# 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, many thanks to our families and loved ones that supported us during the writing of this thesis. Secondly, thank you professors Zhang [4], Annaert[5] and De Ceuster[6] for the valuable insights during courses you have given us in preparation of this thesis, the dozens of assignments using the R language and the many questions answered this year. We must be grateful for the classes of R programming by prof Zhang.

Secondly, we have to thank the developer of the software we used for our thesis. A profuse thanks to Allaire, the founder and CEO of RStudio. Thanks for making data science easier, more accessible and fun. We must also be grateful to the inventors of "Markdown", "Pandoc", "knitr", "bookdown", "thesisdown". Then, we must say thanks to Ulrik Lyngs who made it a bit easier to work together in R with a pre-build template for the university of Oxford, also 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 process of understanding the whole process and doing the analysis for our thesis.

Enjo Faes,
Stephane Mertens de Wilmars,
Filippo Pratesi
Antwerp Management School, Antwerp
27 June 2021

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

# Abstract

The greatest abstract all times

# 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

# 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) explains that the question in academic and business circles is: "To what extent can the past history of a common stock's price be used to make meaningful predictions concerning the future price of the stock?". There are two viewpoints towards the predictability of stock prices. Firstly, some argue that stock prices are unpredictable or very difficult to predict by their past returns (i.e. have very little serial correlation) because they simply follow a Random Walk process (Fama 1970). On the other hand, Lo & MacKinlay mention that "financial markets *are* predictable to some extent but far from being a symptom of inefficiency or irrationality, predictability is the oil that lubricates the gears of capitalism". Furthermore, there is also no real robust evidence for the predictability of returns themselves, let alone be out-of-sample (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). The measure Value at Risk (VaR), developed in response to the financial disaster events of the early 1990s, has been very important in the financial world. Corporations have to manage their risks and thereby include a future risk measurement. The tool of VaR has now 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 5% VaR can be informally defined as the maximum loss of a portfolio, during a time horizon, excluding all the negative events with a combined probability lower than 5% while the Conditional Value at Risk (CVaR) can be informally defined as the average of the events that are lower than the VaR. 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 it is an inconsistency with the evidence empirically available which outlines a more skewed distribution with fatter tails than the normal. This lead to the conclusion that the assumption of normality, which simplifies the computation of VaR, can bring to incorrect numbers, underestimating the probability of extreme events happening.

This paper has the aim to replicate and update the research made by Bali, Mo, et al. (2008) on US indexes, analyzing the dynamics proposed with a European outlook. The main contribution of the research is to provide the industry with a new approach to calculating VaR with a flexible tool for modeling the empirical distribution of returns with higher accuracy and characterization of the tails.

The paper is organized as follows. Chapter **??** discusses at first the alternative distribution than the normal that we are going to evaluate during the analysis (Student's t-distribution, Generalized Error Distribution, Skewed t-distribution, Skewed Generalized Error Distribution, Skewed Generalized t-distribution), then the discrete time GARCH models used (GARCH, IGARCH, EGARCH, GJRGARCH, NAGARCH, TGARCH, TSGARCH or AVGARCH and EWMA) are presented as extensions of the Engle (1982) 's ARCH model. Chapter 1 describes the dataset used and the methodology followed in modeling the volatility with the GARCH model by Bollerslev (1986) and with its refinements using Maximum likelihood estimation to find the distribution parameters. Then a description is given of how are performed the control tests (un- and conditional coverage test, dynamic quantile test) used in the paper to evaluate the performances of the different GARCH models and underlying distributions. In chapter 2, findings are presented and discussed, in chapter 3 the findings of the performed tests are shown and interpreted and in chapter 4 the investigation and the results are summarized.

## 0.1 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 iid, or independently and identically distributed, another is that they are normally distribution. Are these valid assumptions? Below the stylized facts[7] 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), but it is time-varying. Bollerslev (1987) describes it as "rates

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

of return data are characterized by volatile and tranquil periods”. Alexander (2008) says this will have implications for risk models: following a large shock to the market, the volatility changes and the probability of another large shock is increased significantly.

- Returns also exhibit *asymmetric volatility*, in that sense 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 pronounced in equity markets: usually there is a strong negative correlation between equity returns and the change in volatility.
- Returns are *not normally distributed* which is also one of the conclusions by Fama (1965). Returns have tails fatter than a normal distribution (leptokurtosis) and thus are riskier than under the normal distribution. Log returns **can** be assumed to be normally distributed. However, this will be examined in our empirical analysis if this is appropriate. This makes that simple returns follow a log-normal distribution, which is a skewed density distribution. A good summary is given by Alexander (2008) as: “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 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 the density function. What we know from the stylized facts of returns that the normal distribution is not appropriate for returns. In appendix we summarize some alternative distributions (T, GED, ST, SGED, SGT) that could be a better approximation of returns than the normal one.

## 0.2 Volatility modeling

### 0.2.1 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[8]. Engle regards this formulation as the first ARCH model.

### 0.2.2 ARCH model

Autoregressive Conditional Heteroscedasticity (ARCH) models, proposed by Engle (1982), was 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. To fully capture the logic behind GARCH models, the building blocks are examined in the first place. 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), (2) and (3). Returns are written as a constant part ($\mu$) and an unexpected part, called noise or the innovation process. The innovation process is the volatility ($\sigma_t$) 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 not yet assumed. The third component is the variance process or the expression for the volatility. The variance is given by a constant $\omega$, plus the random part which depends on the return shock of the previous period squared ($\varepsilon_{t-1}^2$). In that sense when the uncertainty or surprise in the last period increases, then the variance

[8]For example, for the past month it would then be calculated as the equally weighted average of the squared deviations from the mean (i.e. residuals) 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.

becomes larger in the next period. The element $\sigma_t^2$ is thus known at time $t-1$, while it is a deterministic function of a random variable observed at time $t-1$ (i.e. $\varepsilon_{t-1}^2$).

$$y_t = \mu + \varepsilon_t \tag{1}$$

$$\varepsilon_t = \sigma_t \times z_t, \ \textit{where} \ z_t \overset{iid}{\sim} (0,1) \tag{2}$$

$$\sigma_t^2 = \beta_0 + \beta_1 \times \varepsilon_{t-1}^2 \tag{3}$$

From these components we could look at the conditional moments (or expected returns and variance). We can plug in the component $\sigma_t$ into the conditional mean innovation $\varepsilon_t$ and use the conditional mean innovation to examine the conditional mean return. In equation (4) and (5) 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 (5) is equal to the unconditional mean in the most simple case. But variations are possible using ARMA (eg. AR(1)) processes.

$$\mathbb{E}_{t-1}(\varepsilon_t) = \mathbb{E}_{t-1}(\sqrt{\beta_0 + \beta_1 \times \varepsilon_{t-1}^2} \times z_t) = \sigma_t \mathbb{E}_{t-1}(z_t) = 0 \tag{4}$$

$$\mathbb{E}_{t-1}(y_t) = \mu + \mathbb{E}_{t-1}(\varepsilon_t) = \mu \tag{5}$$

For the conditional variance, knowing everything that happened until and including period $t-1$ the conditional innovation variance is given by equation (6). This is equal to $\sigma_t^2$, while the variance of $z_t$ is equal to 1. Then it is easy to derive the conditional variance of returns in equation (7), that is why equation (3) is called the variance equation.

