Modelling of Financial Data

MM905: Financial Econometrics

Inés Blanco Rivas (202393897)

2023-2024

Contents

List of Figures

List of Tables

Chapter 1

Description of the data

In this report we aim to model and analyse the valuation of stocks of Amazon between the years 2018 and 2024. With such an analysis we can attempt to make accurate predictions of the valuation and return of these stock prices in future dates.

The source for the data was taken from yahoo finance, using the getSymbols() function in the quantmod library, which is able to gather data from online sources. In particular, we have used the 'AMZN' ticker on yahoo finance.

For the selection of the timespan we have used the current date (April 2024) and subtracted 2000 days for the initial date. The data is hence daily data of the days the market is open (approximately 252 days a year).

All data is expressed in dollars.

1.1 Stationarity, Returns and log-returns

The time series for the Amazon stock is the following:

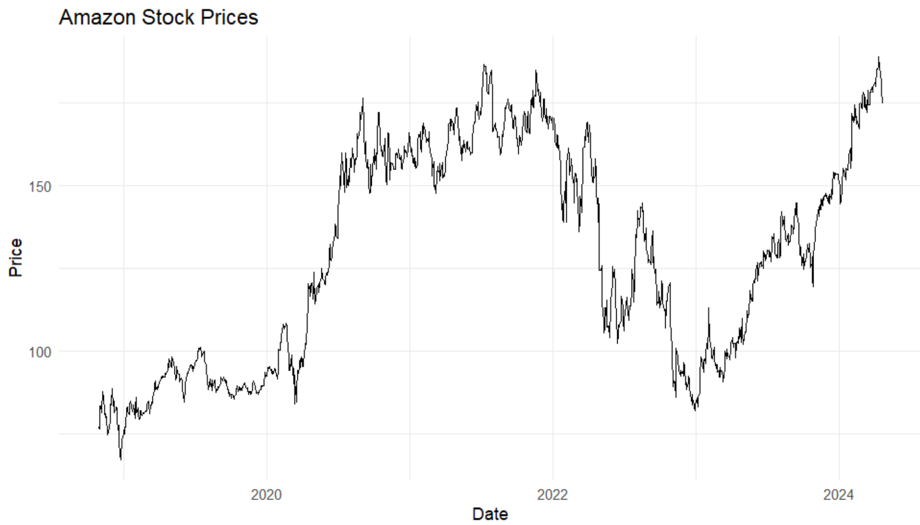

Figure 1.1: Amazon Stock Price per Share from 2018 to 2024 in dollars

Which yield the following returns:

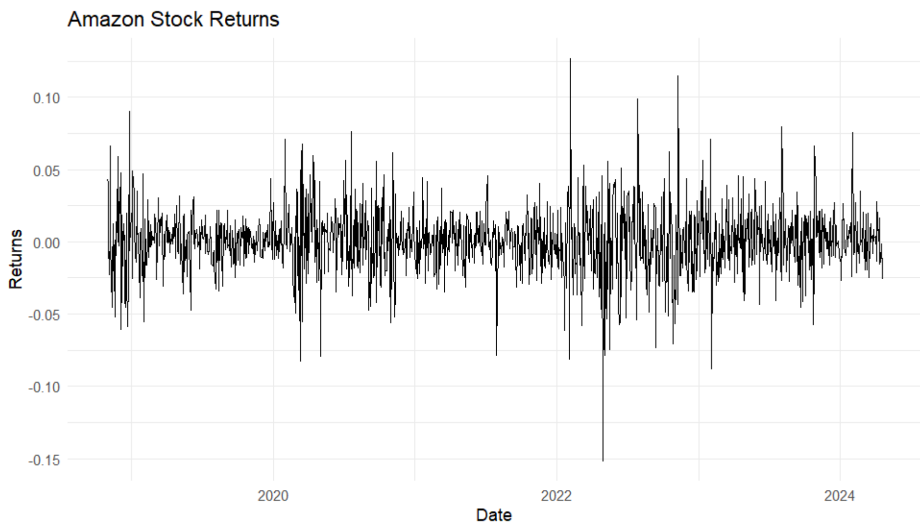

Figure 1.2: Amazon Stock Returns from 2018 to 2024

From these returns we may calculate the mean and autocorrelation function of the time series. We know that rapid decay of the autocorrelation function is a characteristic of time series which present wide-sense stationarity.

