Contents

1 Challenge 1 - Analysing daily mean measurements of the nitrous oxides level

1.1 Introduction

A strong statistical method for examining and analysing data points gathered over time is time series analysis [4]. Researchers can use it to spot patterns, trends, and seasonality in the data as well as forecast what will be observed in the future. Time series analysis is essential for understanding and forecasting air pollution levels in the context of environmental monitoring, which can have a big impact on policy, urban planning, and public health [3].

A significant air pollutant that causes smog, acid rain, and ground-level ozone is nitrous oxide (NOx). High NOx concentrations can harm the environment, aggravate pre-existing medical ailments, and create breathing issues. Therefore, it is crucial to accurately predict nitrous oxide levels in order to put in place effective pollution control procedures and guarantee the welfare of both people and ecosystems [1].

In this report, we will analyse a dataset containing daily mean measurements of nitrous oxide levels (in $\mu g/m^3$) recorded at a location on the A23 Purley Way road in the London Borough of Croydon from 1st February 2017 to 30th September 2017 (inclusive). Our goal is to fit a suitable time series model to describe these data, allowing us to better understand the underlying patterns.

1.2 Data Exploration and Preprocessing

To ensure the dataset is of high quality and appropriate for study, it must be explored and preprocessed before fitting a time series model to the nitrous oxide data. This section will include the dataset's first investigation, including any data pretreatment procedures, visualisations, and summary statistics.

The dataset consists of daily mean measurements of nitrous oxide levels (in $\mu g/m^3$) from 1st February 2017 to 30th September 2017. A sample of the data is shown in the table 1.

Table 1: a23_nox Dataset

Date	**Nitrous Oxide Level (μg/m³)**
01/02/2017	86
02/02/2017	49.5
03/02/2017	55.1
...	...
28/09/2017	69.9
29/09/2017	67.9
30/09/2017	63.8

We created a time series plot of the nitrous oxide levels for the entire period (Figure 1) to better understand the structure and properties of the dataset. With the aid of this visualisation, we are better able to spot any trends, seasonality, or atypical patterns in the data that might affect the time series model we select.

We examined the dataset and discovered that there were no missing values, which makes the preprocessing phase simpler. No additional preprocessing processes, such as addressing missing values, were necessary because the data is already full and in a format that is appropriate for time series analysis.

1.3 Model Fitting and Validation

The next step is to fit an appropriate time series model to the data after examining and preparing the dataset. In this section, we will go through how the selected model was fitted, along with any parameter estimation techniques, and how the model fit was verified.

We wanted to further explore the data's stationarity by studying the autocorrelation function (ACF) and partial autocorrelation function (PACF) plots, even though the time series plot indicated that the data might be weakly stationary. These graphs shed light on the underlying structure of the data and enable us to decide whether any changes, like differencing, are required to achieve stationarity.

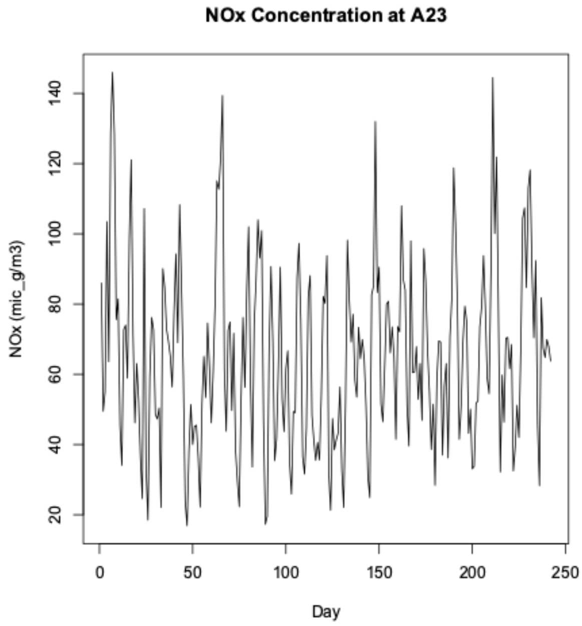


Figure 1: Time Series plot of a23_nox Dataset

The data did not appear to be stationary when we looked at the ACF and PACF plots (Figure 2), as there were strong autocorrelations at delays 7, 14, 21, and so on. This trend raises the possibility that the nitrous oxide levels exhibit weekly seasonality. In order to solve this problem, we chose to do differencing on the data, more specifically, a seasonal differencing with a latency of 7 days. This allowed us to eliminate the weekly seasonality and attain stationarity.