$$var_{t-1}(\varepsilon_t) = \mathbb{E}_{t-1}(\varepsilon_t^2) = \mathbb{E}_{t-1}(\sigma_t^2 \times z_t^2) = \sigma_t^2 \mathbb{E}_{t-1}(z_t^2) = \sigma_t^2 \tag{6}$$

$$var_{t-1}(y_t) = var_{t-1}(\varepsilon_t) = \sigma_t^2 \tag{7}$$

The unconditional variance is also interesting to derive, while this is the long-run variance, which will be derived in equation (11). After deriving this using the law of iterated expectations and assuming stationarity for the variance process, one would get equation (8) for the unconditional variance, equal to the constant $c$ and divided by $1 - \beta_1$, the slope of the variance equation.

$$\sigma^2 = \frac{\beta_0}{1 - \beta_1} \tag{8}$$

This leads to the properties of ARCH models: Stationarity[9] 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 $\varepsilon_t$. The unconditional 4th moment, kurtosis $\mathbb{E}(\varepsilon_t^4)/\sigma^4$ of an ARCH model is given by equation (9). This term is larger than 3, which implicates fat-tails.

$$3\frac{1 - \beta_1^2}{1 - 3\beta_1^2} \tag{9}$$

Another property of ARCH models is that it takes into account volatility clustering. Because we know that $var(\varepsilon_t) = \mathbb{E}(\varepsilon_t^2) = \sigma^2 = \omega/(1 - \alpha_1)$, we can plug in $\beta_0$ for the conditional variance $var_t(\varepsilon_{t+1}) = \mathbb{E}(\varepsilon_{t+1}^2) = \sigma_{t+1}^2 = c + \alpha_1 \times \varepsilon_t^2$. Thus it follows that equation (10) displays volatility clustering. If we examine the RHS, as $\alpha_1 > 0$ (condition for stationarity), when shock $\varepsilon_t^2$ is larger than what you expect it to be on average $\sigma^2$ the LHS will also be positive. Then the conditional

[9] 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.

variance will be larger than the unconditional variance. Briefly, large shocks will be followed by more large shocks.

$$\sigma_{t+1}^2 - \sigma^2 = \alpha_1 \times (\varepsilon_t^2 - \sigma^2) \tag{10}$$

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 (9), 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 (11). This can already be simplified, while we know that $\sigma_{T+1}^2 = \omega + \alpha_1 \times \varepsilon_T^2$ from equation (3).

$$\begin{aligned} \mathbb{E}_T(\varepsilon_{T+k}^2) &= \omega \times (1 + \alpha_1 + ... + \alpha^{k-2}) + \alpha^{k-1} \times \sigma_{T+1}^2 \\ &= \omega \times (1 + \alpha_1 + ... + \alpha^{k-1}) + \alpha^k \times \sigma_T^2 \end{aligned} \tag{11}$$

It can be shown that then the conditional variance in period $T+k$ is equal to equation (12). The LHS is the predicted conditional variance $k$-periods ahead above its unconditional variance, $\sigma^2$. The RHS is the difference current last-observed return residual $\varepsilon_T^2$ above the unconditional average multiplied by $\alpha_1^k$, a decreasing function of $k$ (given that $0 \leq \alpha_1 < 1$). The further ahead predicting the variance, the closer $\alpha_1^k$ comes to zero, the closer to the unconditional variance, i.e. the long-run variance.

$$\mathbb{E}_T(\varepsilon_{T+k}^2) - \sigma^2 = \alpha_1^k \times (\varepsilon_T^2 - \sigma^2) \tag{12}$$

### 0.2.3 Univariate GARCH models

An improvement of the ARCH model is the Generalized Autoregressive Conditional Heteroscedasticity (GARCH)[10]. This model and its variants come in to play because

[10] *Generalized* as it is a generalization by Bollerslev (1986) of the ARCH model of Engle (1982). *Autoregressive,* as it is a time series model with an autoregressive form (regression on itself).

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[11]. An overview (of a selection) of investigated GARCH models is given in the following table.

**Table 1:** GARCH models, the founders

| Author(s)/user(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) | TSGARCH (or AVGARCH) model |
| Morgan Guaranty Trust Company (1996) | EWMA or RiskMetrics model |

## 0.3 ACD models

An extension to GARCH models was proposed by Hansen (1994), the autoregressive conditional density estimation model (referred to as ACD models, 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

*Conditional heteroscedasticity,* while time variation in conditional variance is built into the model (Alexander 2008).

[11] which makes the maximum likelihood estimation explained in part 1.2.1 complex with more parameters that have to be estimated.

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, SGT), or a model that allows to model 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 parameterizations 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 not simple (nonlinear bounding specification of higher moment distribution parameters and interaction). So, are skewness (skewness parameter) and kurtosis ( shape parameters) time varying? The literature investigating higher moments has arguments for and against this statement. In part 1.2.2 the specification is given.

## 0.4 Value at Risk

Value-at-Risk (VaR) is a risk metric developed simultaniously by Markowitz (1952) and Roy1952 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[12], Morgan Guaranty Trust Company (1996) (part of JP Morgan), 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

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

the greatest possible loss in 99% of cases. It can be defined as the threshold value $\theta_t$. Put differently, in 1% of cases the loss would be greater than this amount.It is specified as in (13). Christofferson2001 puts forth a general framework for specifying VaR models and comparing between two alternatives models.

$$Pr(y_t \leq \theta_t | \Omega_{t-1}) \equiv \phi \tag{13}$$

With $y_t$ expected returns in period t, $\Omega_{t-1}$ the information set available in the previous period and $\phi$ the chosen confidence level.

## 0.5 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, [Artzner1998] reject 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 (14).

To calculate $\theta_t$, 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.

$$Pr(y_t \leq \theta_t | \Omega_{t-1}) \equiv \int_{-\infty}^{\theta_t} f(y_t | \Omega_{t-1}) \, \mathrm{d}y_t = \phi \tag{14}$$

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. 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 $CVaR_{97.5}$.

## 0.6 Past literature on the consequences of higher moments for VaR determination

Here comes the discussion about studies that have looked at higher moments and VaR determination. Also a summary of studies that discusses time-varying higher moments, but not a big part, while it is also only a small part of the empirical findings (couple of GARCH-ACD models).

**Table 2:** Higher moments and VaR

| Author | Higher moments |
|---|---|
| Hansen (1994) | Skewness and kurtosis extended ARCH-model |
| \@harvey1999 | 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) extends the ARCH model to include time-varying moments beyond mean and variance. While mean returns and variance 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 an similar way how variance is estimated). Applications on U.S. Treasuries and exchange rates are discussed.

@harvey1999 extends a GARCH(1,1) model to include time varying skewness by estimating it jointly with time varying variance using a skewed t distribution. They find 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 find that including skewness causes the leverage effects of variance to dissapear. They apply their methods on different stock indices (both developed and emerging) at daily, weekly and monthly frequency.

Brooks et al. (2005) proposes a model based on a t-distribution that allows for both the variance and the degrees of freedom to be time-varying, independently from eachother. Their model allows for both assymetric 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 apply their model on different financial assets in the U.S. and U.K. at daily frequency.

*Volatility is unobservable. We can only ever estimate and forecast volatility, and this only within the context of an assumed statistical model. So there is no absolute 'true' volatility: what is 'true' depends only on the assumed model. . .*

*Moreover, volatility is only a sufficient statistic for the dispersion of the returns distribution when we make a normality assumption. In other words, volatility does not provide a full description of the risks that are taken by the investment unless we assume the investment returns are normally distributed.*

— Alexander (2008) in *Market Risk Analysis Practical Financial Econometrics* # Literature review {#lit-rev}

# 1 Data and methodology

## 1.1 Data

We worked with daily returns on the Euro Stoxx 50 Price Index[1] denoted in EUR from 31 December, 1986 to 27 April, 2021. It is 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 we refer to the factsheet (*Calculation guide STOXX ®* 2020).

