Time Series Research Project

Matej Rojec

June 26, 2023

Contents

List of Figures

List of Tables

Abstract

In the project we dealt with two tasks. The first was modeling five time series of hourly-ground-levels of O_3 on an hourly time frame with a `SARIMA-type` model. The data was collected at the following stations of the qualar network: Alfragide, Reboleira, Olivais, Beato and Entrecampos. The second was fitting a `GARCH-type` model to five time of log-returns based on the closing price. The data was taken from Euronext Lisbon from 2022-01-05 to 2022-10-12 and contain daily share related-prices from five companies Altri-SGPS, Mota Engil, Ren, Sonae and The Navigator Company.

In the first task we started by taking a look at the sample ACFs and the sample PACFs of the five time series. This led us to the initial identification of the dependence orders. We found a daily seasonality in the data and that data with a daily seasonal difference and an hourly non-seasonal difference turned the time series into a stationary time series. This allowed us to have a starting model and make other models that were derived from it. The models parameters were estimated with the help of the `arima` function in `R`. When we had the models made we looked at the residual analysis of each model and the AIC of each model and based on that we chose the final model for each of the five time series. We looked at the sample ACF of the standardised residuals, if they form an i.i.d. sequence and at the p values for the Ljung-Box statistic. For Alfragide the model $\texttt{SARIMA}(1,0,1)\times(1,1,1)_{24}$ was chosen, for Reboleira the model $\texttt{SARIMA}(3,0,1)\times(1,0,1)_{24}$ was chosen, for Beato the model $\texttt{SARIMA}(1,0,1)\times(1,0,2)_{24}$ was chosen, for Olivais the model $\texttt{SARIMA}(3,0,1)\times(1,0,1)_{24}$ was chosen and for Entrecampos the model $\texttt{SARIMA}(1,0,1)\times(1,0,1)_{24}$ was chosen. We used the five chosen models to forecast the data into the future up to 5 time periods ahead and calculate 95% prediction intervals for each of the 5 forecasts. We did this with the help the `predict` function in `R`.

In the second task we started by taking a look at the sample ACFs five time series. The sample ACF was negligible at all lags. The sample mean of the data is close to zero whereas the sample variance is of the order 10^{-4} and we noticed asymmetric responses in the volatility. We started by firstly fitting an `ARCH(p)` model. Then we fit `GARCH(p,q)`, `IGARCH(p,q)`, `GARCH-M(p,q)` and `APARCH(p,q)` based on the best choises of `p` for the `ARCH(p)` model. The models parameters were estimated with the help of the `fGarch` and `rugarch` packages in `R`. When we had the models made we looked at the residual analysis of each model and the AIC of each model and based on that we chose the final model for each of the five time series. We looked at the ACF of the standardised residuals, if they form an i.i.d. sequence and at the normal Q-Q plot for the standardised residuals. For the log-return of The Navigator Company, Mota Engil, Sonae and Altri-SGPS the `APARCH(1,1)` was chosen. On the other hand for the log-return of Ren the `APARCH(1,2)` was chosen. The choice of model indicates that there really is an asymmetric responses in the volatility.

1 Modeling a time series of hourly-ground-levels of O_3

1.1 Introduction

Ground-level ozone (O_3) is a type of air pollutant that occurs near the Earth's surface. It is not emitted directly into the atmosphere but is formed through complex chemical reactions between precursor pollutants in the presence of sunlight. The two primary precursor pollutants involved in the formation of ground-level ozone are nitrogen oxides (NO_x) and volatile organic compounds (VOCs).

NOx and VOCs are emitted from various sources, including vehicle emissions, industrial processes, power plants, and chemical solvents. When these pollutants are released into the atmosphere, they undergo chemical reactions in the presence of sunlight and react to form ground-level ozone.

Ground-level ozone is considered a harmful air pollutant and a significant component of smog. High concentrations of ozone can have adverse effects on human health, vegetation, and the environment. It can cause respiratory problems, exacerbate asthma, and contribute to lung inflammation. Ozone exposure can also harm plants, reduce crop yields, and damage ecosystems.

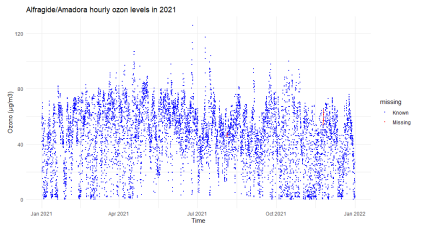
Monitoring and controlling ground-level ozone levels are essential for air quality management and public health protection. Ozone levels are typically monitored using ground-based monitoring stations that measure the concentration of ozone in the ambient air. This data is used to assess air quality, inform regulatory actions, and provide warnings to the public during periods of high ozone concentrations. [1]

Our task will be to fit a `SARIMA-type` model for five time series. The time series are hourly-ground-levels of O3 values, in micrograms per cubic meter ($\mu g/m^3$), collected in 2021 at the following five stations of the qualar network [2]:

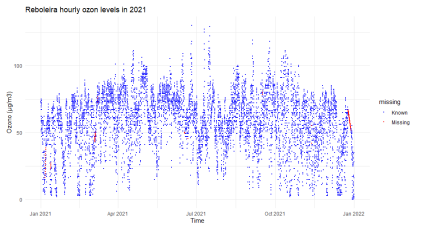
- Alfragide (Amadora)
- Reboleira (Amadora)
- Olivais (Lisboa)
- Beato (Lisboa)
- Entrecampos (Lisboa).

1.2 Examination of the data

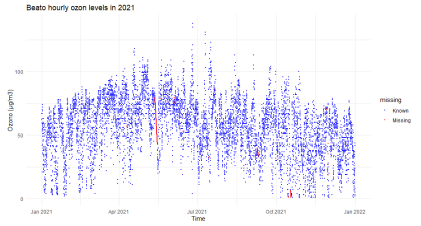
The five data sets include 8760 rows each, one for each hour in the year 2021. Each of the five data sets has missing values. To cope with these values we will use the function `na_interpolation` from the package `imputeTS` in `R`. The function performs linear interpolation to fill in the missing values. Linear interpolation estimates missing values by assuming a straight line between the nearest non-missing values on either side of the missing value.



