MATH70110 Quantitative Risk Management Assessed Coursework II

Harrison Lam CID: 01489877
Guillem Ribas Pescador CID: 06015926
Dominic Rushton CID: 06019954
Mustafa Hamdani CID: 06021930

December 4, 2024

Contents

1 Risk Forecasting with Extreme Value Theory

1.1 Notes on Methodology and Results

In this report, we perform a statistical analysis on the returns of the Tesla (TSLA) stock between 26 November 2012 and 25 November 2022. To understand the risk profile of a long Tesla position over this period, we fit a GARCH(1,1) model to the negative log returns and analyse the resulting standardised residuals against standard normal, Student-t and generalized Pareto distributions.

Subsequently, we forecast the VaR (value at risk) and ES (expected shortfall) risk measures under three different specifications for the strict white noise term in the GARCH(1,1) process.

Full details of Python packages used in our implementation can be found in the references. The full implementation can be viewed in the accompanying Jupyter notebook.

1.2 Preliminary Analysis

We seek to forecast the risk of being long Tesla (subsequently referred to as "the stock") by considering the linearised daily losses

$$\bar{L}_{t+1}^{\Delta} = -r_{t+1} = \log S_t - \log S_{t+1} [4]$$

where S_t represents the close price of the stock on day t. To ensure we have sufficient context on the behaviour of the underlying data to effectively analyse the models we fit to the loss data, we first observe some key features of the return data of the stock.

Figure 1: Tesla price over the period.

From Figure 1 it is clear to see that the distribution of returns should exhibit positive mean given the general upward trend. We also note the substantial volatility in the stock price following the sharp rise in value of 2020, with large losses at several points from 2021 onward.

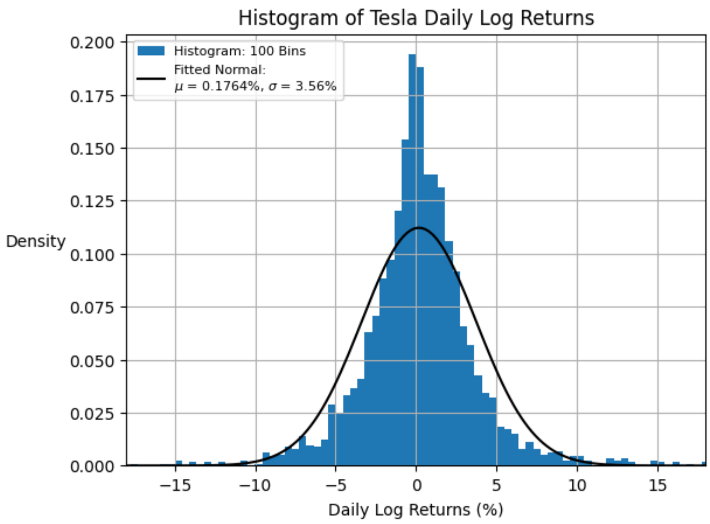

Figure 2: Histogram of log returns with normal distribution overlaid.

Figure 2 demonstrates the heavy tails of the empirical returns as compared to the normal distribution, as well as asymmetry that is consistent with the stylized facts of returns data [4].

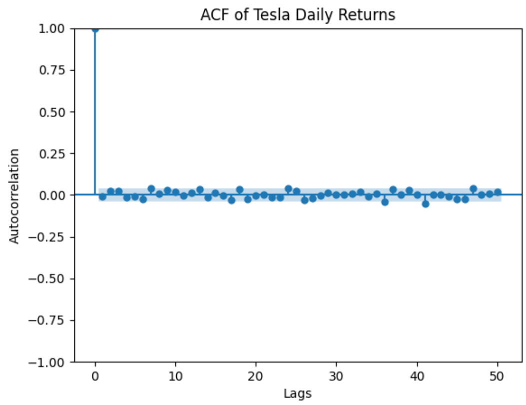

Figure 3: Tesla Returns ACF.

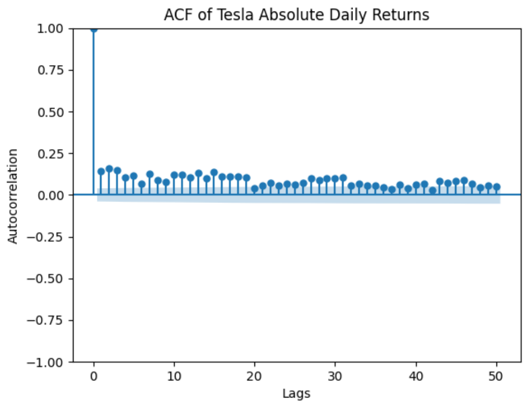

Figure 4: Tesla Absolute Returns ACF.