### 1.1.1 Descriptives

#### Table of summary statistics

Equation 1.1 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 0.017% with a standard deviation of 1.307% and a median of 0.036 which translate to an annualized mean of 4.208% and an annualized standard deviation of 20.748%. The skewness statistic is highly significant and negative at -0.31 and the excess kurtosis is also highly significant

[1] The same analysis has been performed for the INDEX 1, INDEX 2, INDEX 3 and the INDEX 4 indexes with the same conclusions. The findings of these researches are available upon requests.

and positive at 7.208. These 2 statistics give an overview of the distribution of the returns which has thicker 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 19528.62 and a high statistical significance, confirms the non normality feeling given by the Skewness and Kurtosis.

The right column of table 1.1 displays the same descriptive statistics but for the standardizes residuals obtained from a simple GARCH model as mentioned in table 1.1 in Note 2∗. Again, Skewness statistic at -0.633 with a high statistical significance level and the excess Kurtosis at 5.134 also with a high statistical significance, suggest a non normal distribution of the standardized residuals and the Jarque-Bera statistic at NA, given its high significance, confirms the rejection of the normality assumption.

## Descriptive figures

### Stylized facts

As can be seen in figure 1.1 the Euro area equity and later, since 1999 the Euro Stoxx 50 went up during the tech (“dot com”) bubble reaching an ATH of €5464.43. 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 09 March, 2009 of €1809.98. There is an improvement, but then the European debt crisis, with it’s 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.

**Table 1.1:** Summary statistics of the returns

| Statistics | Euro Stoxx 50 | Standardized Residuals |
|---|---|---|
| Arithmetic Mean | 0.0167 | -0.0409 |
| Median | 0.0357 | -0.0193 |
| Maximum | 10.4376 | 5.7126 |
| Minimum | -13.2404 | -11.7732 |
| Stdev | 1.307 | 0.9992 |
| Skewness | -0.31 | -0.6327 |
| | (0***) | (0***) |
| Excess Kurtosis | 7.2083 | 5.134 |
| | (0***) | (0***) |
| Jarque-Bera | 19528.6196*** | 10431.0514*** |

Notes

[1] This table shows the descriptive statistics of the daily percentage returns of EURO STOXX 50 over the period 1987-01-01 to 2021-04-27 (8954 observations). Including arithmetic mean, median, maximum, minimum, standard deviation, skewness, excess kurtosis and the Jarque-Bera test.

[2] The standardized residual is derived from a maximum likelihood estimation (simple GARCH model) as follows:
$R_t = \alpha_0 + \alpha_1 R_{t-1} + z_t \sigma_t$
$\sigma_t^2 = \beta_0 + \beta_1 \sigma_{t-1}^2 z_{t-1}^2 + \beta_2 \sigma_{t-1}^2$,
Where $z$ is the standard residual (assumed to have a normal distribution).

[3] *, **, *** represent significance levels at the 5

In figure 1.2 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.

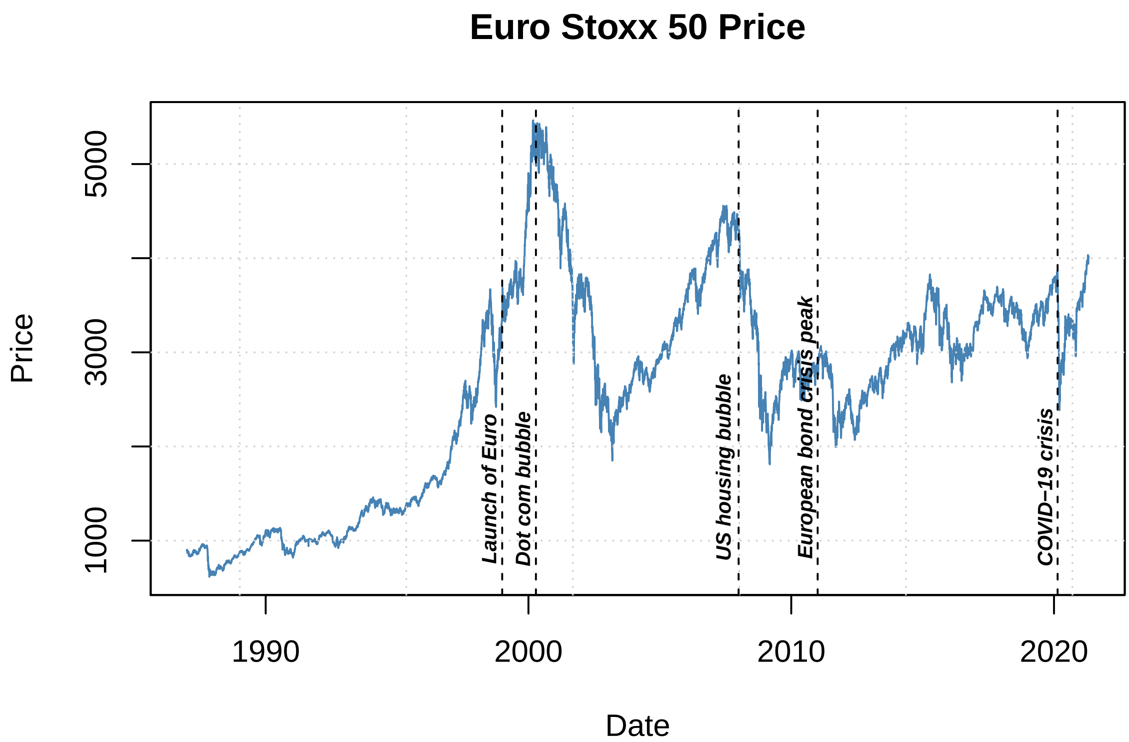


**Figure 1.1:** Euro Stoxx 50 Price Index prices

In figure 1.4 the density distribution of the log returns are examined. As can be seen, as already mentioned in part 0.1, log returns are not really normally distributed. So

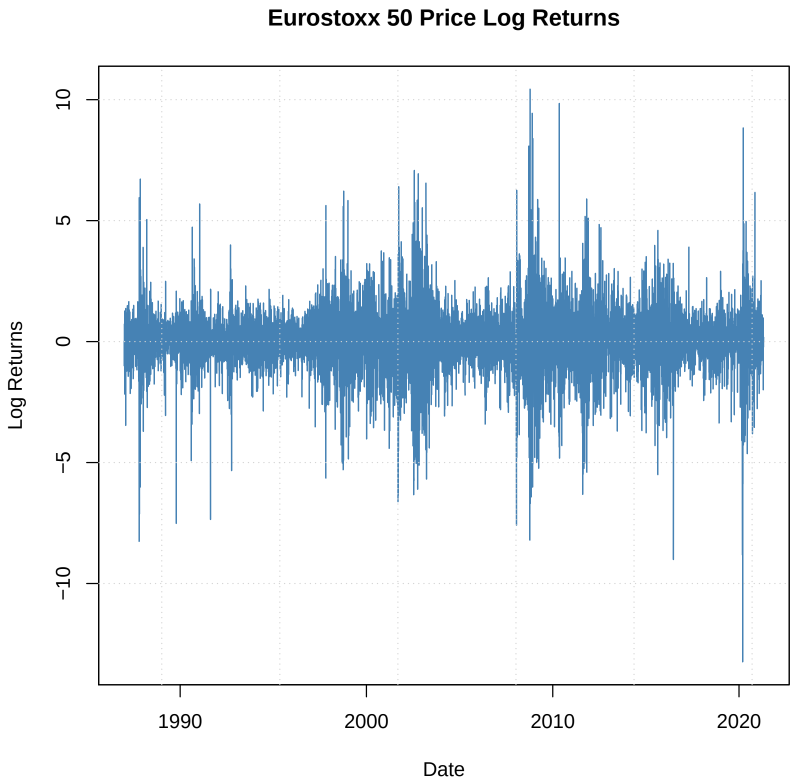


**Figure 1.2:** Euro Stoxx 50 Price Index log returns

ACF plots: to do...