We determine that a lag of less than 0.05 would signify a rapid decay in our time series. In the case of the Amazon stock we observe the time series does not hold wide-sense stationarity.

We may also represent the returns as log returns, as statistical properties of log returns are more easily tractable and their distributional properties are more suitable for statistical inference.

The time plot of the log returns can be seen in 1.3

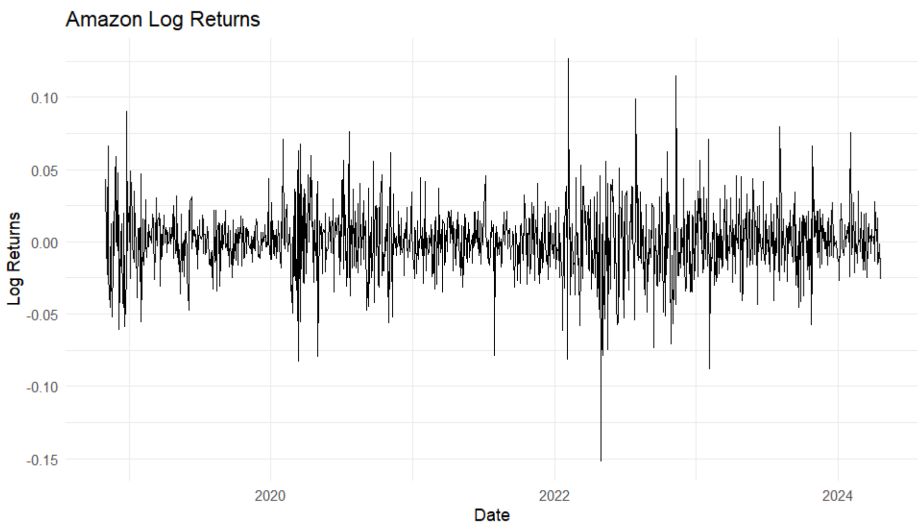

Figure 1.3: Amazon Stock Returns from 2018 to 2024

Chapter 2

Distribution of log returns

We now calculate the empirical distribution function of the log returns. We do this by counting the frequency of each log-return. For visualisation purposes we plot these as a histogram:

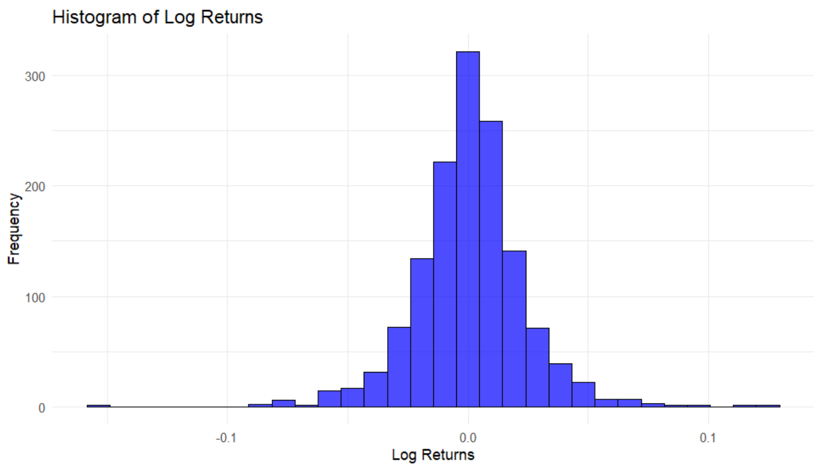

Figure 2.1: Histogram of the log returns distribution

From this distribution we calculate the first four moments of the distribution function:

- Mean: 0.0006

  A mean close to zero shows the log returns are on average close to 0, implying a relatively stable performance.

- Variance: 0.0005

  A low variance shows the log returns values are close to the mean, implying low volatility.

- Skewness: -0.0459

  A negative skewness shows that the distribution function is slightly skewed negatively, implying more negative returns, which could yield a longer left tail.

- Kurtosis: 7.0813

  A high kurtosis shows the tails of the distribution are heavy. For a normal distribution the kurtosis value would be 3. High kurtosis values are quite common in financial market return distributions, as is the case.

2.1 Value at Risk

Although measuring financial risk through the mean-variance framework is a common practice, it is criticised for the lack of distinction between risk and potential.

