Advanced Time Series Econometrics

First Assignment

Umberto Maglione (0001034282)
Luigi Murialdo (0001037355)
Alonso Zúñiga Irigoin (0001026037)

December 2022

1 Introduction

The purpose of this document is to analyze the statistical properties of the United States NASDAQ Stock Exchange index, provide a conditional heteroskedasticity model and propose a Bootstrap method for an ARCH(1) model. The time-series under scrutiny are daily and weekly data, they both span from the 5^{th} February of 1971 (the index inception) to the 15^{th} November of 2022; data are downloaded from Yahoo Finance. More specifically, for the former there are 13,060 observations and for the latter 2,703. Lastly, the weekly prices represent the end of the period price, i.e. Friday's closing price. In Figure 1, 2 and 3 are presented respectively the price levels, the log prices and the log returns of both frequencies.

Figure 1: NASDAQ Price

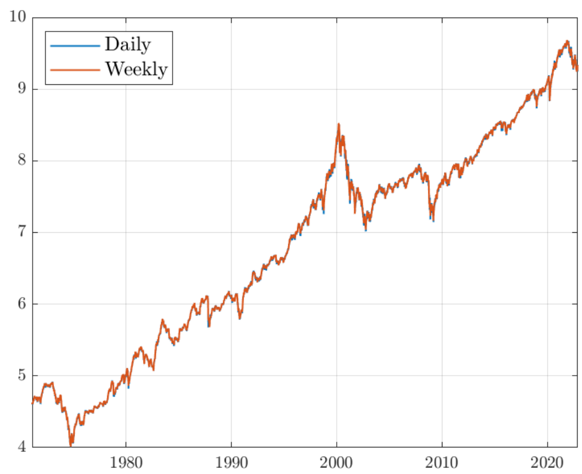

Figure 2: NASDAQ Log Price

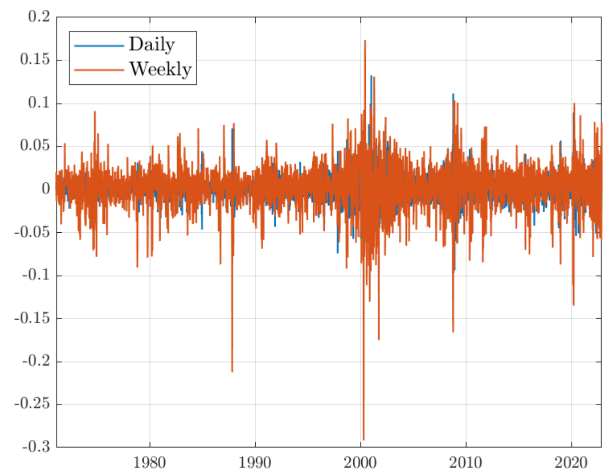

Figure 3: NASDAQ Log Returns

From the first two graphs one can observe that the exponential shape of the prices in levels is cancelled out by the logarithmic transformation, which in turn has a linear trend. At a glance, none of the first two graphs look stationary (nevertheless a formal proof is presented afterwards). With respect to the third graph, the log-returns[1] series looks stationary, although the presence of recurring extreme events suggest that the distribution of the returns might be *leptokurtic*, i.e. there is the presence of heavy tails, especially in the left one.

The rest of the document is divided as follows. Part 2 presents an analysis of some stylized facts showed by these 2 time-series. Among these there are unit root as well as serial correlation tests, and evidences about heavy tails in the distribution of returns. On Part 3 we estimate a GARCH model to measure the conditional volatility of returns. In part 4 we discuss about the Value-at-Risk measure for the NASDAQ. Finally, in Part 5 we switch the focus into how to a Bootstrap procedure for an ARCH(1) model.

2 Stylized Facts

Before starting, it is important noting that from here onward the series under analysis are the log-returns ones, that are our main object of interest. As a first approach to assess the characteristics of both frequencies, observing their distribution is an initial step. As can be seen in Figures 4 and 5, both time frequencies are centered in zero[2] but both show a higher frequency of extreme values than a normal distribution, supporting the idea expressed previously about heavy tails(which will be tested afterwards).

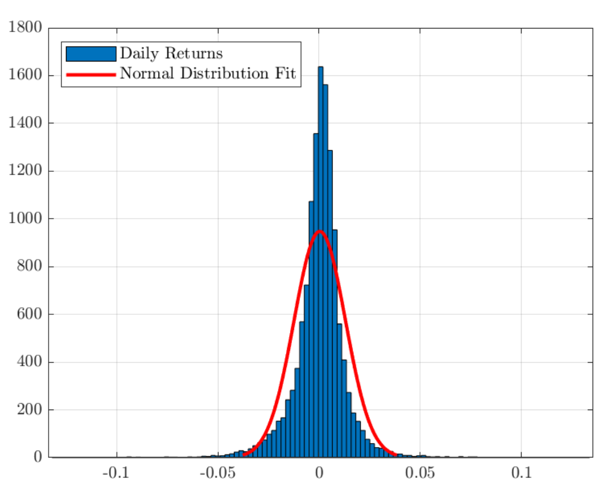

Figure 4: Daily Returns Histogram

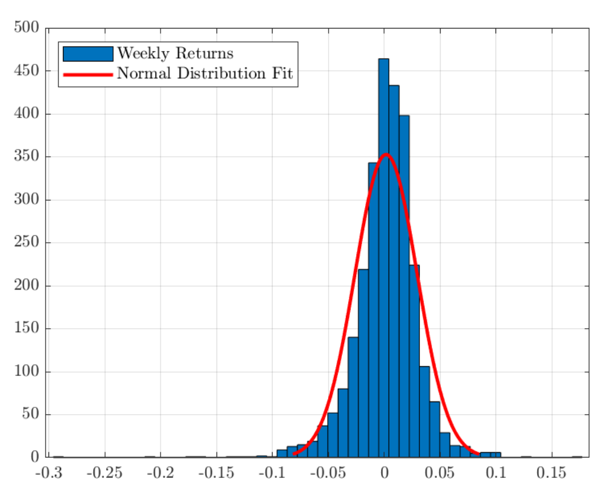

Figure 5: Weekly Returns Histogram

Continuing, a key hypothesis about returns evolves around the existence of a linear relation between past and present realizations. Therefore, one can study the correlogram to check the existence of such relations that could be exploited through an auto-regressive model. Figures 6 and 7 present the correlogram of both time frequencies[3], where one can observe that the autocorrelation is very low and, at best, there is 5% c.a. correlation

[1]In the rest of the document we refer to log-returns as returns indistinctly.
[2]Not exactly zero but very close to it.
[3]The graphs don't have confidence bands in them because for it data must be independent and identically distributed, which we believe it is not.

that could be exploited. This was expected, since other empirical analyses (see Ding and Granger, 1996) report the same absence of correlation. Nevertheless, by observing Figures 8 and 9 one can appreciate the positive correlation of squared returns, which are related to second moments, i.e. volatility of the returns. This positive correlation is what motivates a statistics model, since it can be exploited to predict and forecast future volatility.

More in detail, for the daily frequency we observe that up to the fiftieth lag the sample autocorrelation of returns is higher than 10% peaking at the first lags around 35%. On the other hand, the relation is less strong for weekly returns although it remains higher than 10% until the ninth lag approximately.

The latter result, the positive correlation of the squared returns, suggests the possibility of modelling these relations, conveying the base idea of conditional heteroskedasticity models. As required, such model is computed in the third section of this document.

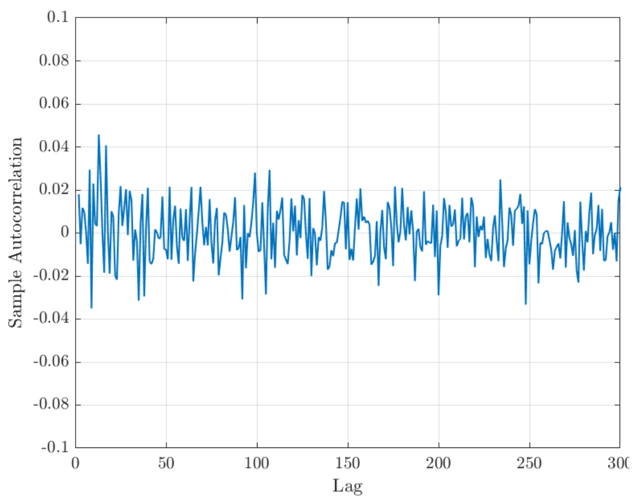

Figure 6: Daily Returns Correlogram

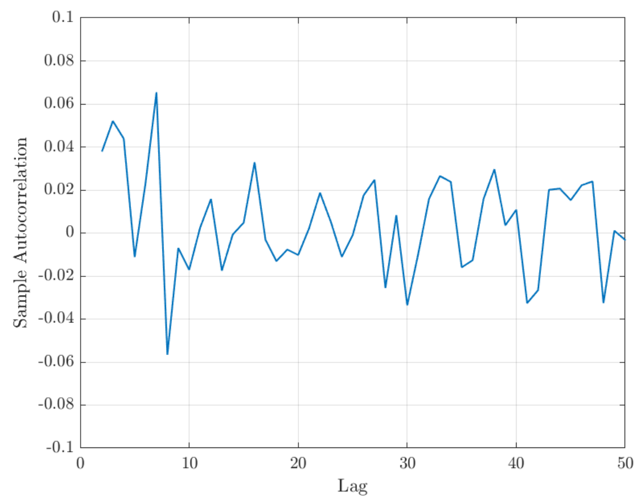

Figure 7: Weekly Returns Correlogram

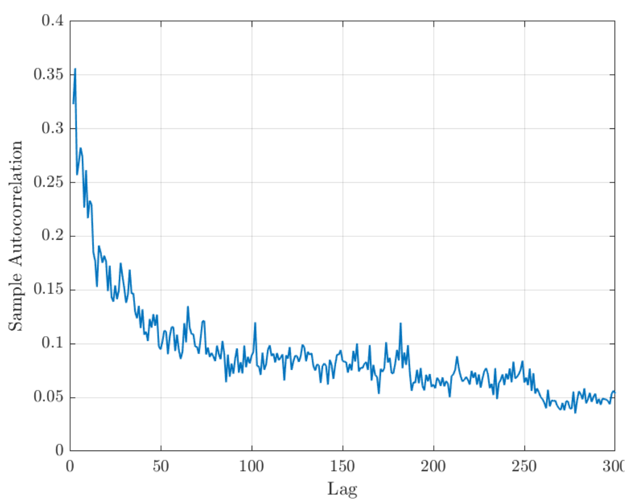

Figure 8: Daily Squared Returns Correlogram

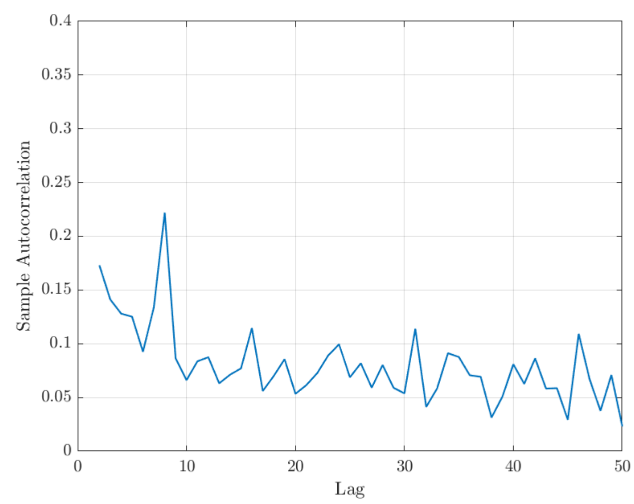

Figure 9: Weekly Squared Returns Correlogram

Given the observed presence of positive correlations within the squared returns and the possibility to model it, in Figure 10[4] are plotted the squared returns of the NASDAQ throughout time. Hence, the next section analyzes some of the stylized facts seen in class.