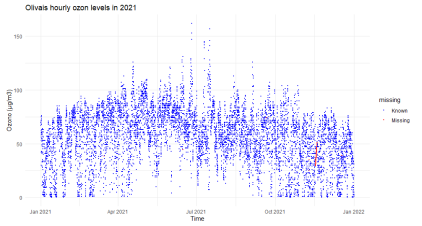
(a) Amadora time series of hourly-ground-levels of ozon in 2021.



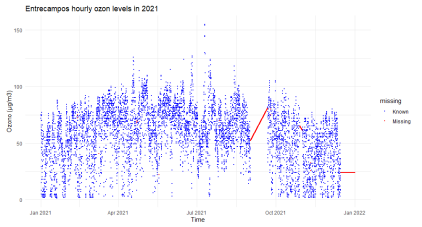
(b) Reboleira time series of hourly-ground-levels of ozon in 2021.



(c) Beato time series of hourly-ground-levels of ozon in 2021.



(d) Olivais time series of hourly-ground-levels of ozon in 2021.



(e) Entrecampos time series of hourly-ground-levels of ozon in 2021.

Figure 1: Plots of the time series of hourly-ground-levels of ozon in 2021.

Figure 1 shows us the five time series and when there was missing data in the time series. It can be observed from figure 1 that Entrecampos has the most missing data. Let's also look at the correlation between the different time series. We can see from table 1 that there seems to be a big correlation between the time series, especially between Beato and Olivais.

This correlation can be further supported by figure 2 that shows all the time series together in the first five days of 2021. The figure also shows that there is a daily periodicity in the data. From the figure we see that the means are not invariant with time. This implies that the time series are not stationary.

Ozone levels often exhibit patterns due to various factors, including meteorological conditions, sunlight intensity, temperature, and the presence of precursor pollutants. In many regions, ozone levels tend to be higher during the summer months when there is increased sunlight and warmer temperatures. These conditions promote the formation of ozone through chemical reactions involving

	Alfragide\Amadora	Reboleira	Entrecampos	Olivais	Beato
Alfragide \Amadora	1.00	0.85	0.81	0.87	0.86
Reboleira	0.85	1.00	0.80	0.86	0.85
Entrecampos	0.81	0.80	1.00	0.87	0.86
Olivais	0.87	0.86	0.87	1.00	0.93
Beato	0.86	0.85	0.86	0.93	1.00

Table 1: Correlation matrix for the hourly-ground-levels of ozon time series.

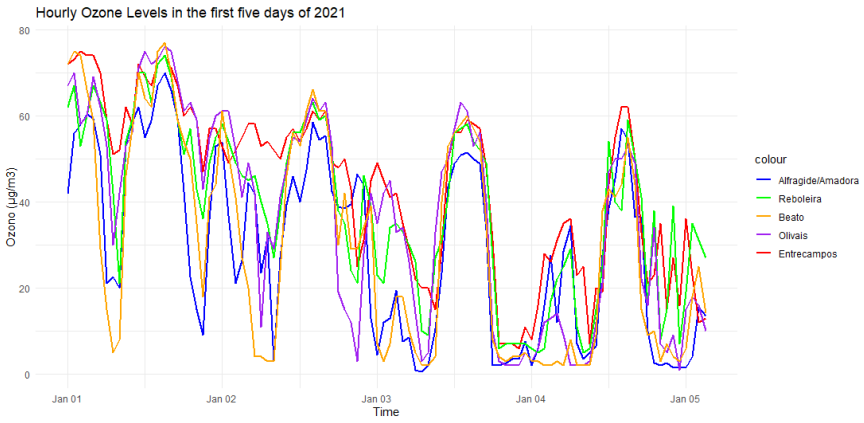


Figure 2: Comparison of the five time series of hourly-ground-levels of ozon in the first five days of 2021.

nitrogen oxides (NO_x) and volatile organic compounds (VOCs). This component can be observed in figure 1 in all five time series. There is also a daily seasonality in the data as seen from figure 2. During the day there are higher levels of O_3 and during the night there are lower levels of O_3.

We can further analyse the seasonal-trend decomposition with the help of the function `STL` in `R`. The decomposition for Alfragide can be seen on figure 3. The figure further supports that there is a seasonal as well as a trend component to the data. The same conclusion follows for the other four time series.

From figure 1 we can notice times with higher and lower peaks. Therefor we need to stabilise the data with some sort of transformation to get a stationary time series. We will look at three possible transformations of the data:

1. logarithmic transformation of the data;
2. 24 hour seasonal difference transformation of the data;
3. 24 hour seasonal and a 1 hour non-seasonal difference transformation of the data.

The three time series can be seen in figure 4. From them we can see that the logarithmic transformation and the the 24 hour seasonal difference transfor-

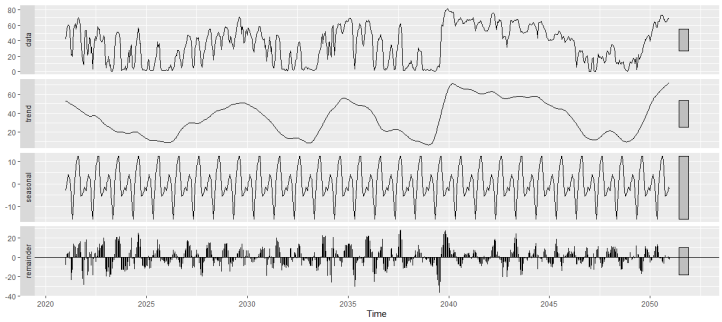
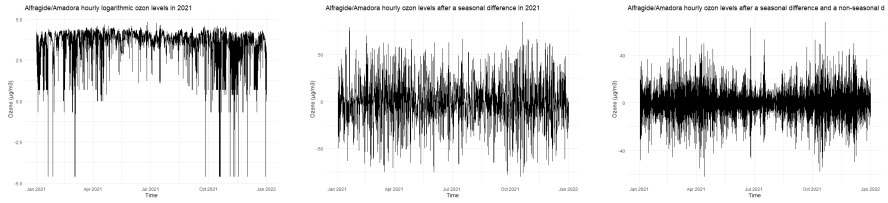


Figure 3: Seasonal-trend decomposition for time series of hourly-ground-levels of ozon in Alfragide for the first month of 2021.

mation of the data does not look like a stationary time series while the 24 hour seasonal and a 1 hour non-seasonal difference transformation of the data seems to resemble a stationary time series.