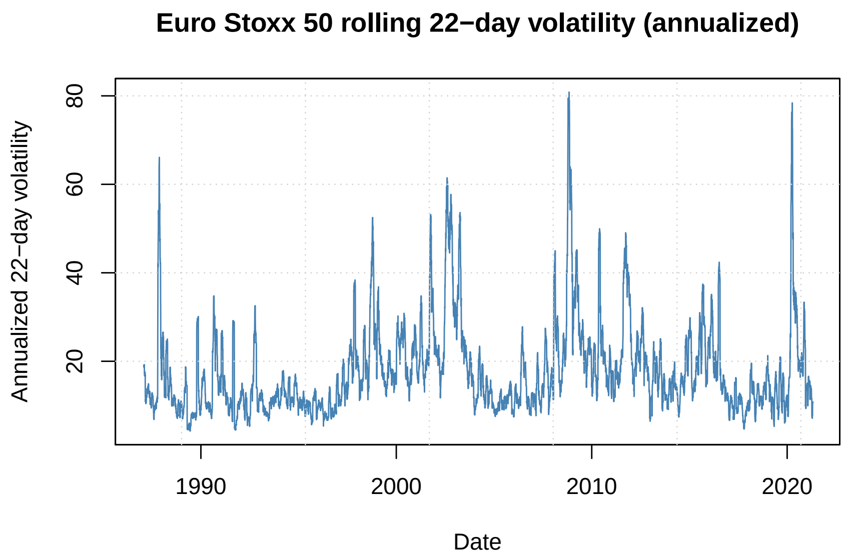


**Figure 1.3:** Euro Stoxx 50 rolling volatility (22 days, calculated over 252 days)

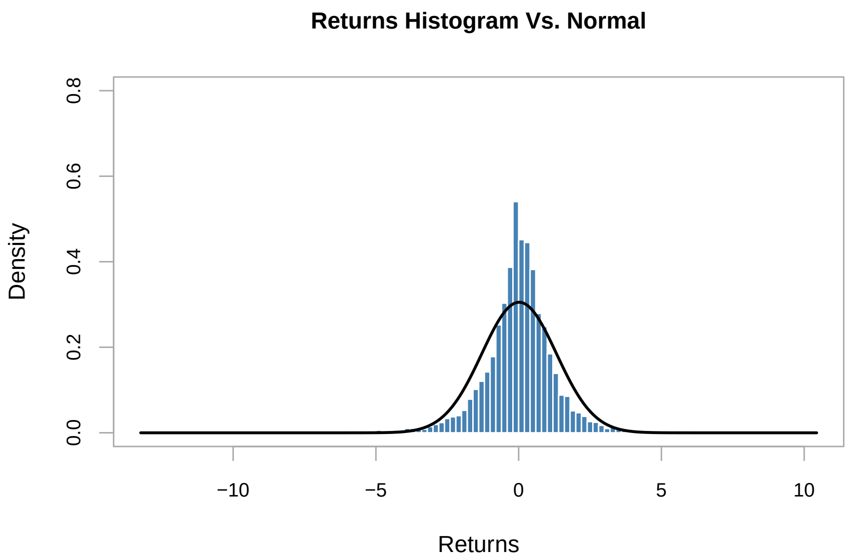


**Figure 1.4:** Density vs. Normal Euro Stoxx 50 log returns)

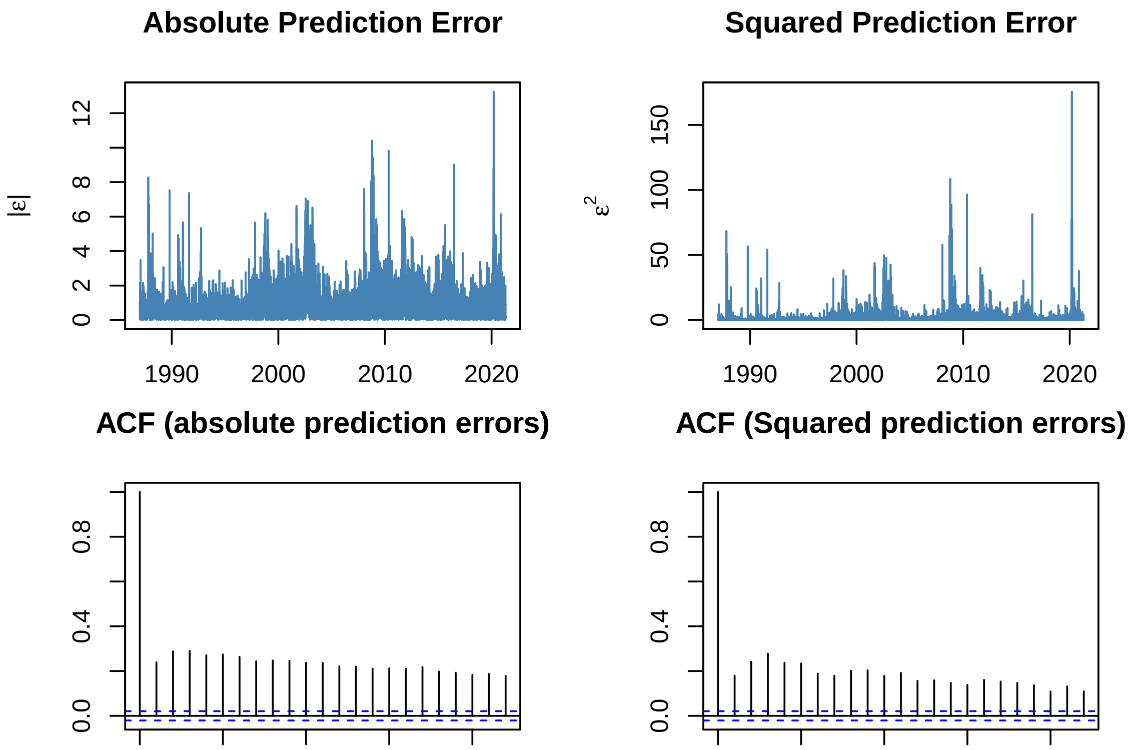


**Figure 1.5:** Absolute prediction errors

## 1.2 Methodology

### 1.2.1 Garch models

As already mentioned in part 0.2.3, GARCH models GARCH, 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 error distribution and the skewed generalized t distribution. They will be estimated using maximum likelihood[2].