[4]The histogram of these two time series is not worth reporting since it is not informative, there is a very high density of observations near zero and a very low one in the rest, nonetheless due to the extreme values it is still presented. Of course, this histogram is nowhere near to be similar to a normal distribution.

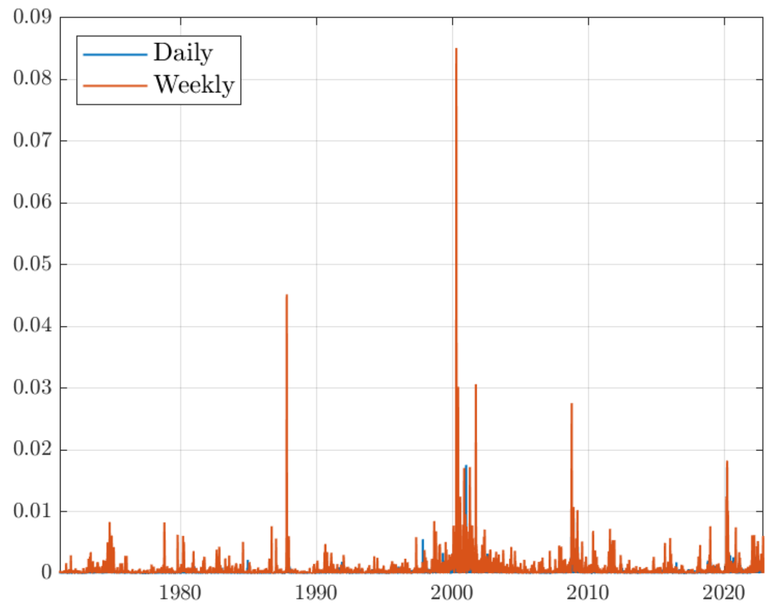

Figure 10: NASDAQ Squared Log Returns

2.1 Unit Roots: Dickey-Fuller Test

To begin, one important property of time series data is stationarity. We will focus on weak stationarity only, which deals with finite and constant first and second moments, additionally it guarantees that the autocovariance of the process depends only on the lags of the process itself and doesn't change through time.

To understand the logic behind the Dickey-Fuller test, let's assume that current returns are explained by the past returns, hence one could think of an autoregressive process and test for the presence of unit roots, which means we could have a random walk type of behaviour:

$$X_t = \alpha X_{t-1} + \epsilon_t \tag{1}$$

Where our returns series is denoted by X_t and we need to test for $\alpha = 1$, which means testing for $\delta = 0$ in the following equation:

$$\Delta X_t = (\alpha - 1)X_{t-1} + \epsilon_t = \delta X_{t-1} + \epsilon_t \tag{2}$$

This is the basis of the Dickey-Fuller test (Dickey and Fuller, 1979), which can be generalized to test for unit roots in higher order autoregressive processes, also including a constant and a time trend. This is called an Augmented Dickey-Fuller (ADF) test and it checks for $\alpha = 0$ in a model of the type:

$$\Delta X_t = \omega + \beta t + \alpha X_{t-1} + \rho \Delta X_{t-1} + \epsilon_t \tag{3}$$

Moving forward, below are reported the outcomes of the ADF test conducted on NASDAQ log-prices and returns. From Figure 2 we can observe that there is a clear linear trend in log-prices, hence we proceeded by using the ADF test with constant and trend. With respect to the other case, from Figure 3 one can appreciate that there is no linear trend and probably no significant constant, thus we ran the ADF test without accounting for these two features. The results of the ADF tests for daily and weekly data are presented in Table 1:

Table 1: ADF Test

Series	P-Value	Null Hypothesis
Daily Prices	0.373	Not Reject
Daily Returns	0.001	Reject
Weekly Prices	0.381	Not Reject
Weekly Returns	0.001	Reject

As could have been expected from the Figures presented before, when working with the first differences of the log-prices the presence of the unit root usually disappears, meaning that the log-prices might be time series integrated of order 1.

As already mentioned, if the data behaves as a random walk, i.e. there is unit root, then the second moments of the time series are infinite and thus they are not weakly stationary. If the test is rejected it is very likely[5] that we are not in the presence of a unit root; nevertheless, the test is uninformative of the homoskedasticity of the series. Thus, even if the presence of a unit root can be rejected, our process could still be heteroskedastic and, as a consequence, not weakly stationary.

2.2 Ljung-Box Test

So far, we have collected evidence that returns might be stationary by discarding, with 95% confidence, the presence of a unit root in an autoregressive process.

Assuming that they actually are weakly stationary[6], from looking at Figure 6 to Figure 9, the question of significant autocorrelation automatically arises. Therefore, to test this hypothesis, one can resort to the Ljung-Box test (Ljung and Box, 1978), that involves analyzing the cumulative correlations through the implementation of the following statistic:

$$Q(\hat{\rho}) = T(T+2)\sum_{i=1}^{L} \frac{\widehat{\rho_i^2}}{T-i} \xrightarrow{d} \chi_L^2 \tag{4}$$

Where in equation (4) T is the sample size, L the number of lags being tested and $\hat{\rho}_i$ is the sample autocorrelation to the i-th lag. The null hypothesis is that there is no significant cumulative serial correlation up to i-th lag. The results of this test for both frequencies are presented in the tables below:

Table 2: Ljung-Box Test for Daily Data

	Returns	**Returns Squared**
Lags	P-Value	P-Value
1	0.039	0.000
2	0.100	0.000
3	0.094	0.000
4	0.106	0.000
5	0.177	0.000
6	0.112	0.000
7	0.003	0.000
8	0.000	0.000
9	0.000	0.000
10	0.000	0.000

Table 3: Ljung-Box Test for Weekly Data

	Returns	**Returns Squared**
Lags	P-Value	P-Value
1	0.049	0.000
2	0.004	0.000
3	0.001	0.000
4	0.002	0.000
5	0.003	0.000

In Table 2 it can be seen that for daily returns, up to the 6^{th} lag, there is no significant cumulative serial correlation. However, from the 7^{th} lag onward, the cumulative correlation turns significant, which is unexpected by observing Figure 6. The same result is observed in Table 3 when analyzing weekly returns. This unforeseen outcome could mean that there probably is one autocorrelation that is driving the test over the critical values, although it does not seem to be the case from seeing the aforementioned Figures. Thus, it makes us question if the assumption of weak stationarity made to implement the test is actually valid, because if it is not then the conclusions of the test are flawed. The latter explanation seems more likely and it is analyzed in the following subsections.

[5] 95% of confidence is the standard.
[6] This is a strong assumption that may or may not be discarded afterwards.

2.3 Tail Indexes

Financial data often have particular characteristics that make working with them more challenging than with other types of data. One feature that has been debated since the sixties is the presence of *leptokurtic* distribution for returns. This means, as seen in Figures 4 and 5, that we could be in presence of heavy tails and have a Kurtosis higher than 3, which is the expected value if returns were normally distributed. In this section we analyze the existence of finite moments for daily and weekly data mainly by computing the tail index of a Pareto distribution with our respect to our returns.

2.3.1 Daily Frequency

To begin with, the first and simplest approach is to observe the sample Kurtosis, that for daily returns is 12.66. Does this mean that, albeit high, we have finite 4^{th} moments? Not really, since in finite sample we will always have finite calculations (they basically would tend to infinite). Additionally, another approach could be using the Jarque-Bera test to see if the distribution has similar 3^{rd} and 4^{th} moments to those of a normal distribution (Jarque and Bera, 1980), nonetheless it would be incorrect because this test is based on the assumption that data is independent and identically distributed; which from Figures 8 and 9 we see that obviously is not the case. Indeed, these two graphs show that the squared return autocorelations possess long-memory features.

However, there exists other approaches to assess the existence of finite moments in our data. In the best case scenario we would have finite 4^{th} moments that imply an autoregressive coefficient smaller than 0.58 in a GARCH-type model and, because of the fewer extreme events, the fitted values would be more precise.

One of these approaches is the estimation of what is called the tail index. This come from the assumption that data in the tails behaves as a Pareto distribution as follows:

$$P(X \geq x) = Cx^{-\theta} \tag{5}$$

Where X is a random variable (in our case the returns), C a given constant and θ is the tail index. For this type of distribution, the θ controls "how fast the function decreases" and it relates to the amount of finite moments the overall distribution has.

The first estimator we have used for the tail index is the the one proposed by Hill in 1975. In his paper, Hill states that the tail index could be estimated the following way:

$$\widehat{\theta_m^{right}} = \left[(m-1)^{-1} \sum_{k=1}^{m-1} \ln \left(\frac{X_{(T-k)}}{X_{(T-m)}} \right) \right]^{-1} \tag{6}$$

Where in our case, T is the sample size and m is the number of observations considered in the tails. $X_{(T-m)}$ refers to the ordered statistic[7], meaning that $X_{(T-m)}$ is the m-highest return of the sample and that we are on the right tail. The same computations are carried on the other tail, considering the respective values to obtain the Hill estimator for the left index.

Using this formula we considered up to the 5% of the observations in each tail[8], meaning that m ranges from 1 up to 653. The graphs of the Hill estimator for each tail are presented in Figures 11, for daily data, and 12, for weekly data.

[7] Ordered returns, from the smallest to the highest.

[8] We considered that, on each side, after the 5% level we are not considering the tail regions anymore and, therefore, the assumption of having a Pareto distribution does not hold anymore. However, also a larger number of observations on each tail have been considered and the conclusions do not change, even if the Hill estimator keeps falling.