(a) Logarithmic transformation of the data.

(b) 24 hour seasonal difference transformation of the data.

(c) 24 hour seasonal and a 1 hour non-seasonal difference transformation of the data.

Figure 4: Transformed time series of hourly-ground-levels of ozon in Alfragide.

1.3 Model fitting

Firstly let us look at the sample ACF and the sample PACF of the initial data and the log transformed data. The sample ACF, as seen from figures 5 and 6, for both the log transform and original data doesn't exhibit the classic exponential drop to 0 displayed by clear `AR(p)` processes. However, we do see a pattern that tends somewhat towards zero, but which may not be as clear-cut due to the daily seasonality we observed. In other words, the seasonality may be masking the exponential drop towards zero since it shows up periodically in the sample ACF plot. In addition, we can see in both sample PACFs that lags after the dashed lines (24 multiples) exceed the confidence bounds. This is in line with our hypothesized daily seasonality. Note that immediately before the dashed lines, we also see some lags that exceed the confidence bounds and slowly fade after 2 days of lag. This corresponds to the last hours of the day (i.e. the

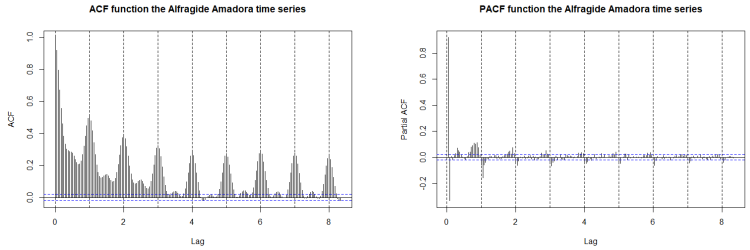
evening) and potentially shows that there might be some correlation during those hours of the day across the board.

If we look at the seasonally differentiated data in figure 7 we can see that sample ACF is decreasing to 0 faster, even though there are still spikes at multiples of 24. The sample PACF is more clear in this case with big spikes at multiples of 24.

If we look at the sample ACF and sample PACF data after a seasonal and a non-seasonal difference as seen in figure 8 we can notice that the time series seems stationary. Furthermore, the sample ACF and the sample PACF now behave as something that we can interpret. Specifically, the sample ACF shows a seasonal spike at lag 24 (as expected), and the sample PACF shows seasonal spikes at multiples of 24, which suggest a seasonal MA component.

In addition, we can see in both sample PACFs that lags after the dashed lines (24 multiples) exceed the confidence bounds. This is in line with our hypothesized daily seasonality. Even though the sample ACF doesn't drop to 0 sufficiently fast as classical examples of `AR(p)` processes, we suspect the daily periodicity maybe masking this and hence we move forward towards formulating our initial model.

Having this in consideration we can start with an initial model as $\texttt{SARIMA}(0,1,0) \times (1,1,0)_{24}$, since we see that the series is stationary and there with the seasonal and non-seasonal transformation. Another model we will consider as an initial model is $\texttt{SARIMA}(0,0,0) \times (1,0,0)_{24}$ as we saw daily seasonality in the data. Further, to start with a model as simple as possible, we can for now ignore any `MA(q)` components in the initial model.

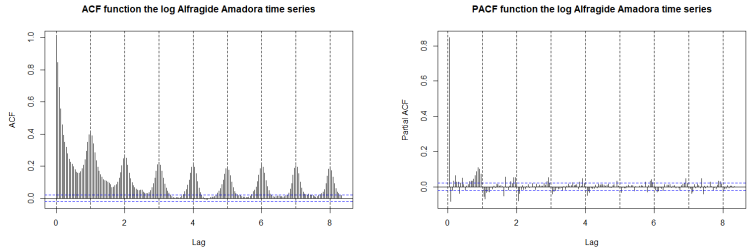


(a) Sample ACF for the Alfragide data. (b) Sample PACF for the Alfragide data.

Figure 5: Sample ACF and sample PACF for the time series of hourly-ground-levels of ozon in Alfragide.

1.4 Model selection and residual diagnostics

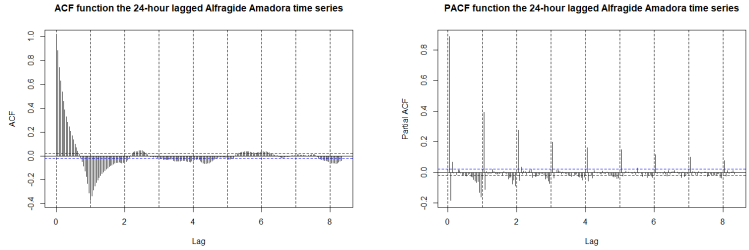
As discussed in chapter 1.3 we can make two initial models to start with. Another model we will take as our initial model is the one produced by the function `auto.arima` in `R`. That being $\texttt{SARIMA}(5,1,0) \times (2,0,0)_{24}$. So the three initial models are therfore:



(a) Sample ACF for the log transformed Alfragide data.

(b) Sample PACF for the log transformed Alfragide data.

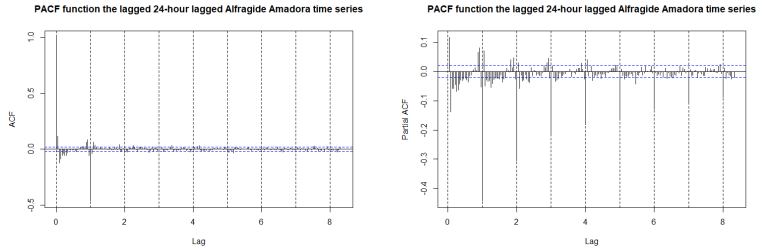
Figure 6: Sample ACF and sample PACF for the for the log transformed time series of hourly-ground-levels of ozon in Alfragide.



