

# Time Series Research Project

Matej Rojec

June 26, 2023

## Contents

<b>1</b>	<b>Modeling a time series of hourly-ground-levels of ozon</b>	<b>4</b>
1.1	Introduction . . . . .	4
1.2	Examination of the data . . . . .	5
1.2.1	Missing values, trend and seasonality . . . . .	5
1.3	The transformations and the initial dependence orders . . . . .	7
1.4	Model selection and residual diagnostics . . . . .	9
1.5	Prediction . . . . .	12
1.6	Conclusion . . . . .	12
<b>2</b>	<b>Modeling a time series of hourly-ground-levels of log-returns</b>	<b>13</b>
2.1	Introduction . . . . .	13
2.2	Examination of the data . . . . .	14
2.3	Model fitting . . . . .	15
2.4	Residual diagnostics and model choice . . . . .	16
<b>A</b>	<b>Appendix A</b>	<b>17</b>

## List of Figures

1	Time series visualisations. . . . .	5
2	Comparison of the five time series of hourly-ground-levels of $O_3$ in the first five days of 2021. . . . .	6
3	Seasonal-trend decomposition for Alfragide in the first month of 2021. . . . .	7
4	Transformed time series data. . . . .	7
5	ACF function for the Alfragide data. . . . .	8
6	ACF function for the log transformed Alfragide data. . . . .	9
7	ACF and PACF function for the seasonally differentiated Alfragide data. . . . .	9
8	ACF and PACF for the seasonally differentiated and normally differentiated Alfragide data. . . . .	9
9	Residual analysis of the $SARIMA(1, 0, 1) \times (1, 1, 1)_{24}$ . . . . .	11
10	Residual analysis of the $SARIMA(1, 0, 1) \times (1, 1, 1)_{24}$ . . . . .	12
11	Time series visualisations. . . . .	13
12	Time series of hourly-ground-levels of $O_3$ visualisation. . . . .	14
13	The sample ACF for the five time series. . . . .	15

## List of Tables

1	Correlation matrix for the hourly-ground-levels of $O_3$ time series. . . . .	6
2	AIC and RMSE of the fitted models for the Alfragide ozon-ground level time series . . . . .	10

3	The parameters and the $p$ -values of the model $\text{SARIMA}(2, 0, 1) \times$ $(1, 1, 1)_{24}$ . . . . .	11
4	Final models for the five time series. . . . .	11
5	AIC of the fitted models for the financial time series of log returns. . . . .	17
6	AIC and RMSE of the fitted models for the Reboleira ozon- ground level time series. . . . .	18
7	AIC and RMSE of the fitted models for the Beato ozon-ground level time series. . . . .	18
8	AIC and RMSE of the fitted models for the Olivais ozon-ground level time series. . . . .	18
9	AIC and RMSE of the fitted models for the Entrecampos ozon- ground level time series. . . . .	19

## Abstract

In the project we analysed the impact of different covariates to understand the impact of them to glycosolated hemoglobin and to diabetes. We conducted a thorough exploratory data analysis, we fitted a gamma regression model, interpreted the model. We concluded that the covariates that have the biggest impact on glycosolated hemoglobin and therefore diabetes are total cholesterol, stabilized glucose, age and waist length in inches.

# 1 Modeling a time series of hourly-ground-levels of ozone

## 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).

$NO_x$  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  $O_3$  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.

### 1.2.1 Missing values, trend and seasonality

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.