We re-examined the ACF and PACF graphs after applying the seasonal differencing (Figure 3). Since the data still doesn't appear to be stationary, we will use the first difference to obtain a stationary process on the new data (Figure 4). We next proceeded to construct a suitable time series model to the stationary data, such as a Seasonal ARIMA (SARIMA) model, which can explain the seasonal trends in the data as well as the autoregressive and moving average components.

The best model was identified as ARIMA$(1, 1, 1)\times(1, 1, 1)_7$ after we used the maximum likelihood estimation approach to estimate the parameters of various SARIMA models. The residuals, which should ideally be white noise, were examined to assess the model's performance after it had been fitted. This result is depicted in figure 5. We checked the residual plots visually, through ACF plots, and through the Ljung-Box test to make sure

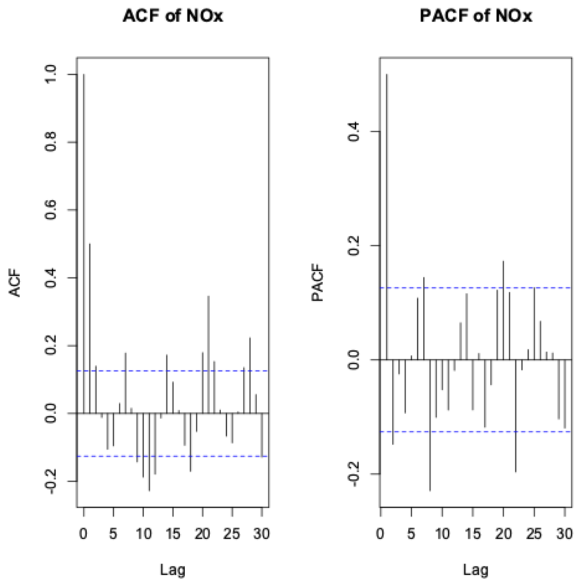
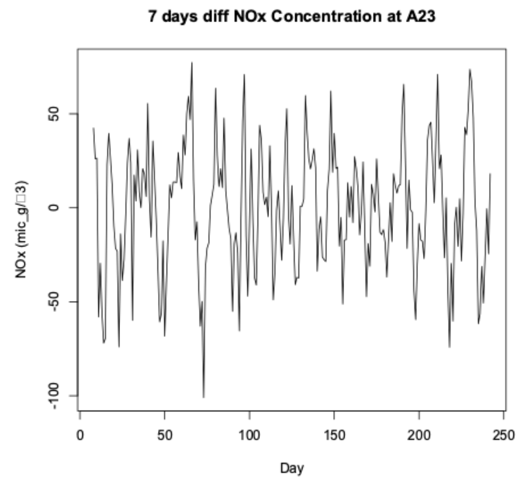
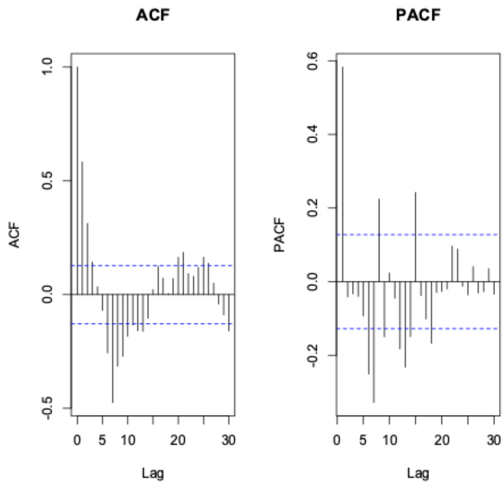


