Thesis title

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 Yihui Xie

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, Annaert and De Ceuster for the valuable insights you have given us in preparation of this thesis and the many questions answered. 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 Gruber for inventing "Markdown", to MacFarlane for creating "Pandoc" which converts Markdown to a large number of output formats, and to Xie for creating "knitr" which introduced R Markdown as a way of embedding code in Markdown documents, and "bookdown" which added tools for technical and longer-form writing.Special thanks to Ismay, who created the "thesisdown" package that helped many PhD students write their theses in R Markdown. And a very special thanks to McManigle, whose adaption of Evans' adaptation of Gillow's original maths template for writing an Oxford University DPhil thesis in "LaTeX" provided the template that Ulrik Lyngs in turn adapted for R Markdown, which we also owe a big thank you. Without which this thesis could not have been written in this format (Lyngs 2019).

Finally, we thank Ghalanos (2020b) 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

Abstract

The greatest abstract all times

Contents

List of Figures

List of Tables

List of Abbreviations

1-D, 2-D . . . One- or two-dimensional, referring in this thesis to spatial dimensions in an image.

Otter . . . . . One of the finest of water mammals.

Hedgehog . . . Quite a nice prickly friend.

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 (. . .). 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 (. . .), 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 its past returns (i.e. have very little serial correlation) because they simply follow a Random Walk process (. . .). 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.

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 iid, or independently and identically distributed, another is that they are normally distribution. Are these valid assumptions? 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.* There is no constant variance, but it is time-varying (homoskedasticity). Bollerslev (1987) describes it as “rates of return data are characterized by volatile and tranquil periods”.

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

- 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.*

- 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.

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. Below we summarize some alternative distributions that could be a better approximation of returns than the normal one.

1.1.1 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). The probability density function (pdf), again following Annaert (2021), is given by equation (1.1). 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).

$$f(x) = \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})\sqrt{\pi n}}(1+\frac{x^2}{n})^{-(n+1)/2} \tag{1.1}$$

As can be seen the pdf depends on the degrees of freedom n. To be consistent with Ghalanos (2020a), the following general equation is used for the pdf (1.2).

$$f(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\beta\pi\nu}}\left(1+\frac{(x-\alpha)^2}{\beta\nu}\right)^{-(\nu+1)/2} \tag{1.2}$$

where α, β and ν are respectively the location, scale and shape (tail-thickness) parameters. The symbol Γ is the Gamma function.

Unlike the normal distribution, which depends entirely on two moments only, the student t distribution has fatter tails (thus it has a kurtosis coefficient), if the degrees of freedom are finite. This kurtosis coefficient is given by equation (1.3). This is useful while as already mentioned, the standardized residuals appear to have fatter tails than the normal distribution following Bollerslev (2008).

$$kurt = 3 + \frac{6}{n-4} \tag{1.3}$$

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, 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 (1.4) following Ghalanos (2020a).

$$f(x) = \frac{\kappa e^{\left|\frac{x-\alpha}{\beta}\right|^{\kappa}}}{2^{1+\kappa^{(-1)}}\beta\Gamma(\kappa^{-1})} \tag{1.4}$$

where α, β and κ are respectively the location, scale and shapeparameters .

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 (1.5) presents the skewed t-distribution.

$$f_\xi(z) \equiv \frac{2\sigma_\xi}{\xi + \xi^{-1}} f_1\left(z_\xi\right), \quad z_\xi \equiv \begin{cases} \xi^{-1}\left(\sigma_\xi z + \mu_\xi\right) & \text{if } z \geq -\mu_\xi/\sigma_\xi \\ \xi\left(\sigma_\xi z + \mu_\xi\right) & \text{if } z < -\mu_\xi/\sigma_\xi \end{cases} \tag{1.5}$$

where $\mu_\xi \equiv M_1\left(\xi - \xi^{-1}\right)$, $\sigma_\xi^2 \equiv \left(1 - M_1^2\right)\left(\xi^2 + \xi^{-2}\right) + 2M_1^2 - 1$, $M_1 \equiv 2\int_0^\infty u f_1(u) du$ and ξ between 0 and ∞. $f_1(\cdot)$ is in this case equation (1.1), the pdf of the student t distribution.

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 (1.5) as the skewed t-distribution but with $f_1(\cdot)$ equal to equation (1.4).

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) is given by equation (1.6) (Bollerslev et al. 1994).

$$f\left[\varepsilon_t\sigma_t^{-1};\kappa,\psi\right]=\frac{\kappa}{2\sigma_t\ \cdot\psi^{1/\kappa}B(1/\kappa,\psi)\cdot\left[1+\left|\varepsilon_t\right|^{\kappa}/\left(\psi b^{\kappa}\sigma_t^{\kappa}\right)\right]^{\psi+1/\kappa}} \tag{1.6}$$

where $B(1/\eta,\psi)$ is the beta function $(=\Gamma(1/\eta)\Gamma(\psi)\Gamma(1/\eta+\psi))$, $\psi\eta>2,\ \eta>0\ and\ \psi>0,\ \beta=[\Gamma(\psi)\Gamma(1/\eta)/\Gamma(3/\eta)\Gamma(\psi-2/\eta)]^{1/2})$, the scale factor and one shape parameter κ.

Again the skewed variant is given by equation (1.5) but with $f_1(\cdot)$ equal to equation (1.6) following Trottier and Ardia (2015).