(a) Sample ACF for the seasonally differentiated Alfragide data.

(b) Sample PACF for the seasonally differentiated Alfragide data.

Figure 7: Sample ACF and sample PACF for the seasonally differentiated time series of hourly-ground-levels of ozon in Alfragide.



(a) Sample ACF for the seasonally differentiated and the non seasonally differentiated Alfragide data.

(b) Sample PACF for the seasonally differentiated and the non seasonally differentiated Alfragide data

Figure 8: Sample ACF and sample PACF for the seasonally differentiated and the non seasonally differentiated time series of hourly-ground-levels of ozon in in Alfragide.

1. $\texttt{SARIMA}(0,1,0) \times (1,1,0)_{24}$,
2. $\texttt{SARIMA}(0,0,0) \times (1,0,0)_{24}$.
3. $\texttt{SARIMA}(5,1,0) \times (2,0,0)_{24}$.

From there on we will iterate through different models. We take the initial models above and try different variations of them and compare them (when relevant) using the AIC measure. However, we can only use AIC within models of the same seasonal difference order. Therefore, we also compute the RMSE of each model to do cross-comparison across models of different seasonal difference orders.

In table 2 we can see the AIC and RMSE for the fitted models.

Model	**AIC**	**RMSE**
$\texttt{SARIMA}(0,1,0) \times (1,1,0)_{24}$	63386	7.7
$\texttt{SARIMA}(0,0,0) \times (1,0,0)_{24}$	76182	18.7
$\texttt{SARIMA}(5,1,0) \times (2,0,0)_{24}$	61144	7.9
$\texttt{SARIMA}(1,0,1) \times (2,1,0)_{24}$	61722	8.3
$\texttt{SARIMA}(3,0,1) \times (1,0,1)_{24}$	59934	7.4
$\texttt{SARIMA}(3,0,0) \times (1,0,0)_{24}$	60834	7.8
$\texttt{SARIMA}(1,0,1) \times (1,0,1)_{24}$	59931	7.4
$\texttt{SARIMA}(2,0,1) \times (1,1,1)_{24}$	59781	7.4
$\texttt{SARIMA}(2,1,1) \times (1,1,1)_{24}$	59812	7.4
$\texttt{SARIMA}(1,0,1) \times (1,0,2)_{24}$	59931	7.4

Table 2: AIC and RMSE of the fitted models for the Alfragide ozon-ground level time series.

Model results can be found in table 2 with the models with the lowest AIC and the lowest RMSE being $\texttt{SARIMA}(2,0,1) \times (1,1,1)_{24}$ and $\texttt{SARIMA}(1,0,1) \times (1,1,1)_{24}$.

For estimation of parameters we have used the `arima` function in `R`. The $\texttt{SARIMA}(2,0,1) \times (1,1,1)_{24}$ model has the following parameters found in table 3. We see that the $p-$value of the parameter ar2 is very high leading to the conclusion that it is not important for the model. As such we will consider the model $\texttt{SARIMA}(1,0,1) \times (1,1,1)_{24}$ instead of the model $\texttt{SARIMA}(2,0,1) \times (1,1,1)_{24}$.

Parameter	ar1	ar2	ma1	sar1	sma1
Estimate	0.8168	0.0542	0.2765	0.0680	-0.9423
p-value	0.0000	25.0836	0.0000	0.0000	0.0000

Table 3: The parameters and the p-values of the model SARIMA$(2,0,1) \times (1,1,1)_{24}$ for the Alfragide ozon-ground level time series

Let us now take a closer look at the model $\texttt{SARIMA}(1,0,1) \times (1,1,1)_{24}$. The analysis of the residuals can be seen on figure 9. From the figure it can be seen

Time Series	Model
Alfragide	$\texttt{SARIMA}(1,0,1) \times (1,1,1)_{24}$
Reboleira	$\texttt{SARIMA}(3,0,1) \times (1,0,1)_{24}$
Beato	$\texttt{SARIMA}(1,0,1) \times (1,0,2)_{24}$
Olivais	$\texttt{SARIMA}(3,0,1) \times (1,0,1)_{24}$
Entrecampos	$\texttt{SARIMA}(1,0,1) \times (1,0,1)_{24}$

Table 4: Final models for the five ozon-ground level time series

that the the standardized residuals are an i.i.d. sequence with mean zero and variance one. From the sample ACF we do not see any patterns of large values. This is also supported by the Ljung-Box test, since the p-values are quite large.

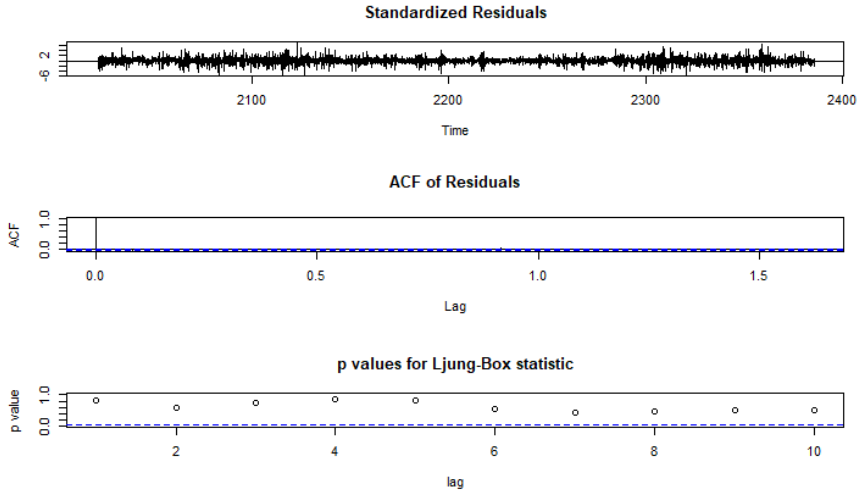


Figure 9: Residual analysis of the SARIMA$(1,0,1) \times (1,1,1)_{24}$ model for the hourly-ground-levels of ozon in Alfragide time series.

