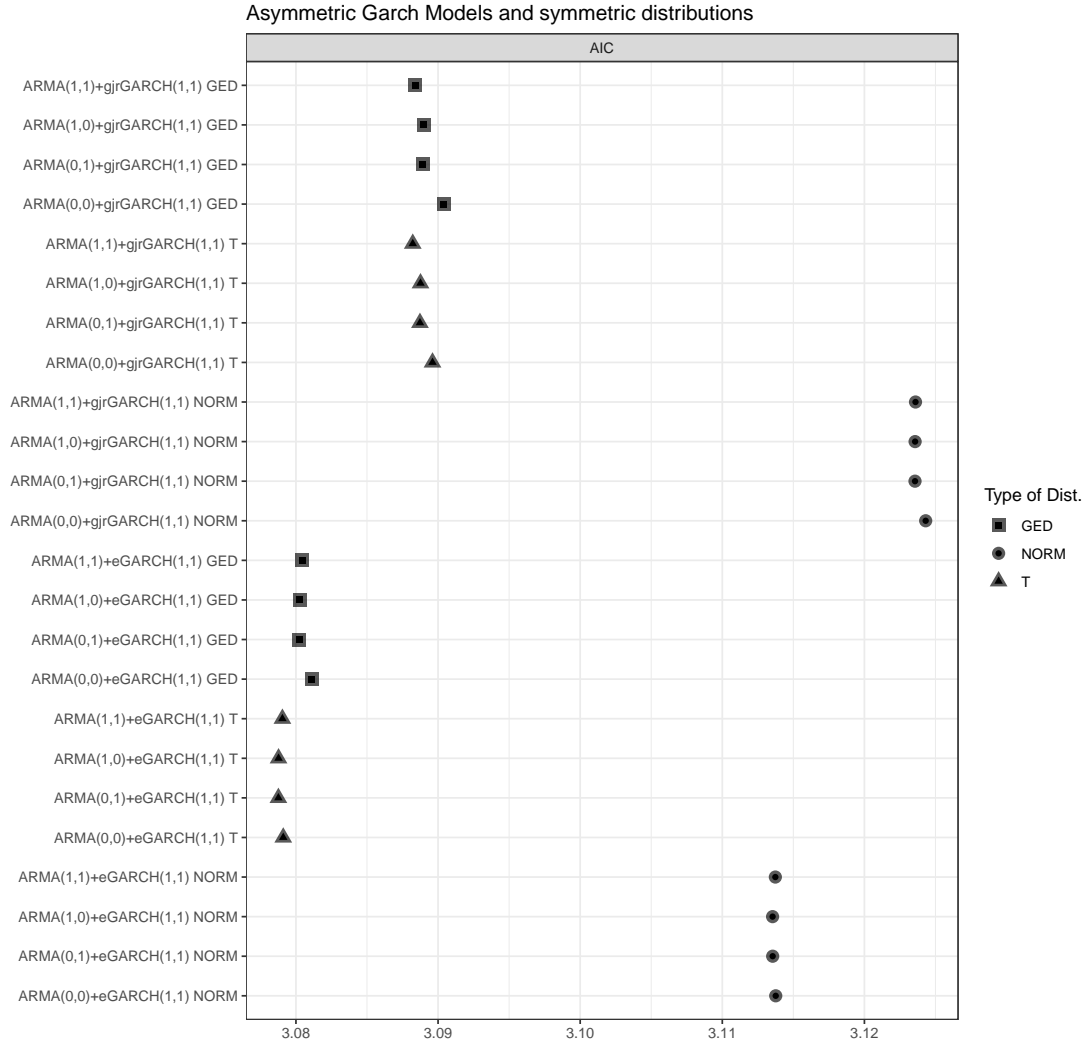

Figure B.4: Goodness of fit asymmetric GARCH and symmetric distributions

B. Appendix to findings

Then, in figure B.5 the same patter arises as in figure B.3, the skewed student's t distribution again seems to be the most optimal one to use. Therefore the ST distribution is chosen as final model for the Euro Stoxx 50 index.
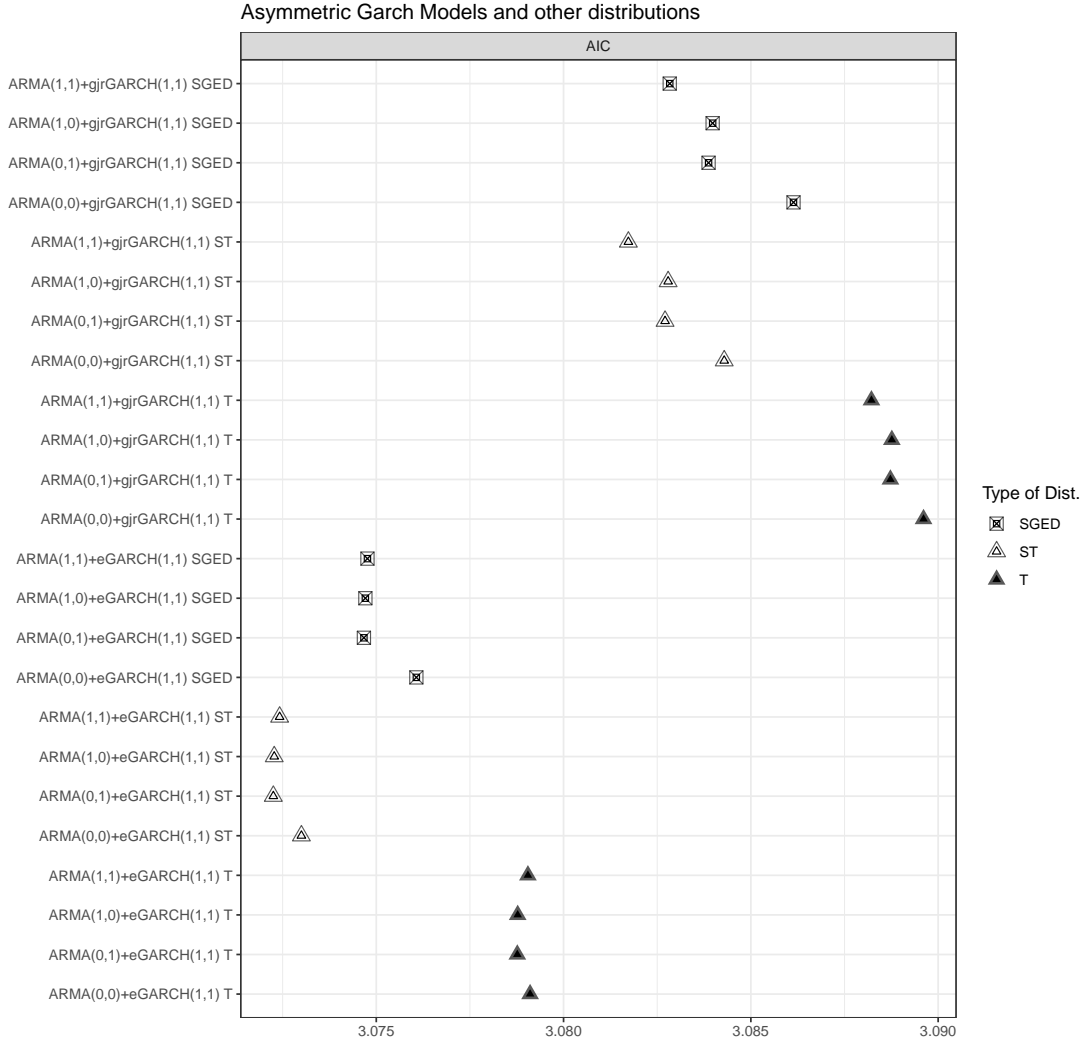

Figure B.5: Goodness of fit asymmetric GARCH and symmetric distributions

In two additional figures the family garch models (TGARCH, NAGARCH and AVGARCH) are examined, the same patterns were observed as above[1].

---

[1]We have to note that for some models like TGARCH and AVGARCH with SGED distribution the AIC was double of other models and therefore these models seem to work very poorly or are misspecified.