From the ACF in Figure 3 we observe that the daily returns of the stock show no significant correlation. However, from Figure 4 it can be seen that the absolute returns exhibit meaningful positive autocorrelation with as much as 20 days of lag. Both of these facts also align well with our intuition on how the returns should behave based on the stylized facts of returns [4].

1.3 Standardised Residuals under a GARCH(1,1) Model

After an initial exploration to gain sufficient context on the structure of the underlying returns data, we proceed with fitting the GARCH(1,1) [2] model with standard normal innovations on the negative log returns with the `arch_model` function in the ARCH package [1]. Recall that the GARCH(1,1) process fitted on the linearised losses of the stock is defined as

$$\bar{L}_t^{\Delta} := \sigma_t Z_t, \quad \sigma_t^2 = \alpha_0 + \alpha_1 \bar{L}_{t-1}^{\Delta 2} + \beta_1 \sigma_{t-1}^2, \quad t \in \mathbb{Z}[4],$$

where (Z_t) is a strict white noise process and $Z_t \sim N(0,1), \forall t \in \mathbb{Z}$ in this case. Note that the model is fitted only on the data up to and including 25 November 2021. This is done in order to use this data as a training set to fit the model parameters, subsequently using the model to forecast and test risk measures from 26 November 2021 onwards. As opposed to recalculating the model parameters at every day using a rolling window scheme, this approach greatly reduces the computational cost of the risk measure forecasting, as the parameters are only fitted once.

Parameter	**Estimated value**
α_0	0.1411
α_1	0.0411
β_1	0.9479

Table 1: Parameter estimated values for GARCH(1,1) process.

Table 1 displays the parameters of the GARCH(1,1) model fitted on the training data. As a sanity check, it is worth noting that α_1 and β_1 satisfy the linear condition $\alpha_1 + \beta_1 < 1$ [4]. In addition to this, the fact that β_1 is significantly larger than α_0 suggests that the model captures volatility persistence, as β_1 is the weight of the previous volatility in the calculation of the next volatility. This aligns with our insights from the previous exploration of the stock, as the absolute returns exhibit statistically significant autocorrelation, as shown in Figure 4.

In order to assess the suitability of the model on the training data, we examine several residual diagnostic plots. The standardised residuals, $(\hat{Z}_t)_{t\in\mathbb{Z}}$, of the fitted data are given by

$$\hat{Z}_t = \frac{\bar{L}_t^\Delta}{\hat{\sigma}_t}, \quad \forall t \in \mathbb{Z}[4],$$

where $\hat{\sigma}_t$ are the fitted volatilities.

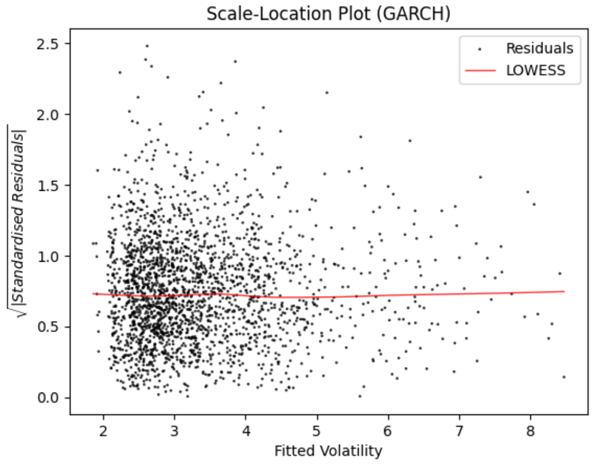

Figure 5: Scale-Location plot for standardised residuals of GARCH(1,1)

The scale-location plot in Figure 5 displays the square root of the absolute residuals, $\sqrt{|\hat{Z}_t|}$, plotted against the fitted volatilities $\hat{\sigma}_t$. If the assumption that $\hat{Z}_t$ in the model is independent of $\hat{\sigma}_t$ holds, there should not be any significant trends in the plot. In order to assess if there is any dependence between $\hat{\sigma}_t$ and $\hat{Z}_t$, we use a locally weighted scatterplot smoothing (LOWESS) line [3], which, by using a sliding window, divides the points into distinct groups and fits a regression line using weighted least squares (WLS) to each group. We make use of the `regplot` function in the SEABORN package [6] to plot it in the figure. As the LOWESS line remains flat throughout all values of the fitted volatility, it suggests that there is no apparent trend in the standardised residuals.

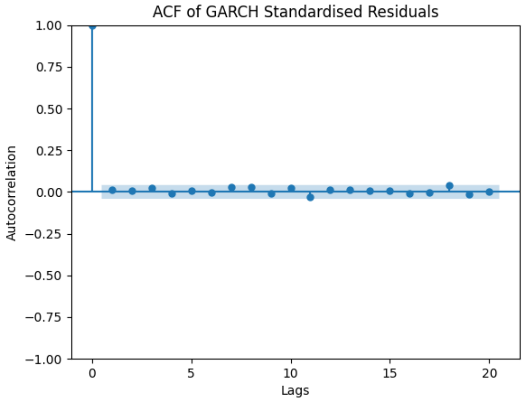

Figure 6: ACF of GARCH(1,1) standardised residuals.