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. 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. Engle regards this formulation as the first ARCH model.

1.2.2 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 in respectively equation (1.7), (1.8) and (1.9). Returns are written as a constant part (μ) and an unexpected part, called noise or the innovation process. 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 from iid, 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 ω, 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 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).

$$y_t = \mu + \varepsilon_t \tag{1.7}$$

$$\varepsilon_t = \sigma_t * z_t, \ where \ z_t \overset{iid}{\sim} (0,1) \tag{1.8}$$

$$\sigma_t^2 = \omega + \alpha_1 * \varepsilon_{t-1}^2 \tag{1.9}$$

From these components 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 (1.10) and (1.11) 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 (1.11) is equal to the unconditional mean in the most simple case. But variations are possible using ARMA (eg. AR(1)) processes.

$$\mathbb{E}_{t-1}(\varepsilon_t) = \mathbb{E}_{t-1}(\sqrt{\omega + \alpha_1 * \varepsilon_{t-1}^2} * z_t) = \sigma_t \mathbb{E}_{t-1}(z_t) = 0 \tag{1.10}$$

$$\mathbb{E}_{t-1}(y_t) = \mu + \mathbb{E}_{t-1}(\varepsilon_t) = \mu \tag{1.11}$$

For the conditional variance, knowing everything that happened until and including period $t-1$ the conditional innovation variance is given by equation (1.12). 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 (1.13), that is why equation (1.9) is called the variance equation.

$$var_{t-1}(\varepsilon_t) = \mathbb{E}_{t-1}(\varepsilon_t^2) = \mathbb{E}_{t-1}(\sigma_t^2 * z_t^2) = \sigma_t^2 \mathbb{E}_{t-1}(z_t^2) = \sigma_t^2 \tag{1.12}$$

$$var_{t-1}(y_t) = var_{t-1}(\varepsilon_t) = \sigma_t^2 \tag{1.13}$$

The unconditional variance is also interesting to derive, while this is the long-run variance, which will be derived in (1.17). After deriving this using the law of iterated expectations and assuming stationarity for the variance process, one would get (1.14) for the unconditional variance, equal to the constant c and divided by $1 - \alpha_1$, the slope of the variance equation.

$$\sigma^2 = \frac{\omega}{1 - \alpha_1} \tag{1.14}$$

This leads to the properties of ARCH models.

- Stationarity condition for variance: $\omega > 0$ and $0 \leq \alpha_1 < 1$.
- Zero-mean innovations
- Uncorrelated innovations

Thus a weak white noise process ε_t

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.

The unconditional 4th moment, kurtosis $\mathbb{E}(\varepsilon_t^4)/\sigma^4$ of an ARCH model is given by equation (1.15). This term is larger than 3, which implicates that the fat-tails (a stylised fact of returns).

$$3\frac{1-\alpha_1^2}{1-3\alpha_1^2} \tag{1.15}$$

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 ω for the conditional variance $var_t(\varepsilon_{t+1}) = \mathbb{E}(\varepsilon_{t+1}^2) = \sigma_{t+1}^2 = c + \alpha_1 * \varepsilon_t^2$. Thus it follows that equation (1.16) displays volatility clustering. If we examine the RHS, as $\alpha_1 > 0$ (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.

$$\sigma_{t+1}^2 - \sigma^2 = \alpha_1 * (\varepsilon_t^2 - \sigma^2) \tag{1.16}$$

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 1.1.1. The serial correlation for squared innovations is positive if fourth moment exists (equation (1.15), this is volatility clustering once again.

The estimation of ARCH model and in a next step GARCH models will be explained in the methodology. However 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 (1.17). This can already be simplified, while we know that $\sigma_{T+1}^2 = \omega + \alpha_1 * \varepsilon_T^2$ from equation (1.9).

$$\begin{aligned} \mathbb{E}_T(\varepsilon_{T+k}^2) &= \omega * (1 + \alpha_1 + ... + \alpha^{k-2}) + \alpha^{k-1} * \sigma_{T+1}^2 \\ &= \omega * (1 + \alpha_1 + ... + \alpha^{k-1}) + \alpha^k * \sigma_T^2 \end{aligned} \tag{1.17}$$

It can be shown that then the conditional variance in period $T+k$ is equal to equation (1.18). 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_T^2 above the unconditional average multiplied by α_1^k, a decreasing function of k (given that $0 \leq \alpha_1 < 1$). The further ahead predicting the variance, the closer α_1^k comes to zero, the closer to the unconditional variance, i.e. the long-run variance.

$$\mathbb{E}_T(\varepsilon_{T+k}^2) - \sigma^2 = \alpha_1^k * (\varepsilon_T^2 - \sigma^2) \tag{1.18}$$

1.2.3 Univariate GARCH models

An improvement on the ARCH model is the Generalised Autoregressive Conditional Heteroscedasticity (GARCH). 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.

All the GARCH models below are estimated using the package rugarch by Alexios 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 does not mathematically allow for a negative output.