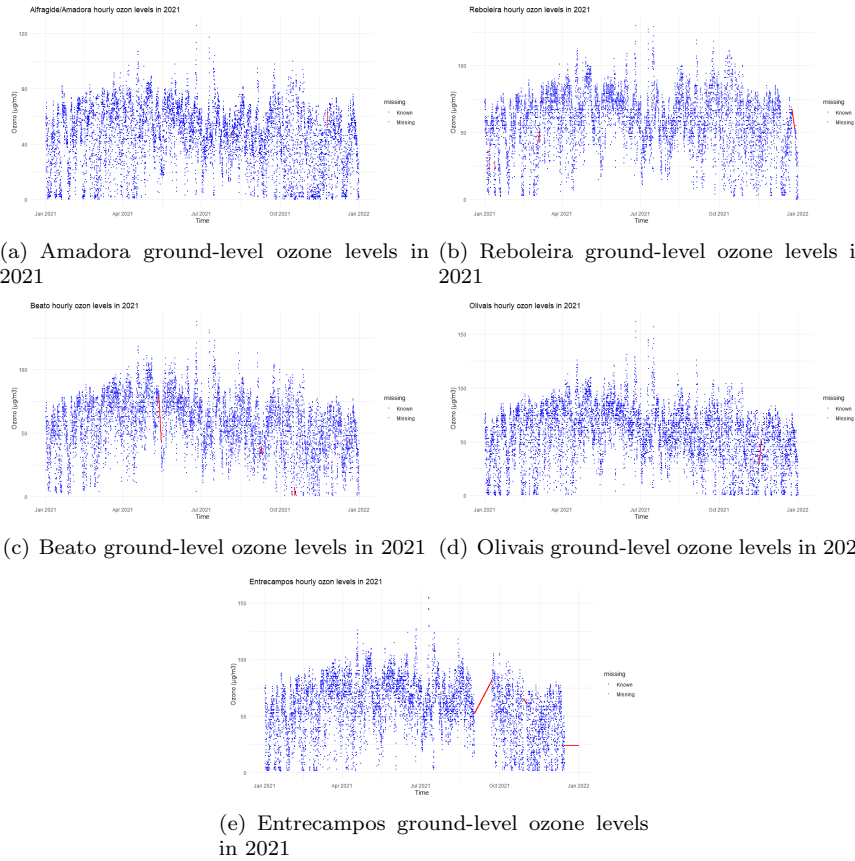


Figure 1: Time series visualisations.

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.

	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 O<sub>3</sub> time series.

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.

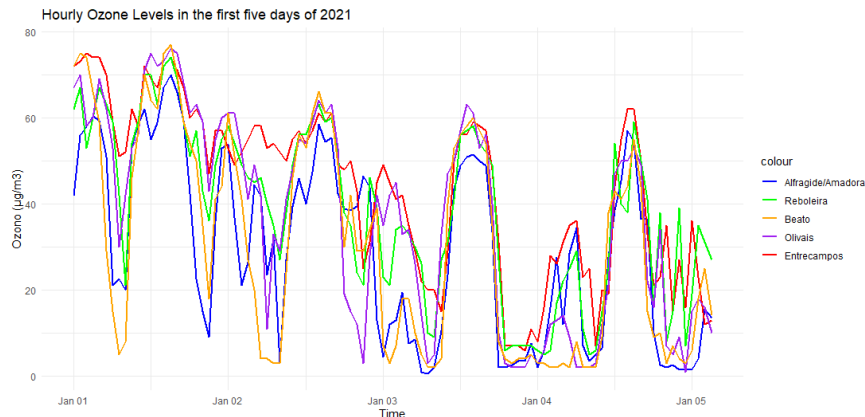


Figure 2: Comparison of the five time series of hourly-ground-levels of O<sub>3</sub> in the first five days of 2021.

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 nitrogen oxides (NO<sub>x</sub>) 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 ozon and during the night there are lower levels of ozon.

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

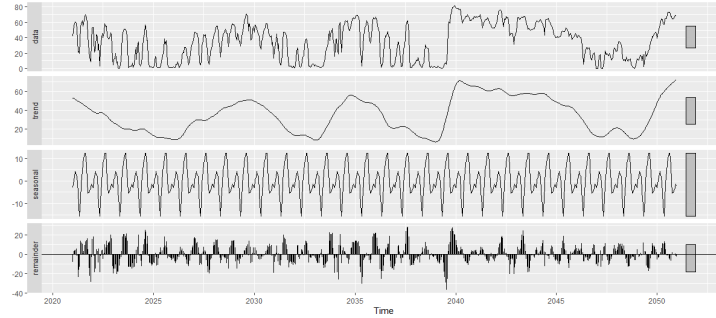


Figure 3: Seasonal-trend decomposition for Alfragide in the first month of 2021.

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 transformation 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.

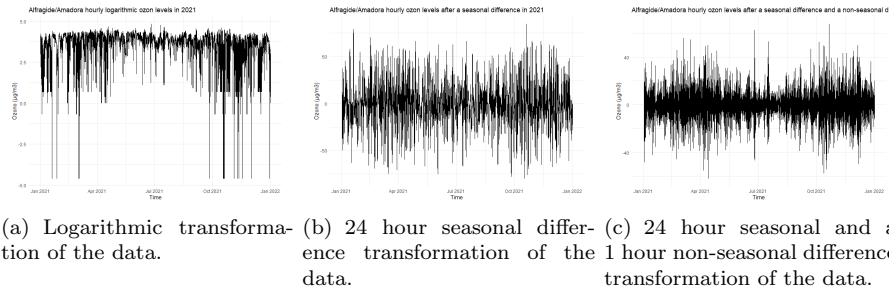


Figure 4: Transformed time series data.