Figure 2: ACF and PACF of a23_nox Dataset



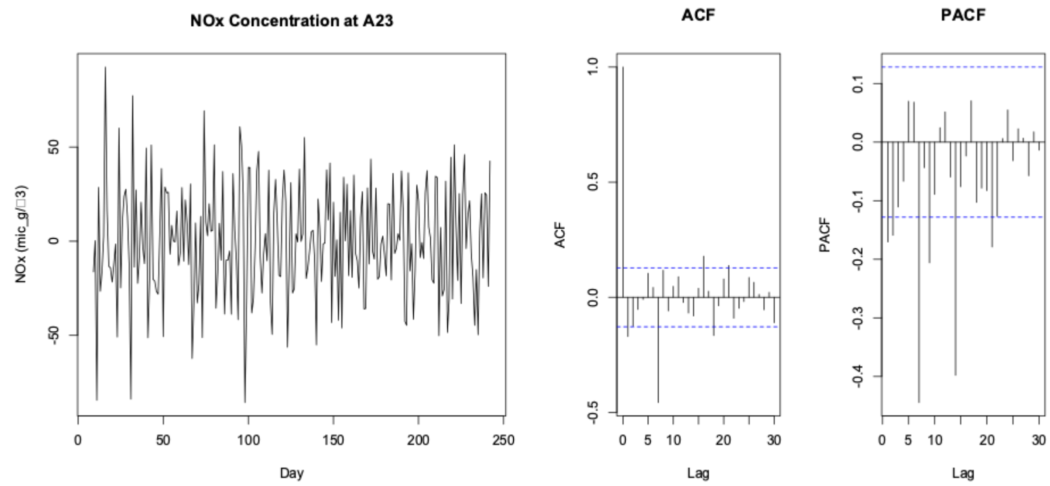
(a) Time plot



(b) ACF and PACF

Figure 3: 7 days differencing of a23_nox Dataset

there were no obvious patterns or trends. This study looked at numerous AR and MA model combinations to find the best model. The results of the residuals' analysis using time plot and ACF are shown in the figures 5. As can be seen in table 2, the findings showed that the ARIMA$(1,1,1)\times(1,1,1)_7$ model was the most appropriate. Refer to the appendix to see the other models that were tested.



(a) Time plot (b) ACF and PACF

Figure 4: differencing of a23_nox Dataset

Table 2: $ARIMA(1,1,1)$x$(1,1,1)_7$ Model Parameters

ar1	ma1	sar1	sma1	$sigma^2$	aic
0.5535	-0.9873	-0.0216	-0.9676	372.3	2084.52

1.4 Conclusion

In order to create this report, we investigated a dataset of daily mean nitrous oxide measurements taken from 1 February 2017 to 30 September 2017 at a position on the A23 Purley Way in the London Borough of Croydon.

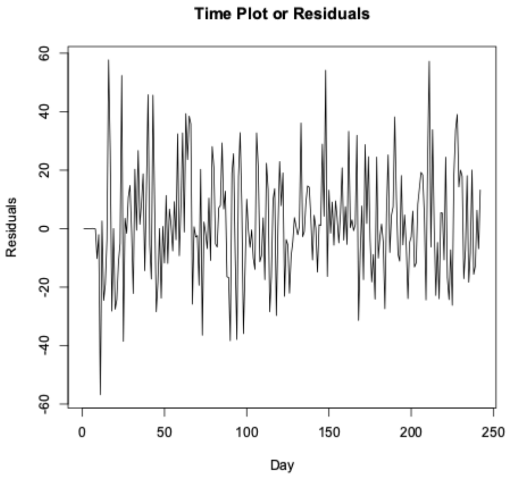
Following data exploration, preprocessing, and model fitting, we came to the conclusion that the $ARIMA(1,1,1)$x$(1,1,1)_7$ model, with the parameters shown in table 2 was the best model for this dataset.

The fitted SARIMA model can be represented by the following equation:

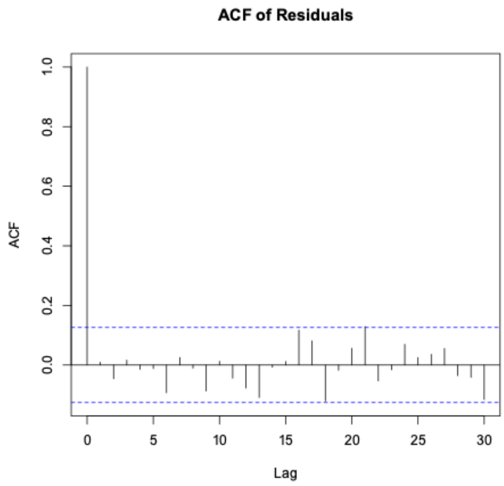
$$X_t - (1+\phi_1)X_{t-1} - \phi_1 X_{t-7} + (1+\Phi_1)X_{t-7} + (1+\phi_1 - \Phi 1 + \phi 1\Phi_1)X_{t-8} + (\phi_1 - \phi 1\Phi_1)X_{t-9}$$

