# Tail Behavior and Extreme Value Analysis of GARCH(1,1) Models

**Hugo Belzil**

**MATH 598: Extreme Value Analysis**

Pr. Johanna Nešlehová

December 2024

# Contents

# Abstract

GARCH models are popular in finance as they capture efficiently the time-dependence of volatility of return series. It is well-known that most of the return series in several fields of finance behave in a heavier-tailed way than the normal distribution or other light-tailed distributions. Understanding how these models behave in the tails is important for practitioners in quantitative finance, especially for those working on risk management. In this project report, we look at the behavior in the tails of GARCH$(1, 1)$ models, specifically by looking at the tail-index of their marginal distribution. We discuss these behaviors as functions of various parameters including the distributions of the innovations, apply them to a dataset, and compare our results with a popular Extreme Value Theory (EVT) technique : the block-maxima method.

Mikosch and Stărică (2000)[1] provide key theoretical results on the regular variation of $\sigma_t$, which are pillars of the results we present. We also refer to the computations of tail-indices $\kappa$ and some other results to Embrechts, McNeil and Frey (2005)[2]

The code used for producing the tables, the data analysis, and the numerical computations is available on this GitHub repository

# 1 Introduction

## 1.1 Motivation for GARCH$(p, q)$ models

Real-life financial data often exhibits time-dependence, as well as non-constant volatility (often referred to as variance). Indeed, this is because financial products may be sensitive to several factors such as geopolitical events, climate, economic policies,... which can come and go with time. As a result, there may be periods of high volatility (during an unstable economic situation for example), but there could also be calmer periods where fluctuations of the price are less important. This can be highlighted by the picture below, which shows the daily log-return series of a stock listed on the NYSE, from November 2020 to May 2021.

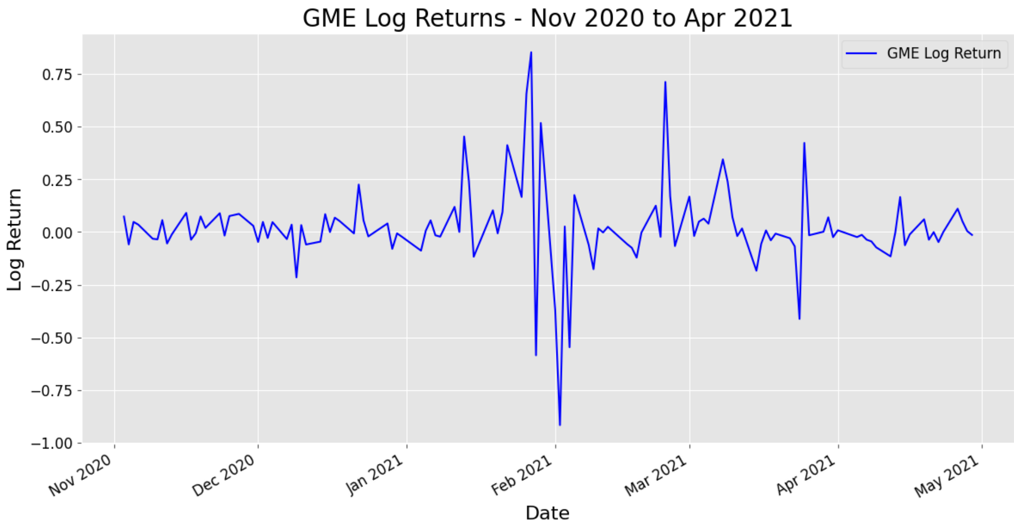



Figure 1: Daily Log-Returns of the GameStop stock between November 2020 and May 2021

On the log-return series above, we can see that the variance (or volatility) of the returns is relatively stable, and seems constant between November and December 2020. On the other hand, a higher amplitude in the returns is observed between January and February 2021, and more importantly, the outbursts in volatility seem to be clustered in time. What we mean here is that one value high in magnitude in the returns usually does not come alone, and most likely depends on the recent past, clearly highlighted by the length of this instability here.
The primary goal of GARCH models is to capture the time-dependence and clustering of variance in time-series. This is argued below in section 1.2.

## 1.2 Definition

This section aims at introducing the GARCH$(p, q)$ model and its vocabulary. GARCH stands for Generalized Auto-Regressive Conditional Heteroskedasticity, meaning that it models the change in conditional variance (heteroskedasticity). In particular, we say it is generalized in the sense that it extends the ARCH model by introducing past conditional variances, and the auto-regressive part comes from the recurrence equation described below. A time-series $\{X_t\}, t \in \mathbb{Z}$, is said to be generated by a GARCH$(p, q)$ process if it is of the form:

$$X_t = \sigma_t Z_t, \qquad \sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i Z_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 \tag{1}$$

where $\{Z_t\}_{t\in\mathbb{Z}}$ is a sequence of i.i.d. random variables, typically centered at 0, with variance 1. $\alpha_0, \alpha_i$'s and $\beta_j$'s are positive parameters. The values of $Z_t$ are referred to as the "innovations", and it is common to study models where $Z_t$ is normally distributed with mean 0 and variance 1. However, financial applications typically require $Z_t$ to be more heavy-tailed than gaussians, as we will discuss below. This is often translated by letting $Z_t$ being Student's $t$-distributed. In this case, we scale the innovations to keep their variance equal to 1 : if $Y \sim t_\nu, \nu > 2$, then $Z_t = \sqrt{\frac{\nu-2}{\nu}}Y \implies Var(Z_t) = 1$. Note that from now on, we omit to write the scaling factor, but all innovations described are scaled (when necessary) such that their variance is 1. We will also refer to $\sigma_t$ as the stochastic volatility.

From the definition of the return series $X_t$ (1), we can see that its variance at time $t$, depends on the past $q$ variances observed at times $t-1, \ldots, t-q$, and also on the $p$ past values of the innovations $Z_{t-1}, \ldots, Z_{t-p}$. Specifically, we have: $Var(X_t|\mathcal{F}_{t-1}) = \sigma_t^2$, where $\mathcal{F}_{t-1}$ denotes the observed values at all times before $t$. This is the argument we elaborated in section 1.1, which explains why GARCH models are efficient at capturing the clustering and time-dependence of variance.

For this project, we focus only on the GARCH$(1,1)$ model, for which the stochastic volatility $\sigma_t$ only depends on the previous innovation, and the previous variance. Hence, $\{X_t\}_{t\in\mathbb{Z}}$ is defined by:

$$X_t = \sigma_t Z_t, \qquad \sigma_t^2 = \alpha_0 + \alpha_1 Z_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \tag{2}$$

Note that $\{X_t\}$ could also not be centered at 0, meaning that we can define $X_t$ as $X_t = \mu + \sigma_t Z_t$, for some $\mu \in \mathbb{R}$.

## 1.3 A look at the marginal distribution of $X_t$

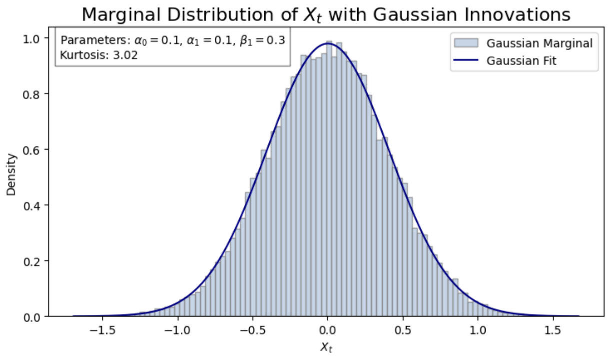
Before going deeper in studying the extreme value behavior of GARCH$(1,1)$ models, we would like to visually inspect their marginal (not conditional!) distribution as a function of their paramaters $\alpha_0, \alpha_1,$ and $\beta_1$. To achieve this, we simulated a few models using the fairly simple recurrence equation described in (2). We used 30,000 points for each simulation, and plotted the marginal distribution for each simulation, with various parameters. For each set of parameters $(\alpha_0, \alpha_1, \beta_1)$, we look at our model using 2 different distributions for the innovations: the first one is using $Z_t \sim N(0,1)$, and the second one using $Z_t \sim t(\nu \in \{4, 8\})$.