GARCH model

The standard GARCH model (Bollerslev 1986) is written consistent with Alexios Ghalanos (2020a) as in equation (1.19) without external regressors.

$$\sigma_t^2 = \omega + \sum_{j=1}^{q} \alpha_j \varepsilon_{t-j}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 \tag{1.19}$$

where σ_t^2 denotes the conditional variance, ω the intercept and ε_t^2 the residuals from the used mean process. The GARCH order is defined by (q, p) (ARCH, GARCH). 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 (1.20).

$$\hat{P} = \sum_{j=1}^{q} \alpha_j + \sum_{j=1}^{p} \beta_j. \tag{1.20}$$

The unconditional variance of the standard GARCH model of Bollerslev is very similar to the ARCH model, but with the Garch parameters (β's) included as in equation (1.21).

$$\begin{aligned}\hat{\sigma}^2 &= \frac{\hat{\omega}}{1-\hat{P}} \\ &= \frac{\hat{\omega}}{1-\alpha-\beta}\end{aligned} \tag{1.21}$$

IGARCH model

Following Alexios 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.

EGARCH model

The EGARCH model or exponential GARCH model (Nelson 1991) is defined as in equation (1.22). 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.

$$\log_e(\sigma_t^2) = \omega + \sum_{j=1}^{q}(\alpha_j z_{t-j} + \gamma_j(|z_{t-j}| - E|z_{t-j}|)) + \sum_{j=1}^{p}\beta_j \log_e(\sigma_{t-j}^2) \tag{1.22}$$

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

GJRGARCH model

The GJRGARCH model (Glosten et al. 1993) models both positive as negative shocks on the conditional variance asymmetrically by using an indicator variable I, it is specified as in equation (1.23).

$$\sigma_t^2 = \omega + \sum_{j=1}^{q}(\alpha_j \varepsilon_{t-j}^2 + \gamma_j I_{t-j}\varepsilon_{t-j}^2) + \sum_{j=1}^{p}\beta_j \sigma_{t-j}^2 \tag{1.23}$$

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).

NAGARCH model

The NAGARCH or nonlinear asymmetric model (Engle and Ng 1993). It is specified as in equation (1.24). 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.

$$\sigma_t^2 = \omega + \sum_{j=1}^{q} \alpha_j(\varepsilon_{t-j} + \gamma_j\sqrt{\sigma_{t-j}})^2 + \sum_{j=1}^{p} \beta_j\sigma_{t-j}^2 \tag{1.24}$$

As before, γ_j represents the *leverage* term.

TGARCH model

The TGarch or threshold model (Zakoian 1994) also models assymetries 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 (1.25).

$$\sigma_t = \omega + \sum_{j=1}^{q} (\alpha_j^+\varepsilon_{t-j}^+\alpha_j^- + \varepsilon_{t-j}^-) + \sum_{j=1}^{p} \beta_j\sigma_{t-j} \tag{1.25}$$

where ε_{t-j}^+ is equal to ε_{t-j} if the term is negative and equal to 0 if the term is positive. The reverse applies to ε_{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 (1.26).

$$\sigma_t = \omega + \sum_{j=1}^{q}(\alpha_j \left|\varepsilon_{t-j}\right|) + \sum_{j=1}^{p}\beta_j\sigma_{t-j} \tag{1.26}$$

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 λ more weight is assigned to recent periods than distant periods. The λ must be less than 1. It is specified as in (1.27).

$$\sigma_t^2 = (1-\lambda)\sum_{j=1}^{\infty}(\lambda^j\varepsilon_{t-j}^2) \tag{1.27}$$

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.

1.3 ACD models

ACD models or Autoregressive Conditional Density models are an extension of the GARCH models. They account for time-varying higher moments.

1.4 Value at Risk

Value-at-Risk (VaR) is a risk metric developed to calculate how much money an investment, portfolio, department or institution such as a bank could lose in a market downturn. According to VaR was adopted in 1998 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 θ_t. Put differently, in 1% of cases the loss would be greater than this amount.It is specified as in (1.28).

$$Pr(y_t \leq \theta_t|\Omega_{t-1}) \equiv \phi \tag{1.28}$$

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

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. 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, the conditional VaR (cVaR) 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 first introduced by (Bertsimas et al. 2004). It is specified as in (1.29).

To calculate θ_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.

$$Pr(y_t \leq \theta_t|\Omega_{t-1}) \equiv \int_{-\infty}^{\theta_t} f(y_t|\Omega_{t-1})\, \mathrm{d}y_t = \phi \tag{1.29}$$

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 $cVaR_{97.5}$.

2

Past literature on the consequences of higher moments for VaR determination

3

Data and methodology

3.1 Data

Here comes text...

3.1.1 Descriptives

Table of summary statistics

Here comes a table and description of the stats

Table 3.1: Summary statistics of the returns

Statistics	Eurostoxx.50	Standardized.Residuals
Arithmetic Mean	-13.2404	-11.7732
Stdev	0.0357	-0.0193
Skewness	0.0167	-0.0409
NA	10.4376	5.7126
NA	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***

Note: This table shows the descriptive statistics of the returns of over the period 1987-01-01 to 2021-04-27. Including minimum, median, arithmetic average, maximum, standard deviation, skewness and excess kurtosis.

