Final Year Project on Economics

Universitat Pompeu Fabra

Statistical and Forecasting Techniques on Financial Markets

Yuanxi He

Tutor: Raúl Merino Fernández

Project code: EME05

Course: 2021/2022

Universitat Pompeu Fabra Barcelona

Abstract

In this project, several steps are taken to analyse and cluster the financial markets' data. The first part is focused on a brief description of the statistical properties of the time series.

Later on, the clustering with the k-Means method, using the rolling four statistical moments, is firstly focused on the NASDAQ index, then to extend the model, several financial products are selected. Moreover, the creation of a strategy based on clusters has been proposed in order to get profit, which has been successful.

Finally, ARCH models backed by statistical tests and criteria, using standardized returns, are used to capture the volatility clustering and forecast it, and the GARCH (1,1) was selected among all since it has the most forecast accuracy.

Keywords: Financial Markets; K-Means Clustering; Forecasting; GARCH

Contents

Chapter 1: Introduction

This project is a brief and modified extension of the research of Horvath, B, Issa, Z, and Muguruza, A (2021)(Horvath, Issa, & Muguruza, 2021) adding some models to forecast volatility. The starting dataset is the returns of NASDAQ, then to generalize the analysis, it is extended in order to see if the models could also apply to these.

1.1 Structure

The project has three parts: i) the statistical analysis, which shows the normality, stationarity and autocorrelations tests; ii) the K-Means clustering applied to the four moments: mean, standard deviation, skewness, and kurtosis of several rolling windows and iii) the volatility forecasting of the return's series, then evaluated through several loss functions, and testing of their predictive ability.

1.2 Motivation

The motivation for this work follows the very recent and innovative based research, with the methodology with clustering on statistical moments of time series. Also, during the last two deep economic and financial depressions, in which the financial returns have been very volatile, this work could be useful and on trend.

1.3 Objectives

The purpose of the project is to make this analysis more intuitive and concise. So, this paper applies several statistical techniques based on R language, to check the data's normality and stationarity. Moreover, clustering of the returns series concerning their rolling volatilities in order to classify the bear and bull markets. Then, the creation of a strategy based on the clustering classification to get profits. In addition, volatility forecasting models will be taken to predict future returns based on the past data.

Chapter 2: Nasdaq's statistical analysis

The purpose of taking the Nasdaq Composite index as the starting point is that it is a well-known index compounded by more than 3,700 stocks, and it is heavily weighted in the technology sector. Some current companies include Apple (12.25%), Microsoft (9.93%), Amazon (7.13%), Tesla (4.79%) and Alphabet Class C and A (7.38%), etc. This could be useful for the first step of the project, and can also be applied to other stock prices or exchange rates, among others.

Nasdaq prices are taken from Yahoo Finance [1] since February 1971. Concerning the stationarity of the data, the log returns are calculated from the lagged differences of prices (details in later sections).

2.1 Statistical Summary

This summary takes into consideration the four moments of the return series, as well as its minimum return and the maximum return with their respective date. See Table 1 below.

Table 1: Brief summary of Nasdaq returns

Description of Moments	**Returns (in percentage)**
Minimum return (2020-03-16)	-13.149
Average return	0.037
Median return	0.107
Maximum return (2001-01-03)	13.255
Daily volatility	1.261
Annualized volatility	20.012
Skewness	-0.388
Kurtosis	12.942

Source: Own elaboration based on data from Yahoo Finance

[1] Source:`https://finance.yahoo.com/quote/%5EIXIC?p=%5EIXIC`.

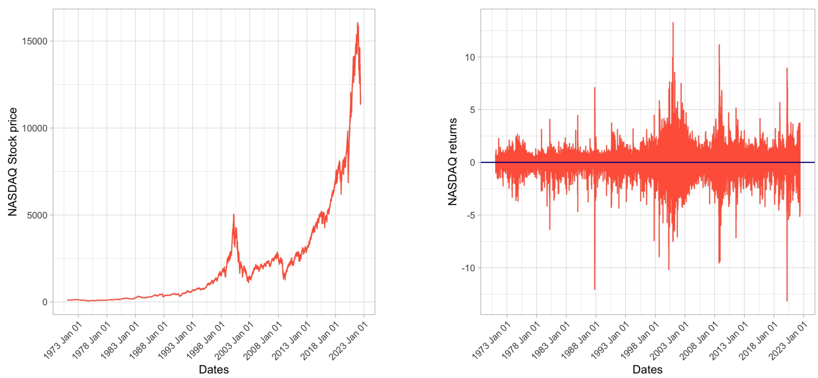
2.2 Data Distribution

First, to have a wider perspective, the prices can be observed as they have evolved over the years. In Figure 1 (a) the data is depicted as it has behaved in 50 years, according to its price. In Figure 1 (b), the returns are shown which are recorded daily with a zero mean, remarked with the blue horizontal line.

Figure 1: Price growth and its returns (9/6/2022)

(a) Nasdaq prices over the years

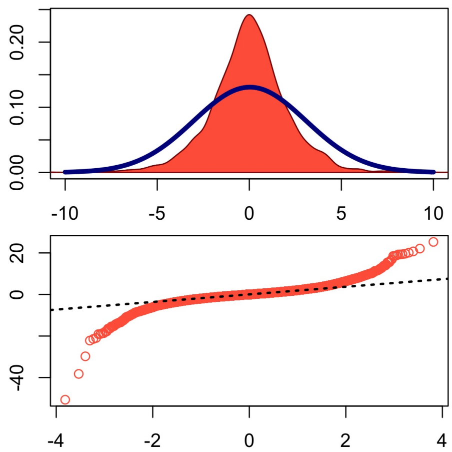
(b) Representation of Nasdaq returns, with line on mean



Source: Own elaboration based on data from Yahoo Finance

It is important to test if the data resembles a normally distributed variable. Figure 2 depicts above the kernel distribution and below the quantile-quantile (QQ) plot of these returns, which does not resemble very much a Gaussian distributed time series.

Figure 2: Kernel (above) and QQ (below) plot of Nasdaq returns (9/6/2022)



Source: Own elaboration based on data from Yahoo Finance

For instance, the kernel representation is helped with the normal blue line, which can be seen that these two do not strictly match. Moreover, considering the QQ plot, in which the fatter tails can be seen, from the perspective of kurtosis, the probability of observing extreme values is much higher than the normal distribution (with kurtosis equal to 3). And from the skewness perspective, the data is a bit left-skewed (compared to the normal distribution) since the value is lower than 0 (the normally distributed skewness). In addition, the median is greater than the mean. These can also be checked in Table 1.

Certainly, this can be tested with the Jarque-Bera test (Jarque & Bera, 1980) parametrized by the equation 2.1.

$$JB = \frac{n}{6}\left(S^2 + \frac{1}{4}(K-3)^2\right) \tag{2.1}$$

where n is the number of the size of the time series, S is the skewness and K the kurtosis.

Under the null hypothesis of normality, the JB statistic tends to a χ^2_2 distribution. The results can be seen in Figure 3 in which the null hypothesis is rejected.

Figure 3: Jarque-Bera test results (9/6/2022)

```
            Jarque-Bera Normality Test

data:  returns
JB = 53520, p-value < 2.2e-16
alternative hypothesis: greater
```

Source: Own elaboration based on data from Yahoo Finance

2.3 Returns Stationarity

The augmented Dickey-Fuller test (Dickey & Fuller, 1981) described by equation 2.2 is useful to prove whether the data acts stationary (see Figure 1) as well as to control for serial dependence. The results are shown in Figure 4, in (a) the null hypothesis ($H_0 : \gamma = 0$) cannot be rejected, implying that prices are non-stationary. However, in (b) once the returns are differentiated, the null is rejected, so it can be concluded that the returns are stationary.

$$\mathrm{DF} = \frac{\hat{\gamma}}{\mathrm{SE}(\hat{\gamma})} \tag{2.2}$$

where $\hat{\gamma}$ is the sample coefficient for the lags of an autoregressive process with p orders as 2.3.

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \delta_1 \Delta y_{t-1} + \cdots + \delta_{p-1} \Delta y_{t-p+1} + \varepsilon_t \tag{2.3}$$

where α is a constant, β the coefficient on a time trend and ε_t the white noise error with mean 0.