### 1.3 The transformations and the initial dependence orders

Firstly let us look at the ACF and the PACF of the initial data and the log transformed data. The 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  $\text{AR}(\text{p})$  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 ACF plot. In addition, we can see in both 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 ACF is decreasing to 0 faster, even though there are still spikes at multiples of 24. The PACF is more clear in this case with big spikes at multiples of 24.

If we look at the ACF and 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 ACF and the PACF now behave as something that we can interpret. Specifically, the ACF shows a seasonal spike at lag 24 (as expected), and the PACF shows seasonal spikes at multiples of 24, which suggest a seasonal MA component.

In addition, we can see in both 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 ACF doesn't drop to 0 sufficiently fast as classical examples of  $\text{AR}(\text{p})$  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  $\text{SARIMA}(0, 1, 0) \times (1, 1, 0)_{24}$ , 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  $\text{SARIMA}(0, 0, 0) \times (1, 0, 0)_{24}$  as we saw daily seasonality in the data. Further, to start with a model as simple as possible, we can for now ignore any  $\text{MA}(\text{q})$  components in the initial model.

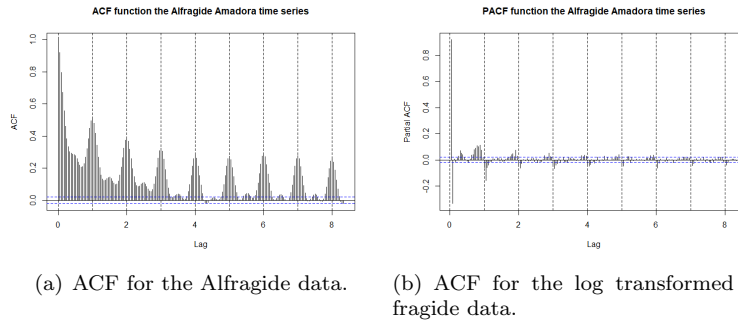


Figure 5: ACF function for the Alfragide data.



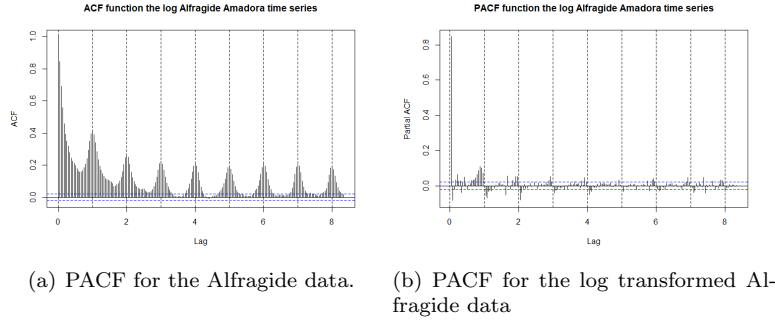


Figure 6: ACF function for the log transformed Alfragide data.

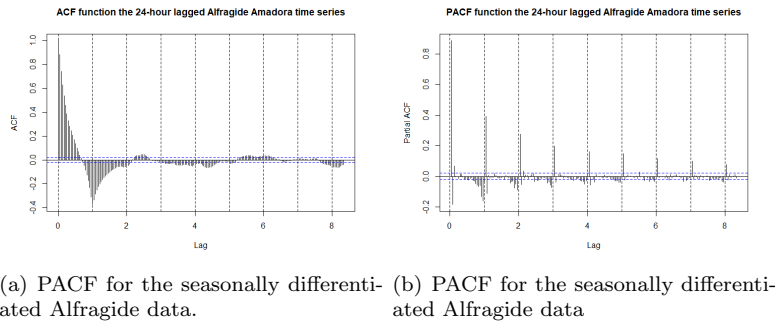


Figure 7: ACF and PACF function for the seasonally differentiated Alfragide data.

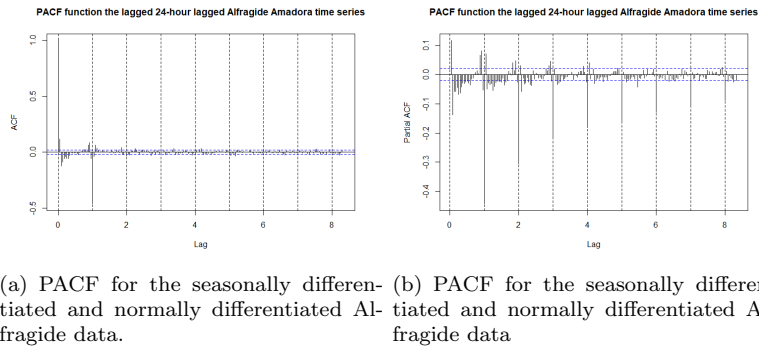


Figure 8: ACF and PACF for the seasonally differentiated and normally differentiated Alfragide data.

