MATH70110 Quantitative Risk Management
Assessed Coursework I

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

November 21, 2024

Contents

1 Part A: Stylized Facts and GARCH Modelling

1.1 Notes on Methodolody and Results

In the first section of this report, we perform a statistical analysis on the *EURO STOXX 50 (SX5E)* stock index to fit appropriate ARMA and GARCH models to the daily log returns of the index. We discuss the distribution of these returns and compare it against historical data before considering the auto-correlation function (ACF) of the log returns, squared log returns and absolute log returns. The distinct ACFs are then used to confirm that the returns follow the stylized facts of financial returns.

Subsequently, a constant mean GARCH(1,1) model with standard normally distributed innovations and an ARMA(1,1)-GARCH(1,1) model with t-distributed innovations are fitted on the data. The fit of these models is finally assessed and compared quantitatively through residual diagnostic plots, including QQ plots, scale-location plots and ACFs of the residuals.

Full details of Python packages used in performing calculations and model construction can be found in the references, and the implementation as described in this report can be found in the accompanying Jupyter notebook.

1.2 Initial analysis of data

The *SX5E* stock index [8] (subsequently referred to as "the index") was initially quoted on the 26th of February 1998 and consists of 50 stocks in the Eurozone. Since inception, the value of the index has varied considerably and in this report we consider fluctuations in the index based on closing price from the 30th of October 2013 to the 27th of October 2023.

In subsequent analysis we focus on the log returns r_t of the index, computed from index value S_t on day t as

$$r_t := \log S_t - \log S_{t-1}[5].$$

To evaluate the distribution of the daily log returns of the index, it is first important to understand the overall trend that the index value has followed over the timeframe being explored.

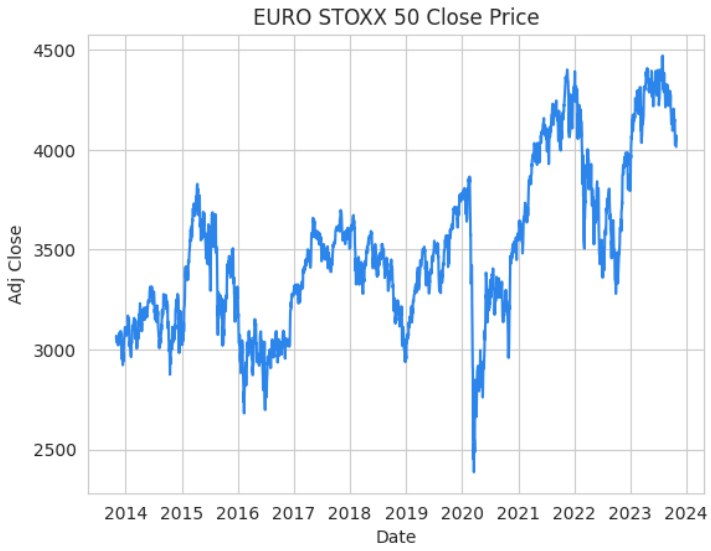

Figure 1: Value of SX5E Index over time.

Figure 1 shows that, alongside the significant fluctuation in value, the index has increased over the chosen timeframe, so it is to be expected that the log returns will exhibit positive mean, as log returns are additive. We also note the sharp declines in the index value over short periods of time, which will prove to be significant in accurately modeling distribution of returns as well as volatility of the index.

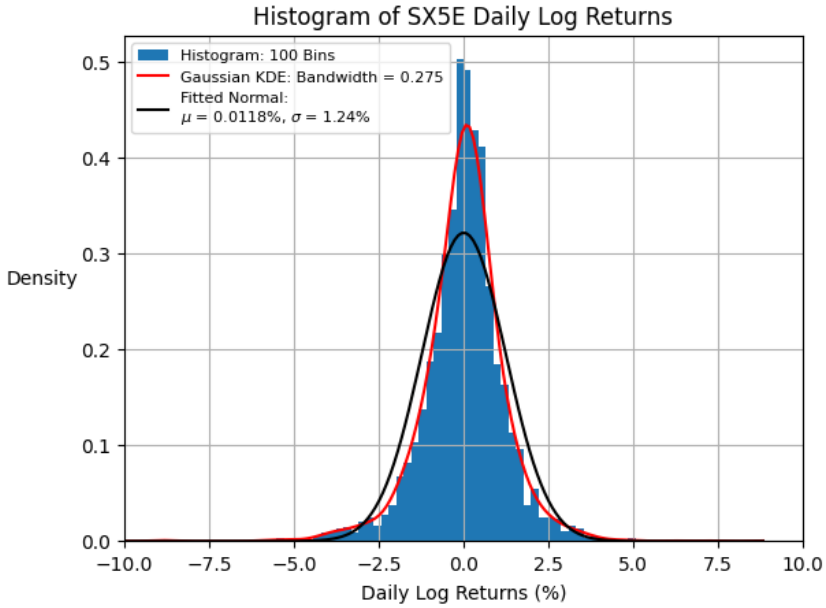

Figure 2: Log returns of the index, Gaussian distribution and Gaussian KDE overlaid.

From Figure 2, we can observe the empirical distribution of log returns. We overlay a normal distribution with mean and variance equal to the empirical equivalents of the log returns, as well as a Gaussian Kernel Density Estimator (KDE) [5] curve to assess their fit to the empirical distribution. The sample statistics of the log return distribution are given in Table 1.

Statistic	Value
Sample Mean, $\hat{\mu}$	0.0118
Sample Standard Deviation, $\hat{\sigma}$	1.24
Sample Skewness, $\hat{s}$	-0.795
Sample Excess Kurtosis, $\hat{\kappa}_3$	7.3

Table 1: Sample Statistics of SX5E Index Log Returns.

As expected, the sample mean is positive, implied by the overall increase in index value between the first and last data point. More interestingly, the sample has a skewness of -0.795 and an excess kurtosis of 7.3, exhibiting asymmetry and a heavy-tail, statistical regularities that coincide with the stylized facts of financial returns [6]. This suggests that over the sample window, there were very frequent small positive returns and a few large negative returns that skewed the overall distribution and elongated the left tail. This could be explained by rapid declines in the index value due to macroeconomic shocks, including the 2015-2016 stock market selloff [12] and the 2020 COVID-19 pandemic [9].

From Figure 2, it is clear to see that the log returns do not follow a normal distribution, as expected from the stylized facts of financial returns [6]. This is exemplified by the normal distribution's inability to capture the large central peak and heavy tails of the empirical distribution. The Gaussian KDE curve provides a much better approximation to the log return distribution, capturing a greater portion of both the central peak and heavy tails.

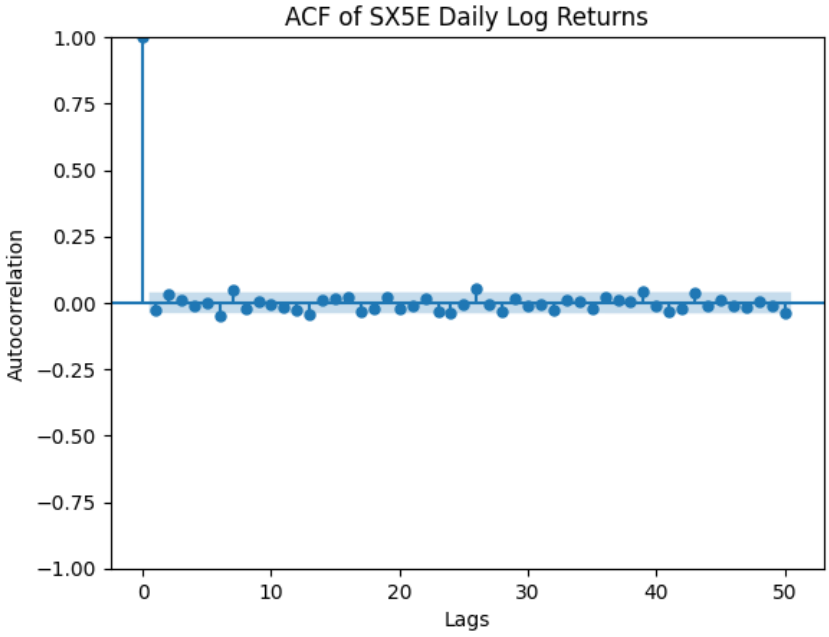

Figure 3: ACF of log returns.

Figure 3 displays the autocorrelation function (ACF) [5] for the daily log returns of the index up to 50 days of lag. From this plot it is clear that the daily log returns are serially

uncorrelated, as one would expect from a time series of returns.

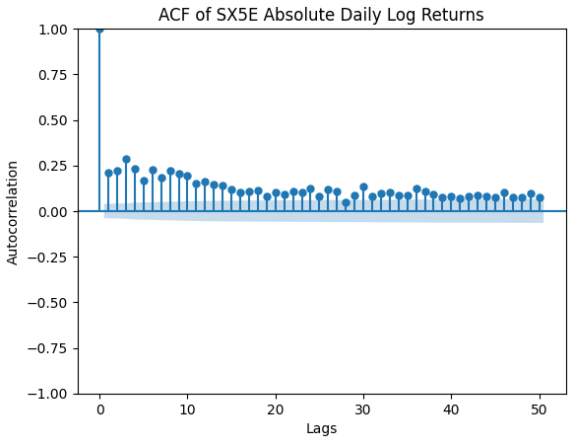

Figure 4: ACF of absolute log returns.