Figure 4: Augmented Dickey-Fuller test results (9/6/2022)

(a) ADF test using Nasdaq prices

```
	Augmented Dickey-Fuller Test

data:  price
Dickey-Fuller = -0.21522, Lag order = 23, p-value = 0.99
alternative hypothesis: stationary
```

(b) ADF test using Nasdaq returns

```
	Augmented Dickey-Fuller Test

data:  returns
Dickey-Fuller = -21.969, Lag order = 23, p-value = 0.01
alternative hypothesis: stationary
```

Source: Own elaboration based on data from Yahoo Finance

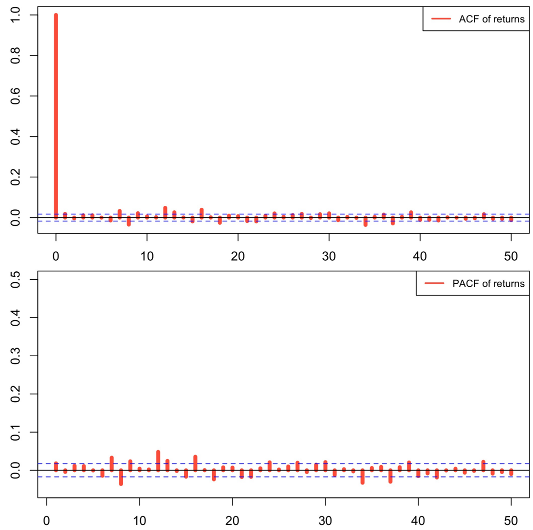
2.4 Autocorrelation Analysis

From Figure 5 it can be seen that the autocorrelations (ACF) and partial autocorrelations (PACF) of the returns are quite dissonant; considering the Ljung-Box test (Ljung & Box, 1978) (see equation 2.4) that under the null hypothesis $H_0 : \rho_1 = 0, \rho_2 = 0, \cdots, \rho_h = 0$ the Q statistic tends to a χ^2_h distribution. The main difference between the ACF and PACF is that the k-lag partial autocorrelation measures the dependence between y_t and y_{t-k} which is not captured by the rest of the lags. The blue dot lines are the confidence intervals $\pm 1.96 * 1/\sqrt{n}$ with a 5% significance level.

$$Q = n\,(n+2) \sum_{k=1}^{h} \frac{\hat{\rho}_k^2}{n-k} \tag{2.4}$$

where n is the number of the length of the time series, $\hat{\rho}_k$ is the sample autocorrelation at lag k and h is the number of lags tested.

Figure 5: Autocorrelation plots of Nasdaq returns, with 50 first lags (9/6/2022)



Source: Own elaboration based on data from Yahoo Finance

The results over the 50 first lags are shown in Figure 6, where it can be seen that these autocorrelations are different from zero, thus the null hypothesis is rejected.

Figure 6: Ljung-Box test results (9/6/2022)

```
        Box-Ljung test

data:  returns
X-squared = 206.86, df = 50, p-value < 2.2e-16
```

Source: Own elaboration based on data from Yahoo Finance

However, the autocorrelations seem negligible as a consequence of the noise's amplification, due to the presence of time-varying volatility. Later on, it can be appreciated that once volatility is adjusted, the ACF of the standardized returns looks harmonic (see Figure 24), meaning that all the autocorrelations are outside the confidence intervals, and the PACF decreases exponentially.

Chapter 3: K-Means Clustering

Concerning the financial volatility, using a concise way to classify the bear and bull stock markets may be useful. One of the best options is taking the k-means method which allows clustering based on their similarities, in this case, the four statistical moments in the rolling windows.

3.1 Rolling windows

The rolling windows are useful to aggregate returns into several periods and then compute their four statistical moments: mean, standard deviation, skewness and kurtosis. For this project, several windows using only the working days are taken, and the intervals are shown below:

- 5 years window: from February 1971 to February 1976.
- 1 year window: from February 1971 to February 1972.
- Half-year window: from February 1971 to August 1971.
- 1 month window: from 1971-02-09 to 1971-03-09.
- 1 week window: from 1971-02-09 to 1971-02-16.

The mean of the four moments is shown in Table 2

Table 2: Mean of the four moments (9/6/2022)

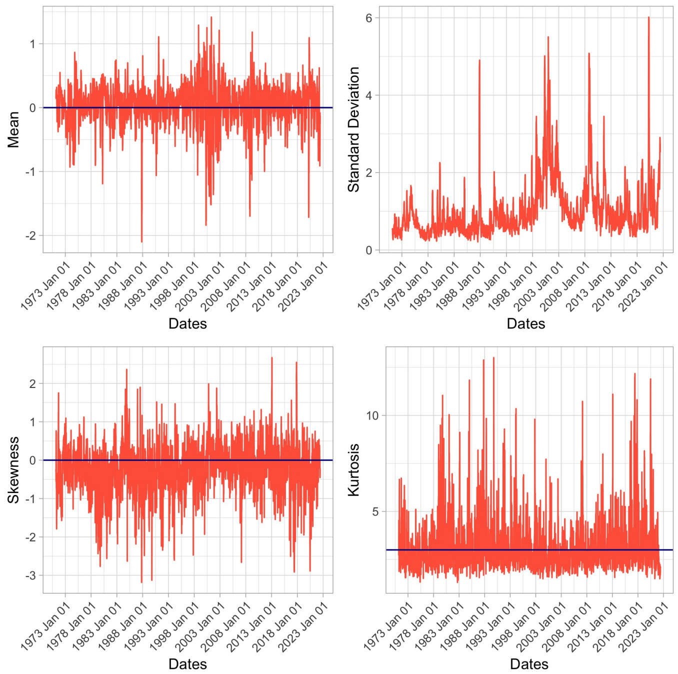
Rolling windows	**Mean**	**Standard deviation**	**Skewness**	**Kurtosis**
5 years	0.0392	1.1727	-0.7583	11.3430
1 year	0.0392	1.1129	-0.5232	5.6725
Half year	0.0378	1.0905	-0.4668	4.7134
1 month	0.0367	1.0304	-0.2338	3.0661
1 week	0.0370	0.9580	-0.0619	2.0081

Source: Own elaboration based on data from Yahoo Finance

It can be seen that the mean and standard deviation don't change a lot between the windows, but the last two moments have significant differences, when the range is wider, the skewness decreases, inversely with the kurtosis which increases exponentially.

To start the analysis, one month window is taken, and the plot of its four moments is shown in Figure 7 below. Applying the ADF test, all moments are stationary, and the null hypothesis of no autocorrelation is rejected and finally, with the JB test, these moments are not normally distributed, which coincides with the previous chapter.

Figure 7: Statistical moments in a monthly rolling window over the years (9/6/2022)



Source: Own elaboration based on data from Yahoo Finance

3.2 K-Means algorithm

K-Means clustering is a simple and well-known unsupervised machine learning algorithm which allows classifying a dataset into given k clusters, depending on their similarities, and set a center, usually named the centroid.

The classification algorithm (Hartigan and Wong, 1979 (Hartigan & Wong, 1979)) is based on minimizing the total within-cluster sum of square (wss), which was defined as the sum of squared Euclidean distances (see equation 3.1) between each data point and their corresponding centroid (see equation 3.2).

$$d(p,q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \cdots + (p_n - q_n)^2} \tag{3.1}$$

where p and q are two separate points and n is the dimension. Notice that if $p = q$, then the distance is 0.

$$wss = \sum_{k=1}^{K} W(C_k) = \sum_{k=1}^{K} \sum_{x_i \in C_k} d(x_i, \mu_k)^2 \tag{3.2}$$

where $W(C_k)$ is the total within-cluster variation, x_i is a data point in the cluster C_k, and μ_k is the centroid of the cluster C_k. Notice that increasing k will decrease the total wss.

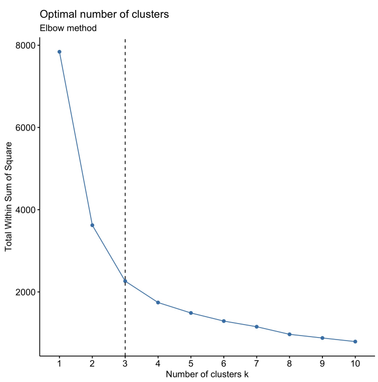
3.3 Clustering

The clustering starts taking only two first moments (mean and standard deviation) of the one-month rolling window. Using the elbow method and its alternative, the average silhouette method, the first computes the previously shown total wss, and the second measures how a clustering algorithm has performed, i.e. its quality. The suggestion of optimal k for the dataset are 2 and 3.