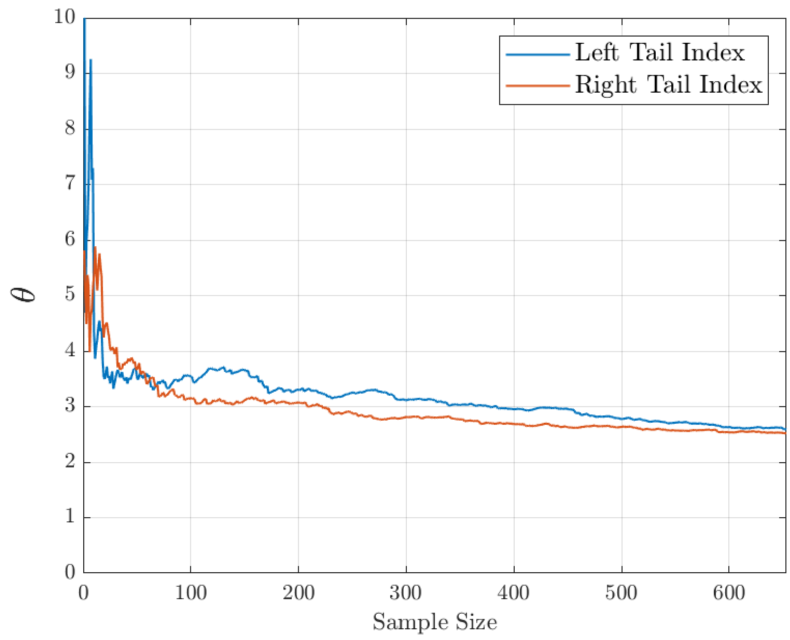

Figure 11: Hill Estimator for the Tail Index of Daily Returns

The Hill estimator does not carry any formal implication, but it can be useful to make an educated guess on the tail index. The estimator displays the usual exponentially decreasing behaviour for both tails, with a progressive convergence to a value of $\hat{\alpha} \approx 2.5$, meaning that the series has finite second moments but infinite third moments, however it keeps decreasing as the sample considered increases. This result relates to the issue of choosing the correct m and why the estimated tail index is just an approximation. Additionally, we know that to obtain a reasonably accurate estimation of the tail index using Hill's estimator we would need around 10,000 observations on the tail (Németh and Zempléni, 2020), which we do not have. Consequently, a better and simpler approach is to use a "rank-size" regression where, assuming that the data behaves as a Pareto distribution on the tails, we have that empirical distribution function will be as follows:

$$\hat{P}(X \geq X_{(T-m)}) = \left(\frac{m+1}{T}\right) = CX_{(T-m)}^{-\theta} \tag{7}$$

Equation (7) suggests that the estimated probability $\hat{P}$ of the returns being higher or equal than the m-highest return is equal to the relative frequency of all the returns higher to this threshold. Linearizing equation (7) we obtain:

$$\ln\left(\frac{m+1}{T}\right) = \hat{\beta}_0 + \hat{\beta}(\ln X_{t-m}) \tag{8}$$

Given the expression in (7), by estimating a regression on each tail using m observations and setting $\hat{\beta} = -\hat{\theta}$, it is possible to obtain an estimate of the tail index. We used this procedure for different sample sizes, up to 5% of the observations, on each tail, the results of the regressions and estimation of the tail indexes are presented in Table 4 below.

Table 4: Point Estimates of the Tail Index for Daily Data

Quantile Left	Observations	θ_{left}	θ_{right}	Observations	Quantile Right
0.001	13	4.53	4.18	13	0.999
0.002	26	3.80	4.25	26	0.998
0.003	39	3.59	3.97	39	0.997
0.004	52	3.53	3.86	52	0.996
0.005	65	3.50	3.74	65	0.995
0.006	78	3.46	3.60	78	0.994
0.007	91	3.44	3.51	91	0.993
0.008	104	3.44	3.44	104	0.992
0.009	118	3.44	3.37	118	0.991
0.01	131	3.46	3.33	131	0.99
0.02	261	3.37	3.12	261	0.98
0.03	392	3.25	2.97	392	0.97
0.04	522	3.13	2.86	522	0.96
0.05	653	3.00	2.79	653	0.95

From the specular Table above we observe that this method yields higher estimates than the ones using Hill's method. Yet, when accounting up to the 3% of the observations, the right tail index estimator goes below 3, again confirming the result of finite second moments but infinite third (and higher) moments in the daily returns distribution.

2.3.2 Weekly Frequency

For the weekly returns the same computations and reasoning are conducted within this section . First, the sample Kurtosis is 11.81, suggesting the presence of heavy-tails, a most probable result given what we analyzed in the daily returns case. Secondly, the Hill estimator for the tail indexes of a Pareto distribution are presented in Figure 12. For this series, 5% of the tail is equal to 135 observations, as can be seen on the graph.

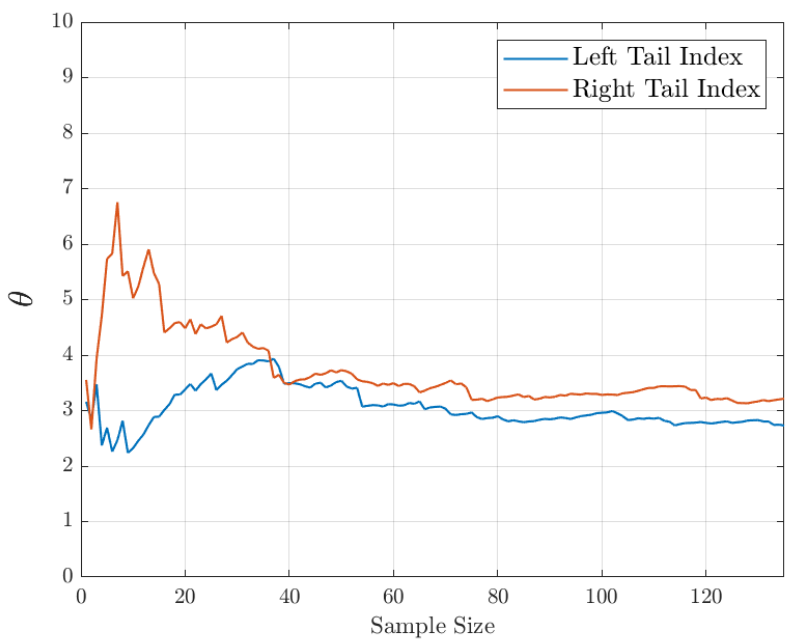

Figure 12: Hill Estimator for the Tail Index of Weekly Returns

From the graph one can observe that for weekly returns we have similar result as for daily returns, the Hill estimators tends to be slightly below 3 for the left tail and a little above for the right tail. These results are in line with the previous ones presented for daily data.

Similarly, the table of the results of the "rank-size" regression are presented in Table 5.

Table 5: Point Estimates of the Tail Index for Weekly Data

Quantile Left	Observations	θ_{left}	θ_{right}	Observations	Quantile Right
0.001	3	2.15	2.14	3	0.999
0.002	5	2.17	2.63	5	0.998
0.003	8	2.12	3.24	8	0.997
0.004	11	2.12	3.59	11	0.996
0.005	14	2.17	3.86	14	0.995
0.006	16	2.22	3.98	16	0.994
0.007	19	2.31	4.01	19	0.993
0.008	22	2.40	4.05	22	0.992
0.009	24	2.45	4.06	24	0.991
0.01	27	2.52	4.09	27	0.99
0.02	54	2.90	3.80	54	0.98
0.03	81	2.91	3.57	81	0.97
0.04	108	2.88	3.45	108	0.96
0.05	135	2.84	3.39	135	0.95

From Table 5, one can appreciate that for the left tail of weekly returns, the tail index oscillates between 2 and 3. Instead, for the right tail, it starts small when considering only 0.1% of the observations, it rises up to 4 for then go back again when using 5%.

From this exercise, we have gathered evidence that, most probably, both daily and weekly returns have finite 2^{nd} moments but not 3^{rd} ones. Out of this result, we can infer that by implementing a GARCH-type model it should be weakly stationary but have a sum of coefficients higher than 0.58, because of the absence of 4^{th} moments.

2.4 CUSUM Test (Loretan and Phillips, 1994)

So far we have the following results:

- Presence of a unit root in prices but not in returns, indicating the possibility of covariance stationarity in the returns.
- According to Ljung-Box test, there is significant correlation among the returns (which is unexpected because of Figures 6 and 7) and in squared returns.
- As reported by the tail index estimator, there should be finite second moments but infinite third moments onward in the returns' distribution.

With these outcomes, one could think that returns might be weakly stationarity, implying constant variance throughout time. However, by observing Figures 8 and 9 some correlation in squared returns is expected, which are related to the returns' volatility. This means that there must be some explanatory power of past volatility to present volatility and, logically, it cannot be constant as expected from a covariance-stationary process. To test this hypothesis, namely constant variance, we base the following exercise on Pagan (1996) and mainly on Loretan and Phillips (1994) papers, where they implement the CUSUM (cumulative sum) test to check for weak stationarity in heavy tailed time-series[9].

To begin with, as stated on the paper, the CUSUM test is based on the following statistic:

$$\psi_T = (T\hat{v}^2)^{-1/2} \sum_{t=1}^{[Tr]} (X_t^2 - \hat{\mu}_2) \tag{9}$$

Where T is the sample size, $\hat{\mu}^2$ is the estimator for the unconditional variance (which we know it must be finite because of the tail index analysis), X_t are the returns and $\hat{v}^2$ is "a kernel-type estimator of the long-run 4^{th} moment" defined in equation (10). In Loretan and Phillips' paper, they make their analysis with continuous time and say that the statistic with finite fourth moments converges to a tied-down Brownian motion[10]. Since it is not our case, because the tail index is between 2 and 3, we still know from the paper that the limit distribution of the CUSUM statistic is a tied-down stable-Lévy bridge and the critical values can still be searched for[11]. One last consideration is that for infinite fourth moments the convergence in distribution exists over the residuals of an autoregressive process and not over the series that is missing the finiteness of the moments, because of this, we calculated a simple AR(1) model and used its residuals as done in the paper.

$$\hat{v}^2 = \hat{\gamma}_0 + 2\sum_{j=1}^{L} \frac{\hat{\gamma}_j (L - j + 1)}{L + 1} \tag{10}$$

In equation (10), $\hat{\gamma}_j$ is the j-th serial covariance of the returns and L a "suitable lag truncation number"[12].

With this, we can plot our CUSUM statistic for daily and weekly returns with the critical values presented in the paper for a tail index of 2.5.

[9] Curiosly, this is almost the exact name of the paper. The authors use 3 different approaches but we used only the CUSUM test which we also saw in class.

[10] Tied-down because the process is pinned down and $t = 0$ and $t = T$

[11] The critical values are reported in the paper (Table 2b), they did it with 50,000 simulations for a sample size of 1,000.

[12] We tried with different lags, e.g. 10, 50, 100, 1000, 3000, all yielding the same conclusion. For daily data we used 3000 lags and for weekly data 500.