On Figure 2, we see six marginal distributions, each one for a model with distinct parameters. Models 1, which is given with relatively "low" values for the parameters $(\alpha_1, \beta_1)$. With $Z_t \sim N(0,1)$, we can see that the histogram of the empirical distribution matches almost perfectly a normal distribution fitted with the mean and variance of the model. We could therefore be tempted to think that the distribution of the innovations determines the distribution of $X_t$. However, we are quickly brought to conclude that this is not true, by looking at model the same set with $t_4$-distributed innovations. For the same set of parameters, the Student-$t$ distributed innovations yield a much more heavy-tailed distribution for $X_t$, this is highlighted by looking at the support of the distribution, and by the excess kurtosis (14.59). Moreover, the p.d.f. of the fitted $t$ distribution does not match with the histogram of the model.

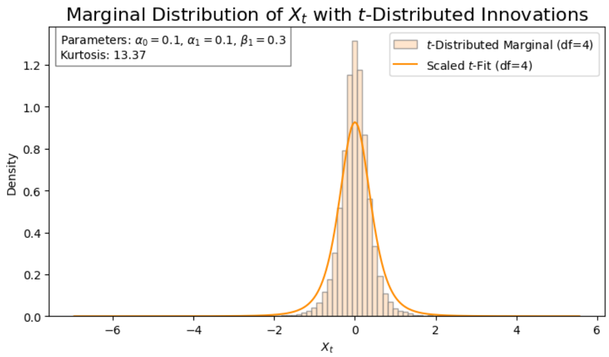
Looking at these figures from top to bottom, we can also conclude that for a fixed set of parameters, increasing the values of $\alpha_1$ and $\beta_1$ tends to increase the kurtosis (or heavy-tailedness) of the model, even when the innovations remain the same. The bell-shape appearance of the fitted distribution that we had for the first model completely fails for the third model, where $(\alpha_1, \beta_1) = (0.8, 0.8)$, and fails even harder for $(\alpha_1, \beta_1) = (0.47, 0.47)$. In fact, we will see in section 2.2.2 that as $\alpha_1 + \beta_1 \to 1$, some interesting things happen in the tails of the models.
Our two (non-rigorous) take-aways from looking at these plots are:

- Heavy-tailedness of $X_t$ for fixed $(\alpha_0, \alpha_1, \beta_1)$ seems to be determined by the heavy-tailedness of the innovations. The heavier the tails of the innovations, the higher the kurtosis of the marginal distribution.

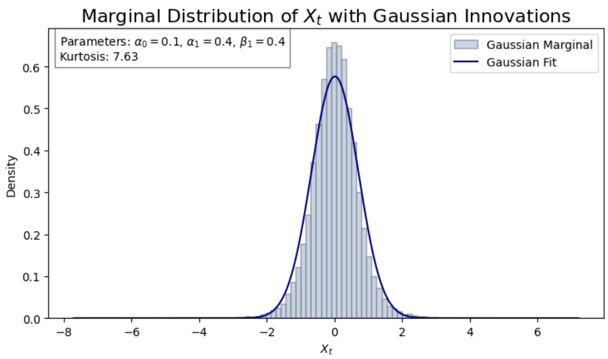
- Heavy-tailedness of $X_t$ for fixed innovations$(\alpha_0, \alpha_1, \beta_1)$ seems to be dependent on $(\alpha_1, \beta_1)$



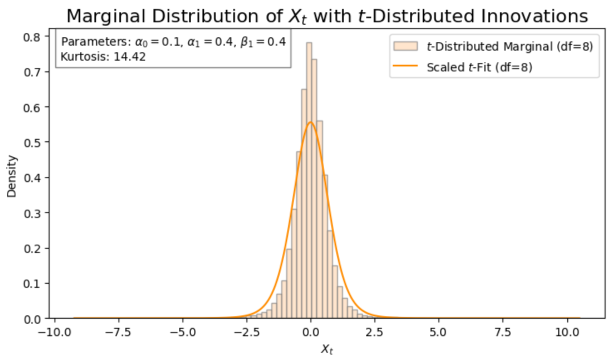

(a) Model 1 ($Z_t \sim N(0, 1)$)



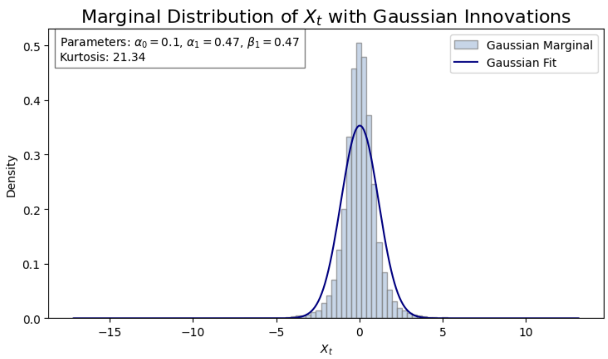

(b) Model 1 ($Z_t \sim t_4$)



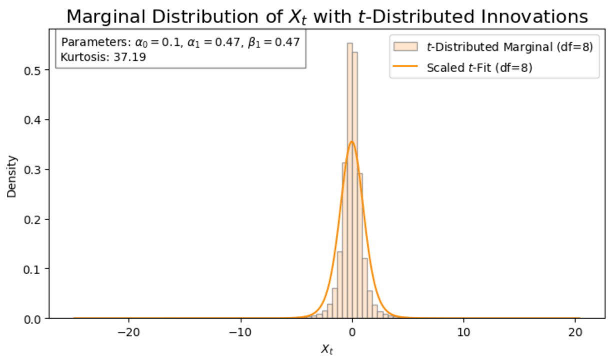

(c) Model 2 ($Z_t \sim N(0, 1)$)




(d) Model 2 ($Z_t \sim t_8$)




(e) Model 3 ($Z_t \sim N(0, 1)$)




(f) Model 3 ($Z_t \sim t_8$)

Figure 2: Marginal distribution of 3 GARCH models. Each row corresponds one model with a fixed set of parameters $(\alpha_0, \alpha_1, \beta_1)$, with 2 different distributions for the innovations (gaussians on the left, Student-$t$ on the right.)

Note that we omit to study in depth the influence of the parameter $\alpha_0$ in our analysis, as it is of lesser importance in the extreme value behavior of GARCH models, unlike $\alpha_1$ and $\beta_1$

# 2 Tail-behavior of $X_t$

Quantitative risk management practitioners frequently use GARCH$(1, 1)$ models and are often interested in understanding their tail behavior for effective risk assessment. In this section, we present a few results and arguments about random variables and GARCH models, which lead us to computing the tail index $\kappa$ of our models

## 2.1 Breiman's Lemma

An important result for diving into the extremal behavior analysis of GARCH models is Breiman's Lemma. It can be found in [2] (Theorem 7.3.5) and stated as follows:

**Breiman's lemma :** Let $X, Y$ be two non-negative and independent random variables such that:

- $Y$ has a regularly varying tail with index $\alpha \in \mathbb{R}^+$
- $\mathbb{E}[X^{\alpha+\epsilon}] < \infty$ for some $\epsilon > 0$

Then, the random variable $Z = XY$ also has a regularly varying tail with index $\alpha$.
As a reminder, a random variable $Z$ with c.d.f. $F$ is said to have a regularly varying with tail index $\alpha$ if its survival function $\bar{F}$ belongs to $\mathcal{R}_{-\alpha}$ (the set of regularly varying functions with index $\alpha$).

## 2.2 Results on Regular Variation of the stochastic volatility $\sigma_t$

In their paper [1], Mikosch and Stărică (2000) explain that under some "relatively weak" conditions on the innovations $Z_t$, the stochastic volatility $\sigma_t$ of a GARCH model can be assumed to have a regularly varying tail. We consider these relatively technical conditions to be weak in the sense that in the context of GARCH models, they quite often hold. We discuss these conditions below, in section 2.2.1.