## 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  $\text{SARIMA}(5, 1, 0) \times (2, 0, 0)_{24}$ . So the three initial models are therefore:

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

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
$\text{SARIMA}(0, 1, 0) \times (1, 1, 0)_{24}$	63386	7.7
$\text{SARIMA}(0, 0, 0) \times (1, 0, 0)_{24}$	76182	18.7
$\text{SARIMA}(5, 1, 0) \times (2, 0, 0)_{24}$	61144	7.9
$\text{SARIMA}(1, 0, 1) \times (2, 1, 0)_{24}$	61722	8.3
$\text{SARIMA}(3, 0, 1) \times (1, 0, 1)_{24}$	59934	7.4
$\text{SARIMA}(3, 0, 0) \times (1, 0, 0)_{24}$	60834	7.8
$\text{SARIMA}(1, 0, 1) \times (1, 0, 1)_{24}$	59931	7.4
$\text{SARIMA}(2, 0, 1) \times (1, 1, 1)_{24}$	59781	7.4
$\text{SARIMA}(2, 1, 1) \times (1, 1, 1)_{24}$	59812	7.4
$\text{SARIMA}(1, 0, 1) \times (1, 0, 2)_{24}$	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 model with the lowest AIC and the lowest RMSE being  $\text{SARIMA}(2, 0, 1) \times (1, 1, 1)_{24}$ .

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

For estimation of parameters we have used the `arima` function in R. The  $\text{SARIMA}(2, 0, 1) \times (1, 1, 1)_{24}$  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  $\text{SARIMA}(1, 0, 1) \times (1, 1, 1)_{24}$  instead.

Let us now take a closer look at the model  $\text{SARIMA}(1, 0, 1) \times (1, 1, 1)_{24}$ . The analysis of the residuals can be seen on figure 9. From the figure it can be seen that the the standardized residuals are an i.i.d. sequence with mean zero and variance one. From the 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.

Parameter / Value	ar1	ar2	ma1	sar1	sma1
Value	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  $\text{SARIMA}(2,0,1) \times (1,1,1)_{24}$ .

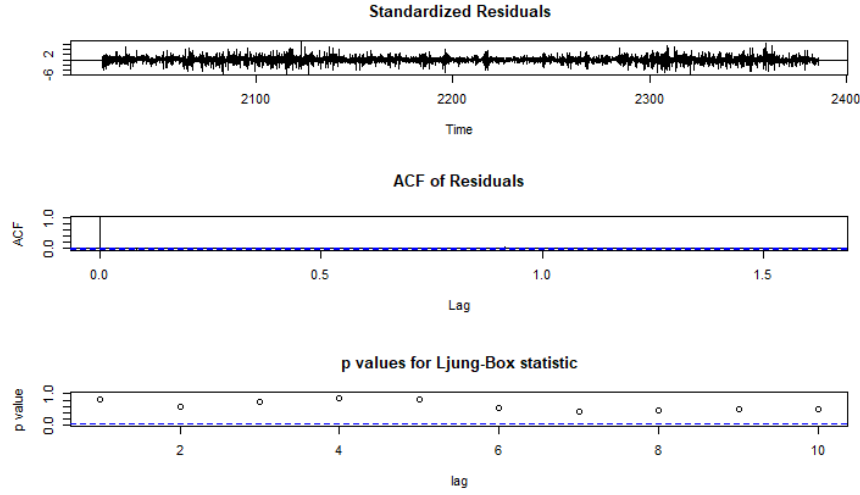


Figure 9: Residual analysis of the  $\text{SARIMA}(1,0,1) \times (1,1,1)_{24}$ .

Let's look also at the residual analysis of another model for comparison. Let's look at the model  $\text{SARIMA}(3,0,0) \times (1,0,0)_{24}$ . The residuals, as seen on figure 10, are also a i.i.d. sequence squance 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

For the other models we preform the same analysis. In table 4 we have the best model for each of the time series.

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

Table 4: Final models for the five time series.