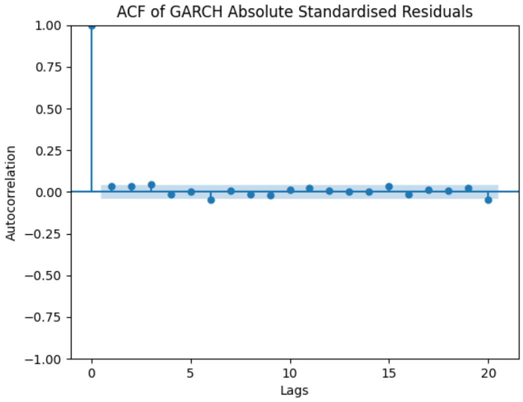

Figure 7: ACF of GARCH(1,1) absolute standardised residuals.

This is further confirmed by Figure 6 and Figure 7, showing that the standardised residuals are serially uncorrelated. This serial uncorrelation, thus, suggests that the standardised residuals are independent and identically distributed, however, the particular distribution that they follow is yet unknown.

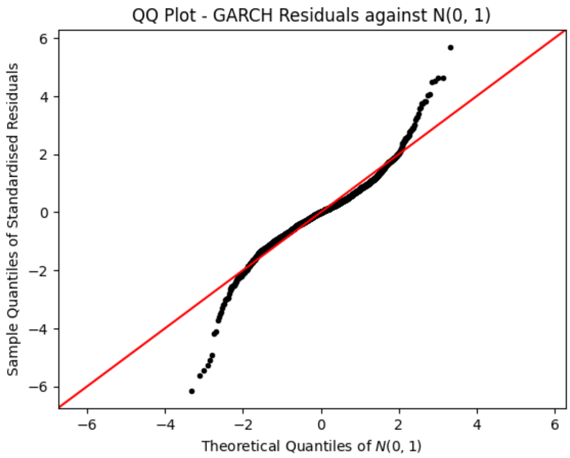

Figure 8: QQ plot of GARCH(1,1) standardised residuals vs N(0,1)

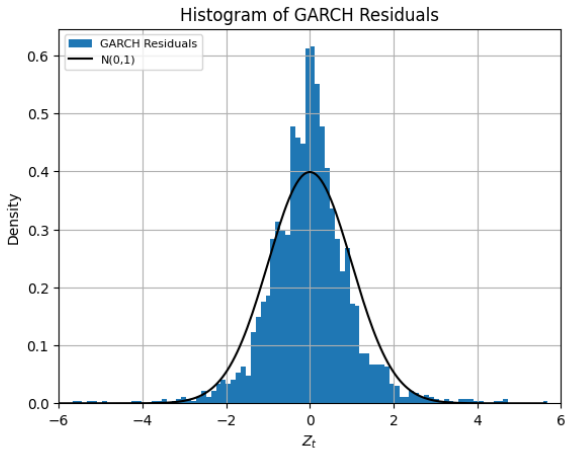

Figure 9: Histogram of standardised residuals vs N(0,1) density

As can be seen in Figures 8 and 9, the tails of the distribution of the standardised residuals are significantly heavier than those of the standard normal distribution. This suggests that a heavier-tailed distribution might provide a better fit to the standardised residuals, motivating the following section.

1.4 Fitting a t-distribution

We explore fitting the standardised residuals with a t-distribution instead, since this theoretically should capture the behaviour of the standardised residuals better than a standard normal distribution, as the t-distribution has heavier tails. In order to fit the t-distribution, we must estimate the value of ν, the degree of freedom of the distribution.

To find an estimator $\hat{\nu}$ for the t-distribution, we use the maximum likelihood estimation (MLE) method. The likelihood function is given by

$$L(\nu; \hat{\mathbf{Z}}) := \Pi_{t=1}^{T} \sqrt{\frac{\nu}{\nu-2}} f_T(\sqrt{\frac{\nu}{\nu-2}} \hat{Z}_t; \nu)[4],$$

where $\sqrt{\frac{\nu}{\nu-2}} f_T(\sqrt{\frac{\nu}{\nu-2}} t; \nu)$ is the probability density function of a normalised t-distribution with ν degrees of freedom and variance 1, and $\hat{\mathbf{Z}} = \{\hat{Z}_1, \ldots, \hat{Z}_T\}$, where T is the total number of standardised residuals. We then maximise the log-likehood function

$$l(\nu; \hat{\mathbf{Z}}) = \ln(L(\nu; \hat{\mathbf{Z}}) = \sum_{t=1}^{T} \ln \left(\sqrt{\frac{\nu}{\nu-2}} f_T(\sqrt{\frac{\nu}{\nu-2}} \hat{Z}_t; \nu) \right) [4].$$