Let's look also at the residual analysis of another model for comparison. Let's look at the model $\texttt{SARIMA}(3,0,0) \times (1,0,0)_{24}$. The residuals, as seen on figure 10, are also a i.i.d. sequence but the $p-$values for Ljung-Box statistic go to zero after lag 5, so that is why we prefer the first model. If we look at the residual analysis of the other models we come to the same conclusion

To chose the final model for the other four time series we perform the same analysis. In table 4 we have the best model for each of the time series. The choices of the final models were based on the same analysis that we did for the final choice of the final model for the time series Alfragide. That being the residuals diagnostics and model choice based on the AIC metrics. The AIC metrics can be found in appendix A in the tables 8, 9, 10 , 11.

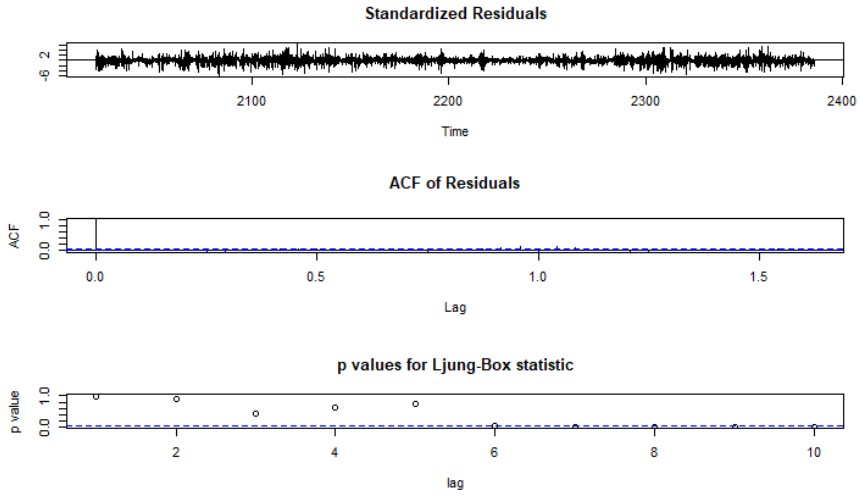


Figure 10: Residual analysis of the SARIMA$(1,0,1) \times (1,1,1)_{24}$ model for the hourly-ground-levels of ozon in Alfragide time series.

1.5 Prediction

We have now decided on the appropriate models and now we will forecast the data into the future up to 5 time periods ahead and calculate 95% prediction intervals for each of the 5 forecasts.

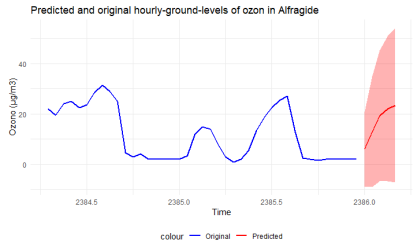
We will do this with the help of `R` with the function `predict`. From this we get predictions for five steps ahead and their respected standard errors. To get the 95% interval for each prediction we apply the following formula:

$$\text{CI}_{95} = [f - 2 \cdot \sigma, f + 2 \cdot \sigma],$$

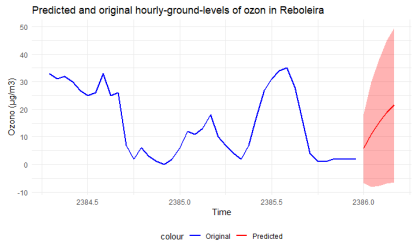
where σ is the standard errors of the forecasted value and f is the forecasted value.

The results for each of the five time series can be seen on figure 11.

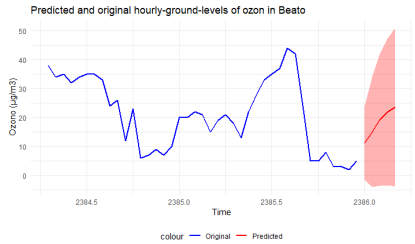
1.6 Conclusion



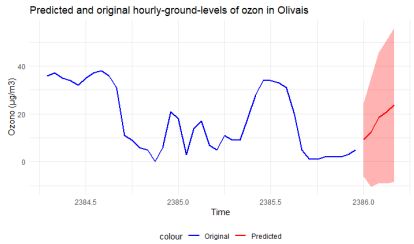
(a) Predicted and original hourly-ground-levels of ozon in Alfragide.



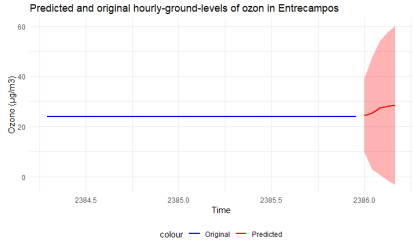
(b) Predicted and original hourly-ground-levels of ozon in Reboleira.



(c) Predicted and original hourly-ground-levels of ozon in Beato.



(d) Predicted and original hourly-ground-levels of ozon in Olivais.



(e) Predicted and original hourly-ground-levels of ozon in Entrecampos.

Figure 11: Predicted and original hourly-ground-levels of ozon in the five qualar networks.

2 Modeling a time series of hourly-ground-levels of log-returns

2.1 Introduction

In finance, log-returns are often used to measure the relative change in the price or value of a financial asset over a certain period of time. The log-return of an asset is defined as the natural logarithm of the ratio of its price at time $t+1$ that is P_{t+1} to its previous price at time t that is P_t, expressed as a decimal or percentage.

The formula for log-returns at time $t+1$ can be expressed as:

$$\text{Log-Return}_{t+1} = \ln\left(\frac{P_{t+1}}{P_t}\right).$$

Log-returns provide several advantages, including the ability to simplify calculations involving asset returns and their statistical properties. Additionally, log-returns can be added over time periods to obtain the cumulative log-return.

We will be modeling five financial time series from 2022-01-05 to 2022-10-12 with the help of log-returns, namely the daily share related-prices from five companies on Euronext Lisbon:

- Altri-SGPS,
- Mota Engil,
- Ren,
- Sonae,
- The Navigator Company.

In our case the log-returns will correspond to the daily closing prices.

2.2 Examination of the data

The five data sets include 199 rows each, one for each day from 2022-01-05 to 2022-10-12 including the data: date, open, high, low, close, number of shares, number of trades, turnover and vwap. In our time series we will have 198 rows each containing the log-returns corresponding to the daily closing prices and the date. In the data there is no missing values so we do not need to interpolate.