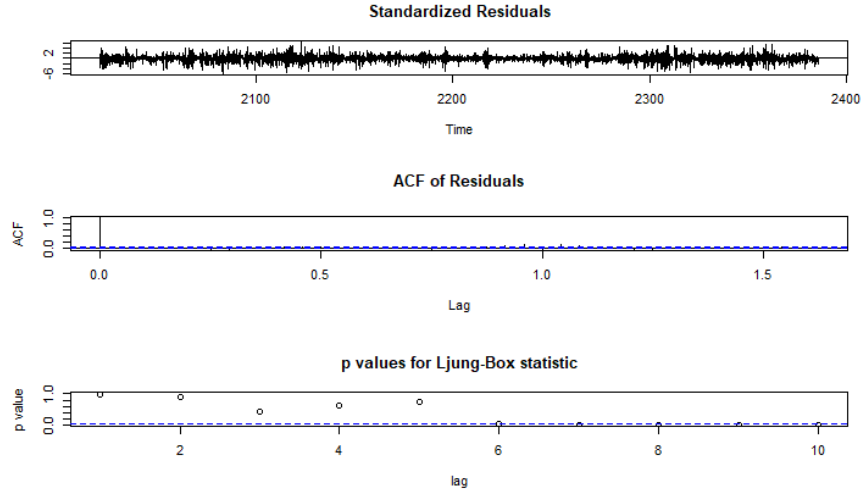


Figure 10: Residual analysis of the  $\text{SARIMA}(1, 0, 1) \times (1, 1, 1)_{24}$ .

## 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.

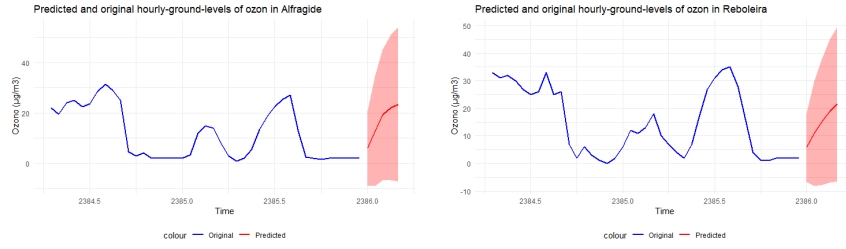
We will do this with the help of R with the function `predict`. From this we get predictions for five steps ahead and their respected standard errors. To get the 95% interval for each prediction we apply the following formula:

$$CI_{95} = [f - 2 \cdot \sigma, f + 2 \cdot \sigma],$$

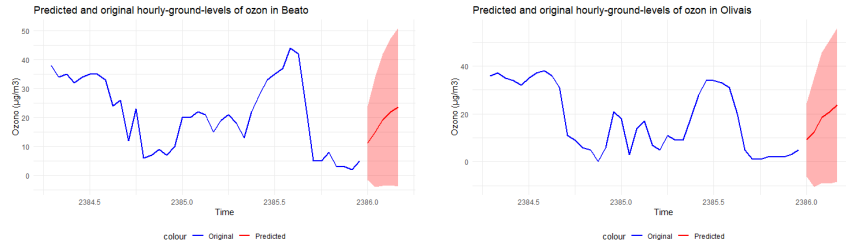
where  $\sigma$  is the standard errors of the forecasted value and  $f$  is the forecasted value.

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

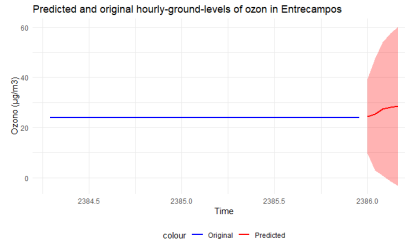
## 1.6 Conclusion



(a) Predicted and original hourly-ground-levels of ozone in Alfragide. (b) Predicted and original hourly-ground-levels of ozone in Reboleira.



(c) Predicted and original hourly-ground-levels of ozone in Beato. (d) Predicted and original hourly-ground-levels of ozone in Olivais.



(e) Predicted and original hourly-ground-levels of ozone in Entrecampos,

Figure 11: Time series visualisations.

## 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 can be expressed as:

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

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.

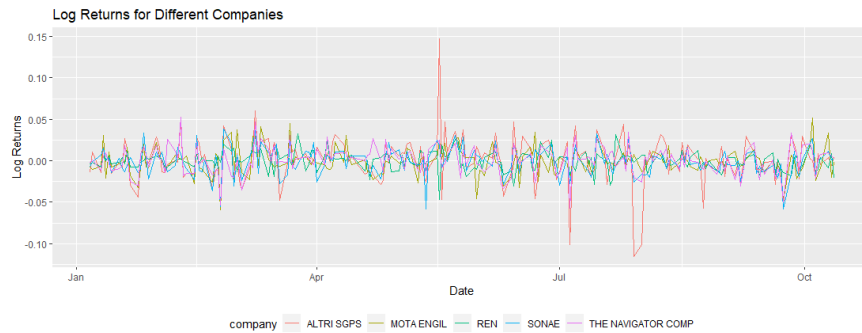


Figure 12: Time series of hourly-ground-levels of  $O_3$  visualisation.