In order to maximize the function, we make use of the `minimize` function present in the SciPy package [5]. Given an initial value, this function minimizes an objective function subject to some linear constraints. We set the initial value to $\nu_0 = 5$ and use this function to minimize the negative log-likehood $-l(\nu; \hat{\mathbf{Z}})$, subject to $\nu > 2$, as the t-distribution has infinite variance otherwise. As $\nu \to \infty$, the t distribution converges to a normal distribution, therefore we expect the MLE estimator to be rather small, motivating our choice of ν_0. We then obtain the MLE estimator $\hat{\nu} = 3.8$.

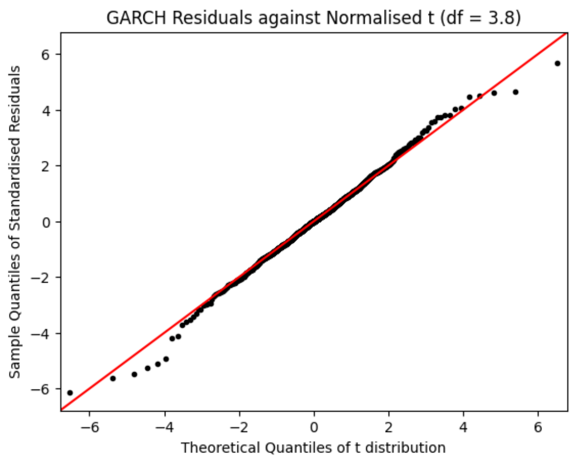

Figure 10: QQ plot of GARCH(1,1) standardised residuals vs t-distribution)

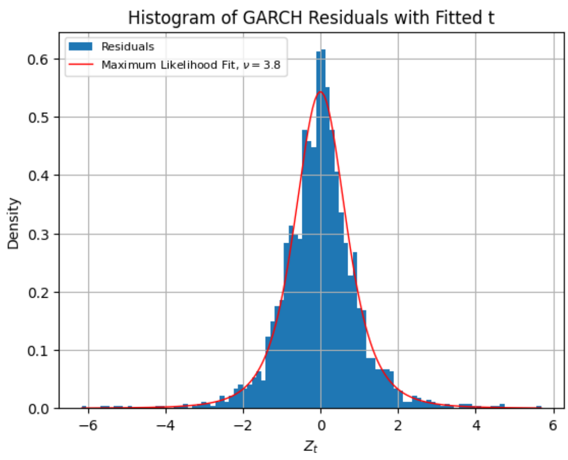

Figure 11: Histogram of standardised residuals with fitted t-distributions overlaid

We see in Figure 11 that the fitted t-distribution captures the large central peak and heavy tails of the standardised residuals. Figure 10 further demonstrates that the t-distribution captures the empirical distribution well, with only small deviations even when comparing the quantiles at the extremes.

Capturing tail behavior when modeling the risk of the stock is crucial in producing robust VaR and ES estimates, and hence we believe that using the t-distribution is a substantial improvement to standard normal innovations.

1.5 Fitting a generalised Pareto distribution

The generalised Pareto distribution (GPD) with parameters $\xi \in \mathbb{R}$ and $\beta > 0$ has cumulative distribution function

$$G_{\xi,\beta}(x) = \begin{cases} 1 - (1 + \xi x/\beta)^{-1/\xi} & \xi \neq 0 \\ 1 - \exp(-x/\beta) & \xi = 0 \end{cases} [7],$$

where ξ dictates the shape and β dictates the scale of the distribution (see appendix for additional constraints on the parameters and support of the distribution). The GPD is effective in modelling tail behaviour of data that follow a variety of distributions, and can provide more robust estimates of risk measures from empirical data.

However, to effectively fit a GPD to a distribution we first need to understand the behaviour of the tails of the distribution in question. To do this, we introduce the excess distribution for a random variable X with cdf F over a threshold $u < x_F$, where x_F is the upper limit of the support of the distribution. The excess distribution is defined as

$$F_u(x) := \mathbb{P}[X - u \leq x | X > u] = \frac{F(x+u) - F(u)}{1 - F(u)} [7],$$

for $0 \leq x \leq x_F - u$, which describes, for any such u, how the distribution behaves in the tail beyond $X = u$. Also of interest in fitting a GPD to the distribution, is the mean excess function of X

$$e(u) := \mathbb{E}[X - u | X > u][7],$$

which describes the mean of the distance between X and u in the tail beyond $X = u$. We also have the following theorem:

Theorem 1. *(Pickands-Balkema-de Haan)*
$F \in \text{MDA}(H_\xi) \iff$ there exists a positive, measurable function $u \to \beta(u)$ such that

$$\lim_{u \to x_F} \sup_{x \in [0, x_F - u]} |F_u(x) - G_{\xi,\beta(u)}(x)| = 0[7],$$

where $\text{MDA}(H_\xi)$ represents the maximum domain of attraction of the generalized extreme value (GEV) distribution with parameter ξ. (See appendix)

In more practical terms, for sufficiently large u, there are parameters ξ and β for which the tail behavior of F (i.e. F_u) can be well approximated by a generalised Pareto distribution. We proceed by assuming that $F_u = G_{\xi,\beta}$ $\forall x \in [0, x_F - u]$, under which it can be shown that the mean excess function follows

$$e(v) = \frac{\xi v}{1 - \xi} + \frac{\beta - \xi u}{1 - \xi} [7],$$

an affine function in v.

Thus, if we are able to find a threshold u beyond which the mean excess of our empirical data exhibits a linear trend, we can fit the parameters ξ and β of the GPD to the empirical observation of F_u to approximate the tail behaviour of the standardised residuals. We define the empirical mean excess function by

$$e_n(v) := \frac{\sum_{i=1}^{n}(X_i - v)\mathbf{1}_{\{X_i>v\}}}{\sum_{i=1}^{n}\mathbf{1}_{\{X_i>v\}}}[7],$$

and observe the results of this computation on the standardised residuals in Figure 12.

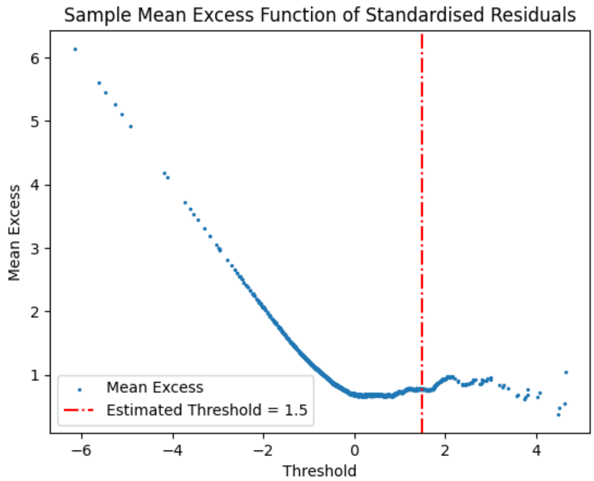

Figure 12: Mean excess of standardized residuals

From Figure 12, we see that the the mean excess of the standardised residuals is roughly linear beyond $u = 1.5$, and so we apply this as our threshold in F_u. Note that we cannot choose $u > 1.56$ since then $F(u)$ would exceed 0.95, and thus we would be unable to obtain a VaR or ES forecast at 95%. We then apply maximum likelihood estimation to fit $F_u = G_{\xi,\beta}$ and report the parameters for the generalised Pareto distribution in Table 2.

Parameter	**Estimated value**
ξ	0.0893
β	0.7007

Table 2: Estimated parameters for fitted GPD

In Figure 13 we plot the empirical excess distribution of the standardised residuals above $u = 1.5$ alongside the fitted generalised Pareto distribution. We observe that any deviations between the empirical and fitted distributions are minor, and the fitted GPD approximates the empirical data well even for large values of x.

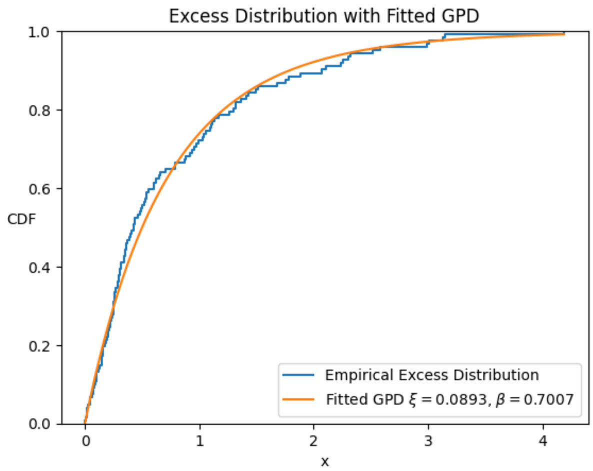

Figure 13: Excess distribution with fitted GPD

2 Forecasting VaR and ES

We now explore how the choice of distribution of the strict white noise term Z_t in the GARCH(1,1) model influences VaR and ES forecasts for the stock. In line with our earlier discussions, we consider Z_t with a standard normal distribution, a Student t-distribution, and a generalised Pareto distribution fitted to the tails of the normal standardised residuals. We also note that the specific parameters used for the t and generalised Pareto distributions were fitted based on the training data, and so through the forecasts, we obtain a practical understanding of how the measures are performing out of sample.