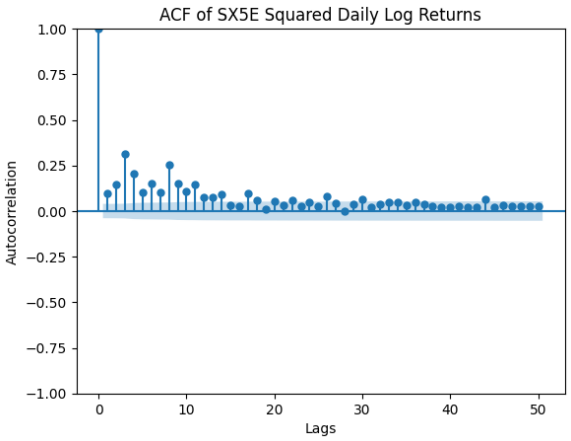

Figure 5: ACF of squared log returns.

Figures 4 and 5 display the ACFs for the daily absolute log returns and daily squared log returns up to 50 days of lag. Contrary to Figure 3, Figures 4 and 5 exhibit serial correlation over short time lags with some persistence over longer time lags. Particularly, in the case of absolute log returns, there is weak correlation for values that are up to 50 days apart. This implies that volatility tends to be clustered and highly persistent, that is, a large shift in the value of the index one day will most likely be succeeded by similarly large changes in subsequent days.

The discrepancy between the persistence of volatility in the absolute and squared daily log returns can be explained by the sensitivity of squared log returns to extreme values. In essence, squaring will amplify large returns disproportionately, introducing more variability and disrupting the persistence of volatility. As a result, this makes the volatility clustering for squared log returns more sporadic and less stable than absolute log returns.

1.3 Fitting the GARCH(1,1) model

Having performed a preliminary analysis of the index returns in the given timeframe, the suggested volatility clustering and persistence imply that fitting a model on the data that captures these properties would be most appropriate. In this initial fitting, a Generalized AutoRegressive Conditionally Heteroskedastic (GARCH) stochastic process is used.

1.3.1 Theory

Introduced by Tim Bollerslev in 1986 [2] as a successor to Robert Engle's ARCH model [7], a GARCH(p, q) process is a stochastic process defined as follows.

Definition 1. 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 $(Z_t)_{t\in\mathbb{Z}} \sim \text{SWN}(0, 1)$,

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

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 [5].

For the empirical application of this process to the sample data, $X_t = r_t$ where r_t will be defined to be the log return at time t. Following this logic, σ_t will be defined as the volatility at time t, which is thus positively dependent on the previous volatilities $\sigma_{t-1}, \ldots, \sigma_{t-q}$, capturing the volatility persistence, and the previous squared returns $X_{t-1}^2, \ldots, X_{t-p}^2$, capturing the volatility clustering. Hence, both of these qualities in volatility will be present in the model if $p, q > 0$.

In practice, GARCH processes with $p, q > 1$ are rarely used [5], motivating the choice of the GARCH(1,1) process for the initial model-fitting. Additionally, the strict white noise chosen follows the standard normal distribution. As a result these modeling choices, reduces the stochastic process to

$$r_t = \sigma_t Z_t, \quad \sigma_t^2 = \alpha_0 + \alpha_1 r_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \quad \text{SWN } Z_t \sim \mathrm{N}(0,1), \quad \forall t \in \mathbb{Z}.$$

To estimate the model parameters α_0, α_1 and β_1, the Maximum Likelihood Estimator (MLE) [5] approach is used.

Proposition 2. *The conditional likelihood function for the GARCH(1,1) process is given by:*

$$L(\alpha_0, \alpha_1, \beta_1; \mathbf{r}, \sigma_0) = \Pi_{t=1}^{T} \frac{1}{\sigma_t} f_Z(\frac{r_t}{\sigma_t}),$$

where $\mathbf{r} := (r_0, \ldots, r_T)$ *[5].*

This function is then maximized by the MLE estimators of α_0, α_1 and β_1. In addition to this, the sample standard deviation $\hat{\sigma}$ is chosen as the initial value σ_0. Nevertheless, in the maximization of the likelihood function, the parameters must be chosen such that the GARCH(1,1) process exists. This motivates the following linear constraint.

Proposition 3. *A covariance-stationary GARCH(1,1) process* $(r_t)_{t\in\mathbb{Z}}$ *exists if and only if*

$$\alpha_1 + \beta_1 < 1 \quad [5].$$

As a result of this, the MLE estimators for α_0, α_1 and β_1 are the values of these parameters that maximize $L(\alpha_0, \alpha_1, \beta_1; \mathbf{r}, \hat{\sigma})$ given that:

$$\alpha_0 > 0, \quad \alpha_1, \beta_1 \geq 0, \quad \alpha_1 + \beta_1 < 1.$$

1.3.2 Initial results

In fitting the empirical returns of the index using a GARCH(1,1) model, we used the Arch package available for Python [1]. The model was fitted using the built-in `arch_model` function within the package [1], automatically calculating the parameters for the model using the aforementioned MLE approach. The specific parameters applied in fitting the model are given in Table 3.

Parameter	Estimated value
α_0	0.0596
α_1	0.1390
β_1	0.8243

Table 2: Parameter estimated values for GARCH(1,1) process.

Clearly, the parameters satisfy the linear constraints imposed by the existence of the covariance-stationary GARCH(1,1) process, however, it is worth noting that α_1 is considerably smaller than β_1. This implies that, in the calculation for the next value of σ_t, the weight attributed to the previous volatility, σ_{t-1}, is much larger than that of the previous squared log return, r_{t-1}^2. This suggests that the model has particularly captured the volatility persistence property.

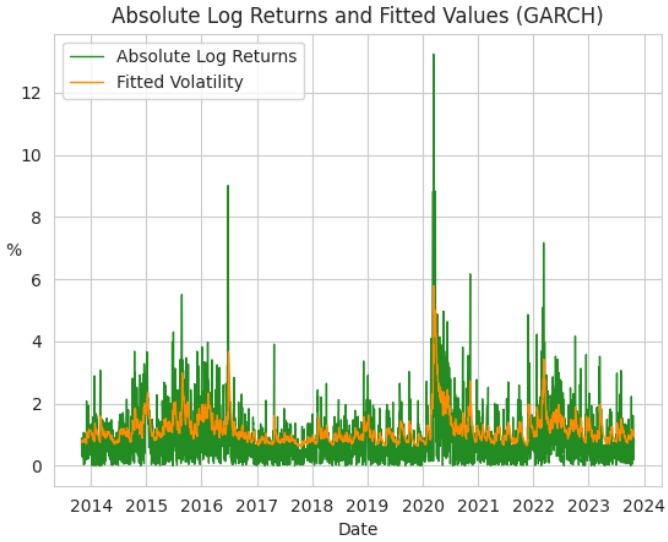

Figure 6: Fitted Volatility with GARCH(1,1) Model.

In Figure 6, we overlay fitted volatility values given by the model on empirical absolute log returns. Indeed, when there are peaks in the absolute log returns, the fitted volatility also experiences a spike, suggesting that the implementation of the model is correct. Furthermore, as theorized by optimizing the values of α_1 and β_1, the fitted volatility values exhibit clear volatility persistence, specifically evident in the volatility spike at the start of 2020. Similarly, volatility clustering is also apparent as seen in the distinct periods of high and low volatility values.

1.3.3 Residual diagnostics

To evaluate the model fit, we first compute the standardized residuals, $(\hat{Z}_t)_{t\in\mathbb{Z}}$, defined by

$$\hat{Z}_t := \frac{r_t}{\hat{\sigma}_t}, \quad \forall t \in \mathbb{Z} \quad [5].$$

From these values, we plot and examine several diagnostic plots to assess the model assumptions. In the scale-location plot, the square root of absolute residuals, $\sqrt{|\hat{Z}_t|}$, are plotted against the fitted volatilities $\hat{\sigma}_t$. Under our model assumptions, Z_t should be independent of $\hat{\sigma}_t$, which means that there should not be any noticeable trends in the scale-location plot.