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

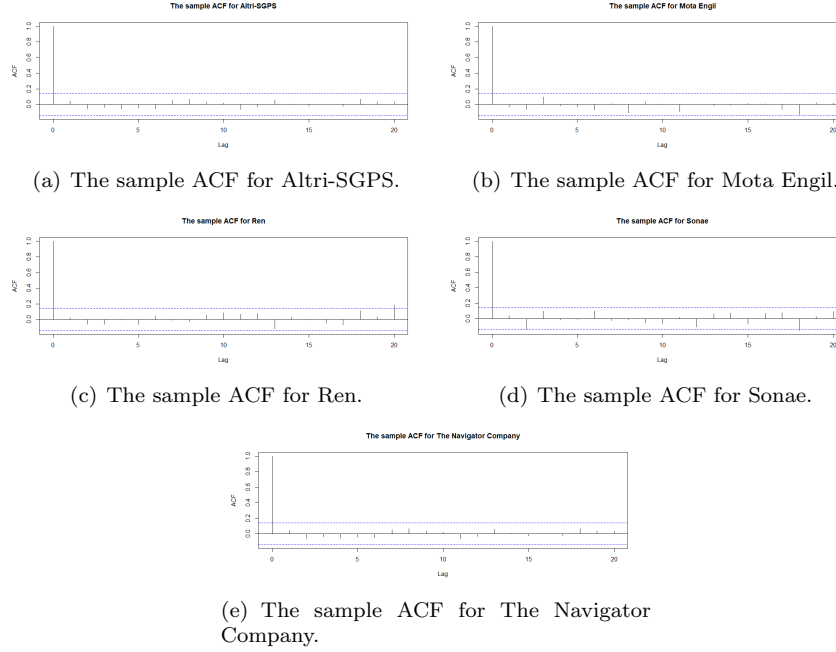


Figure 13: The sample ACF for the five time series.

## 2.3 Model fitting

In this subchapter I will describe the step by step model fitting for the time series Altri-SGPS. From the sample ACF of the data we can see that there 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  $\text{ARCH}(q)$  for  $q = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 40$ . The model with the best AIC turned out to be  $\text{ARCH}(3)$ . 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 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  $\text{GARCH}(p, q)$  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  $\text{GARCH}(3, 1)$  but with a larger AIC of -5.40.

The next step we take is to try to fit a  $\text{IGARCH}(p, q)$  model for the same choices of  $p$  and  $q$ , however the AIC of the best model is worse than the AIC of the best  $\text{GARCH}(p, q)$  model.

Next we fit a  $\text{GARCH-M}(\mathbf{p}, \mathbf{q})$  model for the same choices of  $\mathbf{p}$  and  $\mathbf{q}$ , however the AIC of the best model is worse than the AIC of the best  $\text{GARCH}(\mathbf{p}, \mathbf{q})$  model. Model  $\text{GARCH-M}(1, 1)$  had the best AIC of -5.37.

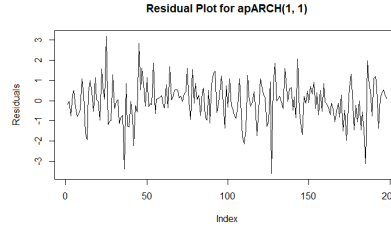
Finally we fit a  $\text{APARCH}(\mathbf{p}, \mathbf{q})$  model for the same choices of  $\mathbf{p}$  and  $\mathbf{q}$ . We find that the  $\text{APARCH}(1, 1)$  model has the best AIC out of all of the mentioned models with an AIC of -5.46. The model has the parameter  $\gamma = 1$  which indicates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

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

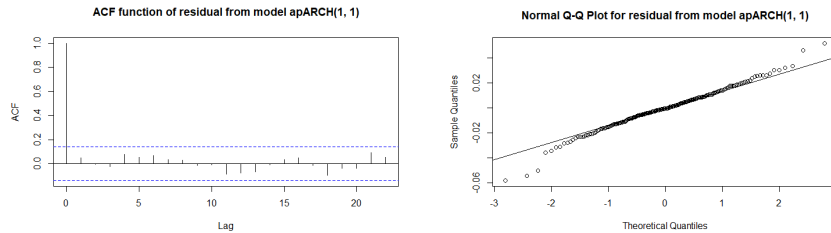
## 2.4 Residual diagnostics and model choice

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 ACF and their Q-Q plot for the model with the lowest AIC, that being the  $\text{APARCH}(1, 1)$  model. This can be seen in figure 2.4. 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 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.