Works Cited

Alexander, Carol. (2008). *Market risk analysis. Volume 2, Practical financial econometrics*. 2nd ed. The Wiley Finance Series. Chichester, England ; Wiley.

Annaert, Jan (Jan. 2021). *Quantitative Methods in Finance*. Version 0.2.1. Antwerp Management School.

Ardia, David, Kris Boudt, and Leopoldo Catania (2019). "Downside risk evaluation with the R package GAS". In: *R Journal* 10.2, pp. 410–421. DOI: 10.32614/RJ-2018-064.

Artzner, Philippe et al. (1996). *Coherent measures of risk*. Tech. rep.

Bali, Turan G., Hengyong Mo, and Yi Tang (Feb. 2008). "The role of autoregressive conditional skewness and kurtosis in the estimation of conditional VaR". In: *Journal of Banking and Finance* 32.2. Publisher: North-Holland, pp. 269–282. DOI: 10.1016/j.jbankfin.2007.03.009.

Bali, Turan G. and Panayiotis Theodossiou (Feb. 22, 2007). "A conditional-SGT-VaR approach with alternative GARCH models". In: *Annals of Operations Research* 151.1, pp. 241–267. DOI: 10.1007/s10479-006-0118-4. URL: http://link.springer.com/10.1007/s10479-006-0118-4.

Basel Committee on Banking Supervision (2016). *Minimum capital requirements for market risk*. Tech. rep. Issue: January Publication Title: Bank for International Settlements, pp. 92–92. URL: https://www.bis.org/basel_framework/.

Bertsimas, Dimitris, Geoorey J Lauprete, and Alexander Samarov (2004). "Shortfall as a risk measure: properties, optimization and applications". In: *Journal of Economic Dynamics and Control* 28, pp. 1353–1381. DOI: 10.1016/S0165-1889(03)00109-X.

Bollerslev, Tim (1986). "Generalized Autoregressive Conditional Heteroskedasticity". In: *Journal of Econometrics* 31, pp. 307–327.

— (1987). "A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return". In: *The Review of Economics and Statistics* 69.3. Publisher: The MIT Press, pp. 542–547. DOI: 10.2307/1925546. URL: https://www.jstor.org/stable/1925546.

— (Sept. 4, 2008). "Glossary to ARCH (GARCH)". In: p. 46. DOI: 10.2139/ssrn.1263250. URL: https://ssrn.com/abstract=1263250.

Brooks, Chris et al. (2005). "Autoregressive conditional kurtosis". In: *Journal of Financial Econometrics* 3.3, pp. 399–421. DOI: 10.1093/jjfinec/nbi018.

Calculation guide STOXX ® (2020). Tech. rep.

Christoffersen, Peter, Jinyong Hahn, and Atsushi Inoue (July 2001). "Testing and comparing Value-at-Risk measures". In: *Journal of Empirical Finance* 8.3, pp. 325–342. DOI: 10.1016/S0927-5398(01)00025-1.

Davidian, M. and R. J. Carroll (Dec. 1987). "Variance Function Estimation". In: *Journal of the American Statistical Association* 82.400. Publisher: JSTOR, pp. 1079–1079. DOI: 10.2307/2289384.

Works Cited

Engle, R. F. (1982). "Autoregressive Conditional Heteroscedacity with Estimates of variance of United Kingdom Inflation,journal of Econometrica, Volume 50, Issue 4 (Jul., 1982),987-1008." In: *Econometrica* 50.4, pp. 987–1008.

Engle, Robert (2001). "GARCH 101: The use of ARCH/GARCH models in applied econometrics". In: *Journal of Economic Perspectives.* DOI: `10.1257/jep.15.4.157`.

Engle, Robert F and Simone Manganelli (Oct. 1, 2004). "CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles". In: *Journal of Business and Economic Statistics* 22.4, pp. 367–381. DOI: `10.1198/073500104000000370`. URL: `https://www.tandfonline.com/doi/full/10.1198/073500104000000370`.

Engle, Robert F. (1990). *Stock Volatility and the Crash of '87: Discussion.* Tech. rep. Issue: 1 Publication Title: The Review of Financial Studies Volume: 3, pp. 103–106. URL: `https://www.jstor.org/stable/2961959%0A`.