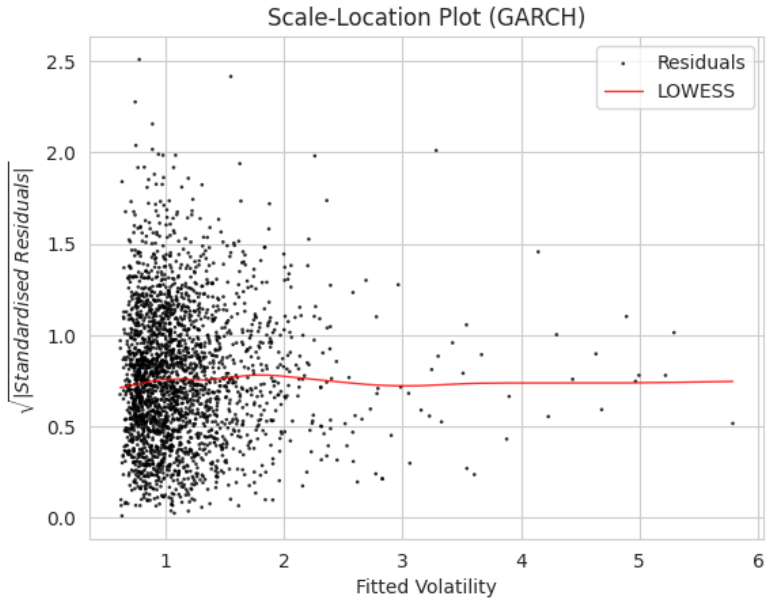

Figure 7: Scale Location plot for GARCH(1,1) residuals.

From Figure 7, although the most extreme residuals seem to originate from points with smaller fitted volatility, this could simply be due to the larger number of such points sampled from the $N(0,1)$ distribution. To investigate if there is a relationship between $\hat{\sigma}_t$ and Z_t, a locally weighted scatterplot smoothing (LOWESS) [4] line is commonly used, which uses a sliding window to divide the points into many groups, then fits a regression line using weighted least squares (WLS) to each group. This is plotted in red in Figure 7 using the `regplot` function in the SEABORN package [15]. We observe that the line is quite flat and does not exhibit any trends, which is consistent with the model assumptions. Also, it takes values between 0.7 and 0.8, which is approximately equal to $\sqrt{\frac{2}{\pi}}$, the theoretical expected deviation $\mathbb{E}[|Z|]$ of a standard normal random variable.

On the other hand, the Quantile-Quantile (QQ) [5] plot displays the empirical quantiles of the residuals against the theoretical quantiles of a $N(0,1)$ distribution. Ideally, the points should be close to the red 45° diagonal line, meaning that the quantiles are approximately equal, and that the sample is indeed very similar to a sample drawn from a $N(0,1)$ distribution.

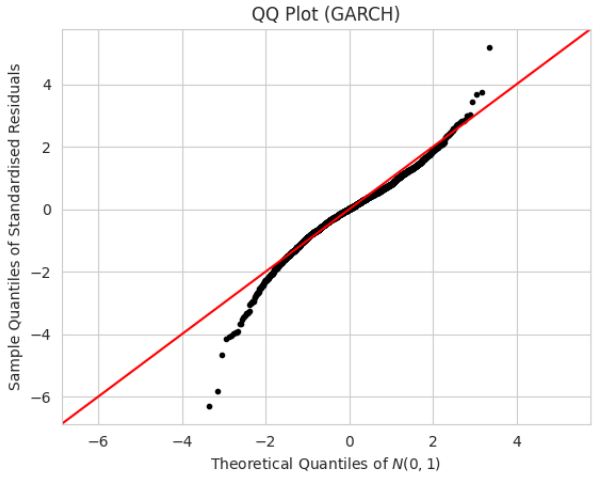

Figure 8: QQ plot for GARCH.

From Figure 8, we can see that the residuals are much more extreme than expected under a $N(0,1)$ distribution, indicated by the deviations of the tail residuals from the diagonal line. This suggests that normal innovations might be inadequate in capturing the behaviour of the time series, motivating the subsequent adjustment in Section 1.4. This was not revealed in the scale-location plot due to the sparsity of these extreme points and the smoothing nature of the LOWESS algorithm.

Lastly, we examine the ACF of the residuals and residuals squared to check for any serial correlation, and should be close to 0 as $(Z_t)_{t\in\mathbb{Z}}$ is a strict white noise.

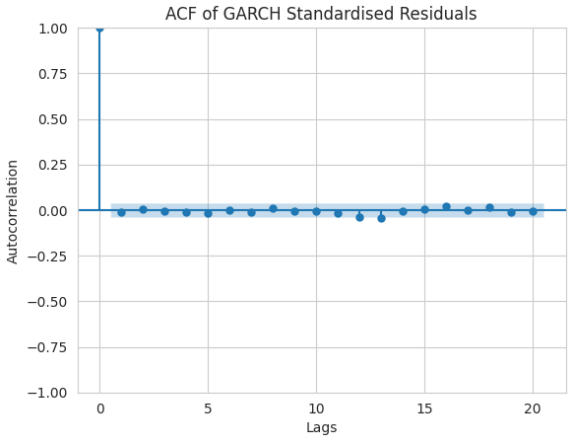

Figure 9: GARCH Residual ACF.

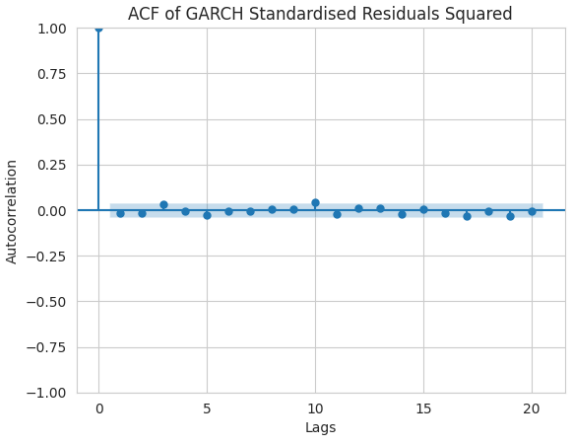

Figure 10: GARCH Squared Residual ACF.

From Figure 9 and Figure 10, we observe that although the serial correlation exhibits a very slight periodic pattern, the correlation values lie within the confidence interval for almost all lags above 1, which aligns with the model assumption.

1.4 Fitting the ARMA-GARCH(1,1)-t model

The residuals diagnostic of the previous model highlighted potential shortcomings of the GARCH(1,1) process with normally distributed innovations. For this reason, we now implement an AutoRe-

gressive Moving Average (ARMA) process with GARCH process innovations based on the student t-distribution.

1.4.1 Theory

The ARMA process, first described in 1951 by Peter Whittle [16], is defined in the following way.

Definition 4. A process $(X_t)_{t\in\mathbb{Z}}$ is a zero-mean ARMA(p,q) process, where $p,q \in \mathbb{N}_0$, if it is *covariance-stationary* and satisfies

$$X_t = \sum_{i=1}^{p} \phi_i X_{t-i} + \varepsilon_t + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} \quad \forall t \in \mathbb{Z},$$

where $\phi_1, \ldots, \phi_p$ and $\theta_1, \ldots, \theta_q$ are the parameters of the process and $(\varepsilon_t)_{t\in\mathbb{Z}} \sim \mathrm{WN}(0, \sigma^2)$. Moreover, $(X_t)_{t\in\mathbb{Z}}$ is an ARMA(p,q) process with mean $\mu \in \mathbb{R}$ if $(X_t - \mu)_{t\in\mathbb{Z}}$ is a zero-mean ARMA(p,q) process [5].

Combining an ARMA mean process with a GARCH volatility process, we obtain the ARMA-GARCH process, defined in the following way.

Definition 5. A process (X_t) is an ARMA-GARCH process if it satisfies

$$\begin{aligned} X_t &= \mu_t + \sigma_t Z_t, \text{ where} \\ \mu_t &= \mu + \sum_{i=1}^{p} \phi_i (X_{t-i} - \mu) + \sum_{j=1}^{q} \theta_i (X_{t-j} - \mu_{t-j}) \text{ and} \\ \sigma_t^2 &= \alpha_0 + \alpha_1 (X_{t-1} - \mu_{t-1})^2 + \beta_1 \sigma_{t-1}^2, \end{aligned}$$

where $\alpha_0, \alpha_1, \beta_1, \phi_1, ..., \phi_p, \theta_1, ..., \theta_q$ are parameters as previously defined [5].

In this case, we model the strict white noise Z_t as a t-distributed random variable. To estimate these parameters, we use the maximum likelihood approach again.

Proposition 6. *The conditional likelihood function for the ARMA(1,1)-GARCH(1,1) process is given by:*

$$L(\alpha_0, \alpha_1, \beta_1, \phi_1, \theta_1; \mathbf{r}, \sigma_0) = \Pi_{t=1}^{T} \frac{1}{\sigma_t} f_Z\left(\frac{r_t - \mu_t}{\sigma_t}\right) \quad [5].$$

Remark 7. *Similar to the GARCH setting, we require these constraints when maximizing the likelihood function,*

$$\alpha_0 > 0, \quad \alpha_1, \beta_1 \geq 0, \quad \alpha_1 + \beta_1 < 1.$$

1.4.2 Initial results

As the Arch [1] package has no functionality for fitting an ARMA-GARCH(1,1) model with t-distributed innovations, we define a custom objective function in Python using the likelihood function mentioned before. In this case, f_Z is the probability density function of a t-distribution with degree of freedom ν, scaled such that it has variance 1,

$$f_Z(t) = \sqrt{\frac{\nu}{\nu-2}} f_T\left(\sqrt{\frac{\nu}{\nu-2}} t; \nu\right) \quad [5],$$

where $f_T(t;\nu)$ is the density function of a t-distribution with degrees of freedom ν.