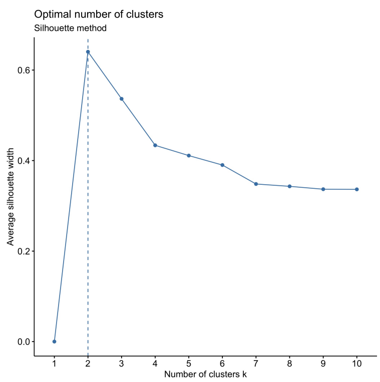
Both methods are shown in Figure 8, with the first method $k = 3$ is optimal since beyond the third, the improvement in the total wss isn't a lot. With the second method, the $k = 2$ returns the maximum quality in clustering, and $k = 3$ is the second best.

Figure 8: Determining the optimal k of clusters (9/6/2022)

(a) Elbow method



(b) Silhouette method

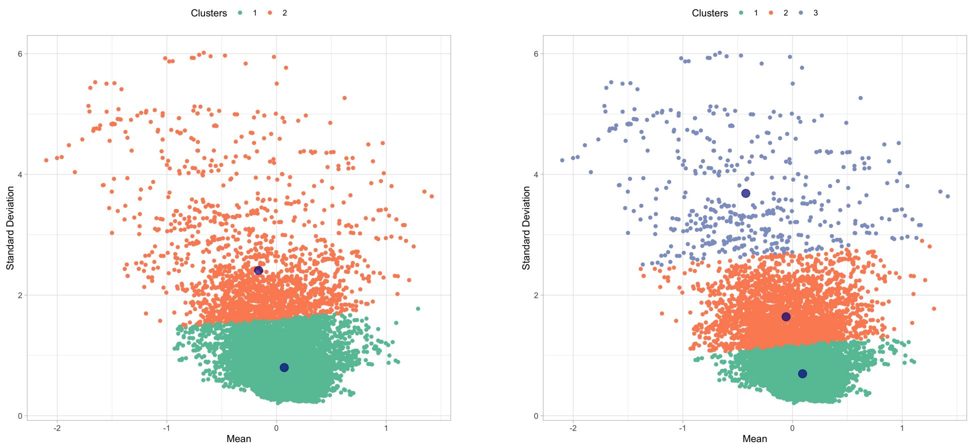


Source: Own elaboration based on data from Yahoo Finance

After setting the model with the optimal number of clusters, the next step is to divide the dataset into two different sets and see their changes adding more variables, which are the skewness and kurtosis. The classification is shown in Figure 9.

Figure 9: Clustering representations with mean and standard deviation (9/6/2022)

(a) With 2 clusters (b) With 3 clusters



Source: Own elaboration based on data from Yahoo Finance

The clustering separates the data for their volatility. As can be seen, the top clusters have more dispersion, compared to the cluster 1 with a centroid with a mean of almost 0, as explained in the stationarity section. And it also can be seen that by increasing the number of clusters, the mean returns seem to decrease. That implies higher volatility tends to get a lower return.

To extend the model, more variables are added, are shown in Figure 25 and 27 and their 2-dimensional representations (Figures 26 and 28). As the table 3 shows, the classification is clear, in the case of 2 clusters, with high volatility, the skewness seems negligible, but in the case of lower volatility, the skewness is negative and significant. But this happens differently when $k = 3$, since between centroid 1 and 2 the standard deviation hasn't a significant difference, but the skewness does.

The same happens when kurtosis is added (see table 4), in the case with 2 clusters, the mean and standard deviation aren't taken into clustering, since the differences are very slight, the clustering is only classified with respect to skewness and kurtosis. Also, these two last moments have a quadratic relationship (see Figure 28). Moreover, with 3 clusters, all four moments are considered. It can be observed that the standard deviation is not directly related with kurtosis, but this is true with skewness.

These results can explain the relevant differences in the rolling windows in the third and fourth moments, the clustering has been also applied to other windows apart from one month, and the statement holds.

3.4 Profits strategy

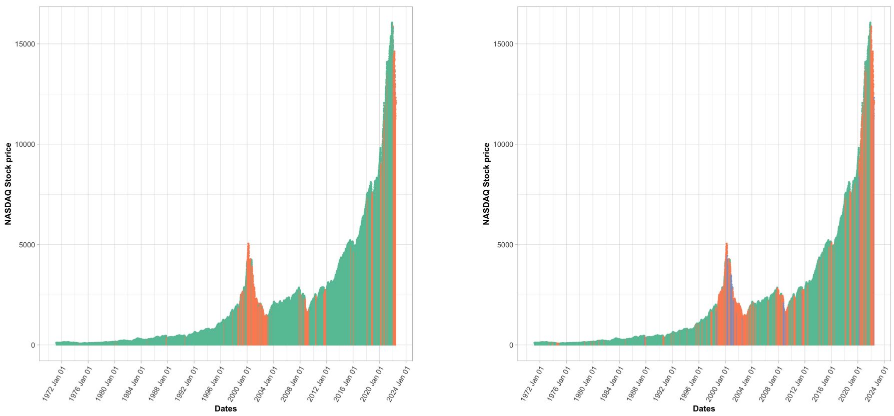
Since the clustering can separate the data well with high and low volatility, also to its moments, the model can be applied to hedge risk, and try to create a strategy to get profit.

3.4.1 Historical evidence

For this purpose, Figure 10 shows the previous Nasdaq price growth separating the periods based on clusters only with mean and standard deviation. The green region is cluster 1, i.e. the region with lower volatility, and others when volatility, the standard deviations are higher. In these plots, several financial crises are clearly identified:

Figure 10: Nasdaq prices over the years by clusters (9/6/2022)

(a) With 2 clusters (b) With 3 clusters



Source: Own elaboration based on data from Yahoo Finance

- The Black Monday in October 1987, in which the Dow Jones index dropped 22.6%.
- The Iraq's Invasion of Kuwait in August 1990, in which the global price of oil raised.
- The Dot-com Bubble started from 1995 to the Nasdaq spikes in March 2000 and followed by a deep fall until October 2002.
- The Great Recession in 2008 until the recovery in mid 2009.

- The crash of 2:45 on 6, May 2010 and also the fall in stock markets on August 2011.
- The stock market losses in 2018, when Nasdaq lost 3.9 percent of its value, then rapidly recovered thanks to the tech giants Apple and Microsoft in 2019.
- The COVID-19 crisis in March 2020, in which the Nasdaq lost 12.3% of its value.
- The Russia's Invasion of Ukraine in February 2022, with losses of nearly 25%.

After the crises' identification, a simple strategy which could test if the model yields profit could be buying and selling when the Nasdaq index changes from one cluster to the others, omitting the transaction cost.

As the algorithm 1 shows, for each data point i which is in cluster 1 (with low volatility, i.e. bull markets), it can be set as buying, and if the data point changes to another cluster, in this case the second (with high volatility i.e. bear markets), the stock will be sold. Another strategy could be to have bought the stock in the first day, i.e. 1971-02-08, then holding this asset until the i day to be sold. Doing this procedure, with the Nasdaq index, would yield a profit of $11153.03 today (2022-06-06), compared to the strategy of holding this asset until i day, which would generate an additional $828.83.