Engle, Robert F. and Victor K. Ng (Dec. 1993). "Measuring and Testing the Impact of News on Volatility". In: *The Journal of Finance* 48.5. Publisher: John Wiley and Sons, Ltd, pp. 1749–1778. DOI: `10.1111/j.1540-6261.1993.tb05127.x`.

Fama, Eugene (1970). *Efficient Capital Markets: A Review of Theory and Empirical Work.* Tech. rep. 2, pp. 383–417. DOI: `10.2307/2325486`.

Fama, Eugene F. (1965). "The Behavior of Stock-Market Prices". In: *The Journal of Business* 38.1, pp. 34–105. URL: `http://www.jstor.org/stable/2350752`.

Fernández, Carmen and Mark F. J. Steel (Mar. 1998). "On Bayesian Modeling of Fat Tails and Skewness". In: *Journal of the American Statistical Association* 93.441, pp. 359–371. DOI: `10.1080/01621459.1998.10474117`. URL: `http://www.tandfonline.com/doi/abs/10.1080/01621459.1998.10474117`.

Ghalanos, Alexios (2016). *racd: Autoregressive Conditional Density Models.* http://www.unstarched.net, https://bitbucket.org/alexiosg/.

— (2020a). *Introduction to the rugarch package. (Version 1.4-3).* URL: `http://cran.r-project.org/web/packages/rugarch/`.

— (2020b). *rugarch: Univariate GARCH models.* R package version 1.4-4.

Giot, Pierre & Sébastien Laurent (Nov. 2003). "Value-at-risk for long and short trading positions". In: *Journal of Applied Econometrics* 18.6, pp. 641–663. DOI: `10.1002/jae.710`. URL: `http://doi.wiley.com/10.1002/jae.710`.

— (June 1, 2004). "Modelling daily Value-at-Risk using realized volatility and ARCH type models". In: *Journal of Empirical Finance* 11.3, pp. 379–398. DOI: `10.1016/j.jempfin.2003.04.003`. URL: `https://www.sciencedirect.com/science/article/pii/S092753980400012X`.

Glosten, Lawrence R., Ravi Jagannathan, and David E. Runkle (Dec. 1993). "On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks". In: *The Journal of Finance* 48.5. Publisher: John Wiley and Sons, Ltd, pp. 1779–1801. DOI: `10.1111/j.1540-6261.1993.tb05128.x`. URL: `http://doi.wiley.com/10.1111/j.1540-6261.1993.tb05128.x`.

Hansen, Bruce E. (1994). "Autoregressive Conditional Density Estimation". In: *International Economic Review* 35.3, pp. 705–730.

Hansen, Lars Peter (1982). "Large Sample Properties of Generalized Method of Moments Estimators". In: *Econometrica* 50.4, pp. 1029–1054. DOI: `https://doi.org/10.2307/1912775`. URL: `https://www.jstor.org/stable/1912775`.

Harvey, Campbell R and Akhtar Siddique (1999). "Conditional Skewness in Asset Pricing Tests". In: p. 41.

Works Cited

Holton, Glyn A (2002). "History of Value-at-Risk: 1922-1998". In: Contingency Analysis Working Paper. URL: http://www.contingencyanalysis.com.

Jorion, Philippe (2007). *Value at Risk: The New Benchmark For Managing Financial Risk.* 3rd ed. McGraw-Hill.

Kupiec, P.H. (1995). "Techniques for Verifying the Accuracy of Risk Measurement Models". In: *Journal of Derivatives* 3.2, pp. 73–84. DOI: 10.3905/jod.1995.407942.

Lee, Ming-Chih, Jung-Bin Su, and Hung-Chun Liu (Oct. 22, 2008). "Value-at-risk in US stock indices with skewed generalized error distribution". In: *Applied Financial Economics Letters* 4.6, pp. 425–431. DOI: 10.1080/17446540701765274. URL: http://www.tandfonline.com/doi/abs/10.1080/17446540701765274.