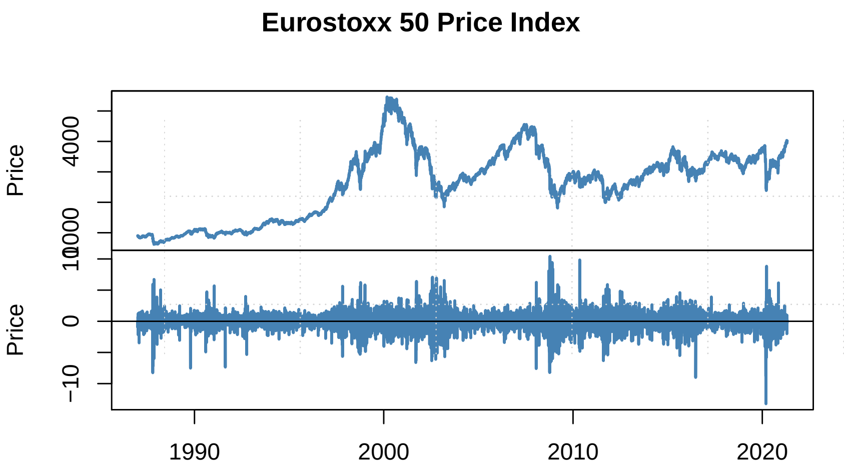


Figure 3.1: Eurostoxx 50 prices and returns

Descriptive figures

As can be seen

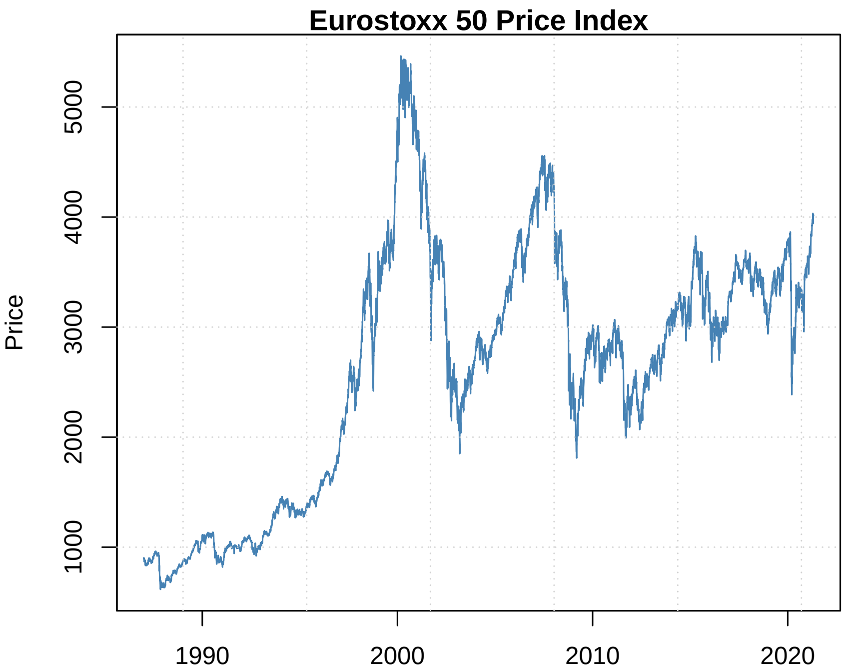


Figure 3.2: Eurostoxx 50 prices and returns

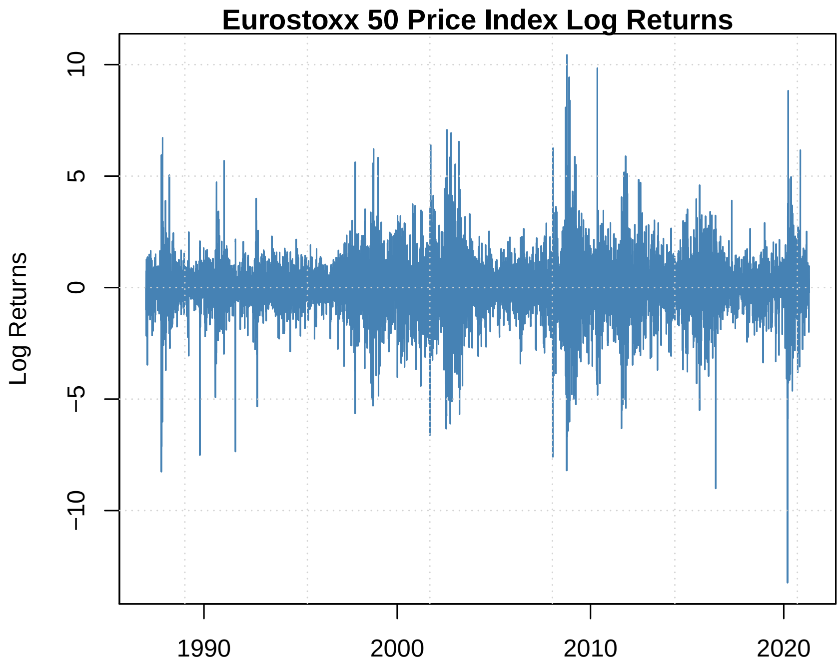


Figure 3.3: Eurostoxx 50 prices and returns

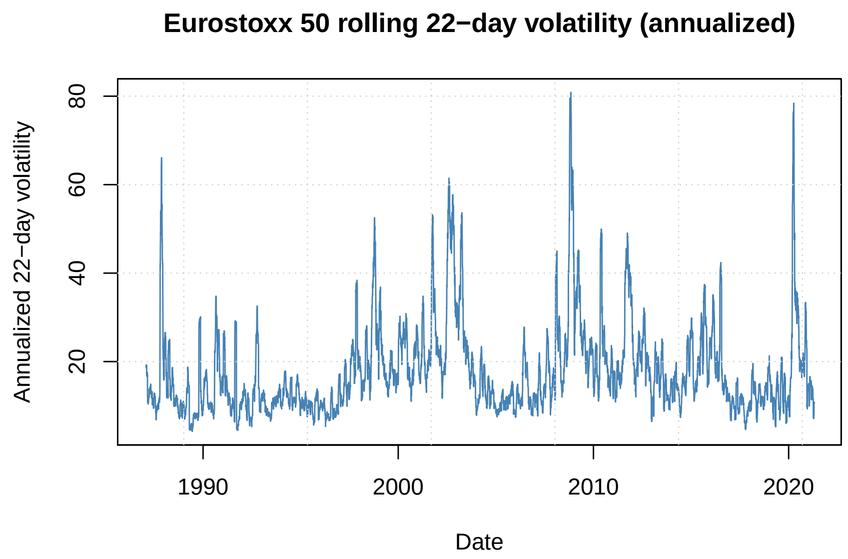


Figure 3.4: Eurostoxx 50 rolling volatility (22 days, calculated over 252 days)