$$+\Phi_1 X_{t-14} + (\Phi_1 - \phi 1\Phi_1)X_{t-15} + \phi 1\Phi_1 X_{t-16} = Z_t + \theta_1 Z_{t-1} + \Theta_1 Z_{t-7} + \theta_1 \Theta 1 Z_{t-8}$$

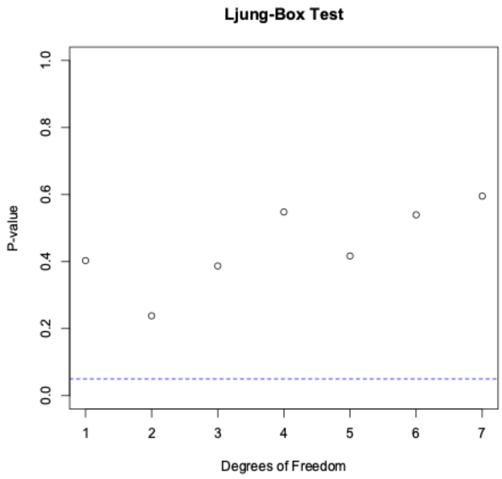
where X_t represents the nitrous oxide level at time t and Z_t is the white noise term at time t. This model effectively captures the underlying patterns and seasonality in the



(a) Time plot



(b) ACF



(c) P-Values

Figure 5: Time Plot, ACF and P-Values of $ARIMA(1,1,1)\text{x}(1,1,1)_7$

nitrous oxide levels, allowing us to make accurate forecasts and better understand the factors influencing air pollution in the area.

2 Challenge 2 - Analysing the number of new cars registered in England

2.1 Executive Summary

In this research, the number of new car registrations in England are analysed each quarter from Q1 2001 to Q3 2022, and the number of new car registrations in Q4 2022 and Q1, Q2, and Q3 of 2023 are also predicted. This study's primary goal is to give industry stakeholders an understanding of the trends and patterns in new car registrations so that they can better plan their sales, marketing, and production efforts.

We started by looking into the underlying trends and patterns in the data. Then, we created the $ARIMA(1,1,1)\text{x}(0,1,1)_4$ time series model. By testing the assumptions of normality, independence, and constant variance for the residuals, the model was proven to be accurate.

The chosen model was used to calculate the anticipated new automobile registration numbers for Q4 2022 and Q1, Q2, and Q3 of 2023, as well as any associated uncertainties. Forecasts show that new automobile registrations are anticipated to rise in Q1 2023, fall in Q2 2023, and then rise once more in Q3 2023.

Figure 6 shows the historical data and the forecasted values for the next few quarters.

In summary, this study offers insightful analysis of the trends and patterns of new car registrations in England as well as precise projections for the upcoming quarters.

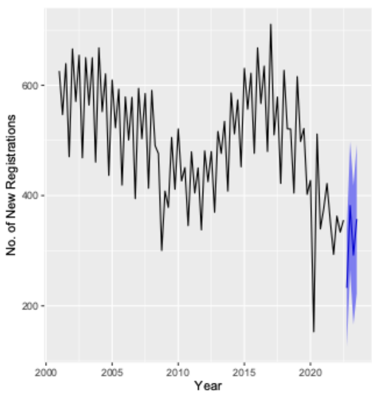


Figure 6: Forecasted New Car Registrations in England

2.2 Introduction

The automobile sector contributes significantly to the economy and is essential to the transportation industry. Making informed judgements about production, marketing, and sales strategies can benefit industry stakeholders aware of the trends and patterns in new car registrations [2]. In this paper, the number of new cars registered in England for each quarter from Q1 2001 through Q3 2022 will be analysed, and the numbers for Q4 2022 and Q1, Q2, and Q3 of 2023 will be projected.

The main objectives of this study are:

(i) To explore the underlying patterns and trends in the dataset, such as seasonality and trends.

(ii) To develop a suitable time series model that can accurately capture the observed patterns and trends in the data.