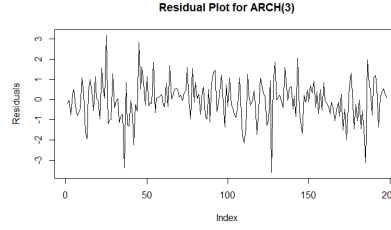
(a) Plot of the standardised residuals for the model  $\text{APARCH}(1, 1)$ .



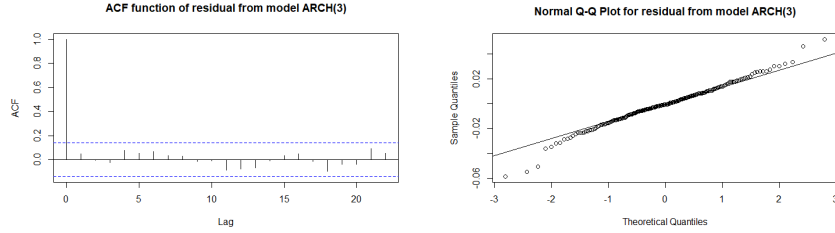
(b) The ACF function of the standardised residuals for the model  $\text{APARCH}(1, 1)$ . (c) Q-Q plot for the standardised residuals for the model  $\text{APARCH}(1, 1)$ .



If we look at the same figures, on figure 2.4, for second best model, that being the  $\text{ARCH}(3)$  we see that the residuals also follow an i.i.d. sequence and appear to have no autocorrelation.



(d) Plot of the standardised residuals for the model  $\text{ARCH}(3)$ .



(e) The ACF function of the standardised residuals for the model  $\text{ARCH}(3)$ . (f) Q-Q plot for the standardised residuals for the model  $\text{ARCH}(3)$ .

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

Model	Altri-SGPS	Mota Engil	Ren	Sonae	The Navigator Com.
$\text{ARCH}(3)$	Value 1	Value 2	Value 3	Value 4	Value 5
$\text{GARCH}(3,1)$	Value 6	Value 7	Value 8	Value 9	Value 10
$\text{GARCH-M}(1,1)$	Value 6	Value 7	Value 8	Value 9	Value 10
$\text{APARCH}(1,1)$	Value 6	Value 7	Value 8	Value 9	Value 10

Table 5: AIC of the fitted models for the financial time series of log returns.

## A Appendix A

Model	AIC	RMSE
SARIMA(0, 1, 0) $\times$ (1, 1, 0) <sub>24</sub>	60454	7.7
SARIMA(0, 0, 0) $\times$ (1, 0, 0) <sub>24</sub>	75627	18.1
SARIMA(5, 1, 0) $\times$ (2, 0, 0) <sub>24</sub>	58194	6.7
SARIMA(1, 0, 1) $\times$ (2, 1, 0) <sub>24</sub>	58874	7.0
SARIMA(3, 0, 1) $\times$ (1, 0, 1) <sub>24</sub>	57103	6.2
SARIMA(3, 0, 0) $\times$ (1, 0, 0) <sub>24</sub>	58032	6.6
SARIMA(1, 0, 1) $\times$ (1, 0, 1) <sub>24</sub>	57103	6.3
SARIMA(2, 0, 1) $\times$ (1, 1, 1) <sub>24</sub>	56962	6.3
SARIMA(2, 1, 1) $\times$ (1, 1, 1) <sub>24</sub>	56962	6.3
SARIMA(1, 0, 1) $\times$ (1, 0, 2) <sub>24</sub>	57426	6.4

Table 6: AIC and RMSE of the fitted models for the Reboleira ozon-ground level time series.