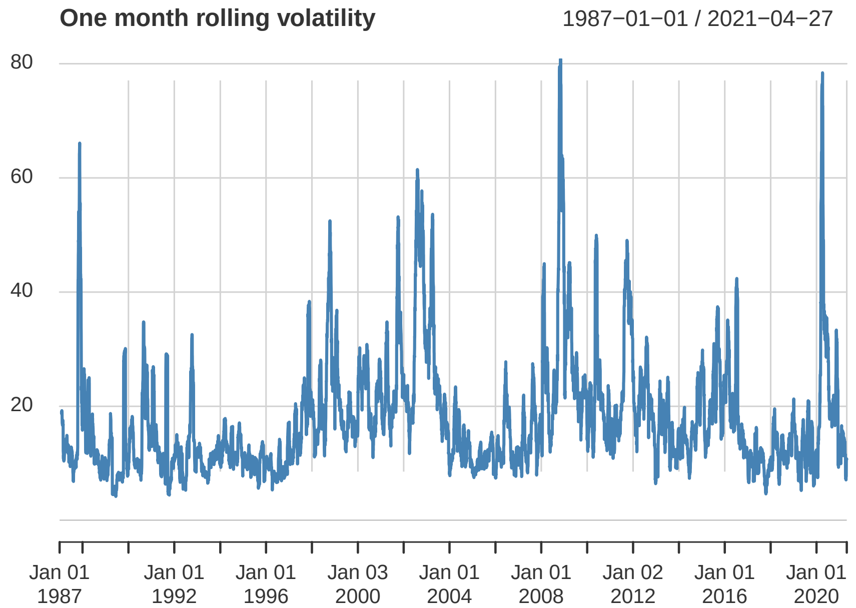


Figure 3.5: Eurostoxx 50 rolling volatility (22 days, calculated over 252 days)

3.1.2 Methodology

Garch models

As already mentioned in . . . , 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.

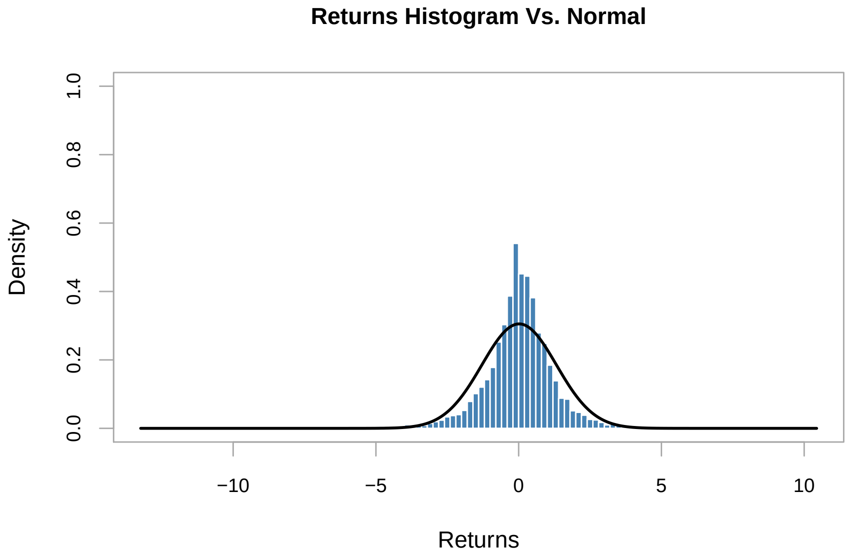


Figure 3.6: Density vs. Normal Eurostoxx 50 log returns)