Algorithm 1 Strategy with 2 clusters

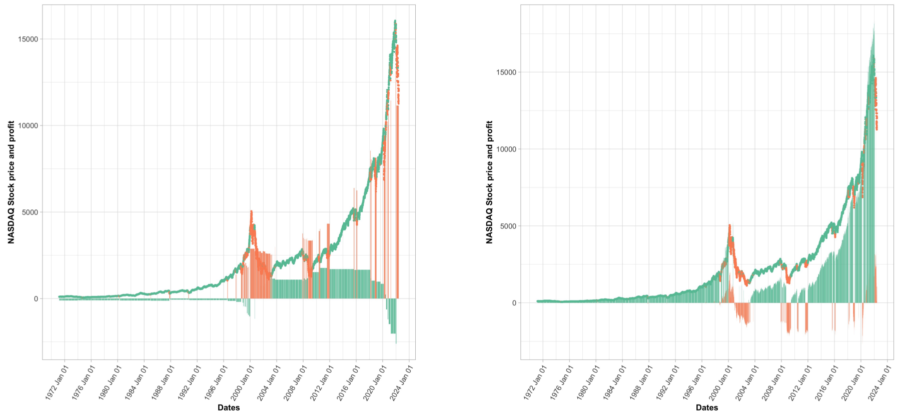
```
for i in 1:length(rollingdf) do
    if i == 1 then
        if cluster[i] == 1 then
            buy=price[i]
        end if
    else if cluster[i] == 2 & cluster[i − 1] == 1 then
        sell+=price[i]
    else if cluster[i] == 1 & cluster[i − 1] == 2 then
        buy+=price[i]
    end if
    profit[i]=sell-buy
    holding[i]=price[i]-price[1]
end for
```

To visualize the changes, see Figure 11. Where the bar plot above the 0 is the profit and below are the losses, distinctly, in cluster 2, the strategy can always generate profits. Comparing holding the asset to the strategy (i.e. holding[i]-profit[i]), it can be observed that on most occasions with cluster 2, that is when the market is more volatile, the strategy is better than holding.

Figure 11: Nasdaq prices (line) and profit (bar plot) (9/6/2022)

(a) Nasdaq prices and profits

(b) Nasdaq prices and the difference on profits



Source: Own elaboration based on data from Yahoo Finance

3.5 Extensions

Since the model works with the Nasdaq index, it would be useful to generalize and extend to other returns series.

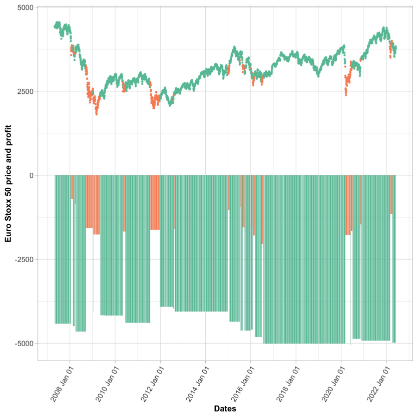
3.5.1 Euro Stoxx 50

Rather than analysing one of the most popular US stock indexes, it can be applied also to the Euro Stoxx 50 index. The index starts from 2007-03-30, and its statistical properties are quite similar to the Nasdaq prices, except that the price and returns are stationary by the ADF test in a 10% significance level.

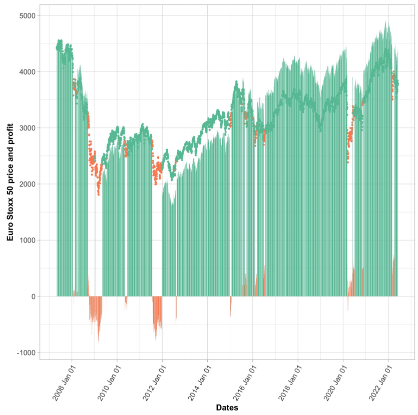
The profit is shown in Figure 12, in this case, the month rolling model with the first two moments generates losses because the data starts in 2007 and follows a recession. But, what is interesting is that the losses during the bear markets are very low compared to the bull markets. Nevertheless, the right picture can be observed that during the bear markets the strategy is slightly better than holding the asset.

Figure 12: Euro Stoxx 50 prices (line) and profit (bar plot) (9/6/2022)

(a) Euro Stoxx 50 prices and profits



(b) Euro Stoxx 50 prices and the difference on profits



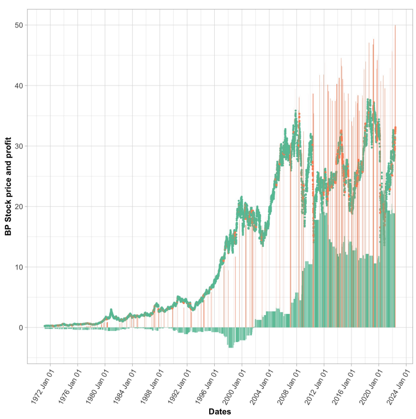
Source: Own elaboration based on data from Yahoo Finance

3.5.2 British Petroleum

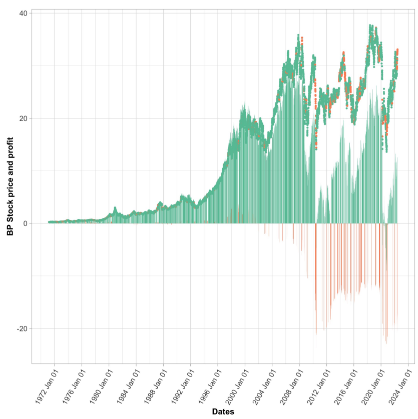
Concerning the last high volatile markets on the petroleum prices, one focus can be on the British Petroleum European company, which is the fifth oil and gas producer worldwide based on revenue in 2021. The first is that the price and returns are stationary by the ADF test, with a 5% significance level. Applying the month window clustering strategy with all moments to the company, it generates $49.98, even more profitable than holding the stock, which can be seen in Figure 13.

Figure 13: BP prices (line) and profit (bar plot) (9/6/2022)

(a) BP prices and profits



(b) BP prices and the difference on profits



Source: Own elaboration based on data from Yahoo Finance

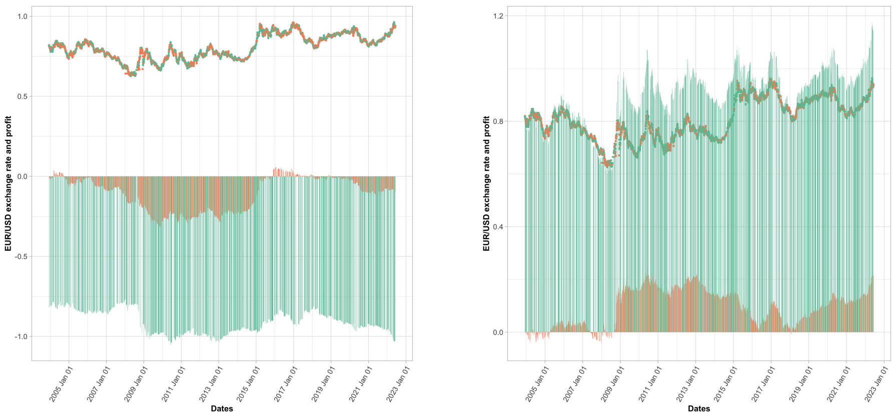
3.5.3 EUR/USD exchange rate

In this case, instead of profit, the concern would be the stability of the exchange rate, that is when the ratio is low. Applying the method to a one-week window with the first three moments, the lowest ratio was achieved (-0.1 Euro per US dollar) which is similar to holding the exchange rate (0.12). See Figure 14 below.

Figure 14: Gold prices (line) and profit (bar plot) (9/6/2022)

(a) EUR/USD exchange rate and profits

(b) EUR/USD exchange rate and the difference on profits



Source: Own elaboration based on data from Yahoo Finance

3.5.4 Bitcoin

A trending point could be to see if the model also works with cryptocurrencies, since they are very volatile. Bitcoin is chosen for its importance, and the data starts from 2014-09-17. Now instead of monthly rolling, one week rolling window clustering still controlling the first two moments is better since the prices are very volatile so when the period is shorter, the model will work better. The model works similarly to the previous analysis, and it generates a profit of $9213.52, shown in Figure 15.