To maximize the objective function, we reduce the problem to a linear optimization problem by altering the objective function. By considering

$$l(\alpha_0, \alpha_1, \beta_1, \phi_1, \theta_1, \nu; \mathbf{r}, \sigma_0) = \ln(L(\alpha_0, \alpha_1, \beta_1, \phi_1, \theta_1, \nu; \mathbf{r}, \sigma_0)),$$

where

$$l(\alpha_0, \alpha_1, \beta_1, \phi_1, \theta_1, \nu; \mathbf{r}, \sigma_0) = -\ln(\sigma_t) + \ln(\sqrt{\frac{\nu}{\nu-2}} f_T(\sqrt{\frac{\nu}{\nu-2}}(\frac{r_t - \mu_t}{\sigma_t}); \nu) \quad [5],$$

we can solve the maximization problem by minimizing the function $-l(\alpha_0, \alpha_1, \beta_1, \phi_1, \theta_1, \nu; \mathbf{r}, \sigma_0)$ using the `minimize` function in the SciPy package [14] with the appropriate constraints. This outputs the following estimated parameters.

Parameter	**Estimated value**
α_0	0.0475
α_1	0.148
β_1	0.834
μ	0.07
ϕ_1	0.911
θ_1	-0.934
ν	4.727

Table 3: Parameter estimated values for ARMA-GARCH-t process.

Similar to the GARCH setting, β_1 is considerably larger than α_1, which suggests relatively high persistence. The rather small fitted value for ν is also consistent with our observation that the residuals are much heavier than a $N(0,1)$ distribution, as the t-distribution converges to a normal distribution as $\nu \to \infty$ [5].

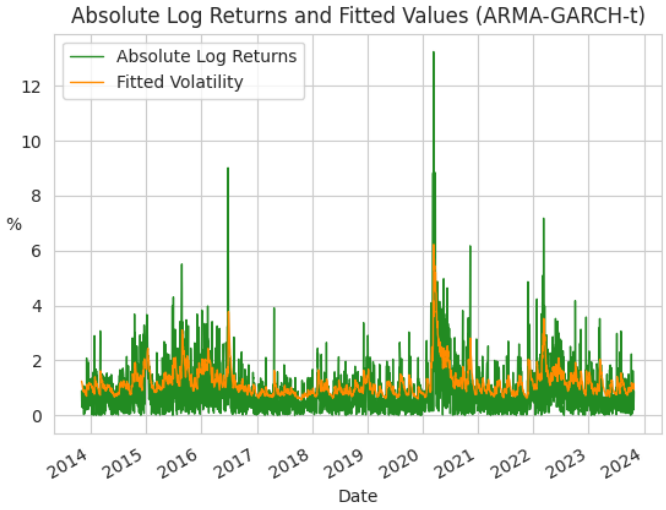

Figure 11: Fitted Volatility with ARMA-GARCH-t Model.

In Figure 11, we observe a very similar pattern to the one in the simpler GARCH model, with the fitted volatility exhibiting clear persistence and volatility clustering. Furthermore, the periods of higher and lower volatility, as well as the volatility spikes, coincide with the previous model, suggesting that the fit on the data is very similar.

1.4.3 Residual diagnostics

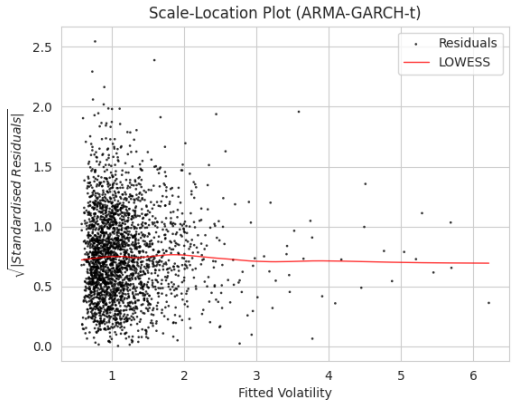

Figure 12: Scale Location plot for ARMA-GARCH-t residuals.

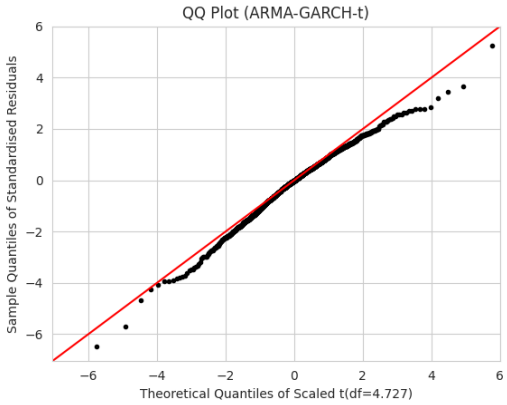

Figure 13: QQ plot for ARMA-GARCH-t.

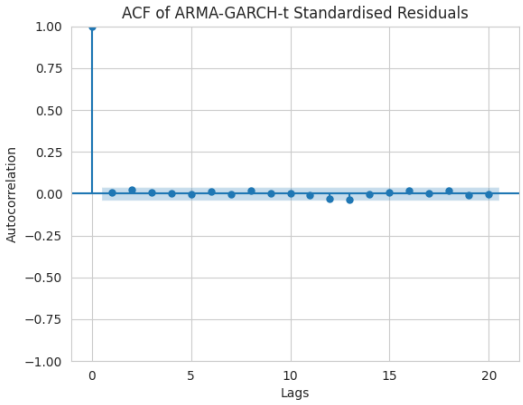

Figure 14: ARMA-GARCH-t residuals ACF.

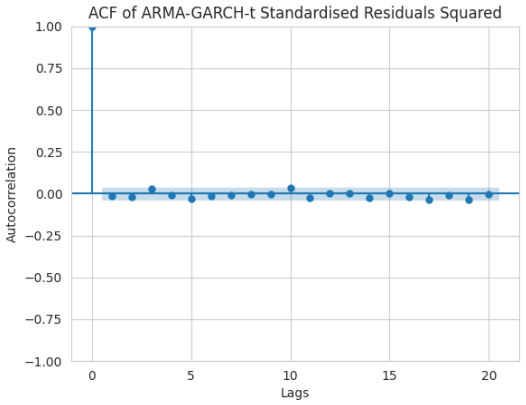

Figure 15: ARMA-GARCH-t squared residuals ACF.

Figure 12 to Figure 15 show the diagnostic plots for the ARMA-GARCH-t residuals. The plots are very similar to the GARCH plots and the respective analysis holds equally for them, with the exception of the QQ plot. From Figure 13, we observe that the standardised residuals are a lot closer to its theoretical distribution, the scaled t-distribution with $\nu \approx 4.727$, compared to the GARCH model with normal innovations. This suggests that using t-distributed innovations offers a better fit to the heavy-tailed behaviour of the time series. The tail characteristics of the QQ plot also show that there is a slight skew in the residuals, as they are more extreme in the left tail and less extreme in the right tail, consistent with the negative skew computed in our initial analysis (Table 1).

1.5 Comparison of models

Having assessed the results for each process fitted onto the data, a comparison between the models is necessary to identify the best fit of the two.

We note that a shortcoming of this approach is that the models we are comparing differ in more than one way, which complicates the task of analyzing how much impact each change to the model made to the fitted results. In a future project, we believe it would be pragmatic to fit the GARCH(1,1) model with standard normal innovations, followed by the GARCH(1,1) model with t-distributed innovations, and finally the ARMA-GARCH(1,1) model with t-distributed innovations. This would allow for a better understanding of the effect of the t-distributed innovations on the fitting of the log return data, as well as the effect of introducing an ARMA component to the GARCH model, without spuriously attributing the impact of each to other changes in the model.

As can be seen by comparing Figure 6 and Figure 11, the t-distributed ARMA-GARCH(1,1) process seems to capture the peaks in volatility with marginally greater precision, as they appear to be more extreme in periods with sudden large absolute log returns, for example, in the early 2020 peaks caused by the Covid-19 pandemic [9]. We attribute this difference to the t-distributed innovations of the ARMA-GARCH process, as the heavier tail of this distribution with respect to the standard normal distribution better captures the extreme values. Nevertheless, this only a very minor improval in the fit of the model and it remains unclear whether the ARMA component of the process has had any effect on the fitting of the model. As a result, it is worth considering a strictly quantitative approach to compare the models.

We can use the Bayesian Information Criterion (BIC) [13] of each model to compare them and choose the best fitting of the two.

Definition 8. Let k be the number of parameters to be estimated in a model, $\hat{L}$ be the maximized value of the likelihood function for the model and n be the number of data points in the sample. Then, the *Bayesian Information Criterion*, denoted BIC, of the model is defined as

$$\text{BIC} = k\ln(n) - 2\ln(\hat{L}) \quad [13].$$

Introduced in 1978 by Gideon E. Schwarz [13], the BIC is a measure of the information lost by using the model to represent the process that generated the data. Hence, the lower the BIC value, the less information the model has lost, and the better it is compared to another model with a higher BIC value. In addition, however, the model penalizes large numbers of parameters, as they will always improve the fit of the model, thus ensuring the model does not overfit the data. This penalization is also scaled by the sample size, penalizing the complexity of the model itself.