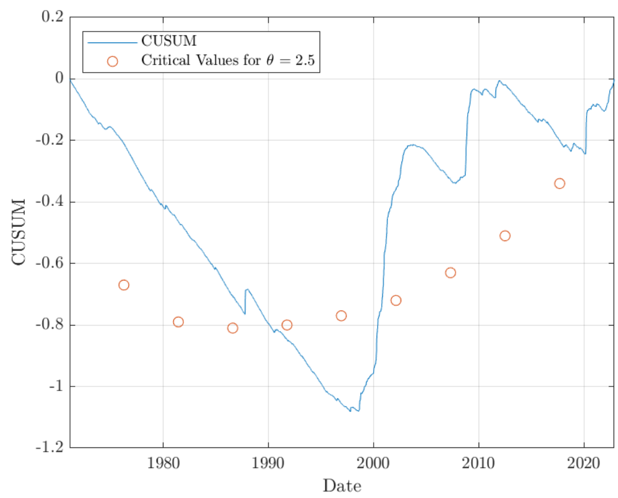

Figure 13: CUSUM Test for Daily Returns

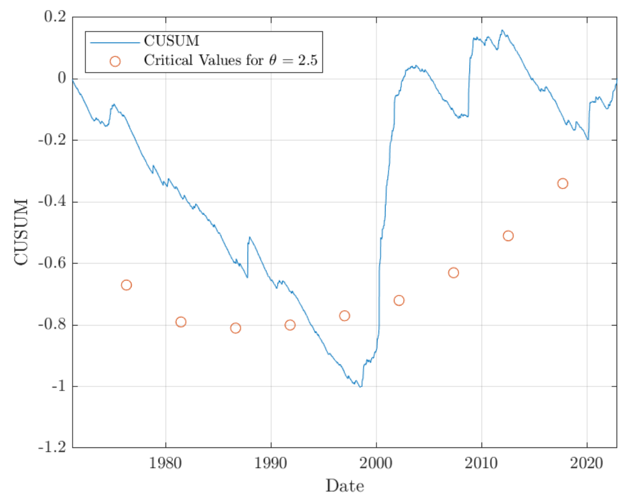

Figure 14: CUSUM Test for Weekly Returns

From Figures 13 and 14 we can see that the CUSUM statistic surpasses the critical values, hence the test is rejected. it follows that there are two possibilities for the variance of the process, either it is infinite or it is time-varying, however we already know because of the tail index analysis that variance is finite, so we can conclude that most probably[13] the variance is time-varying, delivering, together with Figures 8 and 9, the perfect motivation to compute a GARCH-type model to find an adequate representation of the volatility and try to predict it. This will also might help to explain the volatility clustering presented in the graph of returns, where are clear periods of high and low volatility.

3 Conditional Heteroskedasticty Model

3.1 Introduction

For this section we are required to work with the initial two thirds of our data, spanning from the 5^{th} of February of 1971 to the 1^{st} of August of 2005.

Starting from the basic Generalized Autorregressive Conditional Heteroskedasticity (GARCH) model we know that it is defined by the following 2 equations:

$$\begin{cases} X_t = \sigma_t z_t \\ \sigma_t^2 = \omega + \alpha x_{t-1}^2 + \beta \sigma_{t-1}^2 \end{cases} \tag{11}$$

Where, in our case, X_t corresponds to the time series of the NASDAQ returns and σ_t^2 to its variance, which changes in time, last, z_t usually is assumed to be a random variable with a standard normal distribution. The idea of implementing this model is that, as seen in Figure 3, the volatility of the series changes over time and it is related to its past values, as seen in Figures 8 and 9. The next 2 subsections present the results of the exercise for daily and weekly data.

3.2 Daily Data

3.2.1 Model Selection, Misspecification and Forecasts

To open with, there are 2 key questions to analyze so as to correctly choose model, the first one is which type of GARCH model to use (e.g. GARCH, EGARCH, GJR)[14] and the second one concerns the lag selection for the chosen model.

From Figure 3 we can observe that there might an asymmetric reaction to returns, that suggests to use a GJR or EGARCH model, as discussed by Black in his paper of 1976. In order to choose correctly, we performed all the combinations up to 5 lags (i.e. 25 models) for each of the 3 types of models[15] and we considered the Bayesian Information Criteria (BIC) for each regression, the smallest the better. Using the BIC we can discern which model is the best among its type and it also allows us to choose between different models. The candidate models of these 3 groups are:

[13] With 95% confidence.
[14] There are more models but this ones are built-in in MATLAB.
[15] The asymmetric models are estimated with only one leverage coefficient.

- GARCH(1,1)
- GJR(1,1)
- EGARCH(1,1)

Among these three, the one with the smallest BIC is the GJR(1,1) with one leverage coefficient. The model specification is the following one:

$$\begin{cases} X_t = \sigma_t z_t \\ \sigma_t^2 = \omega + \gamma\sigma_{t-1}^2 + \alpha X_{t-1}^2 + \xi I\left[X_{t-1} < 0\right] X_{t-1}^2 \end{cases} \tag{12}$$

Where X_t is our series of returns, γ is the GARCH coefficient (corresponding to past volatility values), α is the ARCH coefficient (corresponding to past values of squared returns) and ξ is the leverage coefficient that accounts for the asymmetry of the response of volatility to negative returns, confirming the leverage effect mentioned before.

The estimation through maximum likelihood of the model yields the following output[16]:

Table 6: GJR(1,1) Regression Output

	Value	Standard Error	T-Statistic	P-Value
Constant	0.00	0.00	1.49	0.14
GARCH	0.94	0.00	257.08	0.00
ARCH	0.05	0.01	8.58	0.00
Leverage	0.01	0.01	1.78	0.08

From the Table above we can see, assuming inference is correctly specified, that the constant is not significant an close to zero while the leverage coefficient is significant at a 10% level. Additionally, the latter has a very small coefficient compared to the other two estimators, implying that the asymmetry does not have a big impact. The following graphs illustrate comparisons among the model's output and real values. Figure 15 plots daily returns for the first two third of our dataset compared to volatility fitted values ((σ_t)) estimated by the model, whereas figure 16 graphs the stock returns for the remaining part of the sample against the Out-of-sample forecasts of the variance using a one step-ahead method. Figure 17 and 19 substantially show similar content except that the computations refer to variance (σ_t^2). Instead, figure 19 and 20 report,respectively for volatility and variance, the forecasts obtained using a five step-ahead method.

As can be seen from Figures 15 to 20, the GJR(1,1) model tracks well the volatility and variance of the returns, for the fitted values as well as in the forecasts' case. nevertheless, whereas it does not forecast accurately shocks as it is often the case with forecats, it unexpectedly fails to capture the extreme events after they occured. However, when comparing the volatility of the one-step out-of-sample forecast (Figure 16) with the 5-step ahead out-of-sample forecast (Figure 19) it looks like the latter does a much better job when predicting the volatility even in more extreme events (like the one of 2020).

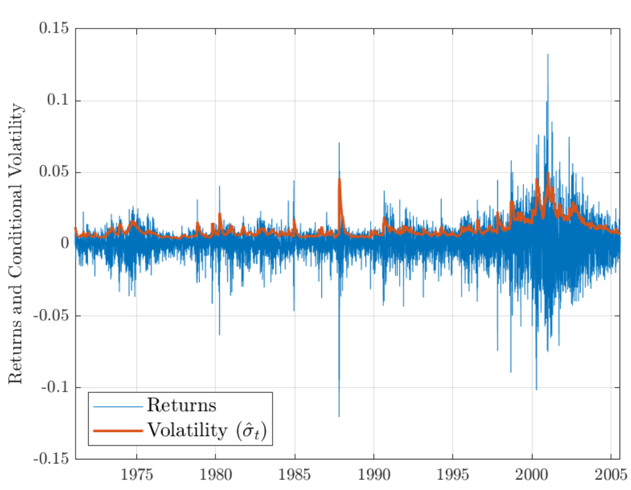

Figure 15: Daily Returns and $\hat{\sigma}_t$ In-Sample

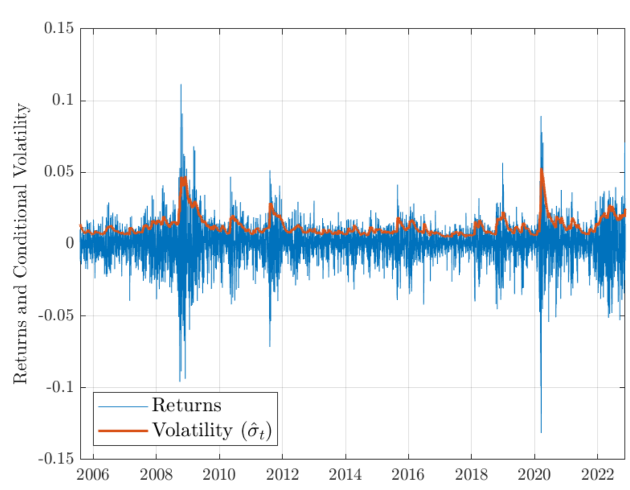

Figure 16: Daily Returns and $\hat{\sigma}_t$ Off-Sample

[16]The estimation is done over a t-student distribution, which has fatter tails than a normal distribution, thus it is closer to our case for returns.