Figure 12 shows the log-return of all five financial time series. We can calculate that the sample mean of the data is close to zero whereas the sample variance is of the order 10^{-4}. From the figure 12 we can see that exceedances of high/low thresholds tend to occur in clusters. We can also conclude from the figure an asymmetric responses in the volatility.

Let us now look at the sample ACF for all of the financial time series and the sample ACF of the squares of the time series. From figure 13 we see that the sample ACF use to be negligible at all lags.

2.3 Model fitting

In this subchapter I will describe the step by step model fitting for the time series of log-returns for The Navigator Company. From the sample ACF of the data we can see that the isn't any linear dependence in the data.

We will start by an initial identification of the dependence orders. We will do this by firstly trying to fit a ARCH-type model to the time series. We fitted models `ARCH(q)` for `q = 1,2,3,4,5,6,7,8,9,10,15,20,40`. The model with the lowest AIC turned out to be `ARCH(3)`. The AIC value was -5.41. The p-value of a_3 in the model was 0.0156 indicating that it might be important for

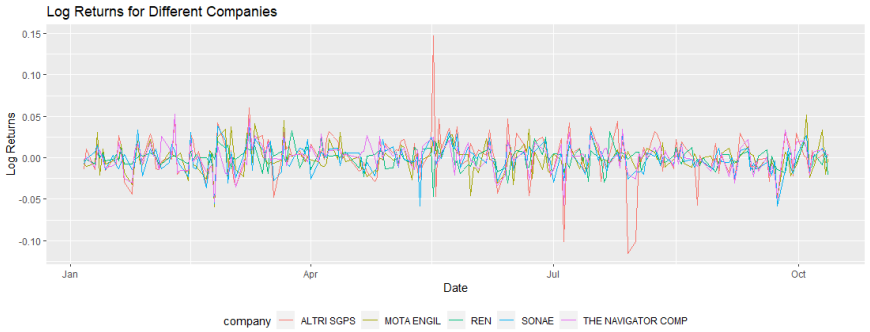
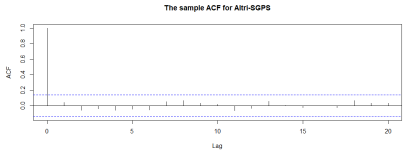
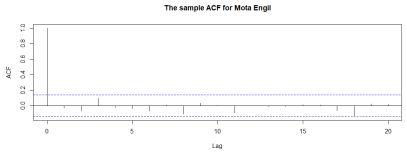


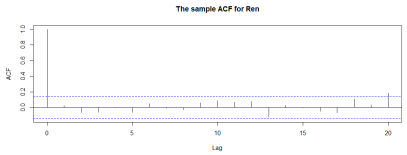
Figure 12: Plot of the log-returns based on the closing prices for the five companies listed on Euronext Lisbon from 2022-01-06 to 2022-10-1.



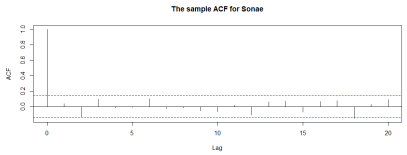
(a) The sample ACF for the log-returns of Altri-SGPS.



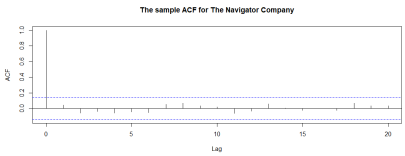
(b) The sample ACF for the log-returns of Mota Engil.



(c) The sample ACF for the log-returns of Ren.



(d) The sample ACF for the log-returns of Sonae.



(e) The sample ACF for the log-returns of The Navigator Company.

Figure 13: The sample ACF for the log-returns of the five companies listed on Euronext Lisbon.

the final model. All of the other a_i in this model (and most of the others) had very huge $p-$values indicating that they might not be significantly influential.

Next we take a look at the `GARCH(p,q)` models hoping to find a better one, where `q = 1,2,3,4,5,6,7,8,9,10,15,20,40` and `p = 1, 3`. The best model

was GARCH(3,1) but with a larger AIC of -5.40.

The next step we take is to try to fit a IGARCH(p, q) model for the same choices of p and q, The model with the lowest AIC was IGARCH(1, 1) with an AIC of -5.38.

Next we fit a GARCH-M(p, q) model for the same choices of p and q, however the AIC of the best model is worst then the AIC of the best GARCH(p,q) model. Model GARCH-M(1, 1) had the best AIC of -5.37.

Finally we fit a APARCH(p, q) model for the same choices of p and q. We find that the APARCH(1, 1) model has the best AIC out of all of the mentioned models with an AIC of -5.46. The model has the parameter $\gamma = 1$ which indicates negative shocks have stronger impact on volatility than positive shocks. This makes sense as we have noticed this in chapter 2.2.

The results are summarised in table 5.

Model	**The Navigator Company**
ARCH(3)	-5.41
GARCH(3,1)	-5.40
IGARCH(1,1)	-5.38
GARCH-M(1,1)	-5.37
APARCH(1,1)	-5.46

Table 5: The AIC of the fitted models for the financial time series of log-returns of The Navigator Company.

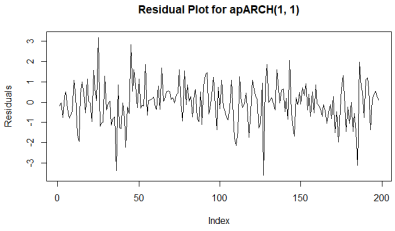
For estimation of parameters we have used the packages fGarch and rugarch.

2.4 Model selection and residual diagnostics

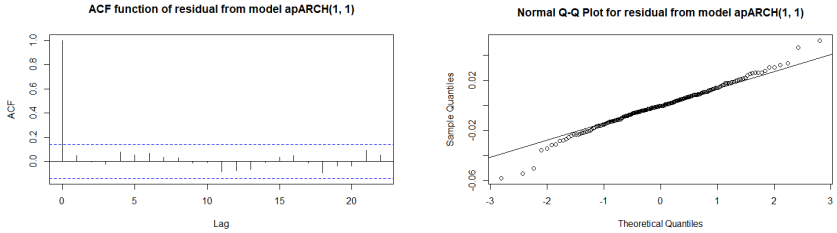
To select the final model we need to analyse the residuals as well as model compare the models.