(iii) To evaluate the model's performance by assessing its fit to the historical data.

(iv) To forecast the number of new car registrations for Q4 2022 and Q1, Q2, and Q3 of 2023, along with associated uncertainties.

Following the discussion of data exploration and preprocessing, data analysis methodology, model selection, and model validation, we will present our predictions and conclusions in the following sections of this report.

2.3 Data Exploration and Preprocessing

The dataset contains quarterly numbers of new car registrations in England (in thousands) from Q1 2001 to Q3 2022. The data is presented in the table 3.

The time series plot we created for the data is shown in figure 7. With a peak in new automobile registrations in Q1 2017 and a low in Q2 2020 that may be related to the COVID pandemic, the data clearly demonstrates a seasonal trend. The data also appears to show a falling trend from 2001 to 2010, an improving trend from 2010 to 2017, and then another decreasing trend starting in 2017.

We did not encounter any missing values in the dataset, and therefore, no imputation was necessary.

Table 3: New Cars Registration Dataset

Year	Quarter	New Car Registrations (in thousands)
2001	Q1	625.9
2001	Q2	546.8
2001	Q3	639.2
...	...	...
2022	Q1	362.7
2022	Q2	333.8
2023	Q3	355.5

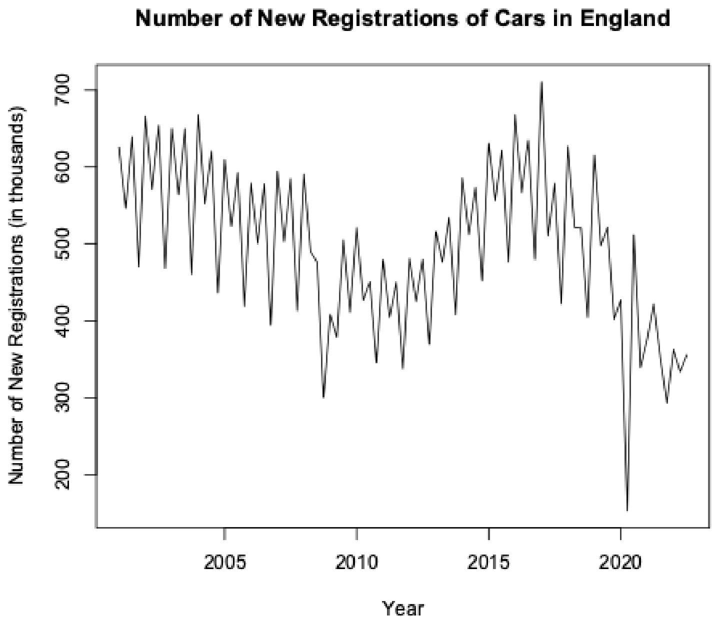
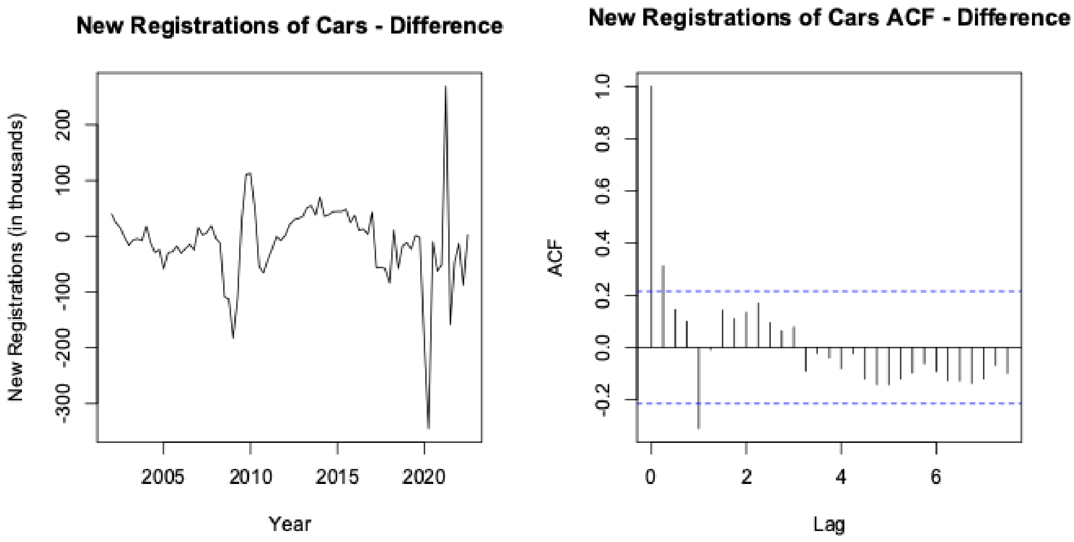