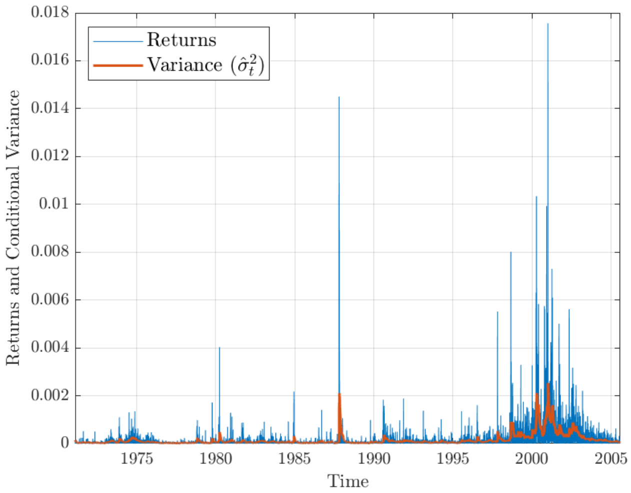

Figure 17: Daily Returns and $\hat{\sigma}_t^2$ In-Sample

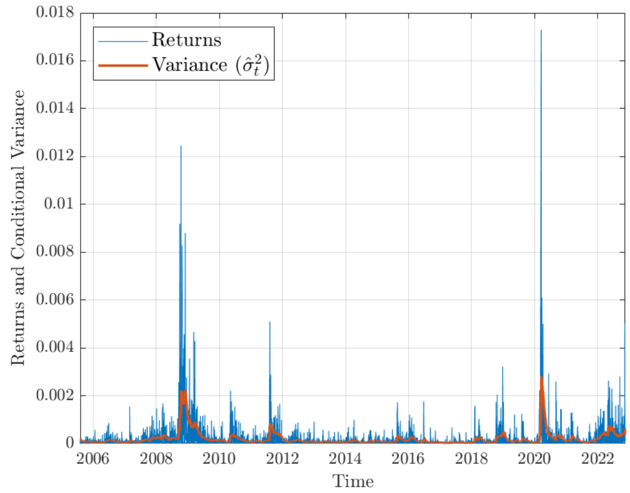

Figure 18: Daily Returns and $\hat{\sigma}_t^2$ Off-Sample

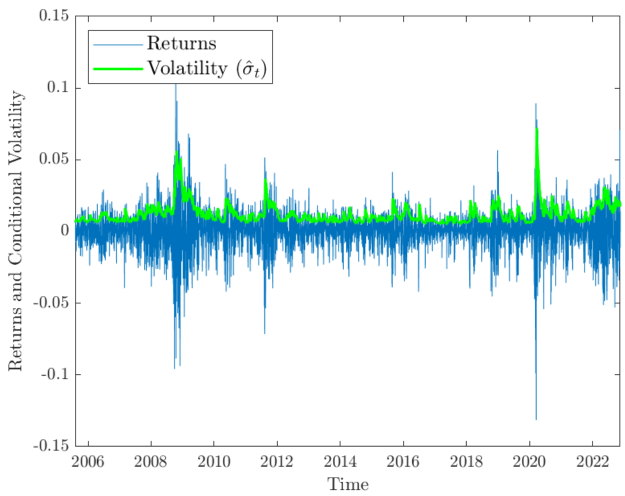

Figure 19: Daily Returns and $\hat{\sigma}_{t+5}$ Off-Sample

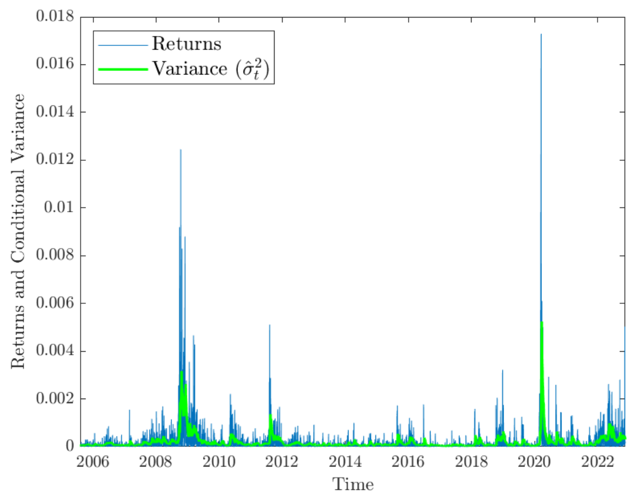

Figure 20: Daily Returns and $\hat{\sigma}_{t+5}^2$ Off-Sample

However, the facets of this model do not end here, indeed there are two additional details worth mentioning. The first one is that although the fitted values estimated by the GJR(1,1) model are the best among the combinations tried, when analyzing the fit of the out-of-sample forecast to the data, the EGARCH(1,1) perfomance is better.

The second detail concerns the misspecification analysis. For this GJR(1,1) the standardized residuals, coming from equation (11), are defined as follows:

$$\hat{z}_t = \frac{X_t}{\hat{\sigma}_t} \tag{13}$$

Where we have used the estimated conditional volatility obtained from the model. The graph of the residuals on the first two third of the data is plotted below along its histogram:

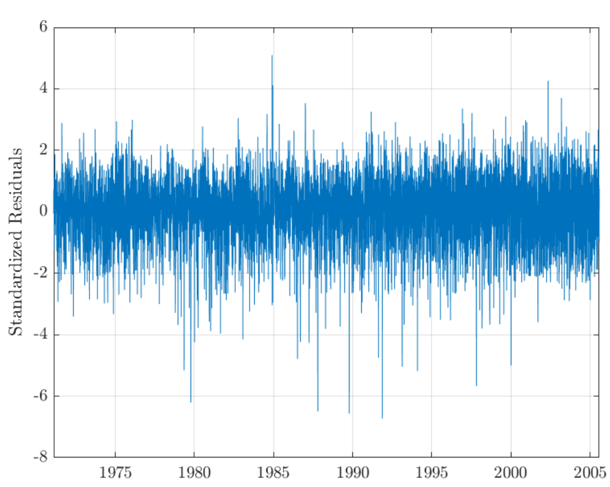

Figure 21: Daily Standardized Residuals

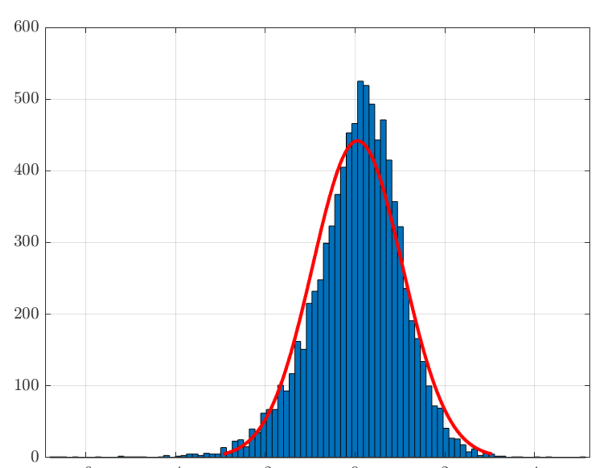

Figure 22: Daily Standardized Residuals Histogram.

As can be seen from the figures above, although the model captures part of the volatility correctly, it still misses many of the extreme events, especially the negative ones. Then, as expected, the Jarque-Bera test is rejected, so it is safe to conclude that the series does not have or symmetry of a normal distribution, a result further supported by figure 22.

Furthermore, to check for misspecification of the model, one of the assumptions made over the errors is that they are independent and identically distributed. To evaluate if this assumption holds in our model, we can observe the sample correlogram:

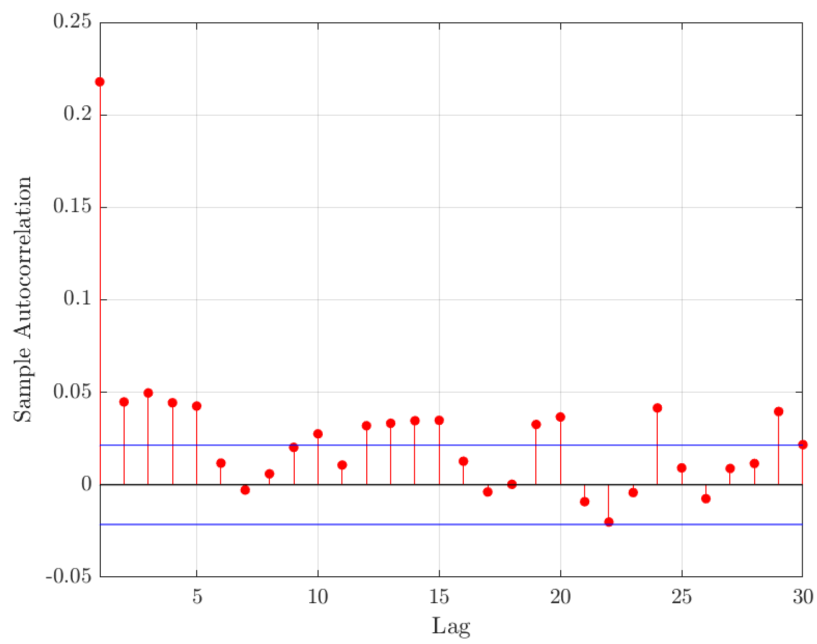

Figure 23: Daily Standardized Residuals Correlogram

In Figure 23 one can observe that, up to the thirtieth lag, most of the times we have significant serial correlation among the residuals. This is also reflected into the Ljung-Box test, which is rejected, confirming the presence of significant serial correlation at any lag tested.

Summing up, although the GJR(1,1) model for daily data should be the best according to the information criterion, it could use some improvements as it fails multiple tests ran within the misspecification analysis. A better model might be out there and to find it there exists different methods from the information criteria, for example out-of-sample mean squared errors, however our analysis is based mainly on information criteria.

3.3 Weekly Data

3.3.1 Model Selection, Misspecification and Forecasts

By performing the same analysis presented in the previous subsection on weekly data, our candidate models, chosen through the minimum BIC, for each one of the three types are equal as before[17]:

- GARCH(1,1)
- GJR(1,1)
- EGARCH(1,1)

However among these 3, the one that has the minimum BIC is the GARCH(1,1) and not the GJR(1,1) as happened in the daily frequency. The model specification is the one of Equation (11). The estimation output is the following:

Table 7: GARCH(1,1) Regression Output

	Value	Standard Error	T-Statistic	P-Value
Constant	0.00	0.00	3.21	0.00
GARCH	0.87	0.02	40.27	0.00
ARCH	0.11	0.02	5.89	0.00

[17]We have used again only one leverage coefficient for the asymmetric models.

As observed on Table 7 above, if inference is correctly computed, all parameters are significant at the 1% level, although the value of the constant is practically zero. The graphs for the in-sample and off-sample volatility(σ_t) and variance (σ_t^2) fitted values are presented in the Figures below:

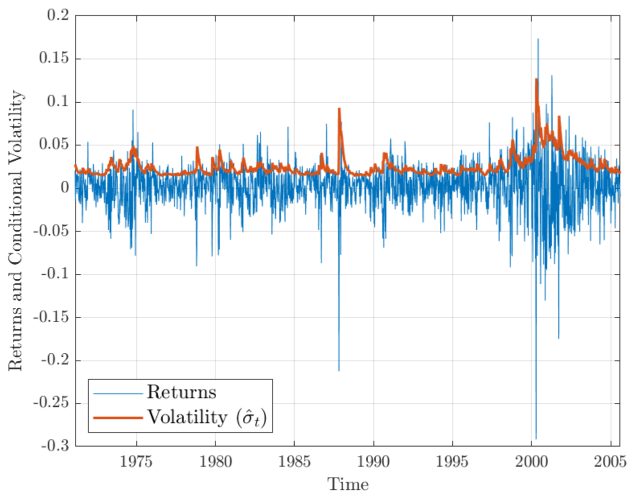

Figure 24: Weekly Returns and $\hat{\sigma}_t$ In-Sample

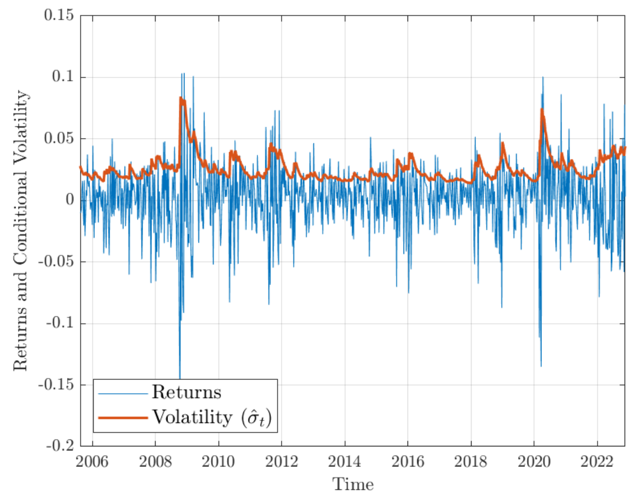

Figure 25: Weekly Returns and $\hat{\sigma}_t$ Off-Sample

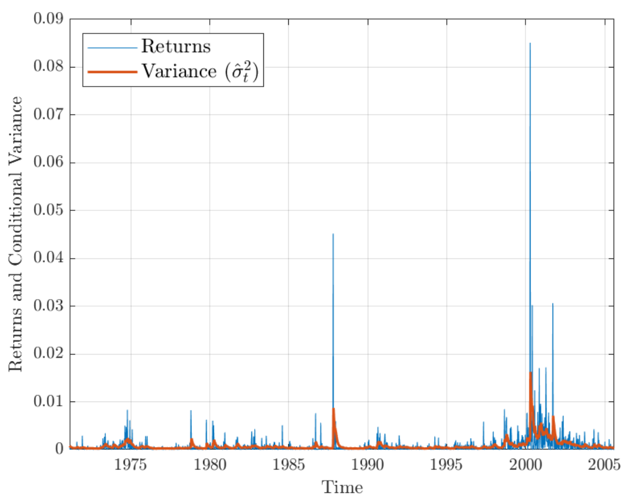

Figure 26: Weekly Returns and $\hat{\sigma}_t^2$ In-Sample

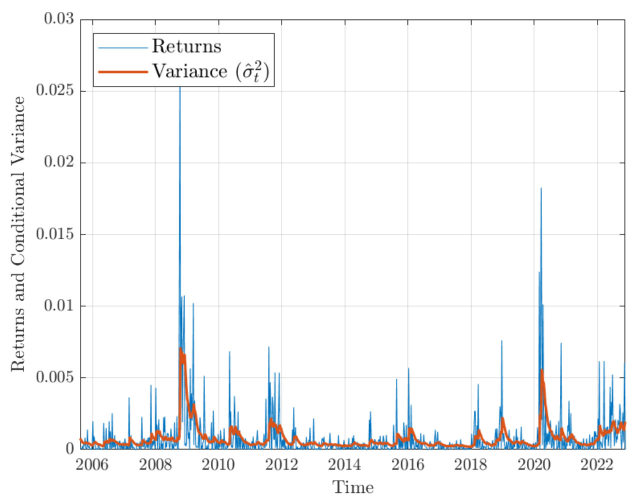

Figure 27: Weekly Returns and $\hat{\sigma}_t^2$ Off-Sample

As can bee seen from the Figures above, the GARCH(1,1) models tracks fairly well the variances, even in the case of more extreme events (first two graphs), differently from the case of daily data. However, to be certain we now conduct the misspecification analysis. The standardized residuals and their histogram are graphed below:

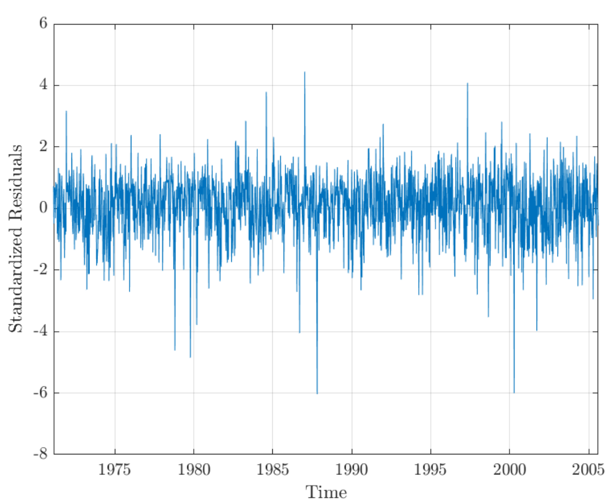

Figure 28: Weekly Standardized Residuals

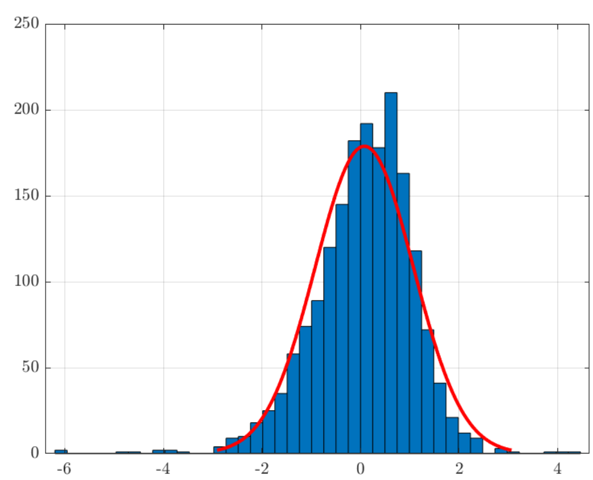

Figure 29: Weekly Standardized Residuals Histogram

From Figures 28 and 29, we see that this model better captures the volatility, although there are still some extreme events where the model does not perform well. Similarly as before, both the Jarque-Bera and Ljung-

Box tests are rejected, confirming that the distribution of Figure 29 is not approximately normal and that there is serial correlation. The sample correlogram is reported in Figure 30:

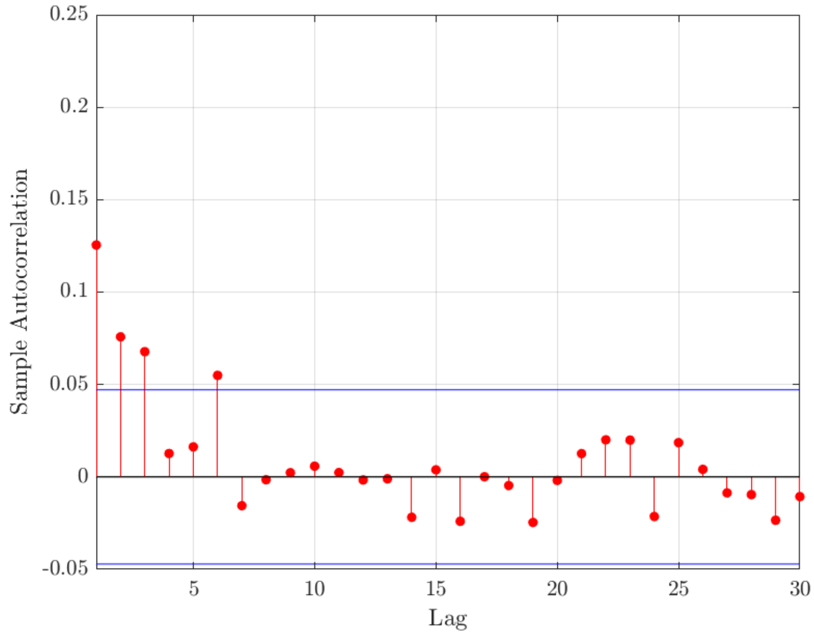

Figure 30: Daily Standardized Residuals Correlogram

We see from the sample correlogram that for weekly data most of the lags are inside the confidence bands and the overall picture is significantly more accurate than the one in Figure 23 (for daily data). However, there is still significant serial correlation, especially in the first lags.

So, the same way as happened with daily data, a method for model selection different from information criteria could be used to select a different model, however, for time-contraint reasons we were not able to do a full analysis of it. Nevertheless, a small peak into what could be done is to select the model that has the smallest Mean Squared Error out-of-sample. This is, training the model in the first 2 thirds and check how it behaves on the remaining part of the data.

4 Value at Risk

In the following section we are required to explore different methods for computing the value at risk (VaR) at the 5% and 1% level and use the calculations to observe the number of exceedances of the returns. This is done for daily data (12,580 observations) and weekly data (2,603), although they yield similar results.

The VaR is a risk measure that helps to quantify the potential losses of investing in the NASDAQ (in our case). The idea is the following: the VaR is the quantile value that, given a distribution, is at the 5% (or 1%) of the cumulative density function. An easy way of calculating the value of the VaR is the one called "variance-covariance method" that is[18]:

$$VaR_t^{95\%} = \alpha \sigma_t \tag{14}$$

Where α is the quantile corresponding to the 5% of a distribution and σ_t the standard deviation of said distribution.

The calculation, evidently, depends on the distribution that is assumed, whence both components to calculate the VaR are obtained. We have used 4 methods to assess the risk of the NASDAQ:

1. The Normal Distribution method;
2. The GJR(1,1) estimates for conditional variance;
3. The Historical Simulation method;
4. Exponential Weighted Moving Average Method (EWMA).

When using Normal Distribution method, we compute the VaR by multiplying the standard deviation of the returns by the z-score at the 0.01 and 0.05 confidence levels. The standard deviation is defined retrospectively over a period of 250 trading days.

The GJR(1,1) estimates use still the z-score of the normal distribution but it uses the conditional standard deviation estimated by the daily model.

Differently, the Historical Simulation method adopts the quantiles of the distribution of past returns (thus avoiding parametric assumptions) and implies that past returns may be informative of future ones and does not use the σ_t parameter since the VaR is already in the desired scale (the scale of the returns).

The EWMA approach is a little bit more complicated, and its main idea is that it assumes that recent returns exert a greater influence on the current returns. This is imposed by applying exponentially decreasing weights. The EWMA variance $\widehat{\sigma}_t^2$ over T periods is defined as follows[19]:

$$\widehat{\sigma}_t^2 = \frac{1}{c} \sum_{i=1}^{T} \lambda^{i-1} y_{t-i}^2 \tag{15}$$

With

$$c = \sum_{i=1}^{W_E} \lambda^{i-1} = \frac{1-\lambda^{W_E}}{1-\lambda} \quad \rightarrow \frac{1}{1-\lambda} \text{ as } T \rightarrow \infty \tag{16}$$

$$\widehat{\sigma}_t^2 \approx (1-\lambda) \left(y_{t-1}^2 + \sum_{i=2}^{\infty} \lambda^{i-1} y_{t-i}^2 \right) = (1-\lambda) y_{t-1}^2 + \lambda \widehat{\sigma}_{t-1}^2 \tag{17}$$

For this case, 5% and 1% of 12,580 is 629 and 126 respectively. As can be observed in Table 8, the number of exceedances (the times that the returns surpassed the VaR calculation) differs considerably depending on the method used. However, the Historical Simulation and the EWMA approaches show rather satisfying results in terms of exceedances. For the former, the VaR(5%) is surpassed in 5.57% of the times and the VaR(1%) 1.41%, while for the latter the first VaR is quite precise and it is exceeded 5.22% and the other 1.44%.