Model	AIC	RMSE
SARIMA(0, 1, 0) $\times$ (1, 1, 0) <sub>24</sub>	60659	7.8
SARIMA(0, 0, 0) $\times$ (1, 0, 0) <sub>24</sub>	75269	17.7
SARIMA(5, 1, 0) $\times$ (2, 0, 0) <sub>24</sub>	58699	6.9
SARIMA(1, 0, 1) $\times$ (2, 1, 0) <sub>24</sub>	59222	7.2
SARIMA(3, 0, 1) $\times$ (1, 0, 1) <sub>24</sub>	57481	6.4
SARIMA(3, 0, 0) $\times$ (1, 0, 0) <sub>24</sub>	58515	6.8
SARIMA(1, 0, 1) $\times$ (1, 0, 1) <sub>24</sub>	57499	6.4
SARIMA(2, 0, 1) $\times$ (1, 1, 1) <sub>24</sub>	57313	6.4
SARIMA(2, 1, 1) $\times$ (1, 1, 1) <sub>24</sub>	57361	6.4
SARIMA(1, 0, 1) $\times$ (1, 0, 2) <sub>24</sub>	57432	6.4

Table 7: 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) <sub>24</sub>	63969	9.4
SARIMA(0, 0, 0) $\times$ (1, 0, 0) <sub>24</sub>	77832	20.5
SARIMA(5, 1, 0) $\times$ (2, 0, 0) <sub>24</sub>	61938	8.3
SARIMA(1, 0, 1) $\times$ (2, 1, 0) <sub>24</sub>	62386	8.5
SARIMA(3, 0, 1) $\times$ (1, 0, 1) <sub>24</sub>	60603	7.6
SARIMA(3, 0, 0) $\times$ (1, 0, 0) <sub>24</sub>	61763	8.2
SARIMA(1, 0, 1) $\times$ (1, 0, 1) <sub>24</sub>	60610	7.7
SARIMA(2, 0, 1) $\times$ (1, 1, 1) <sub>24</sub>	60454	7.6
SARIMA(2, 1, 1) $\times$ (1, 1, 1) <sub>24</sub>	60501	7.7
SARIMA(1, 0, 1) $\times$ (1, 0, 2) <sub>24</sub>	60584	7.7

Table 8: 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) <sub>24</sub>	63315.33	9.1
SARIMA(0, 0, 0) $\times$ (1, 0, 0) <sub>24</sub>	77183	19.8
SARIMA(5, 1, 0) $\times$ (2, 0, 0) <sub>24</sub>	61160	7.9
SARIMA(1, 0, 1) $\times$ (2, 1, 0) <sub>24</sub>	61748	8.2
SARIMA(3, 0, 1) $\times$ (1, 0, 1) <sub>24</sub>	60017	7.4
SARIMA(3, 0, 0) $\times$ (1, 0, 0) <sub>24</sub>	61063	7.9
SARIMA(1, 0, 1) $\times$ (1, 0, 1) <sub>24</sub>	59811	7.4
SARIMA(2, 0, 1) $\times$ (1, 1, 1) <sub>24</sub>	59902	7.4
SARIMA(2, 1, 1) $\times$ (1, 1, 1) <sub>24</sub>	59918	7.4
SARIMA(1, 0, 1) $\times$ (1, 0, 2) <sub>24</sub>	60012	7.4

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

## References

- [1] *Ground-Level Ozone (O3) Pollution* [viewed 31. 05. 2023], can be found at <https://www.azdeq.gov/ground-level-ozone-o3-pollution>
- [2] *Qualar* [viewed 31. 05. 2023], can be found at <https://qualar.apambiente.pt/>
- [3] *Which blood pressure number is important?* [viewed 28. 05. 2023], can be found at <https://www.health.harvard.edu/staying-healthy/which-blood-pressure-number-is-important>
- [4] The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure (2003 Guideline)
- [5] M. Sue Kirkman, Vanessa Jones Briscoe, Nathaniel Clark, Hermes Florez, Linda B. Haas, Jeffrey B. Halter, Elbert S. Huang, Mary T. Korytkowski, Medha N. Munshi, Peggy Soule Odegard, Richard E. Pratley, Carrie S. Swift; Diabetes in Older Adults. *Diabetes Care* 1 December 2012; 35 (12): 2650–2664
- [6] Schorling JB, Roach J, Siegel M, Baturka N, Hunt DE, Guterbock TM, Stewart HL (1997) A trial of church-based smoking cessation interventions for rural African Americans. *Preventive Medicine* 26:92-101.
- [7] *Body Frame Size Measuring Tables* [viewed 28. 05. 2023], can be found at <https://www.disabled-world.com/calculators-charts/body-frame.php>
- [8] Silva, G.L. (2023). *Lecture Notes of Biostatistics*. Lisbon, Instituto Superior Tecnico

- [9] *Blood Glucose (Sugar) Test* [viewed 28. 05. 2023], can be found at <https://my.clevelandclinic.org/health/diagnostics/12363-blood-glucose-test>
- [10] *What is Diabetes?* [viewed 28. 05. 2023], can be found at <https://www.cdc.gov/diabetes/basics/diabetes.html>