**Mikosch and Stărică (2000) :** Under weak assumptions on the innovations $Z_t$ of a GARCH(1,1) model, the stochastic volatility $\sigma_t$ has a regularly varying tail with tail index $\kappa$, where $\kappa$ is the non-negative solution to the equation:

$$\mathbb{E}[(\alpha_1 Z_t^2 + \beta_1)^{\kappa/2}] = 1 \tag{3}$$

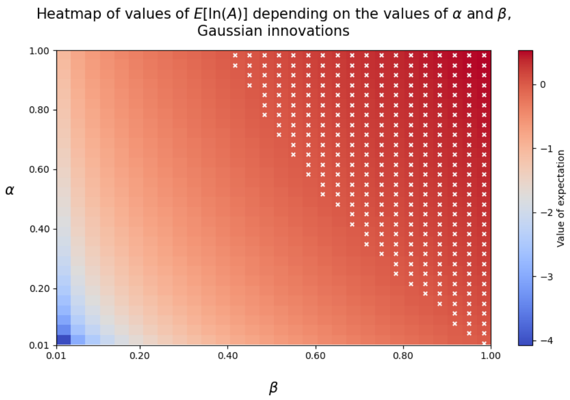
### 2.2.1 Mikosch & Stărică conditions

In [1], Mikosch and Starica describe in Theorem 2.1 the conditions needed for (3) to have a non-negative solution $\kappa$. The following two are necessary but not sufficient conditions for the existence of a solution.

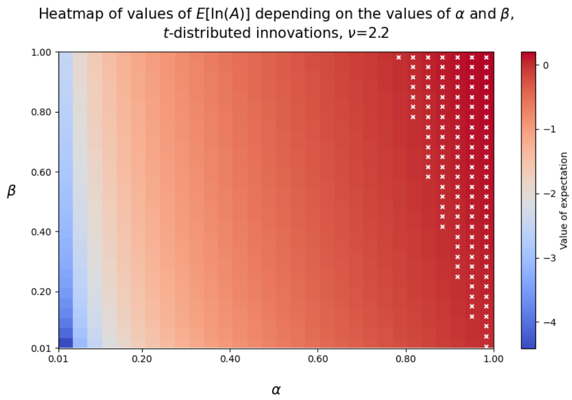
*Let the random variable $A = \alpha_1 Z_t^2 + \beta_1$, for some random innovations $Z_t$ with mean 0 and variance 1. The following must hold true for a non-negative solution $\kappa$ to exist for* (3)

- $\mathbb{E}[\ln A] < 0$ and $\mathbb{P}(A > 1) > 0$
- $\exists\, h_0 \in \mathbb{R}^+$ such that $\mathbb{E}[A^{h_0}] = \infty$ and $\mathbb{E}[A^h] < \infty \quad \forall\, h \leq h_0$

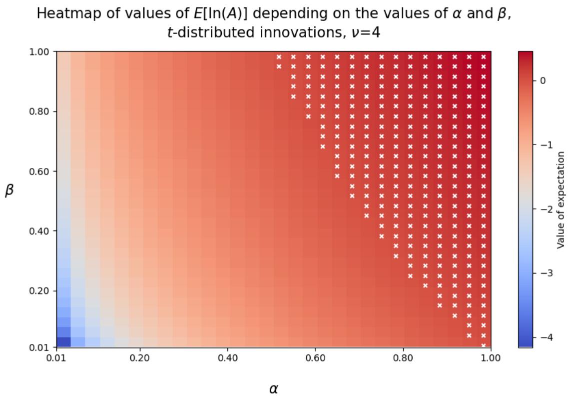
Checking for the first condition, can be done using numerical integration. Figure 3 displays heatmaps representing the value of $\mathbb{E}[\ln(\alpha Z_t^2 + \beta)]$ where $(\alpha, \beta) \in [0.01, 1]^2$. The set $[0, 1]^2$ is approximated with a linear space over $(0.01, 1)$ with $n = 30$ cuts. A white cross, indicates that the expectation is greater than 0, hence meaning a non-negative solution $\kappa$ does not exist for (3).



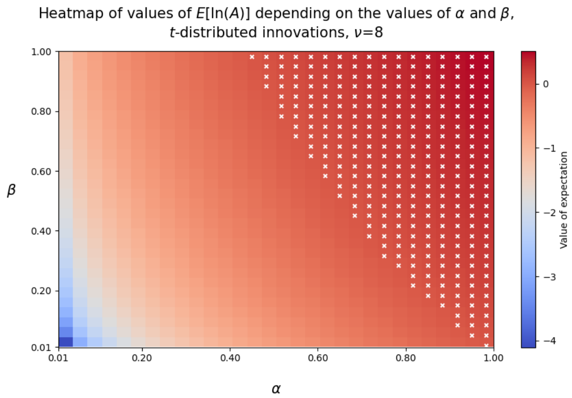

(a) Gaussian innovations




(b) $t_{2.2}$ innovations




(c) $t_4$ innovations




(d) $t_8$ innovations

Figure 3: Heatmaps of $\mathbb{E}[\ln(A)]$ for various innovations, over the parameter space $(\alpha, \beta) = [0.01, 1]^2$

### 2.2.2 Solution when $\alpha_1 + \beta_1 = 1$

A second interesting result due to Mikosch and Stărică is the following:

*If the conditions of Theorem 2.1 in [1] hold, and* $\alpha_1 + \beta_1 = 1$*, then the unique solution of* (3) *is* $\kappa = 2$.

This can be seen on figures 4 and 6, respectively for the set of parameters $(\alpha_1, \beta_1) = (0.05, 0.95)$ and $(\alpha_1, \beta_1, \nu) = (0.1, 0.9, 4)$. This statement above holds for any combination of innovations and parameters satisfying the Mikosch and Stărică conditions.

## 2.3 Tying both results : regular variation and MDA of $X_t$

Breiman's lemma and the result on $\sigma_t$'s regular variation come in together and allow us to claim that under the existence of a solution $\kappa$ to (3) and the finiteness of $\mathbb{E}[Z^{\kappa+\epsilon}]$ for some positive $\epsilon$, the GARCH process $X_t$ is regularly varying with tail index $\kappa$.
**Remark :** Note that $Z_t$ is not a positive random variable, which is a condition for applying

Breiman's Lemma. However, this result still holds, by considering $\tilde{Z} = \max\{Z_t, 0\}$, which lies in the same MDA as $Z_t$

Moreover, with these assumptions holding, Gnedenko's theorem also implies that the GARCH process $X_t$ is in the maximal domain of attraction of the Fréchet distribution with parameter $\kappa$, i.e. $MDA(\Phi_\kappa)$.
The most important result this gives us is that GARCH$(1,1)$ marginals have Pareto-like tails under the assumptions mentioned above. As highlighted in [1], this also indicates us that:

$$\mathbb{P}(\sigma > x) \sim c_0 x^{-\kappa} \text{ and } \mathbb{P}(X > x) \sim \mathbb{E}[|Z|^\kappa]\mathbb{P}(\sigma > x) \implies \mathbb{P}(X > x) \sim c_1 x^{-\kappa} \tag{4}$$

for some positive $c_0, c_1$.