Instead, for the other 2 methods, the VaR computed on GJR[20] estimates and with the Normal distribution method is overly optimistic and characterized by a far greater number of exceedances, especially at 1%, with

[18]This is not what is used for the Historical Simulation method. For this one, no rescaling coefficient (σ) is needed.

[19]We conveniently assumed an infinitely large time window to estimate $\widehat{\sigma}_t^2$.

[20]However, we punctuate that the results from the GJR improve remarkably when the model is estimated on the whole sample and not only trained in the initial two thirds.

the Normal Distribution method reaching more than 2.3%.

Table 8: Value-at-Risk Exceedances

VaR Method:	# of Exceedances at 5%	# of Exceedances at 1%
Normal Distribution	723	302
GJR Estimates	714	250
Historical Simulation	701	177
EWMA	657	181

Furthermore, we illustrate how the different estimations of the VaR(5%) behaved during the year 2000, in which we observed record volatility concurrently with the dot-com bubble and the close relation of the enterprises that constitute the NASDAQ index. We decided to omit the results from the Normal distribution method for clarity, as they are the least precise and reliable.

The following graph shows that the VaR_{GJR} hovers the VaR_{EWMA}, whereas the $\text{VaR}_{Hist.}$ underestimates the potential losses and it is much more constant even though the graph is showing a period of high volatility. The points marked are where one of the VaR(5%) calculations is surpassed by the actual return series. We hypothesize that the Historical Simulation associates a reduced mass of probability to these catastrophic events, as the returns in the past were substantially less volatile thus, past behaviour of returns are not well-suited for the VaR in this context, although its good performance overall as seen in Table 8.

While this could constitute a drawback, the historical simulation does strikingly well at the 99% VaR.

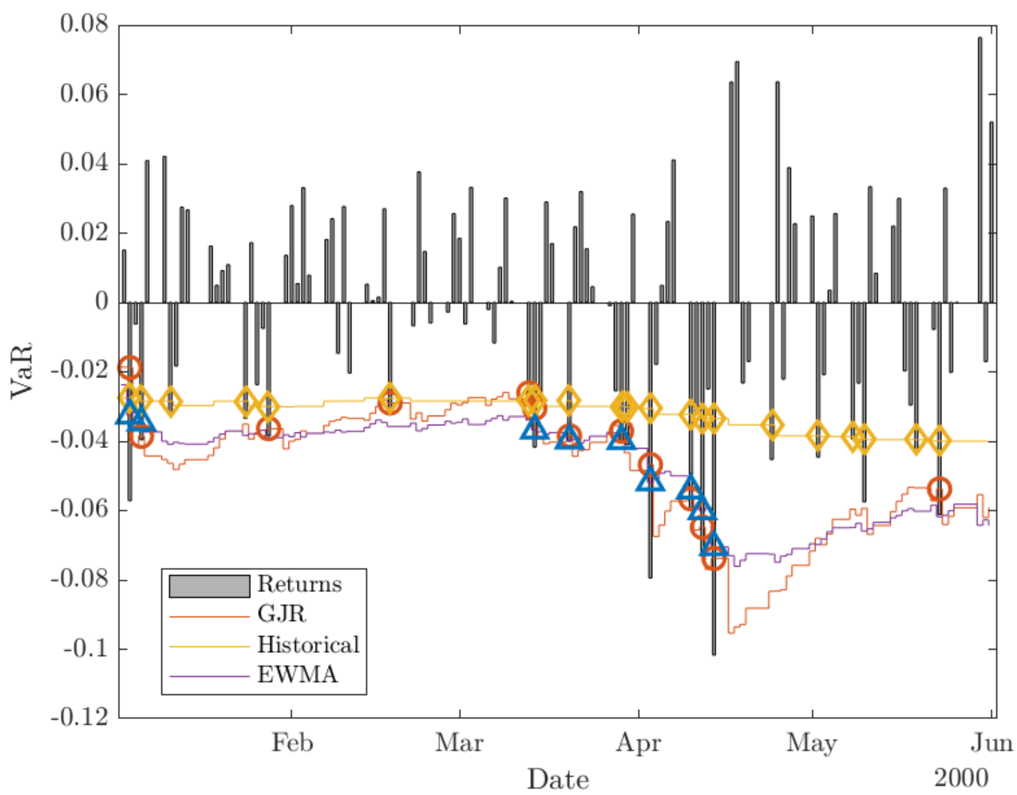

Figure 31: VaR(5%) Violations for Different Models

With respect to weekly data the results are very similar (this time we have 2,603 weekly data points) and the number of exceedances are presented in Table 9.

Table 9: Value-at-Risk Exceedances

VaR Method:	# of Exceedances at 5%	# of Exceedances at 1%
Normal Distribution	145	69
GARCH Estimates	120	45
Historical Simulation	164	66
EWMA	116	36

For this case, 5% and 1% of 2,603 are 130 and 26 respectively. As seen in Table 9, the GARCH estimates is the best VaR of 5% although it could be too conservative (it is surpassed 4.61% of the times). However, for the 1%

VaR the EWMA performs better, being breached 1.38% of the times.

What is important to notice from this exercise is that the EWMA performs actually quite well both for daily and weekly data. This is interesting, because the EWMA is not a model-based measure of the VaR, as done with the GJR and GARCH, so no statistical model underlies and using it instead of building a case for a model might be easier and faster.

In this case, to have an analogous graph to Figure 31, it is a zoom of the year 2000 during the dot-com crisis.

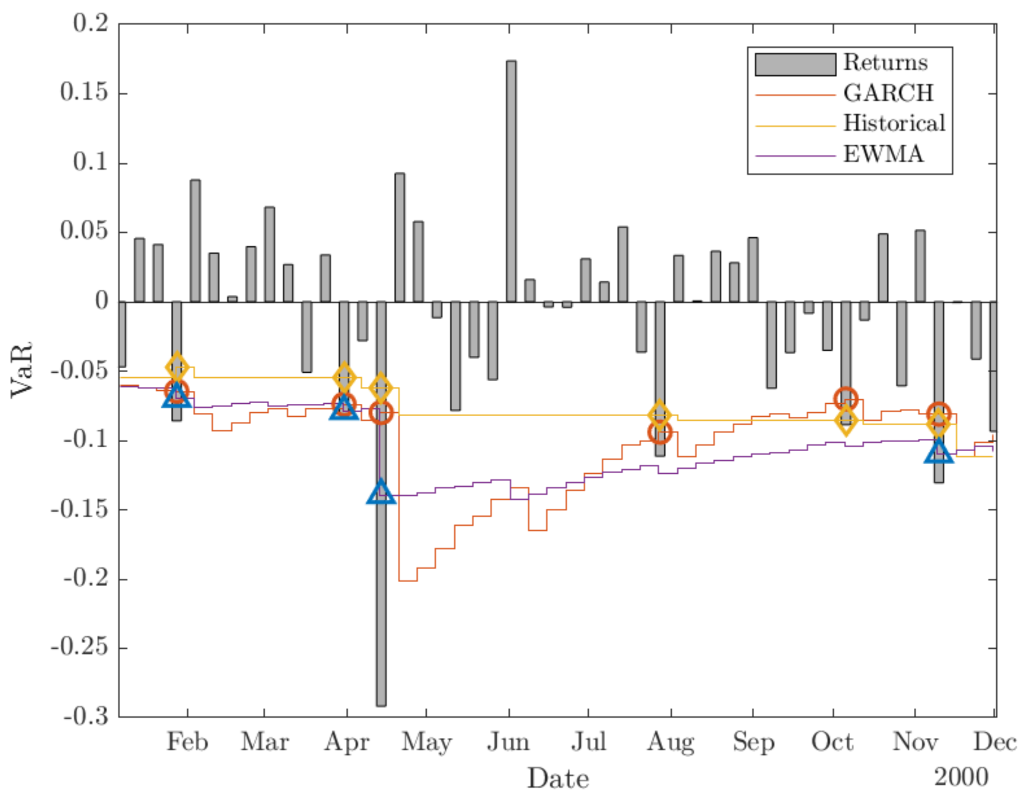

Figure 32: VaR(5%) Violations for Different Models

We see that for this case the EWMA is more conservative, for example, nor in August nor in October it was surpassed by the weekly returns while both of the other models in the graph were.

One last comment about this exercise of the Value-at-Risk is that the model-based one, as its name implies, depends on the goodness-of-fit of the model implemented. As presented before in the misspecification analysis, the GJR(1,1) might not be the most adequate model and it performed worse than the weekly returns to predict volatility. That's probably why in the daily VaR the Historical Simulation yields a better performance than the GJR. On the other hand, in the weekly VaR the GARCH-based one performs much better and makes an improvement over the Historical Simulation. Nevertheless, the EWMA tends to be the method that is accurate and consistent across frequencies (daily and weekly) as, probably, the best VaR measure.

5 Bootstrap

In this part we are required to work with the ARCH(1) model defined as:

$$\begin{cases} X_t = \sigma_t z_t \\ \sigma_t^2 = \omega + \alpha x_{t-1}^2 \\ z_t \sim iid(0,1) \end{cases} \tag{18}$$

Where the parameter space is $\omega \geq 0$ and $\alpha \geq 0$ with true values $\omega_0 > 0$ and $\alpha_0 \geq 0$.

The idea is to design a bootstrap procedure to test through a likelihood ratio test that $\alpha = \bar{\alpha}$:

$$LR_n = -2(l_n^{(0)} - l_n^{(1)}) \tag{19}$$

Where $l_n^{(0)}$ is the log-likelihood with the null imposed ($\alpha = \bar{\alpha}$) and $l_n^{(1)}$ is the log-likelihood under the alternative hypothesis.

Since this is a time series framework, and especially a conditional heteroskedasticity model, the chosen bootstrap procedure must preserve the time dependence of the series and must also maintain the heteroskedasticity we observe in the original sample, since it is the object interested in being modelled. Our group discussed thoroughly and searched for insights and we concluded that some of these options are available, Pairs Bootstrap (Freedman, 1981), the Wild Bootstrap (Liu, 1988) and the Fixed-Volatility bootstrap (Cavaliere, Pedersen and Rahbek, 2018). However, since we are not working with cross-sections and we don't have any assumptions over the stationarity of the process, we should discard the Pairs Bootstrap, that also in it is construction is not imposing the null hypothesis, thus remain the last 2. Generally speaking, the Fixed-Volatility is a special case of a Wild Bootstrap since the bootstrap innovations are drawn from the standardized residuals obtained from the model and the resampling is over the residuals, like the Wild Bootstrap does. But the key point to use the Fixed-Volatility Bootstrap is that since one has the need to test the coefficient of the conditional volatility, having the heteroskedasticity fixed across bootstrap samples may be an improvement.