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 $\theta$ (equation (1.2)). These functions are based on the joint probability distribution of the observed data (equation (1.4)). Subsequently, the (log)likelihood function is maximized using an optimization algorithm (equation (1.6)).

$$y_1, y_2, ..., y_N \sim i.i.d \tag{1.1}$$

$$y_i \sim f(y|\theta) \tag{1.2}$$

$$L(\theta) = \prod_{i=1}^{N} f(y_i|\theta) \tag{1.3}$$

$$\log(L(\theta)) = \sum_{i=1}^{N} \log f(y_i|\theta) \tag{1.4}$$

[2]As already mentioned, fortunately, Ghalanos (2020b) has made it easy for us to implement this methodology in the R language (v.3.6.1) 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.

$$\theta^* = arg\max_\theta[L] \tag{1.5}$$

$$\theta^* = arg\max_\theta[\log(L)] \tag{1.6}$$

### 1.2.2 ACD models

Following Ghalanos (2016), arguments of ACD models are specified as in Hansen (1994). The density function $f(y|\alpha)$ has parameters $\alpha_t = (\mu_t, \sigma_t, \nu_t)$, with equation (1.7), the conditional mean equation. Equation (1.8) as the conditional variance. And $\nu_t = \nu(\theta, x_t)$ the remaining parameters of the distribution like the skewness and kurtosis (shape) parameters.

$$\mu_t = \mu\left(\theta, x_t\right) = E\left(y_t \mid x_t\right) \tag{1.7}$$

$$\sigma_t^2 = \sigma^2\left(\theta, x_t\right) = E\left(\left(y_t - \mu_t^2\right) \mid x_t\right) \tag{1.8}$$

To further explain the difference between GARCH and ACD. The scaled innovations are given by equation (1.9). The conditional density is given by equation (1.10) and related to the density function $f(y|\alpha)$ as in equation (1.2.2).

$$z_t(\theta) = \frac{y_t - \mu\left(\theta, x_t\right)}{\sigma\left(\theta, x_t\right)} \tag{1.9}$$

$$g\left(z \mid \eta_t\right) = \frac{d}{dz} P\left(z_t < z \mid \eta_t\right) \tag{1.10}$$

$$f\left(y_t \mid \mu_t, \sigma_t^2, \eta_t\right) = \frac{1}{\sigma_t} g\left(z_t \mid \eta_t\right) \tag{1.11}$$

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

### 1.2.3 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 (1.12), 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 $\chi^2$-distributed with one degree of freedom or that the probability of failure $\hat{p}$ is equal to the chosen percentile $\alpha$.

$$LR^{uc} = -2\ln\left(\frac{(1-p)^{N-X}p^X}{\left(1-\frac{X}{N}\right)^{N-X}\left(\frac{X}{N}\right)^X}\right) \tag{1.12}$$

**Conditional coverage test of Christoffersen et al. (2001)**

Christoffersen et al. (2001) proposed the conditional coverage test. It is tests for unconditional covrage 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 (1.13).

$$I_t(\alpha) = \begin{cases} 1 & \text{if exceedence occurs} \\ 0 & \text{if no exceedence occurs} \end{cases} . \tag{1.13}$$

It involves a likelihood ratio test's null hypothesis is that the statistic is $\chi^2$-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 $\alpha$. 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 (1999) with the aim to provide completeness to the conditional coverage test of Christoffersen et al. (2001) developed the Dynamic quantile test. It consists in testing some restriction in a ... (work-in-progress).

# 2

# Empirical Findings

## 2.1 Density of the returns

### 2.1.1 MLE distribution parameters

In table 2.1 we can see the estimated parameters of the unconditional distribution functions. They are presented for the Skewed Generalized T-distribution (SGT) and limiting cases thereof previously discussed. Additionally, maximum likelihood score and the Aikake Information Criterion (AIC) is reported to compare goodness of fit of the different distributions. 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 less parameter, which could possibly result in less errors due to misspecification and 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 $q$ 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. We see that except for the normal distribution which is quite off, the theoretical distributions are close to the actual data, except that they are too peaked. This problem is the least present for the SGT distribution.

**Table 2.1:** Maximum likelihood estimates of unconditional distribution functions

| $\theta$ | $\alpha$ | $\beta$ | $\xi$ | $\kappa$ | $\eta$ | $LLH$ | AIC |
|---|---|---|---|---|---|---|---|
| SGT | 0.02 | 1.321 | -0.04 | 1.381 | 3.314 | -13973.01 | 27956.01 |
| | (0.013) | (0.026)*** | (0.013)*** | (0.071)*** | (0.538)*** | | |
| SGED | 0.019 | 1.274 | -0.018 | 0.916 | Inf | -14008.63 | 27956.01 |
| | (0.013) | (0.016)*** | (0.01)*** | (0.017)*** | | | |
| GED | 0.032 | 1.276 | 0 | 0.911 | Inf | -14009.52 | 28025.04 |
| | (0.009)*** | (0.016)*** | | (0.017)*** | | | |
| ST | 0.019 | 1.481 | -0.052 | 2 | 2.793 | -13997.35 | 28002.71 |
| | (0.014) | (0.054)*** | (0.013)*** | | (0.098)*** | | |
| T | 0.056 | 1.494 | 0 | 2 | 1.383 | -14005.14 | 28016.29 |
| | (0.01)*** | (0.056)*** | | | (0.097)*** | | |
| Normal | 0.017 | 1.307 | 0 | 2 | Inf | -15101.73 | 30207.46 |
| | (0.014) | (0.01)*** | | | | | |

*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

[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 GFC (2002-2009) and up to the present Covid-crash (2009-2021). There is no qualitative difference in relative ranking with these subsamples. Results are reported in the appendix.

## 2.2 Constant higher moments

2.2 presents the maximum likelihood estimates for 8 symmetric and asymmetric GARCH models based on the ST distribution with constant skewness and kurtosis parameters ($t$ values are presented in parenthesis). The parameters in the conditional mean equations ($\alpha_0$) are all statistically significant with $t$ values from 6 to 11. The AR(1) coefficient, $\alpha_1$, has parameters going from 2 to 2 with $t$ values ranging from 4 to 5 not suggesting a high significance and indicating slight negative autocorrelation. The GARCH parameters in the conditional variance equations ($\beta_0$) are generally statistically significant with $t$ values ranging from 1 to 11. The results of $\beta_1$ and $\beta_2$ show the presence of significant time-variation in the conditional volatility of the Euro Stoxx 50 Price Index daily returns, in fact, the sum of $\beta_1$ and $\beta_2$ for the GARCH parameters is close to one (from 20 to 34), suggesting the presence of persistence in the volatility of the returns. The parameter $\xi$ is highly significant for all the 8 models tested with values ranging from 12 to 18 confirming the presence of [[[HOW TO BEST INTERPRET??????????]]] Skewness in the returns.

**Table 2.2:** Maximum likelihood estimates of the ST-GARCH models with constant skewness and kurtosis parameters

| | SGARCH | IGARCH | EGARCH | GJRGARCH | EWMA | NAGARCH | TGARCH | AVGARCH |
|---|---|---|---|---|---|---|---|---|
| $\alpha_0$ | 0.049 | 0.049 | 0.026 | 0.028 | 0.053 | 0.02 | 0.023 | 0.019 |
| | (5.278) | (5.192) | (2.747) | (3.022) | (5.852) | (2.148) | (2.404) | (2.03) |
| $\alpha_1$ | -0.018 | -0.018 | -0.008 | -0.008 | -0.02 | -0.005 | -0.005 | -0.006 |
| | (-1.64) | (-1.635) | (-0.795) | (-0.768) | (-1.885) | (-0.485) | (-0.47) | (-0.611) |
| $\beta_0$ | 0.016 | 0.013 | 0.001 | 0.021 | 0 | 0.022 | 0.02 | 0.021 |
| | (5.776) | (5.842) | (0.77) | (7.28) | | (9.811) | (6.219) | (25.122) |
| $\beta_1$ | 0.094 | 0.101 | -0.098 | 0.017 | 0.069 | 0.08 | 0.083 | 0.087 |
| | (12.146) | (13.088) | (-15.524) | (3.021) | (15.02) | (6.286) | (9.717) | (30.759) |
| $\beta_2$ | 0.898 | 0.899 | 0.983 | 0.897 | 0.931 | 0.845 | 0.919 | 0.904 |
| | (115.678) | | (1557.507) | (115.012) | | (86.237) | (107.22) | (365.502) |
| $\xi$ | 0.917 | 0.917 | 0.905 | 0.906 | 0.917 | 0.903 | 0.904 | 0.902 |
| | (68.351) | (67.44) | (67.131) | (67.765) | (73.31) | (67.757) | (67.28) | (67.834) |
| $\kappa$ | | | | | | | | |
| $\eta$ | 6.339 | 5.997 | 6.897 | 6.819 | 7.036 | 6.974 | 6.928 | 6.944 |
| | (15.442) | (16.925) | (14.582) | (14.635) | (18.325) | (14.536) | (14.568) | (14.514) |
| $\gamma$ | | | 0.144 | 0.143 | | | | |
| | | | (15.566) | (10.728) | | | | |
| $shift$ | | | | | | 0.904 | | 0.214 |
| | | | | | | (10.355) | | (9.66) |
| $rot$ | | | | | | | 0.723 | 0.552 |
| | | | | | | | (12.112) | (9.638) |
| $LLH$ | -13066.436 | -13068.628 | -12951.877 | -12973.456 | -13114.375 | -12936.278 | -12934.286 | -12930.492 |