We compute the VaR and ES forecasts at 95% and 99% by first finding the appropriate quantiles. For the normal and Student t-distributions, we compute the quantiles directly from the percentile functions in the SciPy package [5]. We then compute

$$\mathrm{ES}_\alpha(Z_t) = \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{VaR}_u(Z_t)du[4]$$

to arrive at the Z values corresponding to the ES at $\alpha = 95\%$ and $\alpha = 99\%$. In the case of the GPD, we directly use the fact that

$$\widehat{q}_\alpha(Z_t) = u + \frac{\hat{\beta}}{\hat{\xi}}\left(\left(\frac{1-\alpha}{1-\widehat{F}(u)}\right)^{-\hat{\xi}} - 1\right)[7],$$

$$\widehat{\mathrm{ES}}_\alpha(Z_t) = \frac{\widehat{q}_\alpha(Z_t)}{1-\hat{\xi}} + \frac{\hat{\beta}-\hat{\xi}u}{1-\hat{\xi}}[7],$$

for $\widehat{F}$ the empirical cdf, $\hat{\xi}$ and $\hat{\beta}$ the fitted parameters of the GPD. Then, under each scheme we compute

$$\widehat{\mathrm{VaR}}_\alpha(L_t) = \hat{\mu}_t + \hat{\sigma}_t\mathrm{VaR}_\alpha(Z_t)[4],$$

$$\widehat{\mathrm{ES}}_\alpha(L_t) = \hat{\mu}_t + \hat{\sigma}_t\mathrm{ES}_\alpha(Z_t)[4],$$

to produce forecasts for each day t in the test set.