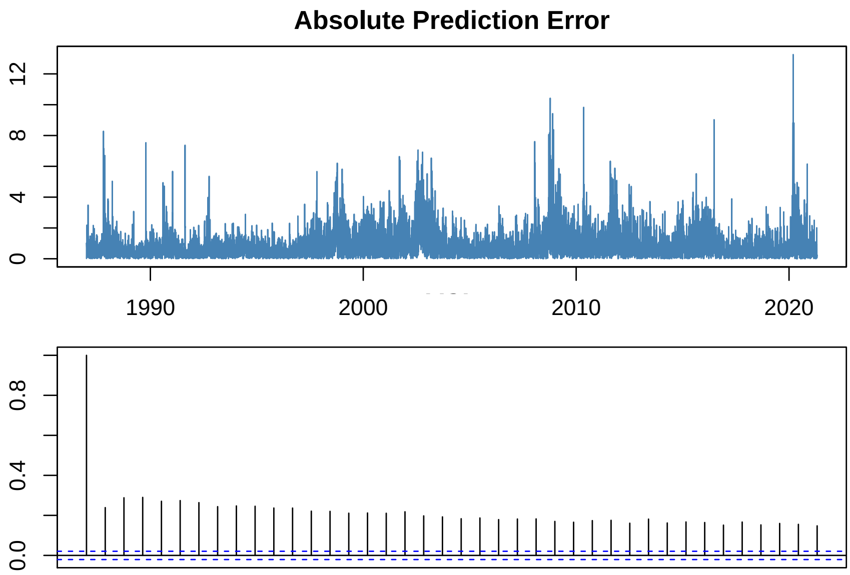


Figure 3.7: Absolute prediction errors

They will be estimated using maximum likelihood. As already mentioned, fortunately, Alexios Ghalanos (2020b) has made it easy for us to implement this methodology in the R language (version 3.6.1) with the package "rugarch" version 1.4-4 (R univariate garch), which gives us a bit more time to focus on the results and the interpretation. Additionally

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 θ that generates the data ((3.2)). These functions are based on the joint probability distribution of the observed data (equation (3.4)). Subsequently, the (log)likelihood function is maximized using an optimization algorithm (equation (3.6)).

$$y_1, y_2, ..., y_N \sim i.i.d \tag{3.1}$$

$$y_i \sim f(y|\theta) \tag{3.2}$$

$$L(\theta) = \prod_{i=1}^{N} f(y_i|\theta) \tag{3.3}$$

$$\log(L(\theta)) = \sum_{i=1}^{N} \log f(y_i|\theta) \tag{3.4}$$

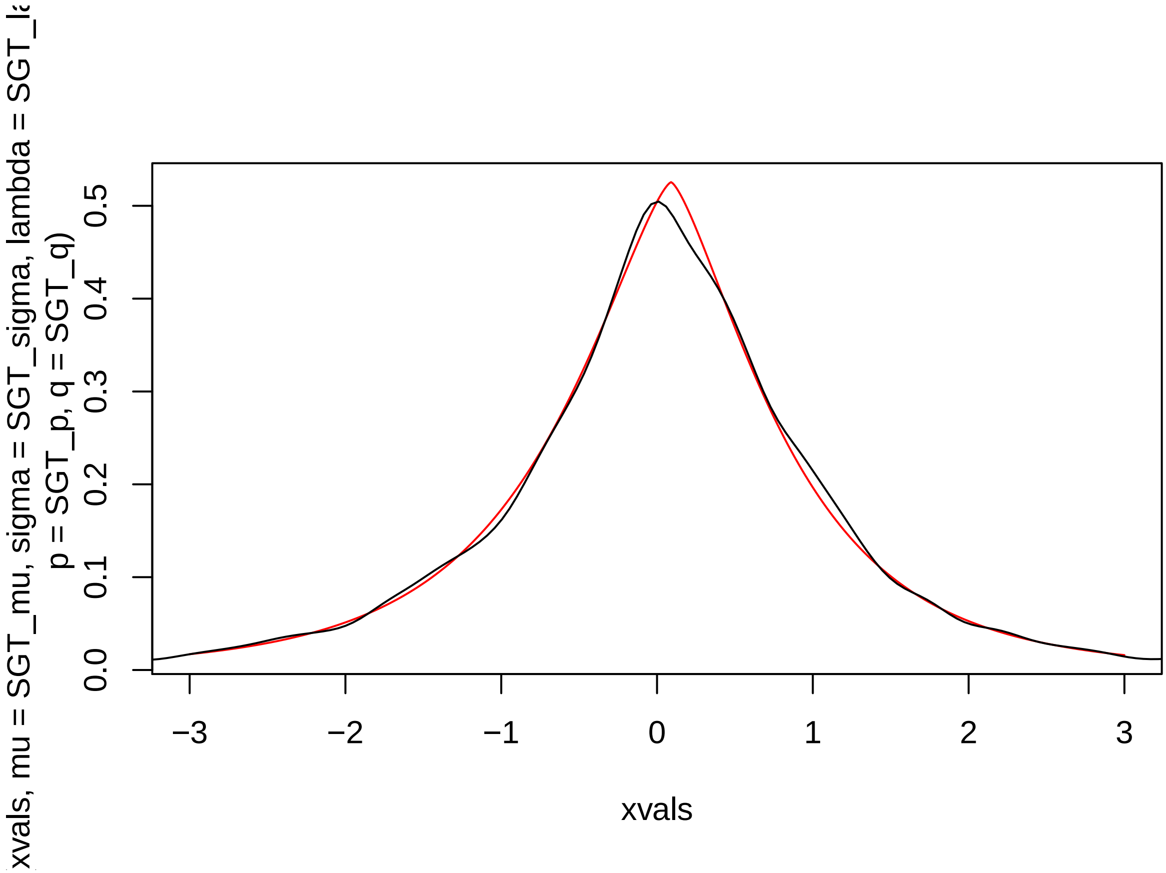
$$\theta^* = arg \max_{\theta}[L] \tag{3.5}$$