The X% value at risk is a value such that there is a (1-X)% chance that the loss is larger than said value.

Table 2.1: Values at Risk of the time series

Percentage	Value at Risk
50%	0.0008
75%	-0.0109
90%	-0.0245
97.5%	-0.0464
99%	-0.0577

Chapter 3

Time series model

As our time series is not stationary, an appropriate model to fit a time series to the AMZN data would be using the AutoRegressive Integrated Moving Average (ARIMA) model.

ARIMA takes three initialisation parameters: p, d, and q representing the number of autoregressive terms, the number of differences needed for stationarity, and the number of lagged forecast errors in the prediction equation, Different values of these parameters will yield different fittings of the time series. To evaluate the best parameters for our data we will make use of Akaike's Information Criterion (AIC). The lower the absolute value of the criterion the better fitted the model is.

3.1 shows several parameter iterations and their respective AIC values

Table 3.1: ARIMA fitting to different values of p, d, q

p	d	q	AIC
1	1	1	-6533.27
2	2	2	-6492.49
3	3	3	-6347.75
2	3	5	-6379.11
1	3	5	-6284.63
2	1	2	-6529.5

Given the tested values and the AIC fitting criterion we decide to implement the ARIMA(1,3,5) model for our data. The model yields the following:

Table 3.2: Coefficients of the ARIMA(1,3,5) model

AR1	MA1	MA2	MA3	MA4	MA5
-0.5738	-2.1914	1.016	0.5288	-0.3295	-0.0234

Table 3.3: Training set errors of the ARIMA(1,3,5) model

ME	RMSE	MAE	MASE	ACF1
-0.0003261907	0.0241712	0.01747119	0.7606767	-0.1076472

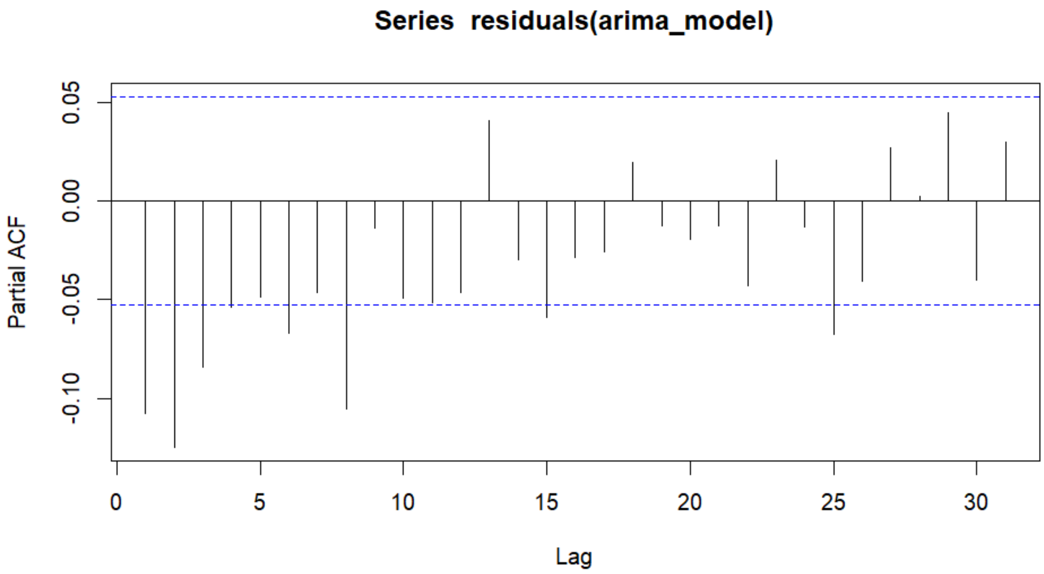

Figure 3.1: Series residuals on partial ACF with ARIMA(1,3,5)

Chapter 4

Model for volatility of returns

To model the volatility of returns we make use of the general ARCH (GARCH) model using the same parameters as we did in the ARIMA model. This yields the following:

Table 4.1: Training set errors of the ARIMA(1,3,5) model

AR1	MA1	omega	alpha 1	beta 1
0.7005031	-0.8919076	0.00001866	0.1336049	0.8392503

The omega coefficient represents the long-term average variance of the residuals, in our case there is low variance.

Alpha 1 indicates the impact of past squared residuals on current variance. The alpha 1 value in this model indicates some level of volatility.