3.5.5 Gold

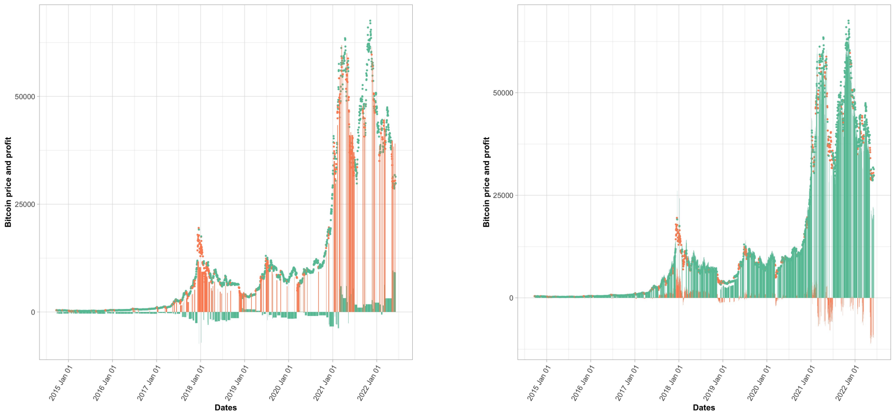
Finally, gold is chosen as a commodity to see if the clustering strategy could generate profits. Applying one year rolling window with mean, standard deviation and skewness, it can gain $701.1. See Figure 16.

To conclude, in almost all cases, the clustering works and could get profits, and

Figure 15: Bitcoin prices (line) and profit (bar plot) (9/6/2022)

(a) Bitcoin prices and profits

(b) Bitcoin prices and the difference on profits



Source: Own elaboration based on data from Yahoo Finance

Figure 16: Gold prices (line) and profit (bar plot) (9/6/2022)

(a) Gold prices and profits

(b) Gold prices and the difference on profits

Source: Own elaboration based on data from Yahoo Finance

in the bear markets, the model works better than in bull markets, in which low volatility appears. But one may need to change the variables and the rolling window depending on the characteristics of each product.

Chapter 4: Volatility Forecasting

To predict the future volatilities, a data split is needed, that means splitting the data into a set for training and a set for testing the model. The criteria are 80%, which goes from the late 70s to 2012 debut, for the in-sample analysis, and the remaining 20% for the out-of-sample forecasting.

4.1 In-Sample Analysis

The autocorrelograms of the returns do not suggest that the data has memory, i.e. it can not be predicted based on the past returns. Moreover, it can be related to the Efficient Market Hypothesis (Fama, 1969 (Fama, Fisher, Jensen, & Roll, 1969)), which states that the expectation of returns conditional on past returns is 0. So, prices reflect all available information, even if markets are inefficient and end up behaving like a random walk, these returns will still be unpredictable.

However, the ACF of the absolute and square returns look persistent, meaning that it can predict the scale of the returns based on the past. Furthermore, the absolute and square returns plots indicate that the scale is not constant, so volatility depends on time, with one evidence being the volatility clustering.

Therefore, the conventional Autoregressive integrated moving average models can not be estimated, because they only take into account the time-varying mean, whereas the conditional variance is constant and equal to the innovation term. Eventually, these kinds of models do not capture the time-varying variance.

4.1.1 Volatility clustering

To assess the existence of volatility clustering in the data, first, the squared and absolute return of data have been plotted (see Figure 29 and 30); clearly, it can be seen the effects of volatility clustering in the data: large (small) returns in the data tend to be followed by large (small) returns.

The next step, before estimating models which can capture these volatility clustering effects, is finding out the presence of AutoRegressive Conditional Heteroskedasticity (ARCH(q)) models (see equation 4.1) effects on our data. To assess this issue, the ARCH-LM test was carried out. This test was introduced by Engle (1982) (Engle, 1982), which estimates the coefficients of the autoregression of the ARCH with q parameters (lags of square returns) by Least Squares.

$$r_t = \sqrt{\sigma_t^2} z_t \tag{4.1}$$

where r_t is the return, z_t has a distribution with mean 0 and variance 1, and the variance σ_t^2 follows to

$$\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \alpha_2 \epsilon_{t-2}^2 + \cdots + \alpha_q \epsilon_{t-q}^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 \tag{4.2}$$

where $\alpha_0 > 0$, $\alpha_i \geq 0$, and $\sum_{i=1}^{q} \alpha_i \leq 1 \; \forall i$.

The null hypothesis is that ARCH has no effects hence $\alpha_i = 0 \; \forall i$ and the test statistic follows to nR^2, where n is the size of the data and R^2 is the R squared coefficient. Under the null hypothesis, it is distributed as χ_q^2.

A shred of strong evidence was found that the data have ARCH effects with 11 lags. The p-value associated with the t-statistic is very small and all the coefficients of the autoregression computed are significant, meaning that past values of square returns can predict future square returns, so the scale of the returns can be predicted.

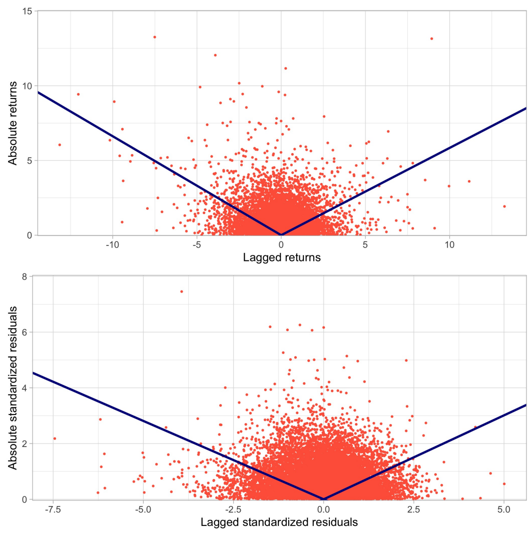
4.1.2 Asymmetric effects

Most of the stocks have in common the presence of asymmetric effects, meaning that negative returns increase the volatility more than positive ones. This effect can be appreciated in Figure 17, where lagged returns have different impacts on absolute returns, depending on the sign. To verify the asymmetry in the data, the Generalized ARCH (GARCH (p,q)) model introduced by Bollerslev (1986) (Bollerslev, 1986) (see equation 4.3) is estimated with p and q lags, in this case, p and q = 1 are used. Also testing if the square standardized residuals are i.i.d, i.e. if they can be predicted based on the past.

$$\begin{aligned}\sigma_t^2 &= \omega + \alpha_1 \epsilon_{t-1}^2 + \cdots + \alpha_q \epsilon_{t-q}^2 + \beta_1 \sigma_{t-1}^2 + \cdots + \beta_p \sigma_{t-p}^2 \\ &= \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2\end{aligned} \tag{4.3}$$

where $\omega > 0$, $\alpha_i \geq 0$, $\beta_i \geq 0$, and $\sum_{i=1}^{q} \alpha_i + \sum_{i=1}^{p} \beta_i < 1\ \forall i$.

Figure 17: The returns (top) and residuals (bottom) for GARCH(1,1) with regression lines (2022-06-09)



Source: Own elaboration based on data from Yahoo Finance

The results of the negative sign and size tests (Engle and Ng, 1993 (Engle & Ng, 1993)) are the following, for the statistics see Figure 18:

- Negative sign test: the null hypothesis is rejected since the coefficient is significant. So, negative returns help to predict square standardized residuals. Also, the GARCH (1,1) model is not the correct specification to model this issue.

- Negative size test: the null is also rejected, so not only the sign, also the size has predictive power.

Figure 18: Summary of negative sign and negative size bias test (2022-06-09)

Asymmetric Effect Tests

	Dependent variable:	
	GARCH(1,1)	
	(1)	(2)
Negative Sign	0.323***	
	(0.035)	
Negative Size		-0.128***
		(0.021)
Constant	0.858***	0.949***
	(0.023)	(0.019)
Observations	12,939	12,939
R^2	0.007	0.003
Adjusted R^2	0.007	0.003
Note:	*p<0.1; **p<0.05; ***p<0.01	

Source: Own elaboration based on data from Yahoo Finance, using stargazer (Hlavac, Marek (2022)(Hlavac, 2022))

In order to take into account the asymmetry, the Threshold ARCH and Exponential GARCH models are very useful, that would be interesting to extend the project in the future.

4.1.3 Parameter estimation and residuals checking

Tables 5 and 6 show the parameter estimates, in the case of ARCH(11) all lagged values are significant, which could indicate that we should include more lags. Also, with the GARCH(1,1) model we get a very high persistence of 0.991, hence a strong autocorrelation.

The analysis of the fitted values and the residuals are shown in Figures 29 - 31. Models seem to fit the conditional variance and capture the volatility clustering. In order to have the correct specifications, the residuals should follow a normal distribution, that is rejected in all the models. However, their JB statistics are much lower, indicating that the standardized residuals are “distributed more normally” than the original data.