Figure 7: Time Series plot of New Car Registrations

2.4 Model Fitting and Validation

The procedure for choosing and validating the time series model for predicting new car registrations in England will be covered in this part. We started by seasonally varying the data at lag 4, and Figure 8 displays a time plot of the seasonally varying data. The data don't seem to be stationary, despite the fact that the seasonality seems to have been eliminated. The sample ACFs for the differenced (lag 4) series were plotted to test this and showed that the data are still not stationary.



(a) Time plot

(b) ACF

Figure 8: Time Plot and ACF of Difference at lag 4

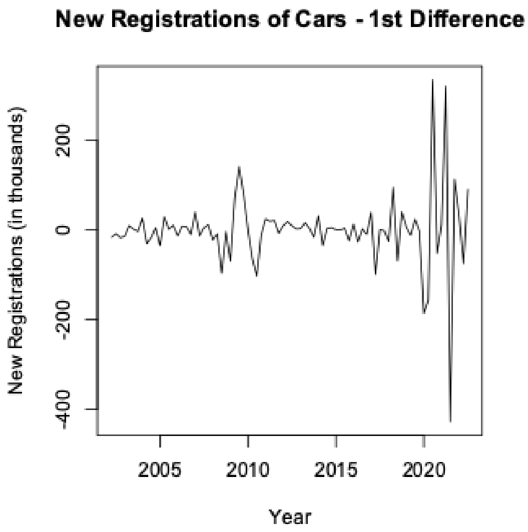
To address this, we took the first difference of the seasonally differenced data, and a time plot of these data is shown in Figure 9. These data certainly appear to be more stationary, with the mean appearing to be constant and equal to zero. We produced a plot of the sample ACF and sample PACF against the lag for these data to examine this further, as shown in Figure 9.

We started by fitting a $ARIMA(0,1,1)$x$(0,1,0)_4$ model to the sample ACF plot. Figure 10 displays the time plot, sample ACF vs lag plot, and Ljung-Box test statistics we created for the residuals. With an aic value of 940.19, it is obvious that the $ARIMA(0,1,1)$x$(0,1,0)_4$ cannot be a reasonable fit.

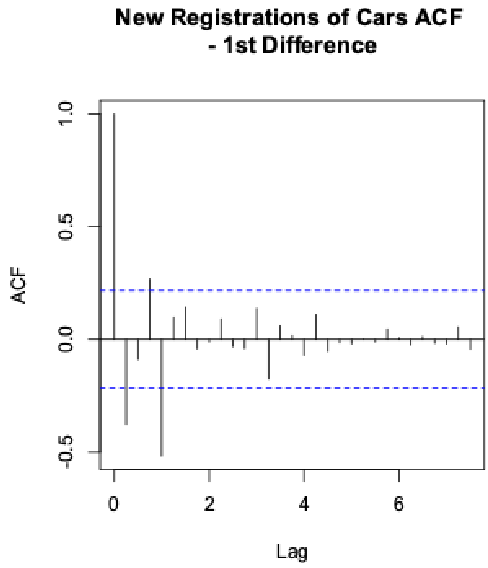
Next, we looked at the $ARIMA(1,1,1)$x$(0,1,0)_4$ model. Residuals, ACF, and Ljung-Box test statistics show that it is still not a good fit.

We then examined the $ARIMA(1,1,1)$x$(1,1,0)_4$ model. Figure 7 shows the plot of residuals, ACF, and Ljung-Box test statistics. This model with an AIC of 917.96 can be a good fit.