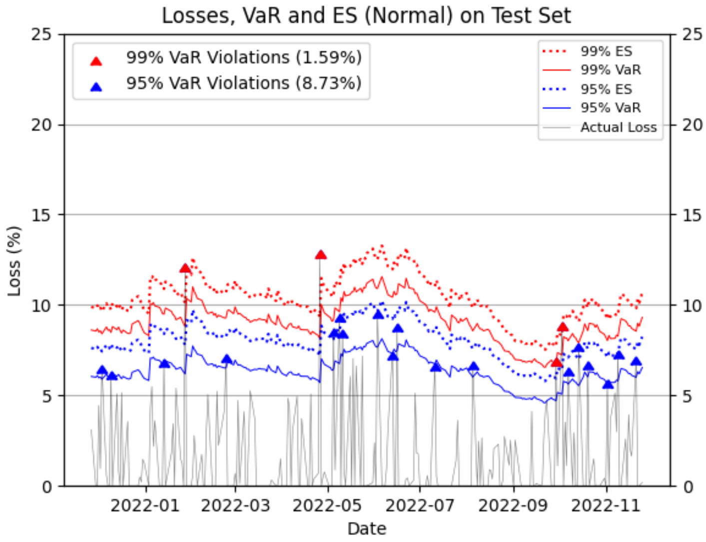

Figure 14: VaR and ES forecast with $N(0,1)$ innovations

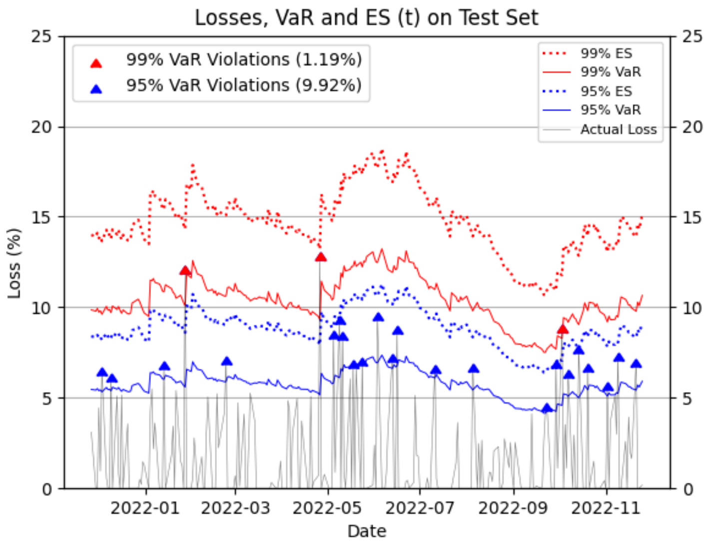

Figure 15: VaR and ES forecast with normalised $t(df = 3.8)$ innovations

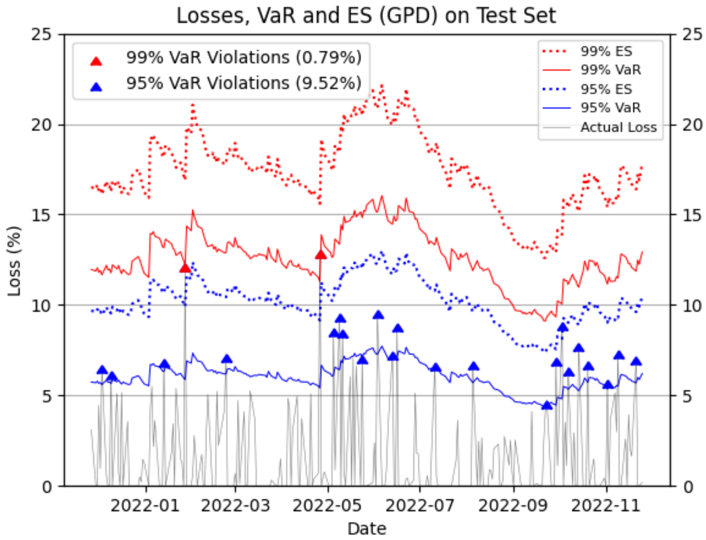

Figure 16: VaR and ES forecast with GPD($\xi = 0.0893, \beta = 0.7007$)-tailed innovations

We immediately notice from Figure 14 that the VaR and ES forecasts are much more closely grouped together when N(0,1) innovations are used as opposed to Student-t or GPD-tailed innovations. This is explained by the inability of the normal distribution to capture the heavy tails of the residuals in the testing data and in general we would expect a high number of violations at the 99% level when using normal innovations. With the returns of the stock exhibiting such high volatility over specific periods, this could result in an underestimation of exposure to material losses.

Further, we see that forecasts in Figures 15 and 16 exhibit much more similarity in their ability to capture tail behaviour of the standardised residuals as demonstrated by the gap between forecasts at the 95% and 99% level. In the case of the GPD, we see noticeably more conservative estimates for VaR and ES at the 99% confidence level, implying that, in general, we would expect to see fewer instances of threshold exceedance at 99% when using the fitted GPD innovations and a more robust estimation of the overall risk level of the stock.

We now report backtest results for VaR using unconditional coverage tests at 95% and 99%.

	$\alpha = .95$			$\alpha = .99$		
Innovations	Violations(Exp)	Test stat.	p-value	Violations(Exp)	Test stat.	p-value
Normal	22 (13)	6.0972	0.0135	4 (3)	0.7451	0.3880
Normalised t	25 (13)	10.1126	0.0015	3 (3)	0.0870	0.7680
GPD-tailed	24 (13)	8.6808	0.0032	2 (3)	0.1166	0.7327

Table 3: Key statistics from unconditional coverage backtests.

From the backtest results for VaR in Table 3, we notice that all three methods produce similar results at 95%. This is within our expectations for how the models should perform, given that this confidence threshold captures events that are relatively more common. In this case, we see that all three distributions are not consistent with the hypothesis that the forecasts are being computed with a correctly specified model. Given the very high degree of volatility of returns present over the testing period as compared to the testing data, we do not believe that this is conclusive evidence that any model outperforms the others at 95% significance, however, all of them fail to provide satisfactory risk measure forecasts.

Moreover, considering the results at 99%, we see that the Student-t and GPD distributions demonstrate more consistency with the hypothesis that the models are correctly specified as compared to the normal distribution. In particular, we see fewer VaR violations and note that from Figures 15 and 16, the losses sustained on threshold exceedances are well within expectations, whereas in Figure 14 there are violations that far exceed even the 99% ES. This aligns with our reasoning that the Student-t and GPD innovations provides much more robust forecasts, and in general will function more effectively as a risk management tool. It is worth noting that the choice of the parameter u when fitting the GPD, adds a layer of flexibility to the model, suggesting that a rolling window recalculation of the GPD using a suitable variable value of u could produce more accurate results. This is left as a possible following improvement on the methodology.

3 Mathematical Appendix

Here we provide some definitions for completeness.

Definition 2 (Simple Return)**.** Let S_t be the price of an asset on (business) day t. The (daily) simple return on this asset is defined as:

$$R_t := \frac{S_t - S_{t-1}}{S_{t-1}}$$

[4].

Definition 3 (Logarithmic Return)**.** Let S_t be the price of an asset on (business) day t. The (daily) logarithmic return is defined as:

$$r_t := \log S_t - \log S_{t-1}$$

where log is the natural logarithm [4].

Definition 4 (Strictly Stationary)**.** A process $(X_t)_{t\in\mathbb{Z}}$ is *strictly stationary* if

$$(X_{t_1}, \ldots, X_{t_n}) \stackrel{d}{=} (X_{t_1+k}, \ldots, X_{t_n+k})$$

for any $t_1, \ldots, t_n \in \mathbb{Z}$, $k \in \mathbb{Z}$, and $n \in \mathbb{N}$. Above, $\stackrel{d}{=}$ indicates equality of joint distributions [4].