In all the models, the null hypothesis of non-autocorrelated residuals from the Ljung-Box test is not rejected. This is an indicator that all the dependence from the original returns' data is eliminated, which makes the standardized residuals i.i.d, and so the squared standardized residuals. However, the ARCH(11) performs worse than GARCH, since its autocorrelogram does not look as clean as the GARCH ones.

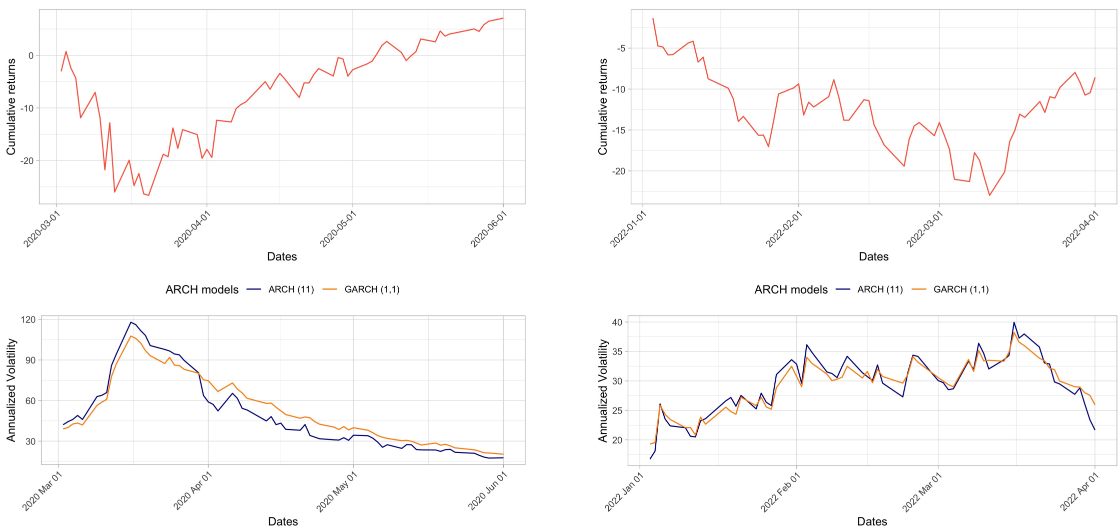
4.1.4 Empirical evidence

Finally, looking into the annualized volatility during the pandemic in terms of the GARCH and ARCH(11) analysis (see Figure 19), one can see that the different models can anticipate financial crises in terms of volatility. Also, taking into account the cumulative returns during the pandemic, one can observe how they increase with the evolution of the pandemic.

Furthermore, analysing the annualized volatility during the Russian invasion of Ukraine (see Figure 19), the previous finding can also be seen. Taking into account the cumulative returns during the Russian war, the downfall right before the war starts can be noticed, and how they are increasing with the course of the war. The same happens also with the dot-com crisis (see Figure 32).

Figure 19: Cumulative returns (top) and annualized volatility (bottom) (2022-06-09)

(a) COVID-19 pandemic (b) Russian invasion of Ukraine



Source: Own elaboration based on data from Yahoo Finance

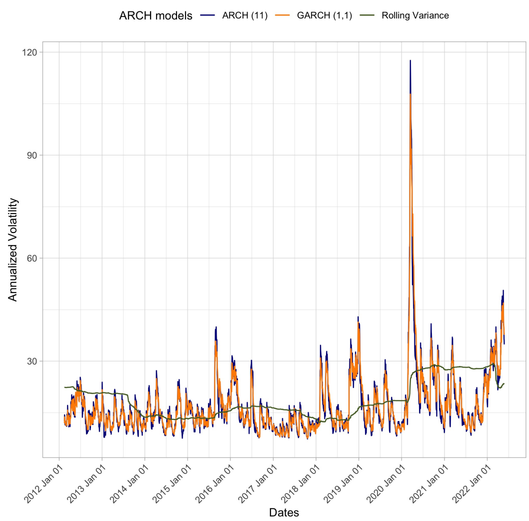
4.2 Out-Of-Sample Analysis

4.2.1 Volatility forecasting

This part is on the remains of the sample (20%), defined by the period: 2012-02-15 to 2022-05-27 and the ARCH, the GARCH as well as the Rolling variance are being considered.

In this last approach of the project, some steps have been taken into forecasting volatility. First, the benchmark is the rolling variance in a 2-year window, which gives the same weights to the past returns, GARCH solves that issue and gives more weight to recent returns. The forecast plot can be seen in Figure 20.

Figure 20: Annualized volatility forecasts among several models with the remains of the sample (20%) (2022-05-27)



Source: Own elaboration based on data from Yahoo Finance

Besides the Rolling Variance, the others show a kind of similar series of volatility. One of the important things to be noted is that since the real conditional variance can not be truly estimated, a proxy of this was used to do the forecast, and then it was compared with the predictions.

However, to choose better, or have a better criterion to compare, the Mean Squared Errors (MSE) and the Quasi-Likelihood (QL) losses were focused on.

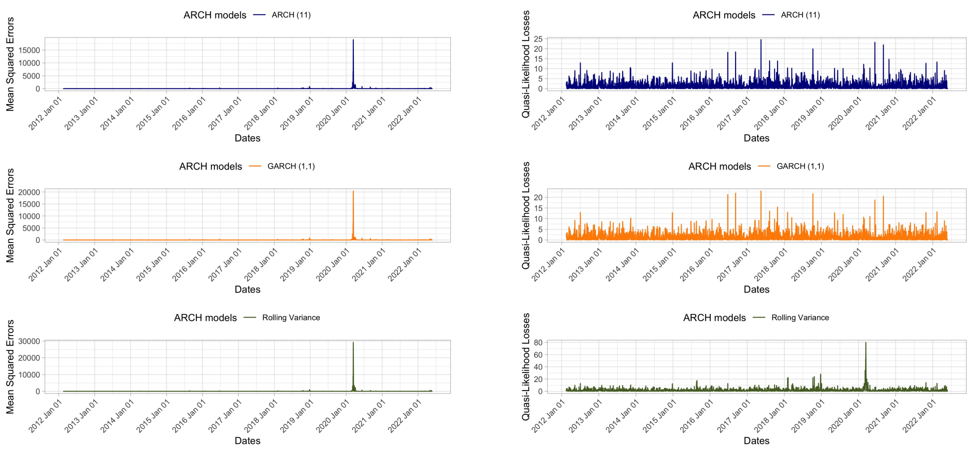
4.2.2 Evaluation

The Rolling is the one that fits the volatility the worst. Whereas the other one's kind of act similarly. Certainly, in the QL plot, a sort of different spikes among the volatility models can be observed.

However, the numerical difference is the one that should be checked after a quick spot on the MSE and QL losses the representation plot (see Figure 21). Figure 22 shows these values. The Rolling variance, in both the MSE and QL, performs worse, in comparison to GARCH and ARCH. Then, the GARCH is the one that has lower QL losses, but a bit higher MSE compared to ARCH. The conclusion is that the GARCH (1,1) is the best model in this NASDAQ volatility forecast.

Figure 21: Losses of several volatility models (2022-05-27)

(a) MSE losses (b) QLIKE losses



Source: Own elaboration based on data from Yahoo Finance

Figure 22: QLIKE and MSE of several models (Volatility Proxy: squared returns) (2022-05-27)

	ARCH (11)	GARCH (1,1)	Rolling Variance
QL	1.551	1.549	1.875
MSE	22.259	22.542	29.254

Source: Own elaboration based on data from Yahoo Finance, using stargazer (Hlavac, Marek (2022)(Hlavac, 2022))

4.2.3 Equal Predictive Ability test

Finally, the Equal Predictive Ability test (Diebold and Mariano, 1995 (Diebold & Mariano, 1995)) was applied to test the ability of the several models, which is the difference of losses (QL and MSE) between the forecasting strategies, in our case the ARCH (11) and GARCH (1,1) compared to Rolling Variance. This can also complement the conclusion that the GARCH (1,1) is the best model for the data analysis (see the values in Figure 23) since one can reject the null hypothesis, that is, the loss differential is 0, according to the DM statistic, which is computed by the average loss differential over the square root of the long-run variance of the difference.