$$\theta^* = arg \max_{\theta}[\log(L)] \tag{3.6}$$

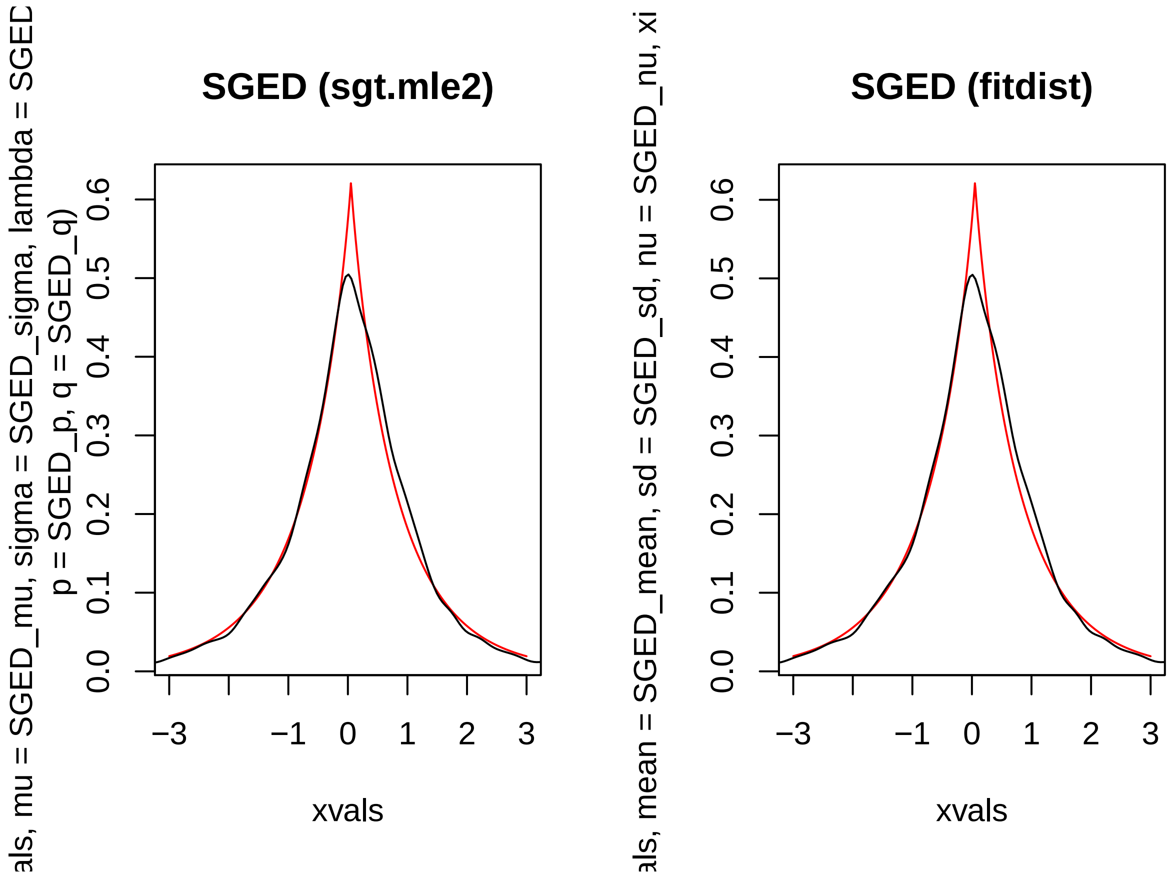
```
##       mean         sd
## 0.01668214 1.30689172
##       mean         sd
## 0.01381119 0.00976596
## [1] -15101.73
##         df        ncp
## 4.31096001 0.03168827
##         df        ncp
## 0.14857777 0.01100453
## [1] -14149.5
##       mean         sd         nu
## 0.03160393 1.27550013 0.91274249
```

```
##         mean          sd          nu
## 0.008555584 0.015772159 0.016622605
## [1] -14009.53
##        mean         sd         nu         xi
## 0.01946361 1.27515748 0.91513166 0.98174821
##         mean          sd          nu          xi
## 0.013176090 0.015786515 0.016652983 0.009638209
## [1] -14008.63
##       mean        sd        nu        xi
## 0.0187729 1.4868913 2.7847974 0.9485825
##        mean         sd         nu         xi
## 0.01375064 0.05550991 0.09972285 0.01270650
## [1] -13997.35
## Skewed Generalized T MLE Fit
## Best Result with BFGS Maximization
## Convergence Code 0: Successful Convergence
## Iterations: NA, Log-Likelihood: -13973.01
##
##           Est. Std. Err.       z  P>|z|
## mu      0.0204    0.0131  1.5574 0.1194
## sigma   1.3214    0.0261 50.5971 0.0000 ***
## lambda -0.0397    0.0126 -3.1583 0.0016  **
## p       1.3818    0.0708 19.5077 0.0000 ***
## q       3.3093    0.5333  6.2058 0.0000 ***
## ---
## Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
## Fitting of the distribution ’ sgt ’ by maximum likelihood
## Parameters :
##           estimate Std. Error
```

```
## mu       0.01974156 0.01263035
## sigma    1.27919321 0.01674109
## lambda -0.03189521 0.01159236
## p        1.09667765        NaN
## q        9.37999498        NaN
## Loglikelihood:  -13984.5   AIC:  27978.99   BIC:  28014.49
## Correlation matrix:
##                 mu       sigma     lambda   p   q
## mu       1.00000000 -0.04998713 0.70347249 NaN NaN
## sigma   -0.04998713  1.00000000 0.04648083 NaN NaN
## lambda   0.70347249  0.04648083 1.00000000 NaN NaN
## p               NaN         NaN        NaN   1 NaN
## q               NaN         NaN        NaN NaN   1
```




```
##       mean         sd         nu         xi
## 0.01946361 1.27515748 0.91513166 0.98174821
```




```
## [1] -14008.18
## [1] -14008.63
```