## 2.4 Computing the tail index $\kappa$

As a conclusion to this section, we finally discuss how to actually compute for $\kappa$, which involves solving equation (3). We address 2 types of distributions for the innovations, Gaussians (standard normal) and Student's $t$ in two different sections as they do not necessit the same level of rigor in the calculations.
The expectation of (3) could be studied either under a discrete or continous aspect:

$$\begin{aligned}\mathbb{E}[(\alpha_1 Z^2 + \beta_1)^{\kappa/2}] &= \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}(\alpha_1 z_i^2 + \beta_1)^{\kappa/2} \\ &= \int_{-\infty}^{\infty}(\alpha_1 z^2 + \beta_1)^{\kappa/2} f_Z(z)dz\end{aligned} \tag{5}$$

However, our study revealed high variance of the discrete sum of (5), even for values of $n$ as large as 2 millions, for certain innovations and parameters $(\alpha_1, \beta_1)$. Thus, we discard this option and proceed with the integral definition of expectation. Another problem now arises; keep in mind that the goal for a given set of parameters $(\alpha_1, \beta_1)$ is to solve (3), i.e. we would like to solve:

$$\int_{-\infty}^{\infty}(\alpha_1 z^2 + \beta_1)^{\kappa/2} f_Z(z)dz = 1 \tag{6}$$

If $Z \sim N(0,1)$ or $Z \sim t_\nu$, this integral is probably not computable by hand, and solving for $\kappa$ is analytically impossible. Fortunately, we can use numerical integration along with Newton-Raphson's method to solve (6). For this purpose, we use the `SciPy` framework, and particularly the `scipy.integrate.quad` method. We also use the `scipy.stats` framework to obtain the p.d.f.'s $f_Z(z)$ for both distributions.

For a given p.d.f. $f_Z(z)$ and an initial guess $\kappa_0 > 0$, an iteration of Newton's method gives us.

$$\kappa_{n+1} = \kappa_n - \frac{L(\kappa_n)}{L'(\kappa_n)} \tag{7}$$

where:

$$L(\kappa) = \int_{-\infty}^{\infty}(\alpha_1 z^2 + \beta_1)^{\kappa/2} f_Z(z)dz - 1 \tag{8}$$

and using Leibniz's rule for differentiation under integrals:

$$\begin{aligned}L'(\kappa) &= \frac{\partial}{\partial\kappa}\left[\int_{-\infty}^{\infty}(\alpha_1 z^2 + \beta_1)^{\kappa/2} f_Z(z)dz - 1\right] \\ &= \frac{1}{2}\int_{-\infty}^{\infty}\ln(\alpha_1 z^2 + \beta_1)(\alpha_1 z^2 + \beta_1)^{\kappa/2} f_Z(z)dz\end{aligned} \tag{9}$$

These iterations must be treated with caution as we will discuss below, especially when dealing with $t$-distributed innovations. The main problems we could encounter are the following:

- Does $L(\kappa)$ exist for all $\kappa \in \mathbb{R}_{++}$ ?
- How should we choose the initial guess $\kappa_0$?
- How can we avoid convergence to the solution $\kappa = 0$, which obviously is not of interest since we assume that $\sigma_t$ is regularly varying with tail-index $\kappa > 0$.

We address these points independently for both innovations distributions below.

### 2.4.1 Gaussian innovations

With Gaussian innovations, the iteration described in (7) does not present particular issues for the existence of $L(\kappa)$ and $L'(\kappa)$. This is due to the fact that the $n$-th moment of $Z_t$ is finite for all $n \in \mathbb{N}$.

This gives us flexibility in choosing the initial guess $\kappa_0$, as we know at least that even for a large value of $\kappa_0$, our iteration should behave correctly and eventually converge to the solution. In the available Python code, we set $\kappa_0 = 40$, which is a very large tail-index, probably not common in most financial applications. Note that convergence at $\kappa = 0$ is avoided by setting $\kappa_0$ to a larger value than the non-zero solution $\kappa$, hence motivating our high intital guess. Note also that all our plots of the function $L(\kappa)$ seem to indicate that $L$ is convex on $[0, 40]$ for the range of parameters that we study, which is a good news for the convergence of our method.

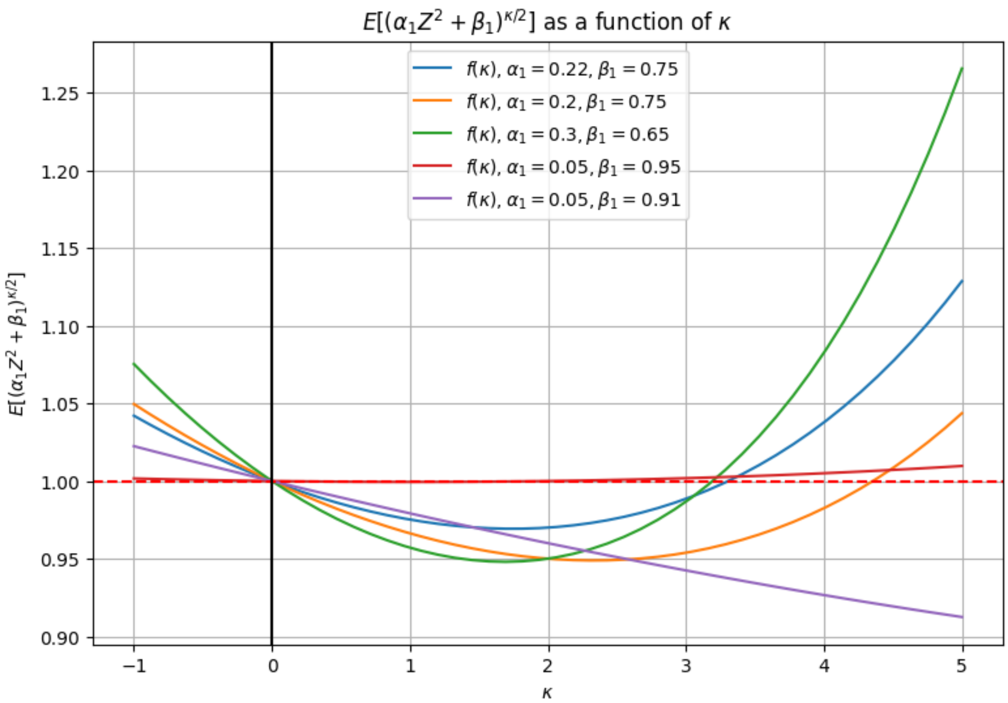



Figure 4: $\mathbb{E}[(\alpha_1 Z_t^2 + \beta_1)^{\kappa/2}]$ for different sets of parameters $(\alpha_1, \beta_1)$, where $Z_t \sim N(0, 1)$

A few clarifications should be brought with this plot, especially for 2 sets of parameters:

- $(\alpha_1, \beta_1) = (0.05, 0.91)$ : the solution $\kappa = 0$ is the only visible solution. However, plotting this function with this exact set of parameter over a larger range shows that another solution is $\kappa = 19.12$ (true value computed with Newton's method), as shown below on figure 5.

- $(\alpha_1, \beta_1) = (0.05, 0.95)$ : Due to the definition of the picture, it is hard to tell but the function intersects the line $y = 1$ both at $\kappa = 0$ and $\kappa = 2$. This is due to the special case $\alpha_1 + \beta_1 = 1$ that we presented in section 2.2.2.