Let's firstly take a look at the the plot of the standardised residuals, their sample ACF and their Q-Q plot for the model with the lowest AIC, that being the APARCH(1, 1) model. This can be seen in figure 14. From the figure it can be noted that the the standardized residuals follow an i.i.d. sequence with mean zero and variance one. If we take a closer look at the sample ACF there is not any patterns of large values. This statement is also supported by the Box-Ljung test with a $p-$value of 0.481. Also on the Q-Q plot we see that the points lie along a straight line, indicating that they follow an i.i.d. sequence as stated before. Taking all this into account we can conclude that the model is a good fit for the given data.

If we look at the same figures, on figure 15, for second best model, that being the ARCH(3) we see that the residuals also follow an i.i.d. sequence and appear to have no autocorrelation. Since both models exibit the same behaviour we will chose the one with the lower AIC, that being APACH(1, 1). The parameters of the model can be found in table **??**.



(a) Plot of the standardised residuals for the model APARCH(1, 1) for the log-returns of The Navigator Company time series.



(b) The ACF function of the standardised residuals for the model APARCH(1, 1) for the log-returns of The Navigator Company time series.

(c) Q-Q plot for the standardised residuals for the model APARCH(1, 1) for the log-returns of The Navigator Company time series.

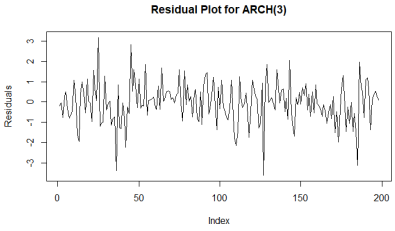
Figure 14: Residual analysis of the APARCH(1, 1) model for the log-returns of The Navigator Company time series.

Parameter	μ	ω	α_1	β_1	γ_1	δ
Estimate	0.001055	0.012674	0.100306	0.768860	1.000000	0.624901
$p-$**value**	0.000000	0.541027	0.110885	0.000006	0.000000	0.034285

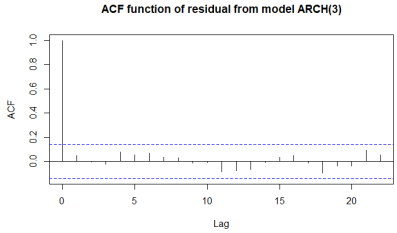
Table 6: Estimates and p-values for the model APACH$(1, 1)$ applied to The Navigator Company time series.

For the other four time series we completed the same exact procedure and the results of the respected models can be found in table

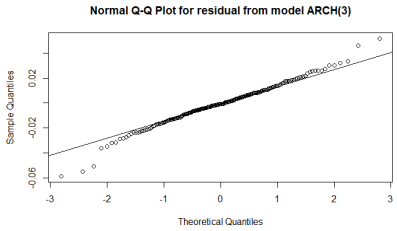
To chose the final model for the other four time series we perform the same analysis. In table 7 we have the chosen model for each of the time series. The choices of the final models were based on the same analysis that we did for the final choice of the model for the log-returns of the Navigator Company time series. That being the residuals diagnostics and model choice based on the AIC metric. The AIC metrics can be found in appendix B in the tables 12 ,13 ,14 and 15. The final models of these models can be found in table 7.



(a) Plot of the standardised residuals for the model ARCH(3) for the log-returns of The Navigator Company time series.



(b) The sample ACF function of the standardised residuals for the model ARCH(3) for the log-returns of The Navigator Company time series.



(c) Q-Q plot for the standardised residuals for the model ARCH(3) for the log-returns of The Navigator Company time series.

Figure 15: Residual analysis of the ARCH(3) model for the log-returns of The Navigator Company time series.

Log-returns time series	Model
The Navigator Company	`APARCH(1,1)`
Mota Engil	`APARCH(1,1)`
Sonae	`APARCH(1,1)`
Ren	`APARCH(1,2)`
Altri-SGPS	`APARCH(1,1)`

Table 7: Final models for the five time series of log-returns.

2.5 Conclusion

A Appendix A: evaluating time series models of hourly ground-level O_3 levels

Model	**AIC**	**RMSE**
$\texttt{SARIMA}(0,1,0)\times(1,1,0)_{24}$	60454	7.7
$\texttt{SARIMA}(0,0,0)\times(1,0,0)_{24}$	75627	18.1
$\texttt{SARIMA}(5,1,0)\times(2,0,0)_{24}$	58194	6.7
$\texttt{SARIMA}(1,0,1)\times(2,1,0)_{24}$	58874	7.0
$\texttt{SARIMA}(3,0,1)\times(1,0,1)_{24}$	57103	6.2
$\texttt{SARIMA}(3,0,0)\times(1,0,0)_{24}$	58032	6.6
$\texttt{SARIMA}(1,0,1)\times(1,0,1)_{24}$	57103	6.3
$\texttt{SARIMA}(2,0,1)\times(1,1,1)_{24}$	56962	6.3
$\texttt{SARIMA}(2,1,1)\times(1,1,1)_{24}$	56962	6.3
$\texttt{SARIMA}(1,0,1)\times(1,0,2)_{24}$	57426	6.4

Table 8: AIC and RMSE of the fitted models for the Reboleira ozon-ground level time series.

Model	**AIC**	**RMSE**
$\texttt{SARIMA}(0,1,0)\times(1,1,0)_{24}$	60659	7.8
$\texttt{SARIMA}(0,0,0)\times(1,0,0)_{24}$	75269	17.7
$\texttt{SARIMA}(5,1,0)\times(2,0,0)_{24}$	58699	6.9
$\texttt{SARIMA}(1,0,1)\times(2,1,0)_{24}$	59222	7.2
$\texttt{SARIMA}(3,0,1)\times(1,0,1)_{24}$	57481	6.4
$\texttt{SARIMA}(3,0,0)\times(1,0,0)_{24}$	58515	6.8
$\texttt{SARIMA}(1,0,1)\times(1,0,1)_{24}$	57499	6.4
$\texttt{SARIMA}(2,0,1)\times(1,1,1)_{24}$	57313	6.4
$\texttt{SARIMA}(2,1,1)\times(1,1,1)_{24}$	57361	6.4
$\texttt{SARIMA}(1,0,1)\times(1,0,2)_{24}$	57432	6.4

Table 9: AIC and RMSE of the fitted models for the Beato ozon-ground level time series.