We calculate the BIC values for the GARCH(1,1) with standard normally distributed innovations and ARMA-GARCH(1,1) with t-distributed innovations to be 7,509.14 and 7,324.96 respectively. Hence, the apparent better fit of the ARMA-GARCH(1,1) process with t-distributed innovations of the data upon the residuals diagnostic, also holds through a quantitative criterion such as the BIC. However the small difference between these values also corresponds to the similarity in fit between the models, as apparent in overlaid volatility plots Figure 6 and Figure 11.

2 Part B: Risk Measures

2.1 Notes on Methodolody and Results

In this section of the report, we implement and evaluate one-day ahead VaR and ES forecasts at the 95% and 99% confidence levels for the loss of the *ProShares Short Dow30* (DOG) exchange-traded fund (ETF) [11] between October 30th 2013 and October 27th 2023.

By using a 500 day rolling window scheme, we first compute the forecasts using a historical simulation (HS) approach. We then employ filtered historical simulation (FHS) with exponentially-weighted moving average (EWMA) with parameter values $\alpha = 0.06$ and $\hat{\mu}_t = 0$ to compute the forecasts. Finally we perform the forecast calculations through an FHS with a GARCH(1,1) model with standard normally distributed innovations.

We approximate the cumulative distribution function (CDF) of the standardized residuals for the model-based approaches by assessing their empirical distributions and perform residuals diagnostic, consisting of QQ plots, scale-location plots and histograms overlaid with the standard normal distribution.

Finally, we perform backtesting for each risk measure and confidence level, with unconditional and joint coverage-independence tests.

2.2 Historical Simulation

The losses, L_t, for the DOG index fund are defined as

$$L_t = -r_t[5],$$

where r_t is the log return of the index. We then use a 500-day rolling window, using the previous 500 values of the loss to forecast the distribution of the next value. Thus, from the sample $L_{t-499}, \ldots, L_t$, the value at risk (VaR) at time $t+1$ can be forecasted as

$$\widehat{\mathrm{VaR}}_\alpha(L_{t+1}) := \hat{q}_\alpha[5],$$

where $\hat{q}_\alpha$ is the empirical α-quantile of the 500 day loss sample. From this approach, the expected shortfall (ES) can be forecasted as

$$\widehat{\mathrm{ES}}_\alpha(L_{t+1}) := \frac{1}{M_{t,\alpha}} \sum_{i=0}^{499} L_{t-i} \mathbf{1}_{\{L_{t-i} \geq \widehat{\mathrm{VaR}}_\alpha(L_{t+1})\}}[5],$$

where $M_{t,\alpha} := \#\{i \in \{0, 1, \ldots, 499\} : L_{t-i} \geq \widehat{\mathrm{VaR}}_\alpha(L_{t+1})\}$ is the number of VaR violations during the previous 500 days [5].

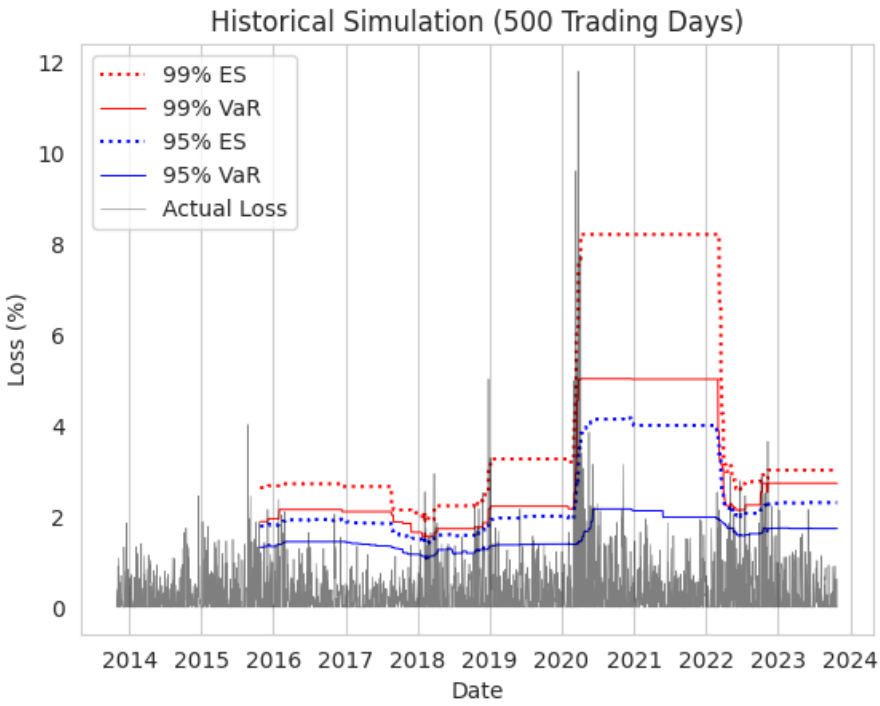

Figure 16: Historical simulation method with 500 day rolling window to forecast VaR and ES

Figure 16 displays the 500-day rolling window HS VaR and ES forecasts at $\alpha = 0.95$ and $\alpha = 0.99$ confidence levels overlaid on the losses throughout the total time period. It is worth noting that the first VaR and ES forecasts begin after the first 500 days due to the rolling window scheme. As can be seen, most of the losses fall below the VaR and ES forecasts, as expected. The proportion of the days at which the actual loss exceeds these risk measure values will be analyzed further in the backtesting section, assessing the reliability of this method.

2.3 Filtered Historical Simulation with EWMA

For the first FHS risk measure forecast, we implement the EWMA model. Originating from the J.P. Morgan RiskMetrics toolset [10], it consists of a recursive function to calculate volatility forecasts, defined as

$$\hat{\sigma}^2_{t+1} = \alpha(L_t - \hat{\mu}_t)^2 + (1-\alpha)\hat{\sigma}^2_t [5],$$

where $\hat{\mu}_t$ is an estimate of the conditional mean, chosen to be 0 for all t, and α is chosen to be 0.06, in line with the recommendations from the RiskMetrics documentation of the method [10]. Non-trivially however, the initial volatility, $\hat{\sigma}_0$, is chosen to be given by the sample standard deviation of the entire loss data, with $\hat{\sigma}_0 = 1.091\%$. This choice is made in order to utilize as much of the data available to obtain a reasonable initialisation.

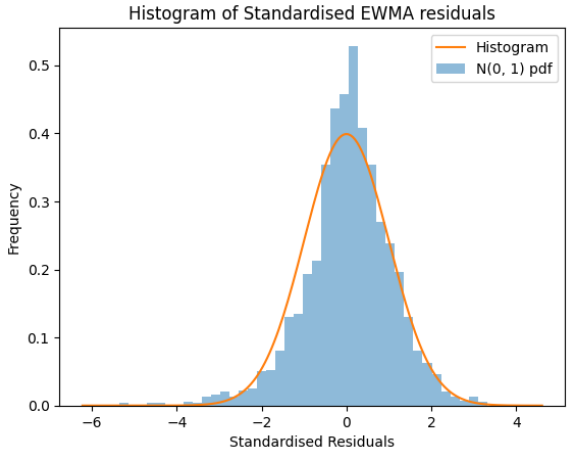

Figure 17: Histogram of EWMA residuals with standard normal distribution overlaid.

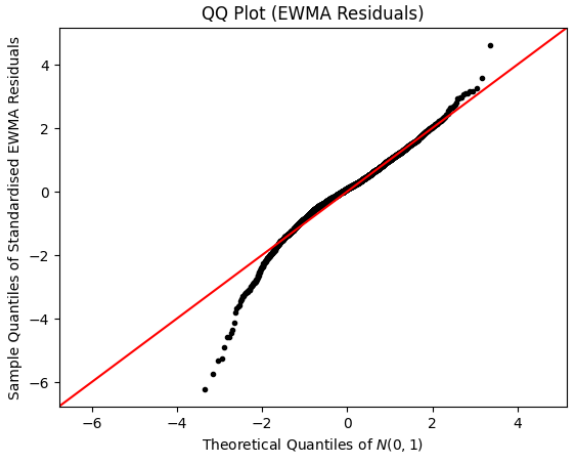

Figure 18: QQ plot of EWMA residuals on the standard normal distribution theoretical quantiles.

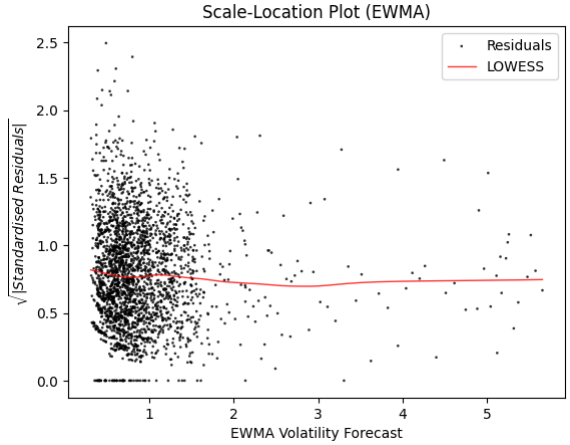

Figure 19: Scale-location plot of EWMA residuals.