*Notes*

[1] This table shows the maximum likelihood estimates of various ST-GARCH models. The daily returns used on the Euro Stoxx 50 Price index cover the periodfrom 01 January, 1987 to 27 April, 2021 (8954 observations).
[2] 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, $\gamma$ is the asymmetry in volatility , $\xi, \kappa$ and $\eta$ are constant and $t$ statistics are displayed in parenthesis.
$LLH$ is the maximized log likelihood value.

As you can see in table **??** the AIC for the skewed student's t-distribution (ST) is the best from all the models. As also shown in appendix part B. The best in all distributions of the GARCH models seems to be the NAGARCH model, but we do not want to overfit our models because of an in-sample estimation. That is why a parsimonious model like the EGARCH (which has the highest maximum likelihood for the standard GARCH models that are considered), but also the model AVGARCH will be examined using the ST distribution while it has the second-best (lowest) AIC.

### 2.2.1 Value-at-risk

As already mentioned 3 candidate models seem to be candidates to check if they perform well using a forecasting technique and backtest. This includes the EGARCH, the NAGARCH and AVGARCH. A simple graph is shown in figure **??** for the EGARCH-ST model. It seems that the VaR model for $\alpha = 0.05$ underestimates the downside events, while the VaR model for $\alpha = 0.01$ overestimates a lot of the downside events.

Let us examine this further using a moving window approach whilst forecasting 1-day ahead results with a window size of 1500. Figure **??** shows that choosing an appropriate forecast period is important, while it includes the decline in 2016 with among which Brexit and the recent COVID-crisis.

As you can see in figure @ref(fig.)

### 2.2.2 Expected shortfall

## 2.3 Time-varying higher moments

# 3
# Robustness Analysis

## 3.1 Backtest

## 3.2 Specification checks

In order to check if the models are specified correctly, some specification 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:

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

### 3.2.1 Figures control tests

Autocorrelation function of the standardized residuals and autocorrelation function of the squared standardized residuals.

Then the density can be examined standardized residuals and compared with the normal distribution.

Also the QQ-plot can be examined.

### 3.2.2 Residual heteroscedasticity

Ljung-Box test on the squared or absolute standardized residuals.

zero-mean unit-variance not skewed no excess kurtosis no serial correlation in the squares no serial correlation in the cubes no serial correlation in the squares

# 4 Conclusion

# Appendices

# A

# Appendix to literature review

## Alternative distributions than the normal

**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.1). As will be seen in 0.2, 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).

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

where $\alpha, \beta$ and $\nu$ are respectively the location, scale and shape (tail-thickness) parameters. $\nu/2$ is equal to the $q$[1] parameter (which we call $\eta$) of the SGT distribution with other restrictions (see part A). The symbol $\Gamma$ is the Gamma function.

[1] Also referred to as $n$ by Theodossiou (1998) or $\eta$ by Bali, Mo, et al. (2008)

Unlike the normal distribution, which depends entirely on two moments only, the student t distribution allows for fatter tails. This kurtosis coefficient is given by equation (A.2) if $\nu > 4$. This is useful while as already mentioned, the standardized residuals appear to have fatter tails than the normal distribution following Bollerslev (2008).

$$kurt = 3 + \frac{6}{\nu - 4} \tag{A.2}$$

**Generalized Error Distribution** The GED distribution is nested in the generalized t distribution by McDonald and Newey (1988) 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, uni-modal 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.3) following Ghalanos (2020a).

$$f(x;\alpha,\beta,\kappa) = \frac{\kappa e^{-\frac{1}{2}\left|\frac{x-\alpha}{\beta}\right|^{\kappa}}}{2^{1+1/\kappa}\beta\Gamma(1/\kappa)} \tag{A.3}$$

where $\alpha, \beta$ and $\kappa$ are respectively the location, scale and shape parameters.

**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.4) presents the skewed t-distribution.

$$f_\xi(z) \equiv \frac{2\sigma_\xi}{\xi + \xi^{-1}} f_1(z_\xi), \quad z_\xi \equiv \begin{cases} \xi^{-1}(\sigma_\xi z + \mu_\xi) & \text{if } z \geq -\mu_\xi/\sigma_\xi \\ \xi(\sigma_\xi z + \mu_\xi) & \text{if } z < -\mu_\xi/\sigma_\xi \end{cases} \tag{A.4}$$

where $\mu_\xi \equiv M_1\left(\xi - \xi^{-1}\right), \quad \sigma_\xi^2 \equiv \left(1 - M_1^2\right)\left(\xi^2 + \xi^{-2}\right) + 2M_1^2 - 1, \quad M_1 \equiv 2\int_0^\infty u f_1(u)du$ and $\xi$ between 0 and $\infty$. $f_1(\cdot)$ is in this case equation (A.1), the pdf of the student t distribution coming to equation (A.5), which has the parameterization following the SGT parameters.

$$\begin{aligned} f_{ST}(x;\alpha,\beta,\xi,\eta) &= \frac{\Gamma\left(\frac{1}{2}+\eta\right)}{\sqrt{\beta\pi\eta}\,\Gamma(\eta)\left(\frac{|x-\alpha+m|^2}{\eta\beta(\xi\,\mathrm{sign}(x-\alpha+m)+1)^2}+1\right)^{\frac{1}{2}+\eta}} \\ m &= \frac{2\xi\sqrt{\beta\eta}\,\Gamma\left(\eta-\frac{1}{2}\right)}{\sqrt{\pi}\,\Gamma\left(\eta+\frac{1}{2}\right)} \end{aligned} \tag{A.5}$$

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

**Skewed Generalized Error Distribution** What also will be interesting to examine is the SGED distribution of Theodossiou (2000) in GARCH models, as in the work of 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.4) as the skewed t-distribution but with $f_1(\cdot)$ equal to equation (A.3). To then get equation (A.6).

$$\begin{aligned} f_{SGED}(x;\alpha,\beta,\xi,\kappa) &= \frac{\kappa e^{-\left(\frac{|x-\alpha+m|}{\nu\beta(1+\xi\,\mathrm{sig}(x-\alpha+m))}\right)^{\kappa}}}{2\nu\beta\Gamma(1/\kappa)} \\ m &= \frac{2^{\frac{2}{\kappa}}\nu\beta\xi\Gamma\left(\frac{1}{2}+\frac{1}{p}\right)}{\sqrt{\pi}} \end{aligned} \tag{A.6}$$

**SGT (Skewed Generalized t-distribution)** The SGT distribution of 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 skewness and leptokurtosis. The density of the generalized t-distribution of McDonald and Newey (1988) can be