Model	AIC	RMSE
SARIMA$(0,1,0)\times(1,1,0)_{24}$	63969	9.4
SARIMA$(0,0,0)\times(1,0,0)_{24}$	77832	20.5
SARIMA$(5,1,0)\times(2,0,0)_{24}$	61938	8.3
SARIMA$(1,0,1)\times(2,1,0)_{24}$	62386	8.5
SARIMA$(3,0,1)\times(1,0,1)_{24}$	60603	7.6
SARIMA$(3,0,0)\times(1,0,0)_{24}$	61763	8.2
SARIMA$(1,0,1)\times(1,0,1)_{24}$	60610	7.7
SARIMA$(2,0,1)\times(1,1,1)_{24}$	60454	7.6
SARIMA$(2,1,1)\times(1,1,1)_{24}$	60501	7.7
SARIMA$(1,0,1)\times(1,0,2)_{24}$	60584	7.7

Table 10: AIC and RMSE of the fitted models for the Olivais ozon-ground level time series.

Model	AIC	RMSE
SARIMA$(0,1,0)\times(1,1,0)_{24}$	63315.33	9.1
SARIMA$(0,0,0)\times(1,0,0)_{24}$	77183	19.8
SARIMA$(5,1,0)\times(2,0,0)_{24}$	61160	7.9
SARIMA$(1,0,1)\times(2,1,0)_{24}$	61748	8.2
SARIMA$(3,0,1)\times(1,0,1)_{24}$	60017	7.4
SARIMA$(3,0,0)\times(1,0,0)_{24}$	61063	7.9
SARIMA$(1,0,1)\times(1,0,1)_{24}$	59811	7.4
SARIMA$(2,0,1)\times(1,1,1)_{24}$	59902	7.4
SARIMA$(2,1,1)\times(1,1,1)_{24}$	59918	7.4
SARIMA$(1,0,1)\times(1,0,2)_{24}$	60012	7.4

Table 11: AIC and RMSE of the fitted models for the Entrecampos ozon-ground level time series.

B Appendix B: evaluating time series models of log-returns

Model	Mota Engil
ARCH(8)	-5.48
GARCH(1,1)	-5.46
IGARCH(1,1)	-5.47
GARCH-M(8,1)	-5.47
APARCH(1,1)	-5.55

Table 12: The AIC of the fitted models for the financial time series of log-returns of Mota Engil.

Model	Sonae
ARCH(1)	-5.44
GARCH(1,1)	-5.44
IGARCH(1,1)	-5.44
GARCH-M(1,1)	-5.44
APARCH(1,1)	-5.47

Table 13: The AIC of the fitted models for the financial time series of log-returns of Sonae.

Model	Ren
ARCH(1)	-6.13
GARCH(1,1)	-6.14
IGARCH(1,1)	-6.14
GARCH-M(1,1)	-6.14
APARCH(1,2)	-6.17

Table 14: The AIC of the fitted models for the financial time series of log-returns of Ren.

Model	Altri-SGPS
ARCH(1)	-4.43
GARCH(1,1)	-4.42
IGARCH(1,1)	-4.40
GARCH-M(1,1)	-4.39
APARCH(1,1)	-4.63

Table 15: The AIC of the fitted models for the financial time series of log-returns of Altri-SGPS.

References

[1] *Ground-Level Ozone (O3) Pollution* [last viewed 31. 05. 2023], can be found at https://www.azdeq.gov/ground-level-ozone-o3-pollution

[2] *Qualar* [last viewed 31. 05. 2023], can be found at https://qualar.apambiente.pt/

[3] *Four-day forecast of ground-level ozone from the Copernicus Atmosphere Monitoring Service (CAMS)* [last viewed 02. 06. 2023], can be found at https://climate-adapt.eea.europa.eu/en/observatory/evidence/projections-and-tools/cams-ground-level-ozone-forecast

[4] Z. Junfeng, W. Yongjie, F. Zhangfu, *Ozone Pollution: A Major Health Hazard Worldwide*, Frontiers in Immunology, October 2019.

[5] C. Yong , H. Ling-Yan, H. Xiao-Feng, *Development of a high-performance machine learning model to predict ground ozone pollution in typical cities of China*, Journal of Environmental Management, December 2021.

[6] *Returns and Log Returns* [last viewed 20. 06. 2023], can be found at `https://gregorygundersen.com/blog/2022/02/06/log-returns/`

[7] *Why we use log returns for stock returns* [last viewed 21. 06. 2023], can be found at `https://medium.datadriveninvestor.com/why-we-use-log-returns-for-stock-returns-820cec4510ba`

[8] *Why Log Returns* [last viewed 22. 06. 2023], can be found at `https://quantivity.wordpress.com/2011/02/21/why-log-returns/`

Euronext Lisbon [last viewed 22. 06. 2023], can be found at `https://live.euronext.com/pt/markets/lisbon`

[9] F. Joseph Magnus, F. Abayie, *Modelling and Forecasting Volatility of Returns on the Ghana Stock Exchange Using GARCH Models*, Munich Personal RePEc Archive, October 2006.

[10] C. Shiyi Chen, Härdle. Wolfgang, J. Kiho, *Forecasting volatility with support vector machine-based GARCH model*, Journal of Forcasting, September 2009.

[11] C. Amélie *Forecasting volatility with outliers in GARCH models*, Journal of Forcasting, September 2008.

[12] Notes after the Senior Lecturer's lectures in the elective course Time Series Analysis at Técnico Lisboa in the academic year 2022/23. [last viewed 24. 06. 2023], can be found at `https://fenix.tecnico.ulisboa.pt/disciplinas/STem/2022-2023/2-semestre/slides`