Lyngs, Ulrik (2019). *oxforddown: An Oxford University Thesis Template for R Markdown.* https://github.com/ulyngs/oxforddown. DOI: 10.5281/zenodo.3484682.

Mandelbrot, Benoit (1963). "The Variation of Certain Speculative Prices". In: *The Journal of Business.* University of Chicago Press 36, p. 394. DOI: 10.1086/294632.

Markowitz, Harry (1952). "Portfolio Selection". In: *Journal of Finance* 7.1, pp. 77–91. DOI: 10.1111/j.1540-6261.1952.tb01525.x.

McDonald, James B. and Whitney K. Newey (Dec. 1988). "Partially Adaptive Estimation of Regression Models via the Generalized T Distribution". In: *Econometric Theory* 4.3, pp. 428–457. DOI: 10.1017/S0266466600013384. URL: https://www.cambridge.org/core/product/identifier/S0266466600013384/type/journal_article.

McNeil, Alexander J. and Rüdiger Frey (Nov. 2000). "Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach". In: *Journal of Empirical Finance* 7.3-4, pp. 271–300. DOI: 10.1016/S0927-5398(00)00012-8.

Morgan Guaranty Trust Company (1996). *RiskMetricsTM—Technical Document.* Tech. rep.

Nelson, Daniel B. (Mar. 1991). "Conditional Heteroskedasticity in Asset Returns: A New Approach". In: *Econometrica* 59.2. Publisher: JSTOR, pp. 347–347. DOI: 10.2307/2938260.

Officer, R. R. (1972). *The Distribution of Stock Returns.* Tech. rep. 340, pp. 807–812.

R Core Team (2019). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing. Vienna, Austria. URL: https://www.R-project.org/.

Roy, A. D. (July 1952). "Safety First and the Holding of Assets". In: *Econometrica* 20.3. Publisher: JSTOR, pp. 431–431. DOI: 10.2307/1907413.

Schwert, G. William (1989). "Why Does Stock Market Volatility Change Over Time?" In: *The Journal of Finance* 44.5, pp. 1115–1153. DOI: 10.1111/j.1540-6261.1989.tb02647.x.

Subbotin, M.T. (1923). "On the Law of Frequency of Error." In: *Matematicheskii Sbornik* 31, pp. 296–301.

Taylor, Stephen J. (1986). *Modelling financial time series.* Chichester: John Wiley and Sons, Ltd.

Theodossiou, Panayiotis (1998). "Financial data and the skewed generalized t distribution". In: *Management Science* 44.12 part 1. Publisher: INFORMS Inst.for Operations Res.and the Management Sciences, pp. 1650–1661. DOI: 10.1287/mnsc.44.12.1650.

Works Cited

Theodossiou, Panayiotis (2015). "Skewed Generalized Error Distribution of Financial Assets and Option Pricing". In: *Multinational Finance Journal* 19.4, pp. 223–266. DOI: 10.17578/19-4-1.

Theodossiou, Peter (2000). "Skewed Generalized Error Distribution of Financial Assets and Option Pricing". In: *SSRN Electronic Journal.* DOI: 10.2139/ssrn.219679. URL: http://www.ssrn.com/abstract=219679.

Trottier, Denis-Alexandre and David Ardia (Sept. 4, 2015). "Moments of standardized Fernandez-Steel skewed distributions: Applications to the estimation of GARCH-type models". In: *Finance Research Letters* 18, pp. 311–316. DOI: 10.2139/ssrn.2656377. URL: https://ssrn.com/abstract=2656377.

Welch, Ivo and Amit Goyal (July 2008). "A Comprehensive Look at The Empirical Performance of Equity Premium Prediction". In: *Review of Financial Studies* 21.4, pp. 1455–1508. DOI: 10.1093/rfs/hhm014. URL: https://academic.oup.com/rfs/article-lookup/doi/10.1093/rfs/hhm014.

Zakoian, Jean Michel (1994). "Threshold heteroskedastic models". In: *Journal of Economic Dynamics and Control* 18.5, pp. 931–955. DOI: 10.1016/0165-1889(94)90039-6.