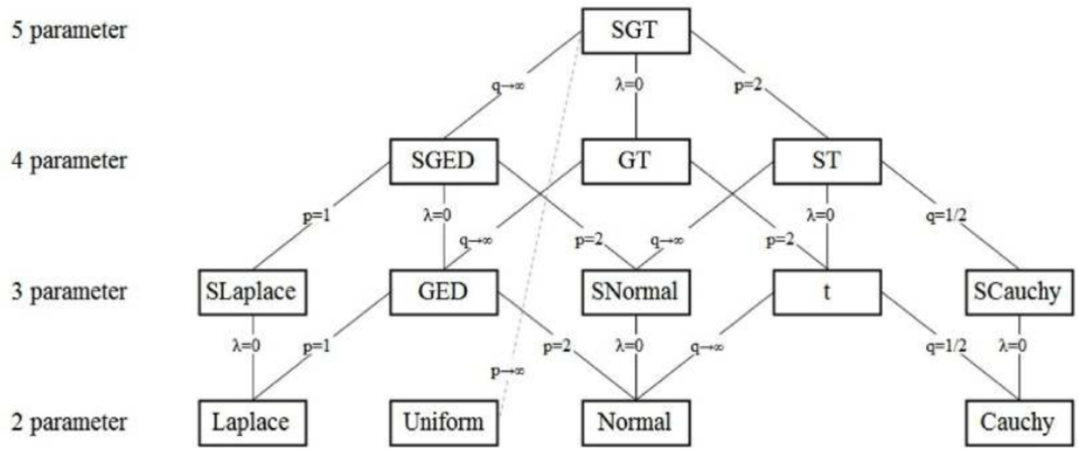


**Figure A.1:** SGT distribution and limiting cases

rewritten as its skew variant following Trottier and Ardia (2015). The pdf of the SGT distribibution is given by eqution (A.7).

$$
\begin{aligned}
f_{SGT}(x;\alpha,\beta,\xi,\kappa,\eta) &= \frac{\kappa}{2v\beta\eta^{1/\kappa}B\left(\frac{1}{\kappa},\eta\right)\left(\frac{|x-\alpha+m|^{\kappa}}{\eta(v\beta)^{\kappa}(\xi\,\mathrm{sign}(x-\alpha+m)+1)^{\kappa}}+1\right)^{\frac{1}{\kappa}+\eta}} \\
m &= \frac{2v\beta\xi\eta^{\frac{1}{\kappa}}B\left(\frac{2}{\kappa},\eta-\frac{1}{\kappa}\right)}{B\left(\frac{1}{\kappa},\eta\right)}
\end{aligned}
\tag{A.7}
$$

Following Theodossiou (1998) however, there are two parameters, $\kappa$[2]and $\eta$[3]) for the shape in the SGT distribution. The $p$ is the peakedness parameter. The $q$ is the tail-thickness parameter. It is equal to the degrees of freedom $\eta$ divided by 2 if $\xi = 0$ and $\kappa = 2$, there is referred to symbol $\nu$ in the tables (although this is not fully statistically correct to interprete this like degrees of freedom at all times). As shown in the following figure[4] A.1 adapted by Carter Davis using , from the SGT the other distributions in the figure are limiting cases of the SGT.

## 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

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

[3]Also referred to as $n$ by Theodossiou (1998) or $\eta$ by Bali, Mo, et al. (2008). This is the $q$ by Carter Davis in the "sgt" packages.

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

restricted so that the variance output always is positive, except for the EGARCH model, as this model does not mathematically allow for a negative output.

**Symmetric (normal) GARCH model** The standard GARCH model (Bollerslev 1986) is written consistent with Ghalanos (2020a) as in equation (A.8) without external regressors.

$$\sigma_t^2 = \beta_0 + \beta_1 \varepsilon_{t-j}^2 + \beta_2 \sigma_{t-j}^2 \tag{A.8}$$

where $\sigma_t^2$ denotes the conditional variance, $\beta_0$ the intercept and $\varepsilon_t^2$ 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.9) for a GARCH model of order $(1, 1)$.

$$\hat{P} = \beta_1 + \beta_2. \tag{A.9}$$

The unconditional variance of the standard GARCH model of Bollerslev is very similar to the ARCH model, but with the Garch parameter ($\beta_2$) included as in equation (A.10).

$$\begin{aligned} \hat{\sigma}^2 &= \frac{\beta_0}{1 - \hat{P}} \\ &= \frac{\beta_1}{1 - \beta_1 - \beta_2} \end{aligned} \tag{A.10}$$

**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 $I_t - j$, it is specified as in equation (A.11).

$$\sigma_t^2 = \beta_0 + \alpha_1 \varepsilon_{t-1}^2 + \gamma_j I_{t-1} \varepsilon_{t-1}^2 + \beta_2 \sigma_{t-1}^2 \tag{A.11}$$

where $\gamma_j$ 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.12). 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.

$$\log_e(\sigma_t^2) = \beta_0 + \beta_1 z_{t-1} + \gamma_1(|z_{t-1}| - E|z_{t-1}|) + \beta_2 \log_e(\sigma_{t-j}^2) \tag{A.12}$$

where $\alpha_j$ captures the sign effect and $\gamma_j$ the size effect.

**NAGARCH model** The NAGARCH or nonlinear asymmetric model (Engle and Ng 1993). It is specified as in equation (A.13). 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.

$$\sigma_t^2 = \beta_0 + \beta_1(\varepsilon_{t-1} + \gamma_1 \sqrt{\sigma_{t-1}})^2 + \beta_2 \sigma_{t-1}^2 \tag{A.13}$$

As before, $\gamma_1$ 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.14).

$$\sigma_t = \beta_0 + \beta_1^+ \varepsilon_{t-1}^+ \beta_1^- + \varepsilon_{t-1}^-) + \beta_2 \sigma_{t-1} \tag{A.14}$$

where $\varepsilon_{t-j}^+$ is equal to $\varepsilon_{t-j}$ if the term is negative and equal to 0 if the term is positive. The reverse applies to $\varepsilon_{t-j}^-$. 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.15).

$$\sigma_t = \beta_0 + \beta_1 \left|\varepsilon_{t-1}\right|) + \beta_2 \sigma_{t-1} \tag{A.15}$$

**EWMA** A 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 $\xi$ more weight is assigned to recent periods than distant periods. The $\xi$ must be less than 1. It is specified as in (A.16).

$$\sigma_t^2 = (1-\xi)\sum_{j=1}^{\infty}(\xi^j \varepsilon_{t-j}^2) \tag{A.16}$$

In practice a $\xi$ (or sometimes referred to as $\lambda$) 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

## Goodness of fit

As already mentioned, next to testing the models in part 2, 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 124 models that were estimated. This can ofcourse lead to overfitting iusing the this dataset 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. First, in figure B.1, symmetric GARCH with symmetric distributions are looked at. As you can see the student's t distribution (T) performs better than general error distribution (GED), that performs better than the normal distribution (NORM) according to both the AIC and BIC. Which is consistent with the literature that found that the assumption of the normal distribution is a rather poor assumption.