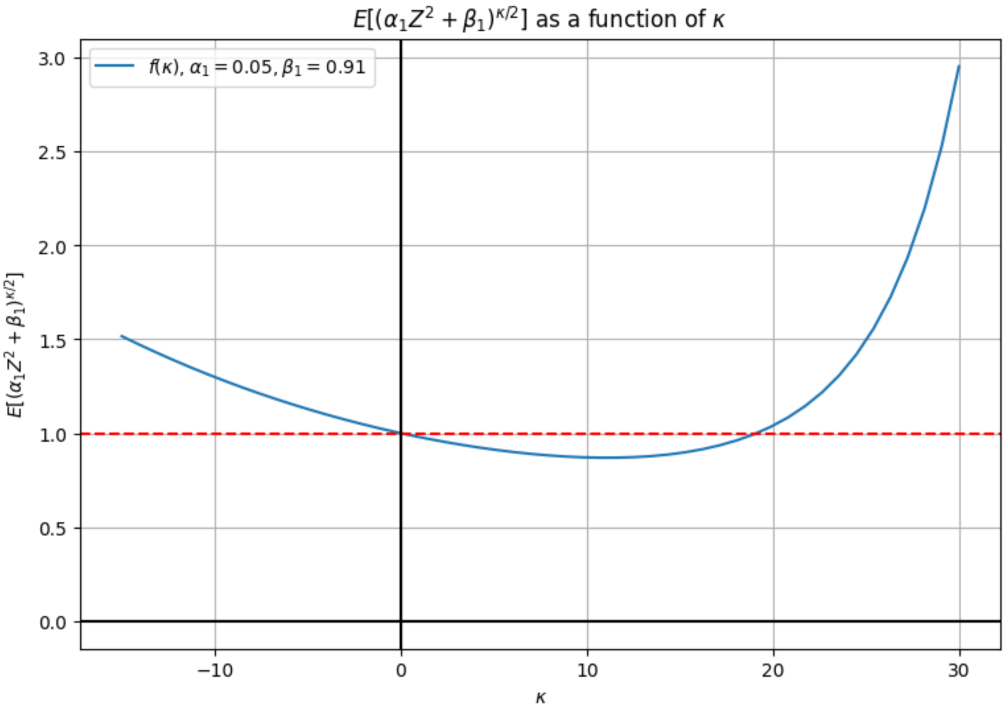



Figure 5: $\mathbb{E}[(\alpha_1 Z^2 + \beta_1^{\kappa/2})]$ for the set of parameters $(\alpha_1, \beta_1) = (0.05, 0.91)$, where $Z_t \sim N(0, 1)$. Note that the range is much largen than for figure 4

### 2.4.2 Student's $t$ distributed innovations

When dealing with $t$-distributed innovations, one has to be careful with the initial guess $\kappa_0$ of Newton's iteration. That is because for a random variable that is $t$-distributed with $\nu$ degrees of freedom, $\mathbb{E}[X^\alpha] = \infty$ for $\alpha \geq \nu$. In this case, the integral $L(\kappa_0)$ in (7) will not converge and the iteration can not be produced. Thus, the initial guess $\kappa_0$ should always be set such that $\kappa < \nu$. However, we do not want to set $\kappa_0$ to low neither, since this will again make our algorithm converge to $\kappa = 0$. Hence, we recommend setting $\kappa_0 = \nu - \delta$ for some small delta such as $\delta = 0.01$ or $\delta = 0.1$. Be careful! A too small value for $\delta$ may cause numerical instability, such as too slowly convergent integrals for `scipy.integrate.quad`.

Below is a plot of $\mathbb{E}[(\alpha_1 Z^2 + \beta_1^{\kappa/2})]$ with various parameters $(\alpha_1, \beta_1, \nu)$.

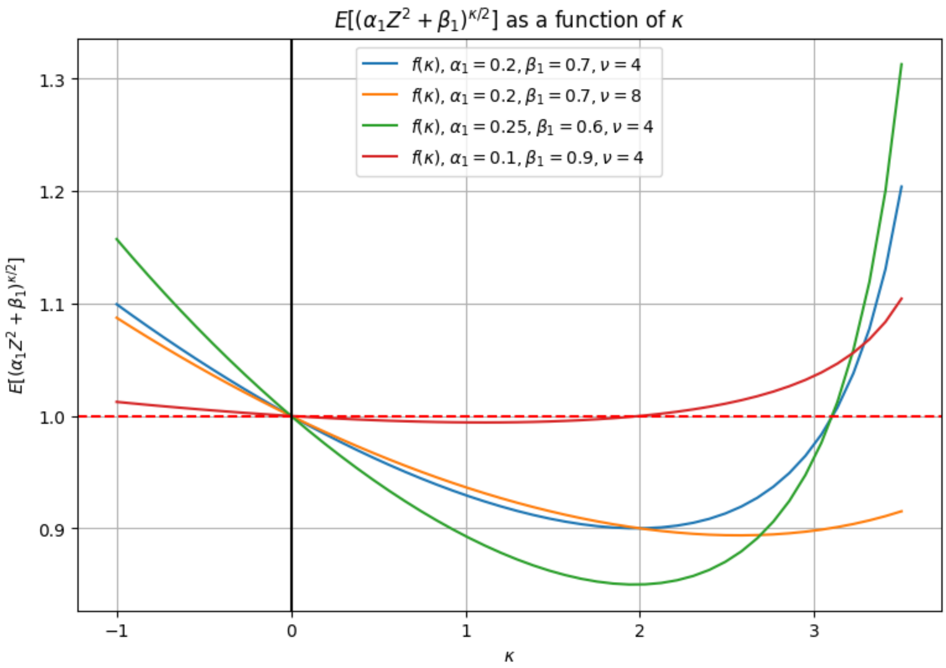



Figure 6: $\mathbb{E}[(\alpha_1 Z_t^2 + \beta_1^{\kappa/2})]$ with various parameters $(\alpha_1, \beta_1, \nu)$. Note that one set of parameters again gives the solution $\kappa = 2$

# 3 Application to the Dow Jones Index data

## 3.1 Fitting the model

In order to apply the tail-index computations that we covered in section 2, we fitted a GARCH$(1, 1)$ model to the Dow Jones data available through the `ismev` package in R. This dataset contains daily closing prices of the Dow Jones index from roughly 1995 to 2000 (1304 points). A few exploratory data analysis plots are given below. Fitting the models is done with the `rugarch` package.

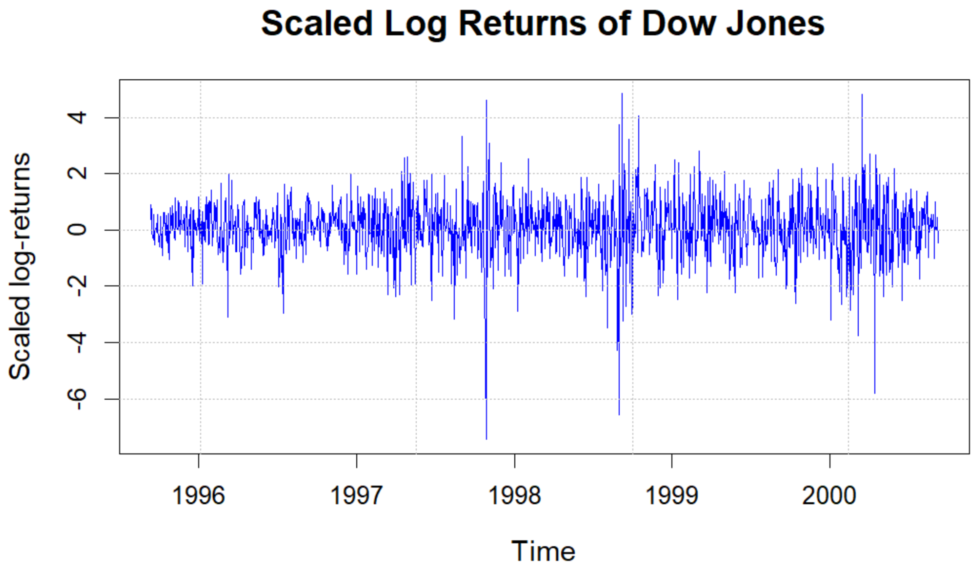



Figure 7: Scaled log-returns of the Dow Jones index

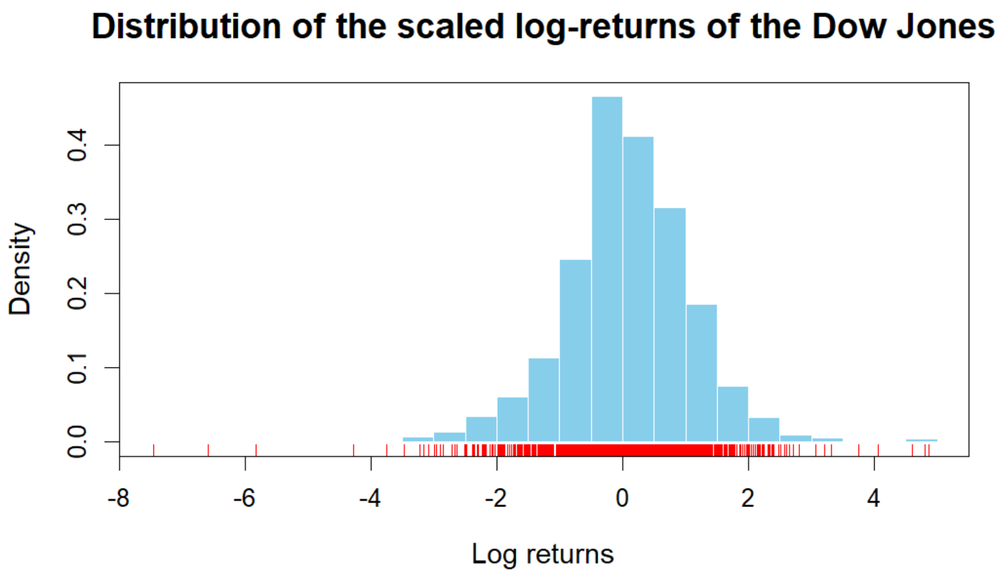



Figure 8: Empirical distribution of the scaled log-returns of the Dow Jones index