Definition 5 (GARCH Process)**.** A *strictly stationary* process $(X_t)_{t\in\mathbb{Z}}$ is a GARCH(p, q) process with, $p, q \in \mathbb{N}_0$ if

1. for some $(Z_t)_{t\in\mathbb{Z}} \sim \mathrm{SWN}(0, 1)$,

$$X_t = \sigma_t Z_t, \quad \sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2, \quad t \in \mathbb{Z},$$

   where $\alpha_0 > 0, \alpha_1, \ldots, \alpha_p \geq 0$ and $\beta_1, \ldots, \beta_p \geq 0$,

2. $(\sigma_t)_{t\in\mathbb{Z}}$ is strictly stationary and positive-valued [4].

Definition 6 (White Noise)**.** A covariance-stationary process $(X_t)_{t\in\mathbb{Z}}$ is called a *white noise* if its autocorrelation function is

$$\rho(h) = \begin{cases} 1, & h = 0, \\ 0, & h > 0. \end{cases}$$

If $(X_t)_{t\in\mathbb{Z}}$ is a white noise with $\mu(t) = 0$ and $\gamma(0) = \sigma^2$, then we write $(X_t)_{t\in\mathbb{Z}} \sim \mathrm{WN}(0, \sigma^2)$. The random variables in a white noise process are called *innovations* [4].

Definition 7 (Strict White Noise)**.** If $(X_t)_{t\in\mathbb{Z}}$ is square integrable and consists of iid random variables, then it is called a *strict white noise.* When $(X_t)_{t\in\mathbb{Z}}$ is a strict white noise with $\mu(t) = 0$ and $\gamma(0) = \sigma^2$, we write $(X_t)_{t\in\mathbb{Z}} \sim \mathrm{SWN}(0, \sigma^2)$. [4]

Definition 8 (Generalised Pareto distribution)**.** The *generalised Pareto distribution (GPD)* with *shape* $\xi \in \mathbb{R}$ and **scale** $\beta > 0$ has the cdf

$$G_{\xi,\beta}(x) = \begin{cases} 1 - \left(1 + \frac{\xi x}{\beta}\right)^{-1/\xi}, & \xi \neq 0, \\ 1 - \exp\left(-\frac{x}{\beta}\right), & \xi = 0, \end{cases}$$

where we require $x \geq 0$ when $\xi \geq 0$ and $0 \leq x \leq -\frac{\beta}{\xi}$ when $\xi < 0$. We denote the pdf of $G_{\xi,\beta}$ by $g_{\xi,\beta}$ [7].

Definition 9 (Maximum Domain of Attraction)**.** We say that F belongs to the *maximum domain of attraction (MDA)* of cdf H, if we can find sequences $(c_n)_{n\in\mathbb{N}} \subset (0,\infty)$ and $(d_n)_{n\in\mathbb{N}} \subset \mathbb{R}$ such that

$$\lim_{n\to\infty} \mathbb{P}\left[\frac{M_n - d_n}{c_n} \le x\right] = \lim_{n\to\infty} F(c_n x + d_n)^n = H(x), \quad x \in \mathbb{R}.$$

[7]

Definition 10 (Generalised Extreme Value Distribution)**.** The *generalised extreme value (GEV) distribution* with *shape* $\xi \in \mathbb{R}$ has the cdf

$$H_\xi(x) := \begin{cases} \exp\left(-\left(1+\xi x\right)^{-1/\xi}\right), & \xi \neq 0, \quad 1+\xi x > 0, \\ \exp\left(-e^{-x}\right), & \xi = 0. \end{cases}$$

[7].

References

[1] Arch. *ARCH Package Documentation*. URL: `https://arch.readthedocs.io/en/latest/`.

[2] Tim Bollerslev. "Generalized autoregressive conditional heteroskedasticity". In: *Journal of Econometrics* 31.3 (Apr. 1986), pp. 307–327. DOI: `10.1016/0304-4076(86)90063-1`.

[3] William S. Cleveland and Susan J. Devlin. "Locally weighted regression: An approach to regression analysis by local fitting". In: *Journal of the American Statistical Association* 83.403 (Sept. 1988), p. 596. DOI: `10.2307/2289282`.

[4] Anthony Coache. *QRM Lecture Slides 2024*. Nov. 2024.

[5] SciPy. *Scipy Package Dcumentation*. URL: `https://docs.scipy.org/doc/scipy/`.

[6] Seaborn. *Seaborn Package Documentation*. URL: `https://seaborn.pydata.org/`.

[7] Harry Zheng. *QRM Lecture Slides 2024*. Dec. 2024.