One last thing to define in the process is if it is Restricted or Unrestricted. As seen in class, if we have an explosive process, thus, a non-stationary ARCH, then the Unrestricted shouldn't be the used. Additionally, the Restricted allows to have better control on the size of the test we are implementing.

1. To compute the LR statistic, in the original sample two models are estimated:
   (a) Estimate $\hat{\omega}^1$ and $\hat{\alpha}^1$ by (Q)MLE and calculate the log-likelihood of this unrestricted model $l_n^{(1)}$.
   (b) Estimate with (Q)MLE a second model but with the null imposed. Here since there is $\alpha = \bar{\alpha}$, the estimation is done only for the other parameter, $\hat{\omega}^2$. Then, calculate the log-likelihood $l_n^{(0)}$.
2. From this original unrestricted model obtain the standardized residuals, defined as:

$$z_t = \frac{X_t}{\widehat{\sigma}_t} \tag{20}$$

3. Now implement the Fixed-Volatility Bootstrap. Resample z_t^*, which are the bootstrap innovations sampled with replacement from the recentered standardized residuals:

$$z_t^* = \frac{\widehat{z}_t - \bar{z_n}}{\sqrt{n^{-1}\sum_{t=1}^{n}(\widehat{z}_t - \bar{z_n})^2}} \tag{21}$$

   Where $\bar{z_n}$ is the sample average of the residuals used to recenter. Then, use this bootstrap residuals to produce the bootstrap sample X_t^*:

$$X_t^* = \widehat{\sigma}_t z_t^* \tag{22}$$

   Where $\widehat{\sigma}_t$ is the conditional volatility from the original sample:

$$\widehat{\sigma}_t^{\,2} = \widehat{\omega} + \bar{\alpha} X_{t-1} \tag{23}$$

   Here is important to notice that the volatility is the one coming from the original sample and not from the bootstrap one. This is what makes it similar to the Fixed-Design Bootstrap seen in class and different from the Recursive Bootstrap. Additionally, differently from what is done in the paper, we propose this special case improvement restricting the autoregressive parameter α by imposing the null hypothesis. This because, we thought that since we don't impose any conditions in the stationarity of the ARCH(1) model, it might be the case that the process is an explosive one and then the Bootstrap procedure fails. Furthermore, this is made so we have full control over the process and we don't loose power when implementing the LR test.

4. To the new bootstrap sample, estimate again the 2 models:
   (a) Estimate $\hat{\omega}^{1*}$ and $\hat{\alpha}^{1*}$ by (Q)MLE and calculate the log-likelihood of this unrestricted model $l_n^{(1)*}$.
   (b) Estimate with (Q)MLE a second model but with the null imposed. Here since there is $\alpha = \bar{\alpha}$, the estimation is done only for the other parameter, $\hat{\omega}^{2*}$. Then, calculate the log-likelihood $l_n^{(0)*}$.
5. Repeat this process for the 999 bootstrap samples.
6. Now, a key feature is that the null is imposed by doing the Restricted Bootstrap, and because of this the bootstrap distribution will be centered around $\bar{\alpha}$ which is the one needed to test the null hypothesis. From this distribution, one can compute quantiles and critical values to evaluate where is the likelihood-ratio statistic of the original sample to reject or not the hypothesis.

Up until now we have discussed a bootstrap procedure to implement the likelihood-ratio test, nevertheless, there are some cases where the bootstrap procedure fails. Some of these cases are autorregressions with unit roots, two-stages least squares with weak instruments and parameters on the boundary of the parameter space. All of these, broadly speaking, make the bootstrap distribution inconsistent.

The case where the true value $\alpha_0 = 0$ is one of the cases mentioned above, because the parameter is right on the boundary. We know that the conditions for the bootstrap distribution consistency are:

- Weak (and uniform) convergence of the underlying statistic when $X_i \sim G_n$ for all distributions G_n in a neighborhood of the true distribution G;
- A continuous mapping from the distribution G_n to the asymptotic of the statistic.

Thus, under these conditions, it holds that for a generic T statistic the bootstrap consistency.

$$sup_{x\in\mathbb{R}}|\widehat{\mathrm{G_n}}(x) - \mathrm{G_n}(x))| \xrightarrow{P} 0 \iff sup_{x\in\mathbb{R}}|\,\mathrm{G_n}(x) - G(x)| \xrightarrow{P} 0 \iff T_n^* \xrightarrow{d^*}_p T_\infty \tag{24}$$

Where

- $\widehat{G_n}(x) = P^*(T_n^* \leq x)$, is the bootstrap cdf.
- $G_n(x) = P(T_n^* \leq x)$, is the sample cdf.
- $G = P(T_\infty \leq x)$ is a continuous cdf.

However, when estimating our $ARCH(1)$ we restrict the parameter space. More precisely, the MLE estimator of α is $\hat{\theta^*} := max\{\hat{\alpha}, 0\}$

$$T_n = n^{1/2}\left(\hat{\theta}_n - \theta\right) \xrightarrow{d} \begin{cases} Z & \text{if } \alpha > 0 \\ \max\{Z, 0\} & \text{if } \alpha = 0 \end{cases} \quad \text{as } n \to \infty, \text{ where } Z \sim N(0,1) \tag{25}$$

We can immediately notice that the asymptotic distribution for T_n (as $n \to \infty$) is not continuous around $\theta = 0$, but rather a stepwise function.

In regard to this, the paper of Andrews (2000) is quite pertinent. We take the bootstrap estimator $\hat{\theta}_n^* :=\max\{\hat{\alpha^*}, 0\}$ and assume $\alpha = 0$. When considering a sequence of n $\{n_k : k \geq 1\}$ such that $\hat{\alpha}_{n_k}(\omega) \leq -c$, with $c > 0$, it follows that:

$$\begin{aligned} T_n^* &= n_k^{1/2}\big(\ \hat{\theta}_{n_k}^* - \hat{\theta}_{n_k}(\omega)\big) \\ &= \max\left\{n_k^{1/2}\ \hat{\alpha}_{n_k}^*, 0\right\} - \max\left\{n_k^{1/2}\hat{\alpha}_{n_k}(\omega), 0\right\} \\ &= \max\{n_k^{1/2}\left(\hat{\alpha}_{n_k}^* - \hat{\alpha}_{n_k}(\omega)\right) + n_k^{1/2}\hat{\alpha}_{n_k}(\omega), 0 - \max\left\{n_k^{1/2}\hat{\alpha}_{n_k}(\omega), 0\right\} \\ &\leq \max\left\{n_k^{1/2}\left(\hat{\alpha}_{n_k}^* - \hat{\alpha}_{n_k}(\omega)\right) - c, 0\right\} \\ &\xrightarrow{d^*}_p \max\{Z - c, 0\} \text{ as } k \to \infty \\ &\leq \max\{Z, 0\} \end{aligned} \tag{26}$$

Andrews (2000) explains that this holds for all the subsequences $\omega \in A_c := \left\{\liminf_{n\to\infty} n^{1/2}\hat{\alpha}_n < -c\right\}$ and also applies to cases of $\hat{\alpha}_{n_k}(\omega) \geq c$, meaning that the bootstrap is asymptotically incorrect both for negative and positive values of $\hat{\alpha}_{n_k}(\omega)$ when $\alpha = 0$. More precisely, the distribution of T_n^* is shifted to the left with respect to G_n, making it inconsistent.

Among the possible solutions to this, Andrews proposes to implement a rescaled bootstrap, which uses bootstrap samples of size $m < n$.

One can show that:

$$\begin{aligned} & m^{1/2}\left(\hat{\theta}_m^* - \hat{\theta}_n\right) \\ & = \max\left\{m^{1/2}\left(\hat{\alpha}_m^* - \hat{\alpha}_n\right), -m^{1/2}\hat{\alpha}\right\} \\ & = \max\left\{m^{1/2}\left(\hat{\alpha}_m^* - \hat{\alpha}_n\right) + m^{1/2}\left(\hat{\alpha}_n - \alpha\right), -m^{1/2}\alpha\right\} - m^{1/2}\left(\hat{\alpha}_n - \alpha\right) \\ & = \max\left\{m^{1/2}\left(\hat{\alpha}_m^* - \hat{\alpha}_n\right) + o(1), -m^{1/2}\alpha\right\} + o(1) \\ & \xrightarrow{d^*}_p \begin{cases} Z & \text{if } \alpha > 0 \\ \max\{Z, 0\} & \text{if } \alpha = 0 \end{cases} \quad \text{as } n \to \infty \end{aligned} \tag{27}$$

Although this may solve the problem of the inconsistency of the bootstrap, the author remarks that this and the other methods are inconsistent when the true α has a form of $\alpha_n = \alpha/n^{1/2}$. There aren't free lunches after all.

6 References

1. Andrews, Donald. (2000). Inconsistency of the Bootstrap when a Parameter Is on the Boundary of the Parameter Space. Econometrica. 68. 399-406.

2. Cavaliere, Giuseppe & Pedersen, Rasmus & Rahbek, Anders. (2018). The Fixed Volatility Bootstrap for a Class of Arch(q) Models. Journal of Time Series Analysis. 39.

3. Dickey, D. A., & Fuller, W. A. (1979). Distribution of the Estimators for Autoregressive Time Series With a Unit Root. Journal of the American Statistical Association, 74(366), 427–431.

4. Ding, Zhuanxin and Granger, Clive, (1996), Modeling volatility persistence of speculative returns: A new approach, Journal of Econometrics, 73, issue 1, p. 185-215.

5. Flachaire, Emmanuel and Davidson, Russell, (2001), The Wild Bootstrap, Tamed At Last, No 1000, Working Paper, Economics Department, Queen's University.

6. Hill, B. M. (1975). A Simple General Approach to Inference About the Tail of a Distribution. The Annals of Statistics, 3(5), 1163–1174.

7. Jarque, Carlos M. and Bera, Anil K., (1980), Efficient tests for normality, homoscedasticity and serial independence of regression residuals, Economics Letters, 6, issue 3, p. 255-259.

8. G. M. Ljung, G. E. P. Box, On a Measure of Lack of Fit in Time Series Models, Biometrika, Volume 65, Issue 2, August 1978, Pages 297–303.

9. Loretan, M., & Phillips, P. C. (1994). Testing the covariance stationarity of heavy-tailed time series: An overview of the theory with applications to several financial datasets. Journal of empirical finance, 1(2), 211-248.

10. Németh, L., Zempléni, A. Regression Estimator for the Tail Index. J Stat Theory Pract 14, 48 (2020).

11. Pagan, A. (1996). The econometrics of financial markets. Journal of empirical finance, 3(1), 15-102.