Beta 1 indicates the impact of past variance on current variance. In our case the model indicates a high relation.

Furthermore the log-likelihood of our model is -19423.97, which indicates the model is not likely to be a good fit for the time series.

We can also see further into our autocorrelation function and distribution of returns through our GARCH model.

4.1 shows our returns have significant serial correlation, while 4.2 shows significant autocorrelation, yielding the existence of some white noise in our data.

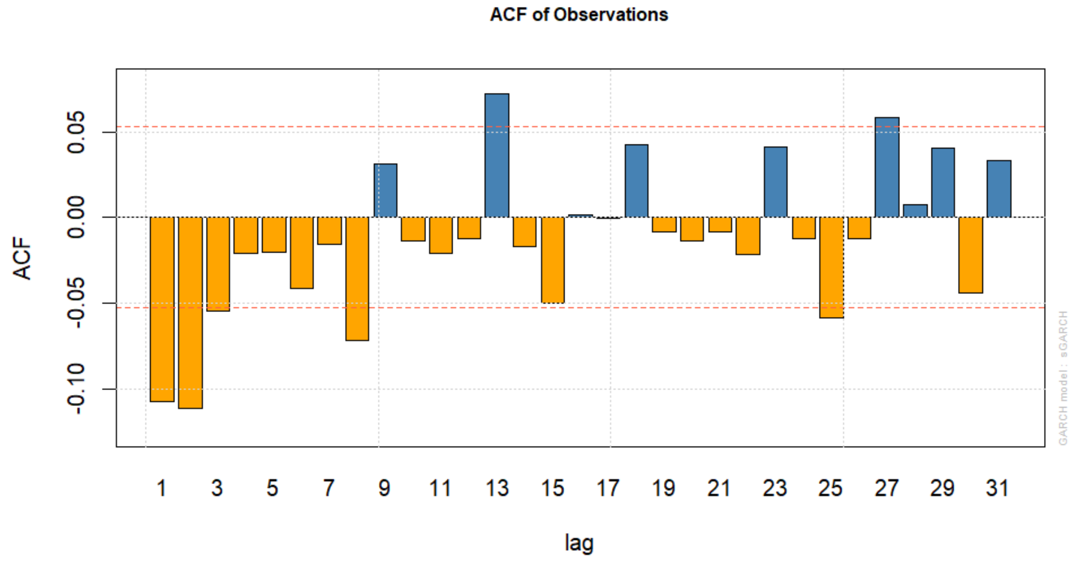

Figure 4.1: ACF of returns

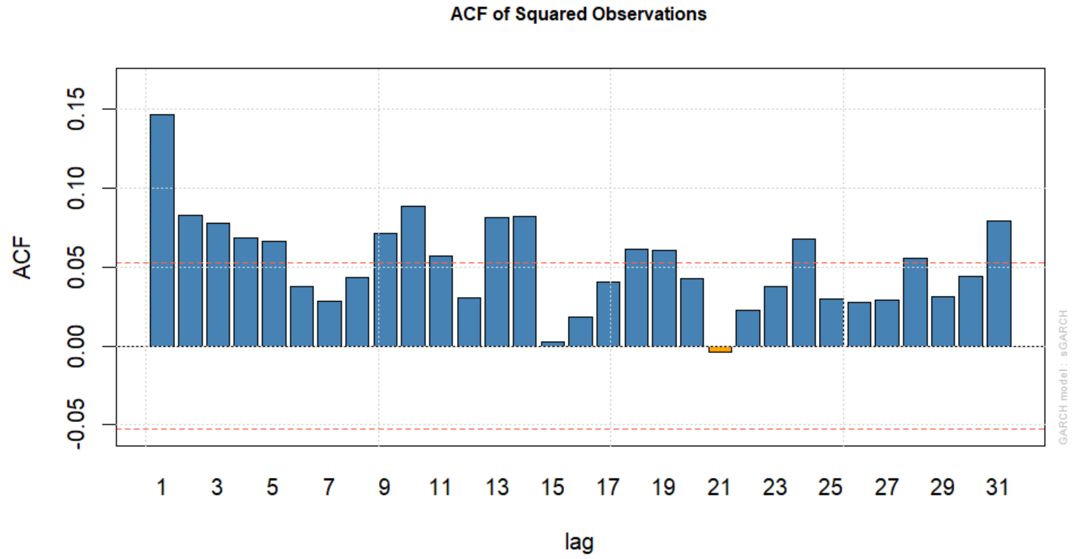

Figure 4.2: ACF of squared returns

In terms of the distributions, the QQ normal scores plot shows the distribution has a heavier tail than a normal distribution.