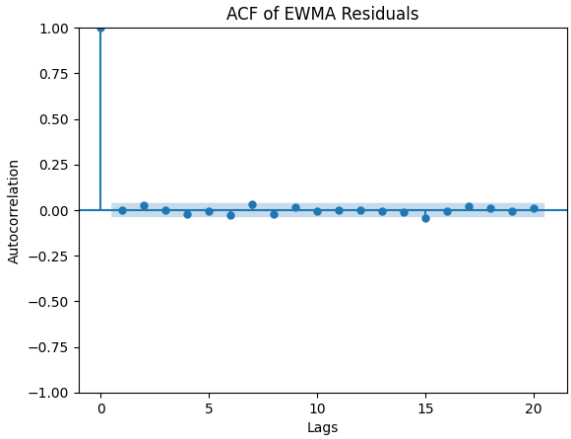

Figure 20: ACF of EWMA residuals.

As can be seen in Figure 17, the distribution of the standardised residuals of the EWMA model is slightly negatively skewed with a heavier left tail, however, aside from this fact, it resembles a standard normal distribution on the right tail of the distribution. This is corroborated by the QQ plot in Figure 18, where the standardised residuals' quantiles match the $\mathrm{N}(0,1)$ distribution except for the left tail. This can be explained by the stock market crashes of 2018 and 2020 [9], which resulted in large gains for the DOG index fund, causing extreme negative values for the losses. Furthermore, the smoothness of the LOWESS line on the scale-location plot in Figure 19 and the ACF of the EWMA residuals in Figure 20 convey a lack of correlation.

Having calculated the standardised residuals for the process, we now apply historical simulation to the residuals and calculate the 500-day rolling window VaR and ES forecasts given by

$$\widehat{\mathrm{VaR}}_\alpha(L_{t+1}) = \hat{\mu}_{t+1} + \hat{\sigma}_{t+1}\hat{q}_\alpha(\hat{F}_Z)[5], \tag{1}$$

$$\widehat{\mathrm{ES}}_\alpha(L_{t+1}) = \hat{\mu}_{t+1} + \hat{\sigma}_{t+1}\mathrm{ES}_\alpha(\hat{F}_Z)[5], \tag{2}$$

where $\hat{F}_Z$ is the empirical distribution of the previous 500 days of standardised residuals.

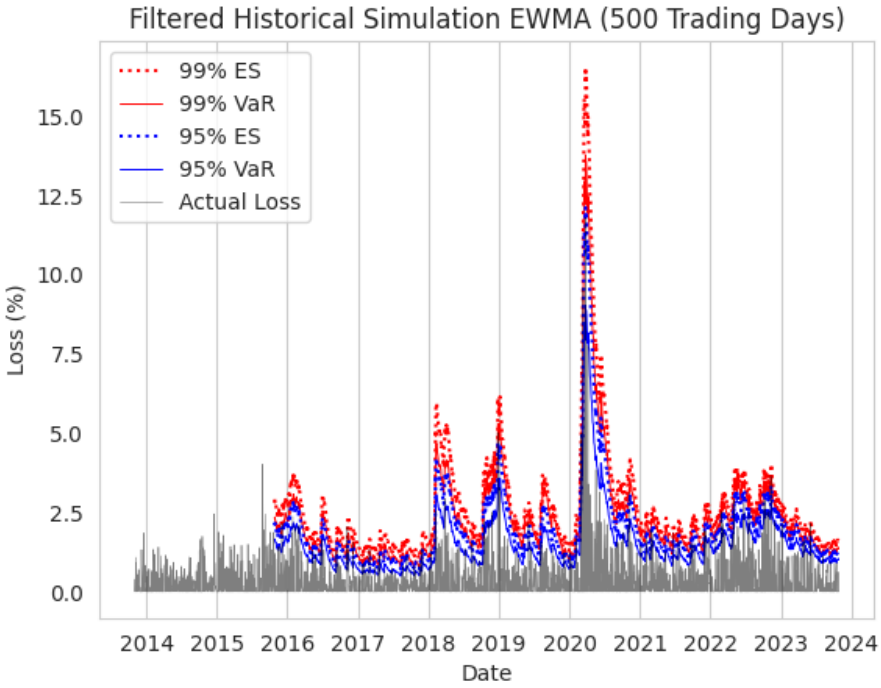

Figure 21: Using filtered historical simulation with EWMA to forecast VaR and ES.

Figure 21 displays the 500-day rolling window EWMA-FHS VaR and ES forecasts at $\alpha = 0.95$ and $\alpha = 0.99$ confidence levels overlaid on the losses throughout the total time period. In comparison with the HS approach, the VaR and ES forecasts for the EWMA-FHS method vary more noticeably with the losses and "hug" the loss data much more closely. This is due to the fact that the VaR forecast depends directly on the previous day's fitted volatility, hence, if this is an extreme value, the VaR forecast will increase greatly, as opposed to the HS case, where a previous day's large volatility would not change the VaR forecast as much.

2.4 Filtered Historical Simulation with GARCH

The last FHS risk measure forecast utilizes the previously introduced GARCH(1,1) process with standard normal innovations to forecast the volatilities of the losses and use these forecasts to calculate the respective VaR and ES values using FHS. The GARCH process is fitted on the data using the `arch_model` function from the ARCH package [1].

The implementation is given by the following iterative approach:

1. Fit the GARCH(1,1) process with standard normal innovations, using the `arch_model` function, to the values $\{L_i, \ldots, L_{i+499}\}$, where $i = 1$ on the first iteration.

2. Use the fitted model parameters to forecast the value of $\hat{\sigma}^2_{i+500}$, given by

$$\hat{\sigma}^2_{i+500} = \alpha_{0,i} + \alpha_{1,i} Z^2_{i+499} + \beta_{1,i} \hat{\sigma}^2_{i+499},$$

where Z_t is the standardised residual of day t. It is worth noting that the fitted model parameters may vary on each iteration.

3. The values of $\widehat{\text{VaR}}_\alpha(L_{i+501})$ and $\widehat{\text{ES}}_\alpha(L_{i+501})$ are calculated using equations (1) and (2) respectively, where $\hat{\mu}_{t+1}$ is given by $\hat{\mu}_t$, the GARCH model estimate of the parameter, and $\hat{F}_Z$ is the empirical distribution of the 500 standardised residuals.

4. The rolling 500-day window is shifted by adding 1 to the index and the method is repeated.

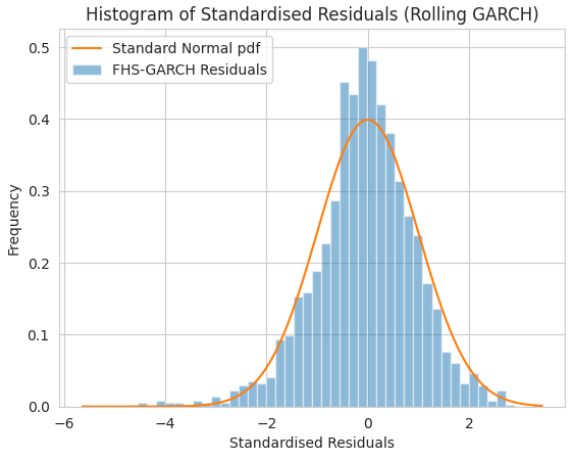

Figure 22: Histogram of GARCH residuals with standard normal distribution overlaid.

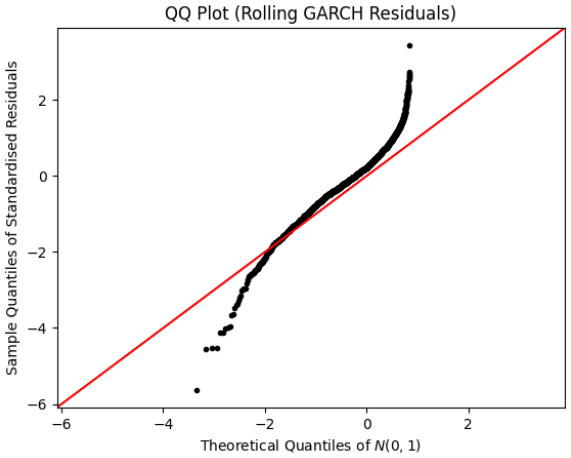

Figure 23: QQ plot of GARCH residuals on the standard normal distribution theoretical quantiles.

As can be seen from the histogram in Figure 22 and the QQ plot in Figure 23, the residuals for the GARCH(1,1) FHS model do not follow the standard normal distribution. It is especially apparent in the QQ plot that the tails of the standardised residual distribution are much heavier. This suggests that a GARCH model with innovations following a heavier-tailed distribution, such as the t-distribution, would fit the data better.