Figure 23: Equal Predictive Ability Tests, compared to Rolling Variance (2022-05-27)

	ARCH (6)	GARCH (1,1)
DM stat (QL)	-4.003	-4.092
p_val_QL	0.0001	0.00004
DM stat (MSE)	-1.026	-1.129
p_val_MSE	0.305	0.259

Source: Own elaboration based on data from Yahoo Finance, using stargazer (Hlavac, Marek (2022)(Hlavac, 2022))

Chapter 5: Results and Conclusions

After checking the statistical summary of the Nasdaq returns, the data is stationary, but not a Gaussian distributed and neither autocorrelated. Nevertheless, once returns are standardized, they appear strongly autocorrelated.

Taking several rolling windows to cluster the returns' series with their respective four conditional moments, it can be seen that the K-Means method does an excellent job. The clusters are clear in respect to their centroids, with the characteristics of the product.

Furthermore, the strategy to yield profit was achieved, in the majority of cases, profit is higher when the market is more volatile. Also, compared to the second strategy, which is to hold the asset until the i day, the first is better in the bear markets, even though profits are lower in many cases. One suggestion for future investigation is to try with more clusters and see if the profit changes. Moreover, implementing a more sophisticated clustering method with a Wasserstein distance in R will be helpful.

Finally, the volatility clustering suggests the GARCH(1,1) model forecast and anticipates the annualized volatility. Concerning the asymmetric effect, which appears with ARCH and GARCH models, one possible extension to the future project is to implement the TARCH or EGARCH models, which will deal very well, and test if they can predict better.

References

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*(3), 307-327. Retrieved from https://www.sciencedirect.com/science/article/pii/0304407686900631 doi: https://doi.org/10.1016/0304-4076(86)90063-1

Dickey, D. A., & Fuller, W. A. (1981, June). Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root. *Econometrica*, *49*(4), 1057-1072. Retrieved from https://ideas.repec.org/a/ecm/emetrp/v49y1981i4p1057-72.html

Diebold, F., & Mariano, R. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, *13*(3), 253-63. Retrieved from https://EconPapers.repec.org/RePEc:bes:jnlbes:v:13:y:1995:i:3:p:253-63

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, *50*(4), 987–1007. Retrieved 2022-06-07, from http://www.jstor.org/stable/1912773

Engle, R. F., & Ng, V. K. (1993). Measuring and testing the impact of news on volatility. *The Journal of Finance*, *48*(5), 1749–1778. Retrieved 2022-06-09, from http://www.jstor.org/stable/2329066

Fama, E. F., Fisher, L., Jensen, M. C., & Roll, R. (1969). The adjustment of stock prices to new information. *International Economic Review*, *10*(1), 1–21. Retrieved 2022-06-07, from http://www.jstor.org/stable/2525569

Hartigan, J. A., & Wong, M. A. (1979). Algorithm as 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *28*(1), 100–108. Retrieved 2022-06-07, from http://www.jstor.org/stable/2346830

Hlavac, M. (2022). stargazer: Well-formatted regression and summary statistics tables [Computer software manual]. Bratislava, Slovakia. Retrieved from https://CRAN.R-project.org/package=stargazer (R package version 5.2.3)

Horvath, B., Issa, Z., & Muguruza, A. (2021). *Clustering market regimes using the wasserstein distance.* arXiv. Retrieved from https://arxiv.org/abs/2110.11848 doi: 10.48550/ARXIV.2110.11848

Jarque, C. M., & Bera, A. K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters*, *6*(3), 255-259. Retrieved from https://www.sciencedirect.com/science/article/pii/0165176580900245 doi: https://doi.org/10.1016/0165-1765(80)90024-5

Ljung, G. M., & Box, G. E. P. (1978, 08). On a measure of lack of fit in time series models. *Biometrika*, *65*(2), 297-303. Retrieved from https://doi.org/10.1093/biomet/65.2.297 doi: 10.1093/biomet/65.2.297

Appendix A: Additional Figures

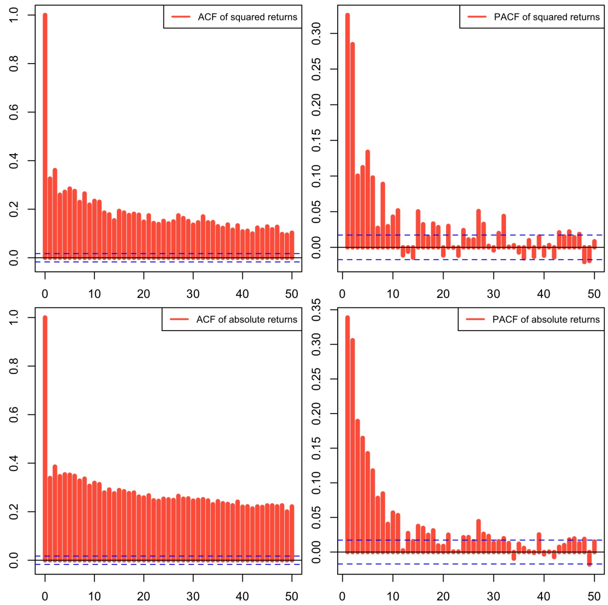
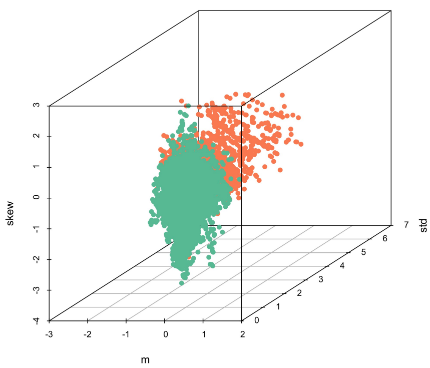
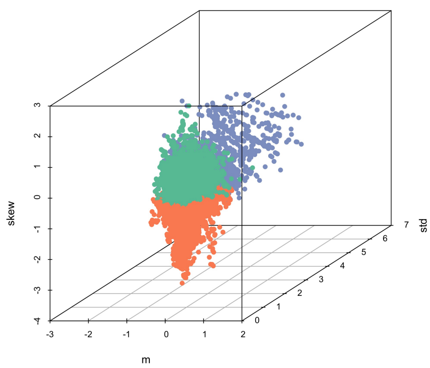


Figure 24: Autocorrelation plots of standardized Nasdaq returns, with 50 first lags