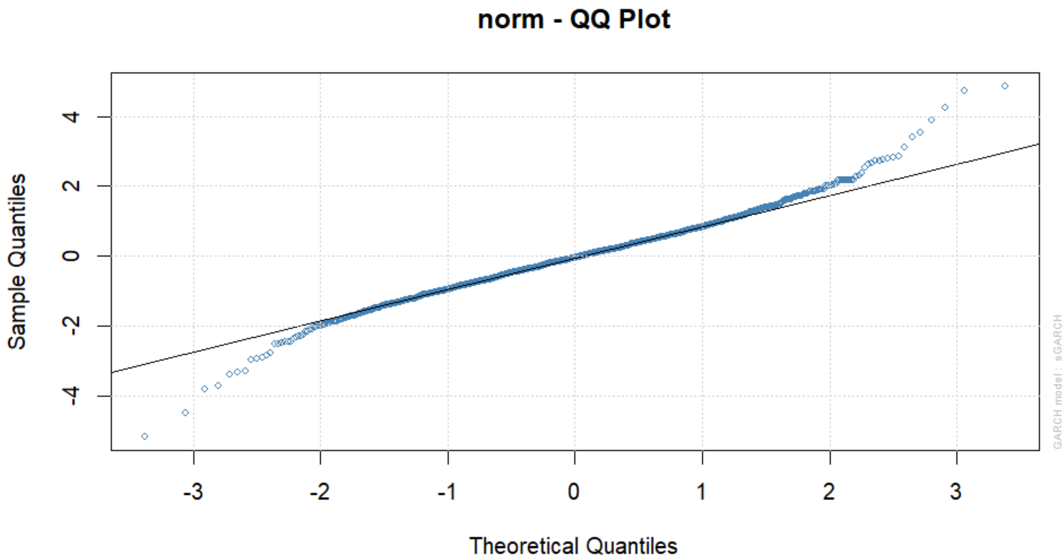

Figure 4.3: QQ plot of the return distribution

Chapter 5

Conclusions and implications

Through this analysis of the time series of data of AMZN data we are able to further understand the tendencies and volatility of the stock data.

We can further use our fitted models to gain an understanding of the expected future returns of AMZN for the next time periods. In particular using ARIMA(1,3,5) we get the following forecast predictions:

Table 5.1: ARIMA(1,3,5) Forecast on next 10 periods

Point	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
1376	-0.01196325	-0.04297373	0.01904724	-0.05938970	0.03546320
1377	-0.01356750	-0.04542166	0.01828667	-0.06228424	0.03514924
1378	-0.01295029	-0.04619755	0.02029698	-0.06379759	0.03789702
1379	-0.01407030	-0.04822562	0.02008503	-0.06630636	0.03816577
1380	-0.01422331	-0.04957668	0.02113005	-0.06829162	0.03984499
1381	-0.01496092	-0.05137428	0.02145245	-0.07065036	0.04072852
1382	-0.01539287	-0.05297854	0.02219281	-0.07287520	0.04208946
1383	-0.01602995	-0.05476154	0.02270164	-0.07526481	0.04320491
1384	-0.01657910	-0.05651275	0.02335456	-0.07765235	0.04449416
1385	-0.01720846	-0.05835617	0.02393924	-0.08013845	0.04572152

The GARCH model yields the following volatility predictions for the next 10 periods

Table 5.2: ARIMA(1,3,5) Forecast on next 10 periods

Point	Volatility
1376	0.01696779
1377	0.01728461
1378	0.01758735
1379	0.01787695
1380	0.01815426
1381	0.01842004
1382	0.01867497
1383	0.01891969
1384	0.01915476
1385	0.01938072

Knowing these forecasts of both returns and volatilities may help investors in determining whether to invest on the AMZN stock, predict price movement, and overall gain an understanding of the tendencies of AMZN shares.

Appendix A

Code Details

Formatting of the code in the document looked illegible, so I am including a link to my github repository of the code:

`https://github.com/inesxblanco/AMZNtimeseries`