We fitted two GARCH$(1, 1)$ models to this data, one assuming standard normal innovations, and the other one assuming $t_\nu$-distributed innovations, where $\nu$ is a parameter. We get the following results

Table 1: Estimated Parameters with $N(0,1)$ Innovations (Model A)

| **Parameter** | **Estimate** | **Std. Error** | **$t$-value** | **$p$-value** |
|---|---|---|---|---|
| $\mu$ | 0.093 | 0.025 | 3.680 | 0.000 |
| $\alpha_0$ | 0.023 | 0.009 | 2.460 | 0.014 |
| $\alpha_1$ | 0.089 | 0.016 | 5.460 | $< 10^{-6}$ |
| $\beta_1$ | 0.895 | 0.019 | 47.310 | $< 10^{-6}$ |

Table 2: Estimated Parameters with $t_\nu$-Distributed Innovations (Model B)

| **Parameter** | **Estimate** | **Std. Error** | **$t$-value** | **$p$-value** |
|---|---|---|---|---|
| $\mu$ | 0.105 | 0.024 | 4.364 | $1.3 \times 10^{-5}$ |
| $\alpha_0$ | 0.021 | 0.009 | 2.274 | 0.023 |
| $\alpha_1$ | 0.066 | 0.015 | 4.267 | $2.0 \times 10^{-5}$ |
| $\beta_1$ | 0.917 | 0.019 | 49.418 | $< 10^{-6}$ |
| $\nu$ | 6.318 | 1.085 | 5.823 | $< 10^{-6}$ |

## 3.2 Model diagnostics

For both models, we obtain satisfactory results, with rather significant estimated parameters, especially for the ones of interest in our analysis ; $\alpha_1, \beta_1,$ and $\nu$ (applicable to model B only). The Akaike and Bayes Information Criteria (AIC and BIC) reported below, tend to lead us to believe that model B provides a more accurate fit of the data

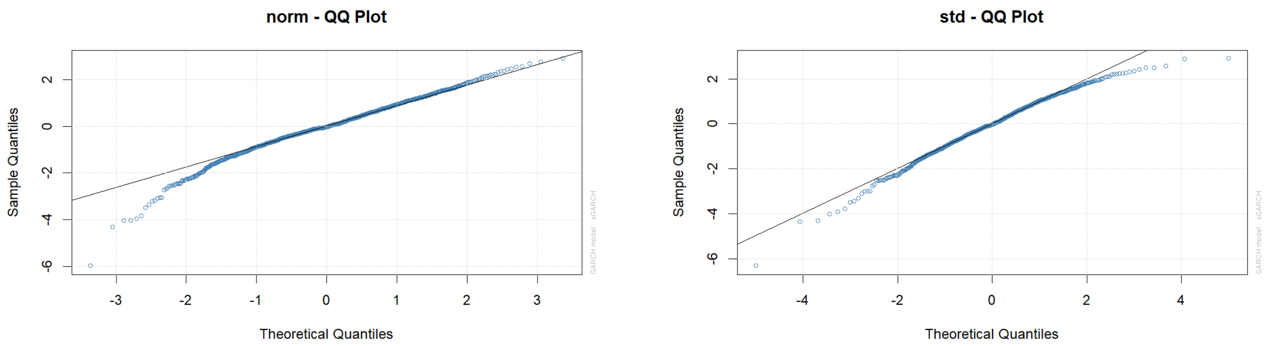
Table 3: Model Comparison using AIC and BIC

| Model | AIC | BIC |
|---|---|---|
| A | 2.83 | 2.85 |
| B | 2.79 | 2.81 |

However, a quick look at the QQ-Plot of the residuals of both models show that they do not perfectly capture the tail behavior of the data.

**QQ-Plot of the normal residuals :** The residuals of the model match well with the standard normal in the center of the distribution. Surprisingly, the upper-tail of the residuals is also well approximated using gaussian innovations here, but we can see that the lower-tail is really not a good match. The empirical residuals lying under the line tell us that the residuals have heavier tails than our gaussian distribution.

**QQ-Plot of the $t$-distributed residuals :** Unlike model A, the $t$-distributed innovations do not capture well the behavior of the data in the upper-tail (right side of the plot). The points lying under the line indicate us that the $t$-distributed innovations are too heavy-tailed for the data. This aligns with the QQ-Plot of the first model, which tshowed us that the data was not heavy-tailed since it matched very well the normal residuals. However, for the lower-tail, we see that the points lying beneath the line indicate that the thickness of the tails of $t_{6.318}$ is not enough to match the heavy-tailed nature of the data.




(a) Model A

(b) Model B

Figure 9: QQ-Plots of the residuals for each model

We now proceed with the tail index computations of both models, given the fitted parameters.

## 3.3 Tail-index computation of each model

We run Newton's algorithm described in section 2.4 for both models in order to compute their tail index.

### 3.3.1 Model A

The fitted parameters obtained above in section 3.1 are $(\hat{\alpha_0}, \hat{\alpha_1}, \hat{\beta}_1) = (0.02256, 0.0890, 0.895)$. As discussed above, we start the iteration at $\kappa_0 = 40$. For this set of parameters, and assuming $Z_t \sim N(0, 1)$, our iteration converges to $\kappa = 5.6731$. A plot of $\mathbb{E}[(\hat{\alpha_1} Z_t^2 + \hat{\beta}_1)^{\kappa/2}]$ as a function of $\kappa$ is displayed below:

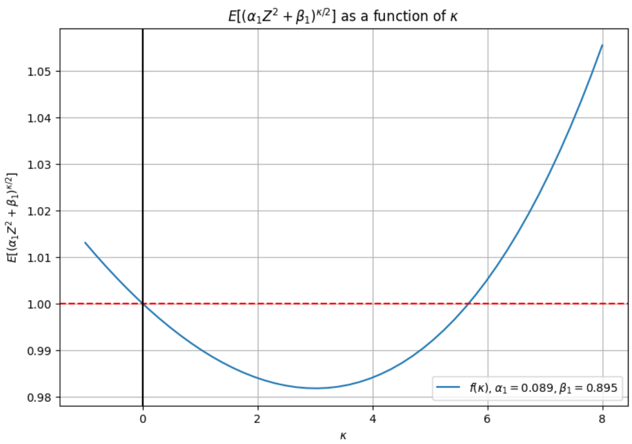



Figure 10: $\mathbb{E}[(\hat{\alpha}_1 Z_t^2 + \hat{\beta}_1)^{\kappa/2}]$ (Dow Jones fit parameters, gaussian innovations)

### 3.3.2 Model B

The fitted parameters obtained in section 3.1 assuming $t_\nu$-distributed innovations are

$$(\hat{\alpha_0}, \hat{\alpha_1}, \hat{\beta_1}, \hat{\nu}) = (0.021463, 0.065670, 0.916909, 6.318297)$$

For the set of parameters above, our Newton method converges to the solution $\kappa = 4.85$, as can be seen on figure 11. Notice that this tail-index value is lower than the one obtained with model A above. This would indicate us that under the assumptions that the innovations are $t$-distributed, the returns of the Dow-Jones would be more heavy-tailed. In fact, this was to be expected, as we mentioned in sections 1.3 and 2.3, as the heavy-tailed nature of the innovations drives the heavy-tailed nature of the marginal.