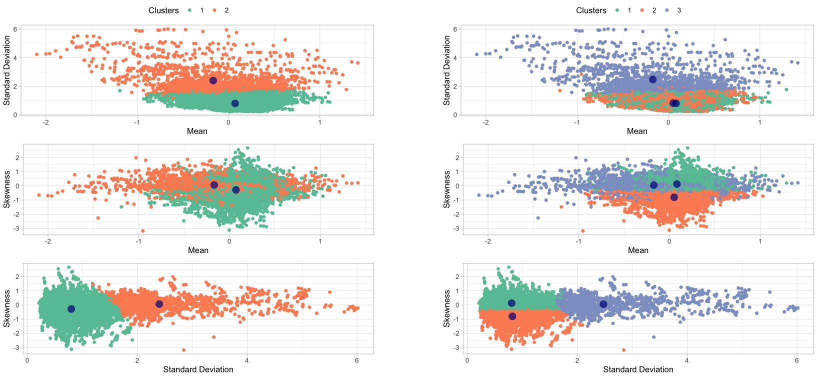


(a) With 2 clusters



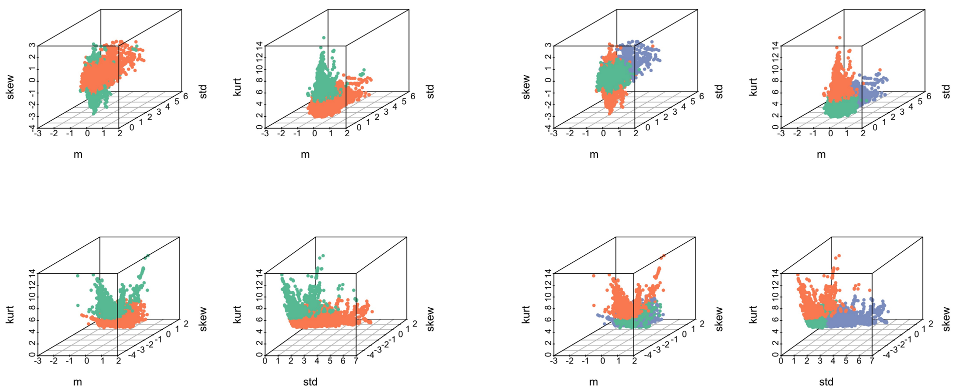
(b) With 3 clusters

Figure 25: Clustering 3D representations with first three moments



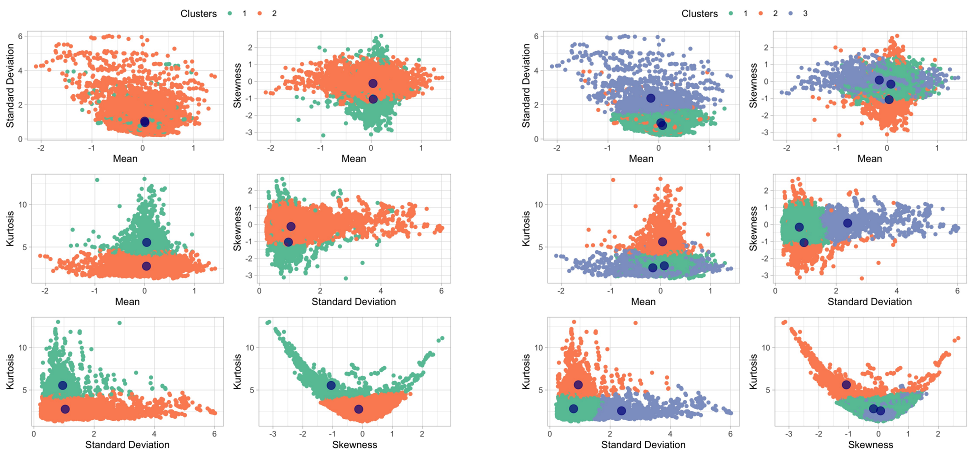
(a) With 2 clusters (b) With 3 clusters

Figure 26: Clustering 2D representations with first three moments



(a) With 2 clusters (b) With 3 clusters

Figure 27: Clustering 3D representations with all four moments



(a) With 2 clusters (b) With 3 clusters

Figure 28: Clustering 2D representations with all four moments

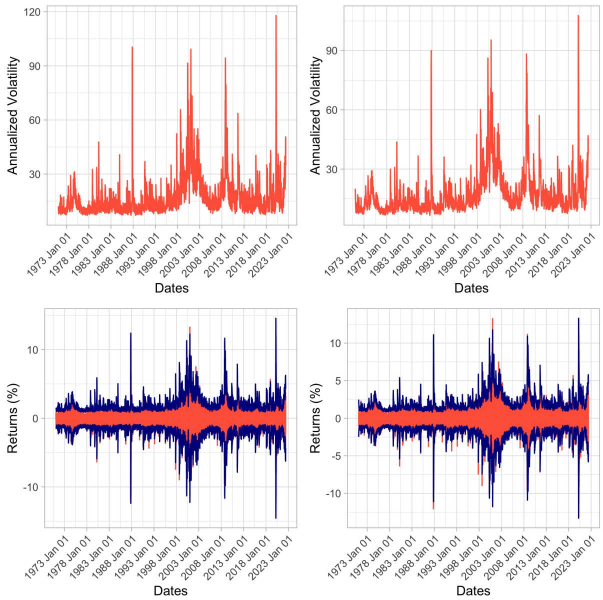
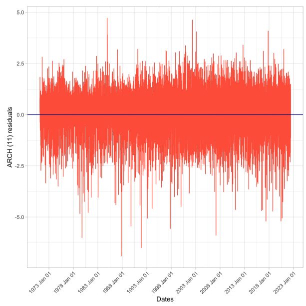
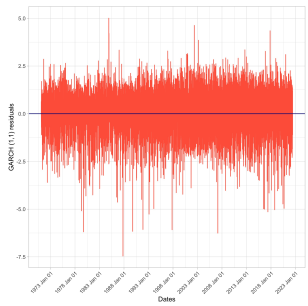


Figure 29:
Annualized volatility (top) and returns with 5% confidence intervals (bottom) of ARCH(11) (left) and GARCH(1,1) (right)



(a) ARCH (11) residuals



(b) GARCH (1,1) residuals

Figure 30: Volatility models residuals

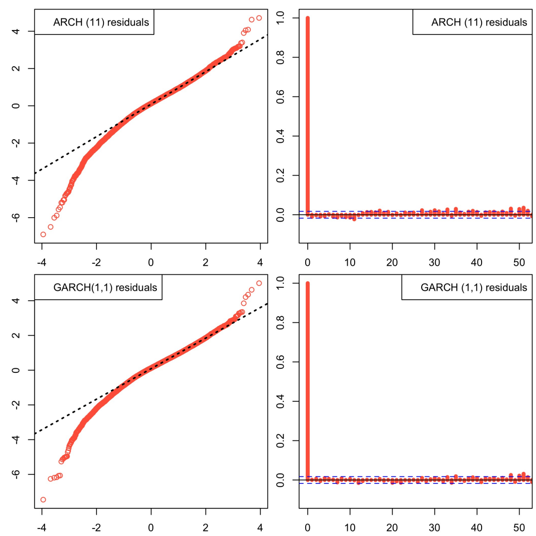


Figure 31: QQ plot (left) and ACF plots (right) of volatility models residuals (left)

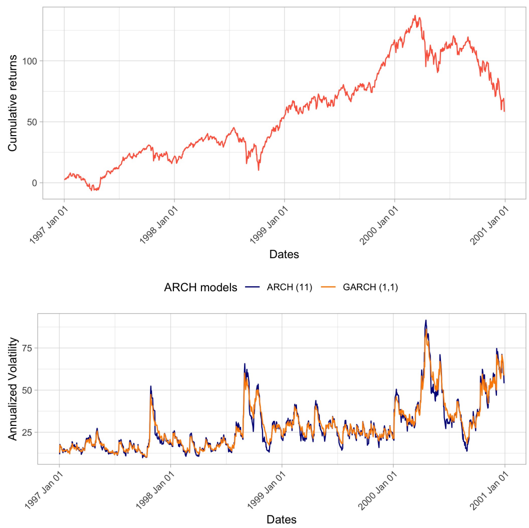


Figure 32:
Cumulative returns (top) and annualized volatility (bottom) during the Dot-com crisis

Appendix B: Additional Tables

Table 3: Centroids adding Skewness (9/6/2022)

Clusters	Mean	Standard deviation	Skewness
$k = 2$	0.0714	0.7993	-0.2838
	-0.1682	2.3998	0.0617
$k = 3$	0.0821	0.8052	0.1310
	0.0510	0.8186	-0.8039
	-0.1725	2.4763	0.0500

Source: Own elaboration based on data from Yahoo Finance

Table 4: Centroids adding Skewness and Kurtosis (9/6/2022)

Clusters	Mean	Standard deviation	Skewness	Kurtosis
$k = 2$	0.0412	0.9558	-1.0512	5.5392
	0.0362	1.0404	-0.1279	2.7455
$k = 3$	0.0734	0.7866	-0.1686	2.7940
	0.0391	0.9427	-1.0783	5.6086
	-0.1573	2.3794	0.0683	2.5528

Source: Own elaboration based on data from Yahoo Finance

Table 5: ARCH (11) parameters

	Coefficients	Standard Errors	p-value
α_0	0.177	0.007	<0.001
α_1	0.134	0.007	<0.001
α_2	0.122	0.009	<0.001
α_3	0.102	0.008	<0.001
α_4	0.107	0.009	<0.001
α_5	0.077	0.010	<0.001
α_6	0.081	0.009	<0.001
α_7	0.053	0.008	<0.001
α_8	0.061	0.008	<0.001
α_9	0.053	0.008	<0.001
α_{10}	0.059	0.008	<0.001
α_{11}	0.057	0.008	<0.001

Table 6: GARCH (1,1) parameters

	Coefficients	**Standard Errors**	**p-value**
ω	0.016	0.001	$<$0.001
α	0.113	0.003	$<$0.001
β	0.878	0.003	$<$0.001
Persistence ($\alpha + \beta$)	0.991		