First, in figure B.2, symmetric GARCH with the best symmetric distribution (T) and the other distributions (SGED, ST) are looked at. As you can see consistent with Giot and Laurent (2003)

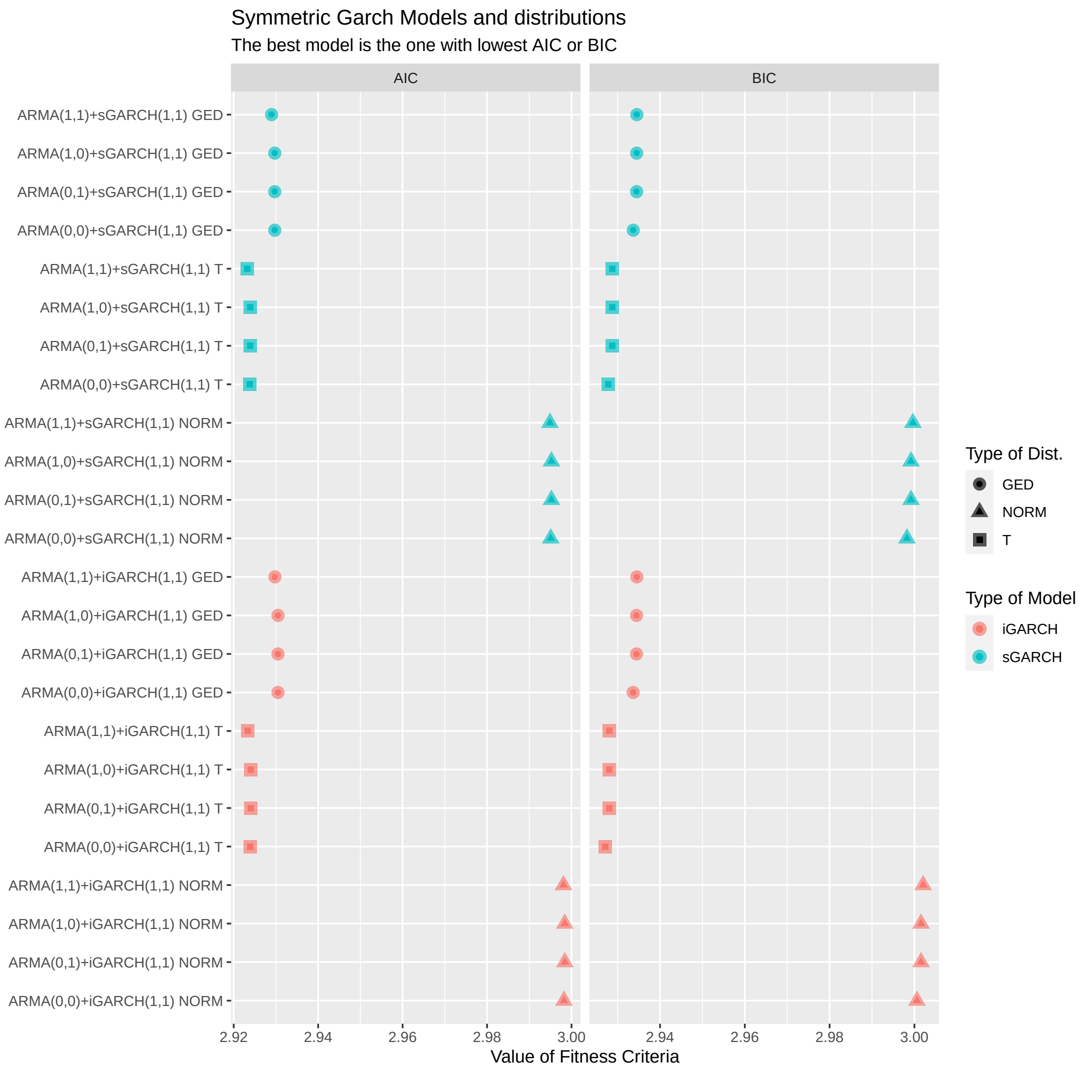


**Figure B.1:** Goodness of fit symmetric GARCH and distributions

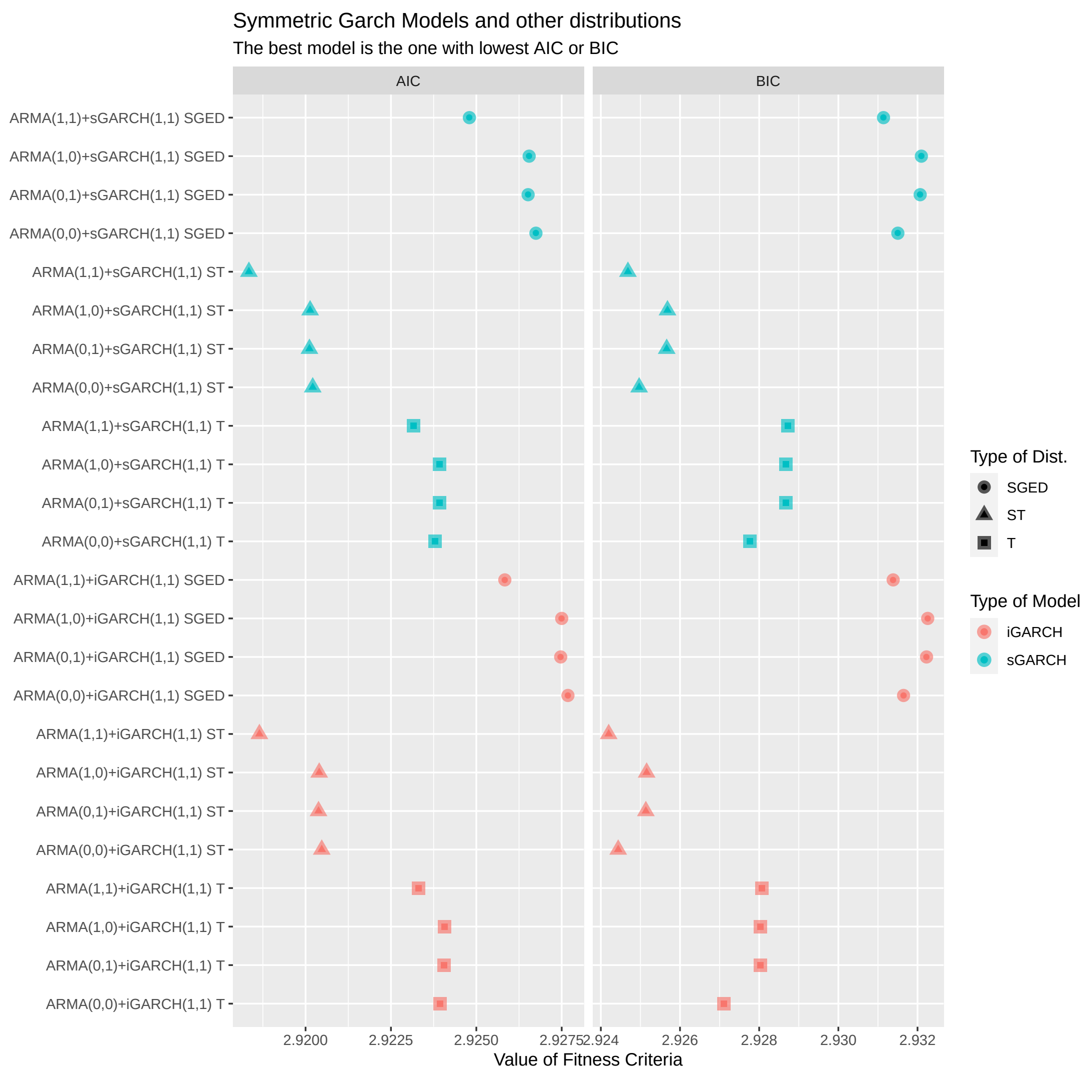


**Figure B.2:** Goodness of fit symmetric GARCH and other distributions

# 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.

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.

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. (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 S. Manganelli (1999). *CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles.* Tech. rep. San Diego: UC San Diego. URL: http://www.jstor.org/stable/1392044.

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 and 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.

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.

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.

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.

— (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.