However, our analysis of residuals above could lead us to think that the normal innovations were in fact more representative to the true data. For this reason, we will compare the results with another distribution of the innovations, and by using Block Maxima method.

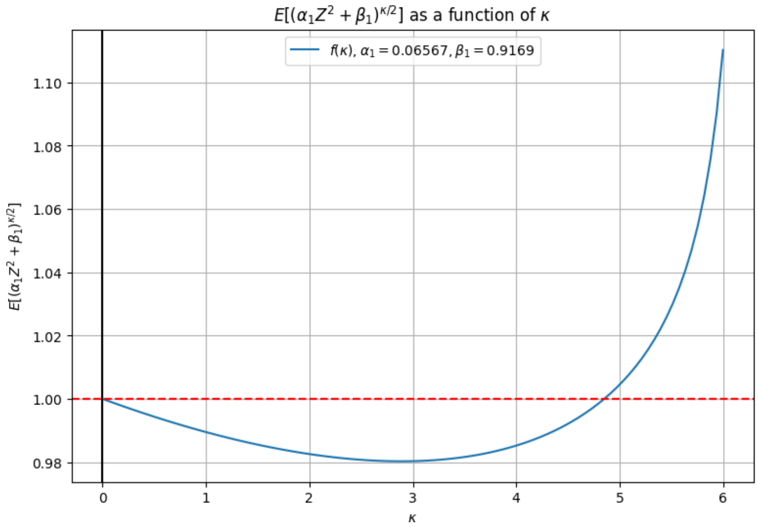



Figure 11: $\mathbb{E}[(\hat{\alpha_1} Z_t^2 + \hat{\beta_1})^{\kappa/2}]$ (Dow Jones fit parameters, $t$ innovations)

## 3.4 Investigating Generalized Hyperbolic innovations

Our data analysis revealed that one distribution for the innovations stood out when fitting a GARCH$(1,1)$ model to the Dow Jones data: the Generalized Hyperbolic (GH) distribution. Its density function in the $(\alpha, \beta, \delta, \mu)$ parameterization, is given by:

$$f(x;\alpha,\beta,\delta,\mu) = \frac{(\alpha^2-\beta^2)^{\lambda/2}}{\sqrt{2\pi}\delta^\lambda K_\lambda(\delta\sqrt{\alpha^2-\beta^2})} \cdot e^{\beta(x-\mu)} \cdot \frac{K_{\lambda-1/2}(\alpha\sqrt{\delta^2+(x-\mu)^2})}{\left(\sqrt{\delta^2+(x-\mu)^2}\right)^{1/2-\lambda}},$$

where:

- $\alpha > 0$ and $\beta$ ($|\beta| < \alpha$) control the shape and skewness.
- $\delta > 0$ is a scale parameter.
- $\mu \in \mathbb{R}$ is the location parameter.
- $\lambda$ is a shape parameter.
- $K_\lambda(\cdot)$ is the modified Bessel function of the second kind, and $x \in \mathbb{R}$

The rugarch package implements the standardized Generalized Hyperbolic distribution using the $(\zeta, \rho)$ parameterization for parameter estimation, which needs converted to the more standard $(\alpha, \beta, \delta, \mu)$ parameterization, as SciPy uses the latter. The process of translating the $(\zeta, \rho)$ parameterization is described in the rugarch package documentation ([3]), ensuring a mean of 0 and a variance of 1. Below is the output of the model:

| **Parameter** | **Estimate** | **Std. Error** | **t value** | **p-value** |
|---|---|---|---|---|
| $\mu$ | 0.0769 | 0.0255 | 3.015 | 0.00257 |
| $\alpha_0$ | 0.0197 | 0.00857 | 2.296 | 0.0217 |
| $\alpha_1$ | 0.0664 | 0.0145 | 4.596 | $< 10^{-5}$ |
| $\beta_1$ | 0.917 | 0.0172 | 53.220 | $< 10^{-5}$ |
| **skew**$(\rho)$ | -0.960 | 0.0201 | -47.678 | $< 10^{-5}$ |
| **shape**$(\zeta)$ | 0.250 | 0.138 | 1.810 | 0.0703 |
| **ghlambda**$(\lambda)$ | -3.97 | 0.884 | -4.489 | $< 10^{-6}$ |

Table 4: Optimal parameter estimates for the GARCH$(1, 1)$ model with a generalized hyperbolic distribution.

This model yields an AIC equal to 2.7803 and a BIC equal to 2.8081, both slightly lower than models A and B (see table 3). The QQ-plot of the residuals let us think that the GH-distributed innovations capture more adequately the tail behavior of the residuals :

ghyp_qqplot.png

Figure 12: QQ-Plot of the residuals for the GH innovations model

Note that the GH hyperbolic distribution is not symmetric (it is skewed by the parameter $\beta \neq 0$). In fact, we claim that the reason it does so well on the data is that its large number of parameters allows to capture the asymmetry of the returns in both tails. As we mentioned earlier, figure 9 was implicitly showing the asymmetry of the returns in the tails, since both the normal and the $t$ distributions are symmetric. Note also that as mentioned in the QRM textbook ([2],example 7.3.7), the GH distribution does not have power tails.
Finally, after reparameterizing, we adapted our Newton's iteration for this distribution, and solved for $\kappa$ as we previously did in the report. The method converges to $\kappa = 4.459$. Thanks to its low AIC and BIC, we will consider this model to be the most accurate one in the section below when comparing our results with the Block-Maxima method. However, this does not mean that we consider our model to be perfect. Indeed, we can see that the shape parameter $\hat{\zeta}$ of our model is not statistically significant at the 95% confidence level, but the model does well enough to act as our "baseline" model.

# 4 Block Maxima method

## 4.1 Fitting the models

We wish to compare all of our results with the block maxima method. This method relies on the Fischer-Tippett theorem, which states that if some data $X_1, ..., X_n$ are truncated in $M$ blocks of size of size $N$: $\underbrace{X_{11}, ..., X_{1M}}_{M \text{ points}}, X_{21}, ..., X_{2M}, ..., X_{N1}, ..., X_{NM}$, and we take for each block its maximum value $M_i = \max\{X_{i1}, ..., X_{iM}\}$ i.e. the block maximum, then the block maxima can be approximated by the Generalized Extreme Value distribution (GEV). If the data $X_1, ..., X_n$ come from a distribution with c.d.f. $F$ such that $F \in MDA(H_\xi)$, we have that the block maxima follow the GEV distribution with parameters $\sigma, \xi, \mu$:

$$H_{\mu,\sigma,\xi}(x) = \exp\left[\left[-1 + \xi(\frac{x-\mu}{\sigma})\right]^{-1/\xi}\right]$$

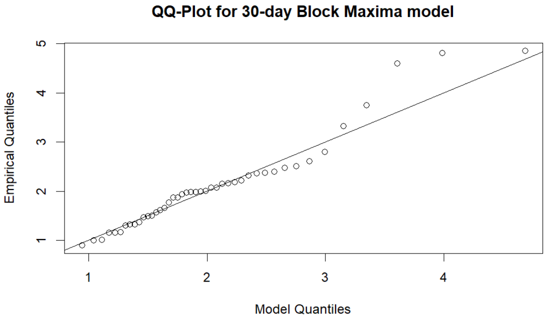
for $-1 + \xi\left(\frac{x-\mu}{\sigma}\right) > 0$.

To conduct the analysis, we consider 4 different values of $M(M \in \{30, 50, 75, 100\})$, and investigate the outputs of the models. The fitting is done with the functions `blockMaxima(.)` and `fevd` from the `extReme` package.