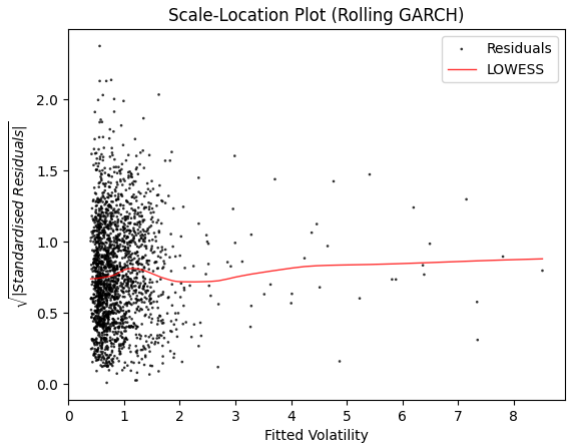

Figure 24: Scale-location plot of GARCH residuals.

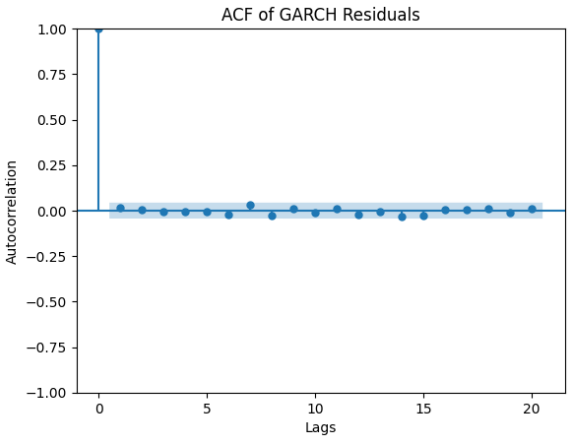

Figure 25: ACF of GARCH residuals.

Similar to the EWMA model, the residuals do not show any sign of correlation, as seen on the scale-location plot and ACF in Figures 24 and 25 respectively.

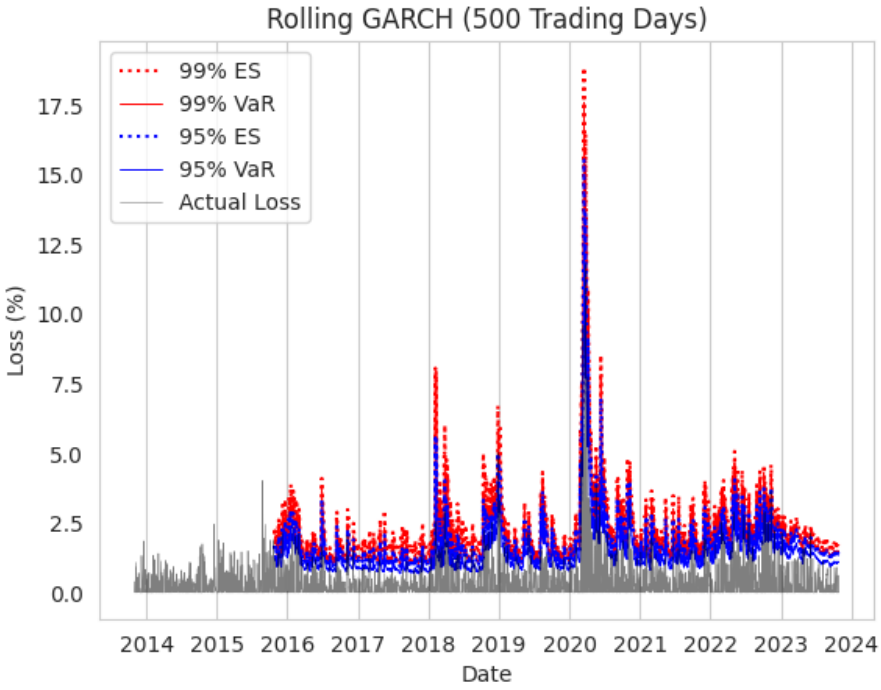

Figure 26: GARCH-FHS VaR and ES forecasts, 500-day rolling window.

Figure 26 displays the final 500-day rolling window GARCH-FHS VaR and ES forecasts at $\alpha = 0.95$ and $\alpha = 0.99$ confidence levels overlaid on the losses throughout the total time period. As the FHS-GARCH VaR and ES forecasts also depend on the previous day's fitted volatility, the VaR and ES forecasts "hug" the loss data much more closely than the HS forecasts. In this case, however, the GARCH volatility forecasts are calculated using different specific fitted parameters for every iteration which results in a higher degree of sensitivity to trailing values and thus a more jagged curve. In addition to this, it seems that the forecasts spike around

periods of high volatility, but quickly mean revert which may result in an inability to capture periods of persistent high volatility as accurately as the EWMA model.

2.5 Backtesting

We now assess the results of the forecasts that the methods mentioned above produced for one-day ahead VaR and ES. We compute p-values at 95% and 99% confidence level with our findings tabulated below.

For all tests we apply a significance level of 5% regardless of whether the test is conducted at a 95% or 99% confidence level. This is to protect against Type II errors, which could result in an unsuitable model being accepted with the potential of realizing material bad outcomes. On the other hand, increasing the rate of Type I errors has a much more limited impact since it does not meaningfully influence our ability to select a model that is consistent with our null hypotheses.

2.5.1 VaR unconditional backtesting

To perform the unconditional coverage hypothesis test, we define the test statistic to be, LR_{uc}, given by

$$\text{LR}_{\text{uc}} = -2\ln\frac{L(1-\alpha; I_1, \ldots, I_T)}{L(\hat{\pi}; I_1, \ldots, I_T)}, [5]$$

where α is the level of the test, for any $t = 1, .., T$ we define $I_t := \mathbf{1}_{\{L_t > \widehat{\text{VaR}}_\alpha(L_t)\}}$ [5], $L(\pi; I_1, \ldots, I_T)$ is the likelihood function, defined as $L(\pi; I_1, \ldots, I_T) := \Pi_{t=1}^{T}(1-\pi)^{1-I_t}\pi^{I_t}$ [5] and $\hat{\pi}$ is the MLE estimator for π, defined as $\hat{\pi} := \frac{1}{T}\sum_{t=1}^{T} I_t$ [5].

With the null hypothesis H_0 that the VaR forecasts are calculated using a correctly-specified model, the test statistic LR_{uc} follows the $\chi^2(1)$ distribution asymptotically (as $T \to \infty$) [5]. Hence, the p-value, p, is calculated as $p = 1 - F_{\chi^2(1)}(\text{LR}_{\text{uc}})$ [5].

	$\alpha = .95$			$\alpha = .99$		
Method	Violations(Exp)	Test stat.	p-value	Violations(Exp)	Test stat.	p-value
HS	99 (101)	0.0322	0.8576	30 (20)	4.2283	0.0398
FHS (EWMA)	99 (101)	0.0322	0.8576	23 (20)	0.3894	0.5326
FHS (GARCH)	101 (101)	0.0007	0.9796	29 (20)	3.4566	0.0630

Table 4: Key statistics from unconditional coverage backtests.

At the 95% confidence level, all three models show a high degree of consistency with H_0 with p-values above 85%, making it unclear which model performs best. However, at the 99% level we notice that the results from the FHS (EWMA) model are significantly more consistent with H_0 than the results from the other models. While it seems that the FHS (EWMA) performs best, the unconditional coverage test is at a disadvantage as it does not detect whether the VaR violations occur in clusters [5]. In practice, clustered losses are a source of existential risk to individual institutions as well as systemic risk to the broader market, motivating the use of the joint coverage-independence test to better understand the robustness of our forecasts.

2.5.2 VaR joint backtesting

Introduced by Christoffersen in 1998 [3], the independence test is based on the assumption that $I_1, \ldots, I_T$ form a first-order Markov chain. This implies that the Markov chain has two parameters;

$$\pi_{11} := \mathbb{P}[I_t = 1 | I_{t-1} = 1], \quad \text{and} \quad \pi_{01} := \mathbb{P}[I_t = 1 | I_{t-1} = 0].$$

These then define the transition matrix for the Markov chain, given by

$$\Pi := \begin{pmatrix} 1 - \pi_{01} & \pi_{01} \\ 1 - \pi_{11} & \pi_{11} \end{pmatrix} [5].$$

Thus, taking the initial value I_1 as known, the likelihood function is given by

$$L(\Pi; I_2, \ldots, I_T) := (1 - \pi_{01}^{T_{00}})\pi_{01}^{T_{01}}(1 - \pi_{11})^{T_{10}}\pi_{11}^{T_{11}}[5],$$

where T_{ij} is the number of observations with $i \in \{0, 1\}$ followed by $j \in \{0, 1\}$ in the violation sequence. Then, the MLE estimators for π_{01} and π_{11} are

$$\hat{\pi}_{01} = \frac{T_{01}}{T_{00} + T_{01}}, \quad \text{and} \quad \hat{\pi}_{11} = \frac{T_{11}}{T_{10} + T_{11}}[5],$$

thus, denoting the matrix $\hat{\Pi}$ as the MLE estimator for Π. In our test, however, the null hypothesis, H_0, states that $I_1, \ldots, I_T$ are independent, implying that $\pi_{01} = \pi_{11}$. This yields the MLE transition matrix under H_0, defined as

$$\hat{\Pi}_0 := \begin{pmatrix} 1 - \hat{\pi} & \hat{\pi} \\ 1 - \hat{\pi} & \hat{\pi} \end{pmatrix},$$

where $\hat{\pi} = \frac{1}{T}\sum_{t=1}^{T} I_t$.