$A\theta + B = 0$

####Control Tests

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 exceedence 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 (3.7), with p the probability of an exceedence for a confidence level, N

the sample size and X the number of exceedence. 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.

$$LR^{uc} = -2\ln\left(\frac{(1-p)^{N-X}p^X}{\left(1-\frac{X}{N}\right)^{N-X}\left(\frac{X}{N}\right)^X}\right) \tag{3.7}$$

#####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 (3.8).

$$I_t(\alpha) = \begin{cases} 1 & \text{if exceedence occurs} \\ 0 & \text{if no exceedence occurs} \end{cases}. \tag{3.8}$$

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.

Dynamic quantile test \ engle2004 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

4 Empirical Findings

4.1 Results of GARCH with constant higher moments

4.2 Results of GARCH with time-varying higher moments

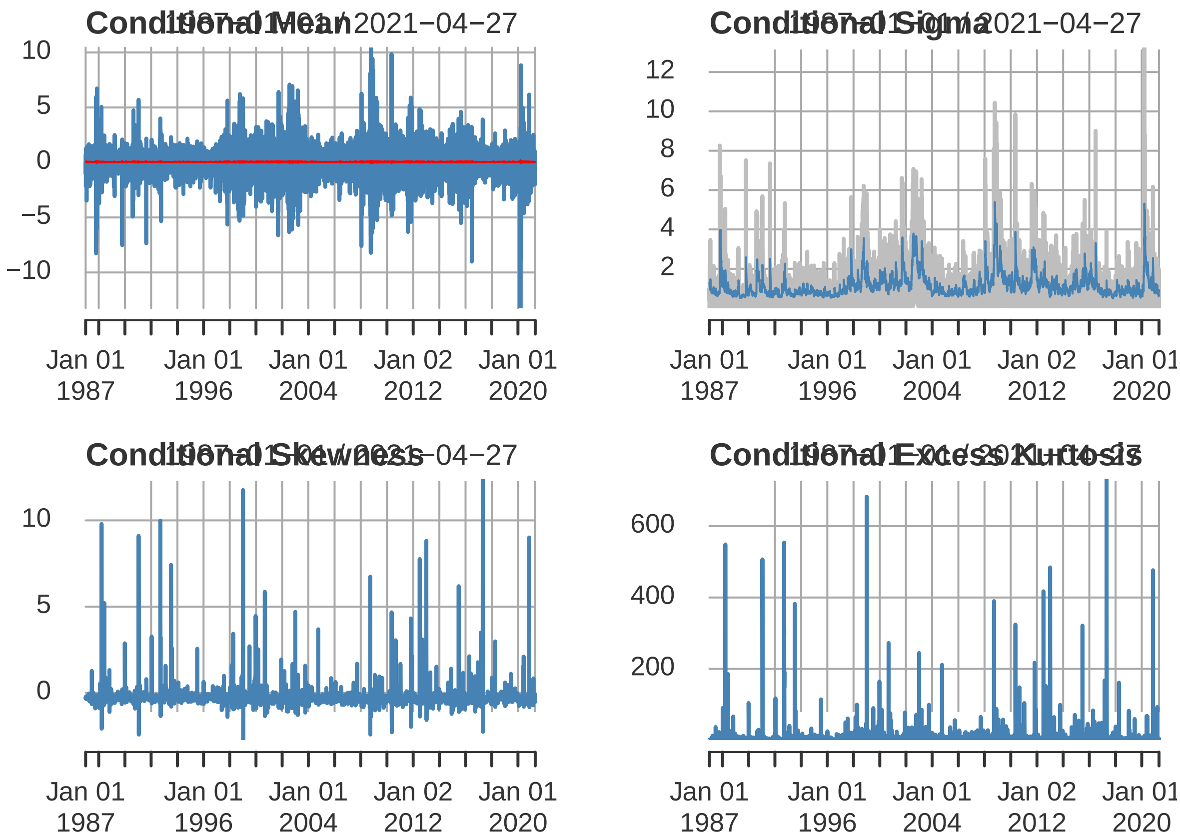


Figure 4.1: Dynamics of the ACD model

5
Robustness Analysis

5.1 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:

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

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

5.1.1 Eye-balling econometrics

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.

5.1.2 GMM test

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

Conclusion

Appendices

A
Appendix

Works Cited

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. URL: www.elsevier.com/locate/econbase.

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: Available%20at%20SSRN:%20https://ssrn.com/abstract=1263250.

Bollerslev, Tim, Robert F. Engle, and Daniel B. Nelson (Jan. 1994). "Chapter 49 Arch models". In: *Handbook of Econometrics* 4. Publisher: Elsevier, pp. 2959–3038. DOI: 10.1016/S1573-4412(05)80018-2.

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. URL: https://linkinghub.elsevier.com/retrieve/pii/S0927539801000251.

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. 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. URL: http://doi.wiley.com/10.1111/j.1540-6261.1993.tb05127.x.

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

— (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.

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.

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.

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.

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.

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. URL: http://pubsonline.informs.org.https//doi.org/10.1287/mnsc.44.12.1650http://www.informs.org0025-1909/98/4412/1650$05.00.

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.