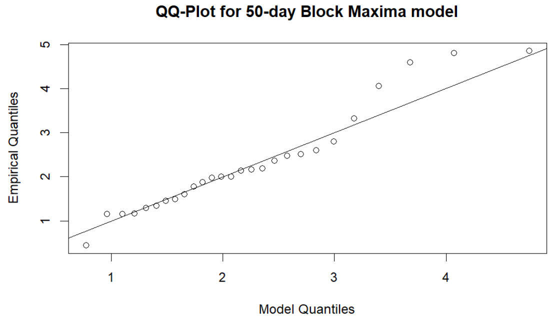
| **Model** | $\hat{\mu}$ **(Location)** | $\hat{\sigma}$ **(Scale)** | $\hat{\xi}$ **(Shape)** | **AIC** | **BIC** |
|---|---|---|---|---|---|
| $M = 30$ | 1.661 | 0.591 | 0.150 | 106.137 | 111.490 |
| $M = 50$ | 1.762 | 0.841 | 0.040 | 83.272 | 87.160 |
| $M = 75$ | 1.990 | 0.715 | 0.192 | 54.792 | 57.464 |
| $M = 100$ | 2.164 | 1.201 | -0.166 | 53.012 | 54.929 |

Table 5: Estimates of $\mu$ (location), $\sigma$ (scale), and $\xi$ (shape), along with AIC and BIC values for the block maxima models. Standard errors can be found in the code.

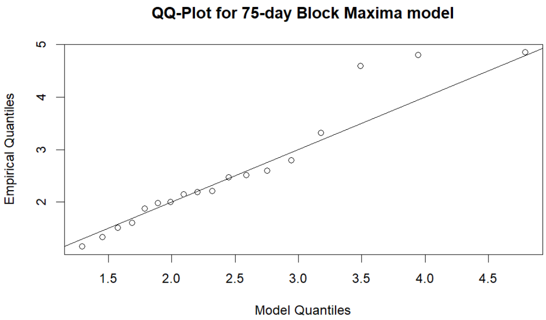
We get the following QQ-plots for the 4 models:



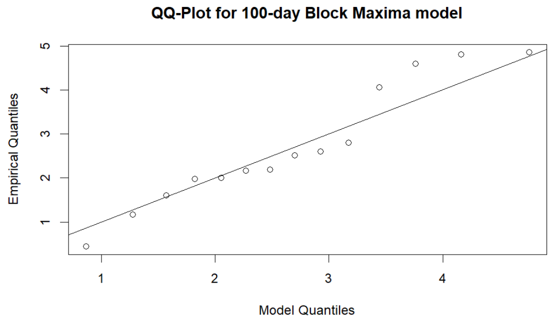

(a) QQ-Plot of the GEV fit for 30-day blocks




(b) QQ-Plot of the GEV fit for 50-day blocks




(c) QQ-Plot of the GEV fit for 75-day blocks




(d) QQ-Plot of the GEV fit for 100-day blocks

Figure 13: QQ-Plots of the GEV fit for various size of blocks $M$. Unfortunately, all the fits exhibit unsatisfactory QQ-Plots, especially in the upper tail.

Table 6: Estimates of $\xi$ and $1/\xi$ (tail index) for the block maxima models.

| **M** | $\xi$ **(Shape)** | $1/\xi$ **(Tail Index** $\kappa = \alpha$**)** |
|---|---|---|
| **30** | 0.150 | 6.667 |
| **50** | 0.040 | 25.000 |
| **75** | 0.192 | 5.208 |
| **100** | -0.166 | N/A |

As mentioned in the lecture notes, the fitted values $\hat{\xi}$ allow us to estimate the tail index of $F$. Specifically, the tail index $\alpha$, or $\kappa$ as we used in this report, is given by $\alpha = \frac{1}{\xi}$, and in particular, $F_{X_t} \in MDA(\Phi_\kappa)$. However, some important points arise looking at table 6 and comments are required about the values of $\xi$ obtained:

1. **Variability :** A high variance in the shape parameters can be observed. We claim that this is due to the fact that an important assumption for using the Block-Maxima technique is that the i.i.d. setting is required. Also, when using large blocks, the model only keeps the maximum value in each block, not taking into account the potential other extremes lying close to it. Hence, the model loses "valuable information". Unfortunately, it is very likely the case since financial data exhibits strong time-dependence and clustering in high exceedances, as we mentioned in the introduction.

Such a variance in the values of $\hat{\xi}$ indeed confirms that all these models do not perform well and that the Block-Maxima method is not appropriate for this type of data. This also highlighted by the QQ-Plots.

2. **Presence of a negative value :** The model using $M = 100$ blocks yielded a negative value for $\xi$, indicating that the marginal distribution would in fact lie in the Weibull Domain of Attraction. This does not corroborate our analysis in section 2.3, where we derived that the marginal of the GARCH model should in fact lie in the Fréchet Domain of Attraction with parameter $\alpha = \kappa = 1/\xi$. This model can most surely be discarded, especially regarding its very low block size which increases the variance.

3. **Shape value near zero :** Our model with $M = 50$ yields a value of $\hat{\xi} = 0.040$, with standard error 0.22, which does not allow us to reject the hypothesis that $\xi = 0$. However, this would once again contradict the theory in section 2.3, since a value of 0 would mean that the marginal of the GARCH$(1, 1)$ model lies in the Gumbel Domain of Attraction. Also, this model would involve a tail index which would be much higher than the ones we observed in sections 3.3.1 and 3.3.2

## 4.2 Comparison with tail-index computations

The Block Maxima results described above were deceiving in the sense that the poor performance of the models does not allow us to make precise estimations of the behavior in the tails of the GARCH models. However, the results of tail-index computations in sections 3.3.1, 3.3.2 and 3.4 were more consistent between each other, and we believe that the GARCH(1,1) model with Generalized Hyperbolic distributed innovations on the Dow Jones data seems to be an accurate way to represent the daily scaled log-returns. This model led us to the conclusion that the marginal distribution $X_t$ of the Dow Jones data behaves as

$$\mathbb{P}(|X_t| > x) = c_1 x^{-4.4589} \quad \text{as } x \to \infty$$

for some positive constant $c_1$.

# 5 Conclusion

In this report, we investigated the tail behavior and some extreme value properties of GARCH(1,1) models, focusing on their applications in financial data analysis. We emphasized the ability of GARCH$(p, q)$ models to capture time-dependent volatility and clustering, and showed through simulations the heavy-tailed nature of the marginal distribution depends on the innovation distribution and the model parameters.

Using theoretical results, including Breiman's Lemma and the regular variation properties of the stochastic volatility, we computed the tail index $\kappa$ for different innovation distributions (Gaussian, Student's $t$, and Generalized Hyperbolic) using numerical methods. Applying these results to the Dow Jones daily log-return data, we found that the GARCH(1,1) model with GH-distributed innovations provided the best fit. This model indicated Pareto-like tails with $\kappa = 4.459$, which also happends to be consistent with the Fréchet domain of attraction.

We also compared our findings with the Block Maxima method, and saw that this method is not appropriate for financial data like log-returns due to the strong time-dependence.

These results highlight the importance of choosing appropriate innovation distributions and parameter estimation techniques when modeling financial extremes. Future investigations could explore extensions to higher-order GARCH models, or alternatives to GARCH$(p, q)$ models, such as EGARCH or COGARCH.

# References


[1] Mikosch, T., Stărică, C. (2000). **Limit theory for the sample autocorrelations and extremes of a GARCH(1,1) process.** The Annals of Statistics, 28(5), 1427-1451.

[2] Embrechts, P., McNeil, A. J., Frey, R. (2005). **Quantitative Risk Management: Concepts, Techniques and Tools.** Princeton University Press.

[3] Ghalanos, A. (2023). **Introduction to the rugarch package**. R package vignette, version 1.4-9. Retrieved from `https://cran.r-project.org/web/packages/rugarch/vignettes/Introduction_to_the_rugarch_package.pdf`