This allows for the construction of the test statistic, LR_{ind}, again a likelihood ratio, defined as

$$\text{LR}_{\text{ind}} := -2\ln\frac{L(\hat{\Pi}_0; I_2, \ldots, I_T)}{L(\hat{\Pi}; I_2, \ldots, I_T)}[5],$$

which under the hypothesis H_0 also follows the $\chi^2(1)$ distribution asymptotically [5]. As a result, the p-value can be calculated as $p = 1 - F_{\chi^2(1)}(\text{LR}_{\text{ind}})$ [5]. Finally, we consider the joint test statistic, LR_{joint}, given by

$$\text{LR}_{\text{joint}} := \text{LR}_{\text{uc}} + \text{LR}_{\text{ind}}[5],$$

which under the null hypothesis follows the $\chi^2(2)$ distribution asymptotically [5]. Its p-value is thus given by $p = 1 - F_{\chi^2(2)}(\text{LR}_{\text{joint}})$ [5].

	$\alpha = .95$			$\alpha = .99$		
Method	Violations(Exp)	Test stat.	p-value	Violations(Exp)	Test stat.	p-value
HS	99 (101)	4.7544	0.0928	30 (20)	7.2840	0.0262
FHS (EWMA)	99 (101)	0.2128	0.8990	23 (20)	0.9208	0.6310
FHS (GARCH)	101 (101)	2.6129	0.2708	29 (20)	0.8480	0.6544

Table 5: Key statistics from joint coverage-independence backtests.

Once again, we observe that the FHS-EWMA model produces a high p-value at both 95% and 99%, suggesting that the results are also consistent with the null hypothesis that violations do not occur in clusters. We also notice the drop in p-values at 95% of the other methods with the inclusion of the independence test. This points to relatively less consistency of the results from the HS and FHS-GARCH models with the hypothesis that VaR violations occur independently.

At the 99% confidence interval, we noted that there are no consecutive VaR violations under FHS-EWMA or FHS-GARCH, and as a result these terms were ignored in the calculation of the test statistics and p-values. The absence of these instances makes it challenging to draw strong conclusions on the results at 99%, and indicates that a larger time horizon may be needed to accurately assess the independence of violations under the models at 99%.

2.5.3 ES backtesting

In backtesting our forecasts for expected shortfall, we take a slightly different approach as compared to the calculations for VaR. This is because ES is an expectation on realized losses when the VaR threshold is breached: when backtesting we are only able to observe a single shortfall for each individual breach of VaR and so we must aggregate over all of these observations to understand the overall performance of our forecasts.

We have that $\mathrm{ES}_\alpha(L_t) = \dfrac{\mathbb{E}(L_t I_t)}{1-\alpha}$ [5], where I_t are the indicators for realized breaches of VaR and $\mathrm{ES}_\alpha(L_t)$ the expected loss in these cases. Taking

$$\xi_t := (L_t - \widehat{ES_\alpha}(L_t))I_t \text{ [5]},$$

and using the fact that the following exhibits asymptotic normality as $T \to \infty$

$$Z := \frac{\Sigma_{t=1}^T \xi_t}{\sqrt{\Sigma_{t=1}^T \xi_t^2}} \text{ [5]},$$

and is a test statistic for null hypothesis

$$H_0 : \widehat{\mathrm{ES}}_\alpha(L_t) = \mathrm{ES}_\alpha(L_t) \ , \ \widehat{\mathrm{VaR}}_\alpha(L_t) = \mathrm{VaR}_\alpha(L_t) \text{[5]},$$

we find p-values at 95% and 99% as tabulated below [5].

	$\alpha = .95$		$\alpha = .99$	
Method	Test stat.	p-value	Test stat.	p-value
HS	1.1248	0.1303	1.7796	0.0376
FHS (EWMA)	0.2488	0.4017	-0.8105	0.7912
FHS (GARCH)	0.8091	0.2092	-0.3091	0.6214

Table 6: Key statistics from ES backtests.

We observe that the FHS-EWMA is the most consistent with H_0, and that the FHS-GARCH model also exhibits a high degree of consistency with H_0. This aligns with the results for the VaR backtests and once again demonstrates that the FHS models are able to capture the key features

of stylized facts of returns. Namely, in this case, periods of higher volatility are persistent in the forecasts which results in a smaller deviation when compared to realized losses. The HS model, as expected, is not sensitive to recent returns/volatility and so exhibits a worse fit over time.

2.5.4 Conclusions

From our hypothesis tests on the each of the forecast methods, we conclude that the FHS-EWMA is the most suitable among the three models. This is because the model is highly consistent with our null hypotheses across all tests conducted. Therefore, we would recommend that Bob uses the FHS-EWMA risk measure forecasting method.

In the HS model, we observe p-values that lead us to believe that our null hypothesis should be rejected in some cases and from Figure 16 can directly observe that the forecasts do not capture variations in volatility and returns over time.

The FHS-GARCH model produces suitable forecasts for VaR and ES, but is relatively less consistent with the tested null hypotheses. Particularly, the difference in p-values between FHS-EWMA and FHS-GARCH is apparent for the joint coverage-independent test. This highlights limitations in the GARCH approach, as it performs slightly worse when volatility is highly clustered. The FHS-EWMA model performs better in this regard, as the volatility forecasts are more sensitive to market movements yielding VaR violations that appear independent across the time period.

Thus, though all three models demonstrate some degree suitability for producing VaR and ES forecasts, we find that the HS model is the weakest while the FHS-EWMA is the most accurate in forecasting VaR and ES.

3 Mathematical Appendix

Here we provide some definitions for completeness.

Definition 9 (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:

$$R_t := \frac{S_t - S_{t-1}}{S_{t-1}} [5].$$

Definition 10 (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 [5].

Definition 11 (Strictly Stationary)**.** A process $(X_t)_{t\in\mathbb{Z}}$ is *strictly stationary* if

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

for any $t_1, \ldots, t_n \in \mathbb{Z}$, $k \in \mathbb{Z}$, and $n \in \mathbb{N}$ [5]. Above, $\overset{d}{=}$ indicates equality of joint distributions.

Definition 12 (Kernel Density Estimator)**.** Let $(X_t)_{t\in\mathbb{Z}}$ be iid samples drawn from some univariate distribution with an unknown density f at any given point x. Its *Kernel Density Estimator* is

$$\hat{f}_h(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - x_i) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right),$$

where K is the *kernel* — a non-negative function — and $h > 0$ is a smoothing parameter called the *bandwidth* [5].

Definition 13 (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* [5].

Definition 14 (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 $(X_t)_{t\in\mathbb{Z}} \sim \mathrm{SWN}(0, \sigma^2)$ [5].

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] Peter F. Christoffersen. "Evaluating interval forecasts". In: *International Economic Review* 39.4 (Nov. 1998), p. 841. DOI: 10.2307/2527341.

[4] 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.

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

[6] R. Cont. "Empirical properties of asset returns: Stylized facts and statistical issues". In: *Quantitative Finance* 1.2 (Feb. 2001), pp. 223–236. DOI: 10.1080/713665670.

[7] Robert F. Engle. "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation". In: *Econometrica* 50.4 (July 1982), p. 987. DOI: 10.2307/1912773.

[8] *EUROSTOXX50*. Feb. 1998. URL: https://stoxx.com/index/sx5e/.

[9] Mieszko Mazur, Man Dang, and Miguel Vega. "Covid-19 and March 2020 stock market crash. evidence from SP1500". In: *SSRN Electronic Journal* 38 (Jan. 2021). DOI: 10.2139/ssrn.3586603.

[10] J.P. Morgan. *Riskmetrics Technical Document*. URL: https://www.msci.com/documents/10199/5915b101-4206-4ba0-aee2-3449d5c7e95a.

[11] ProShares. *DOG: Short DOW30*. URL: https://www.proshares.com/our-etfs/leveraged-and-inverse/dog.

[12] Kit Rees. *European stocks suffer worst 1-day fall in nearly 4 years*. 2015. URL: https://www.reuters.com/article/2015/08/21/markets-stocks-europe-idUSL5N10W2KW20150821/.

[13] Gideon Schwarz. "Estimating the dimension of a model". In: *The Annals of Statistics* 6.2 (Mar. 1978). DOI: 10.1214/aos/1176344136.

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

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

[16] Peter Whittle. "Hypothesis testing in time series analysis." In: *Journal of the Royal Statistical Society. Series A (General)* 114.4 (1951), p. 579. DOI: 10.2307/2981095.