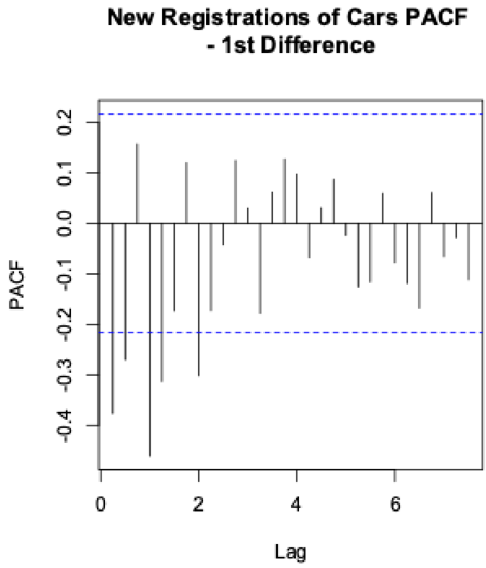
Next, we examined the $ARIMA(1,1,1)$x$(0,1,1)_4$ model. Figure 12 shows the plot of residuals, ACF, and Ljung-Box test statistics. The AIC drops to 903.31, residuals re not



(a) Time plot



(b) ACF



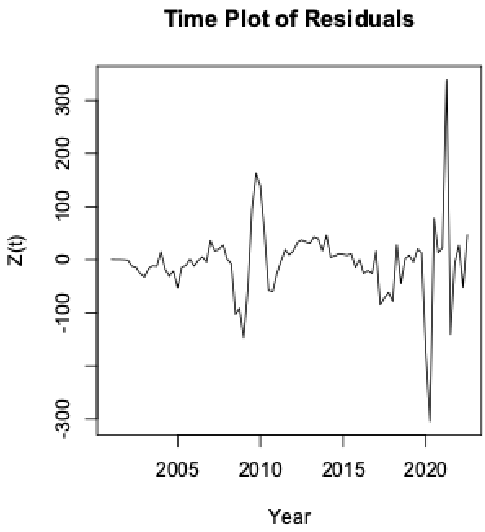
(c) PACF

Figure 9: Time Plot, ACF and PACF of 1st Difference

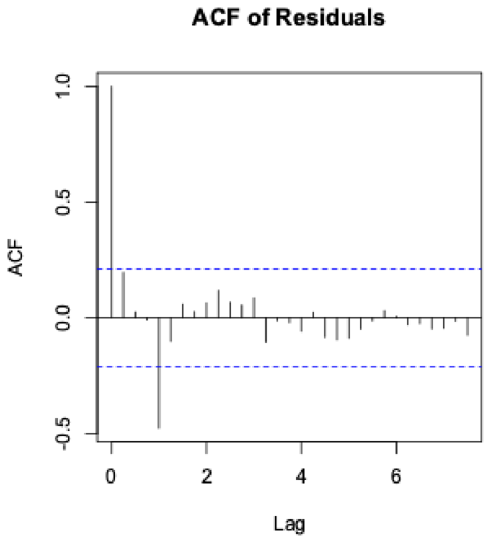
correlated to each other and P-Values are fairly large, hence this can be a better model than the previous one.

Finally, we examined the $ARIMA(1,1,1)\text{x}(1,1,1)_4$ model. Figure 13 shows the plot of residuals, ACF, and Ljung-Box test statistics. Although the AIC (905.31) is near the previous model, as this model has one more parameter, we will stay with the previous model which is $ARIMA(1,1,1)\text{x}(0,1,1)_4$.

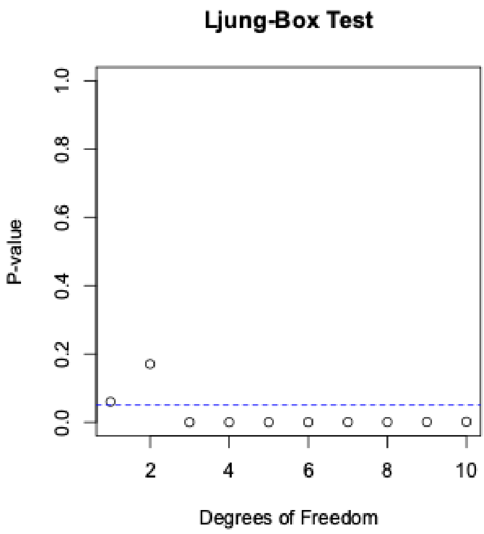
As a result, we decided to use the $ARIMA(1,1,1)\text{x}(0,1,1)_4$ model as our final projec-



(a) Time plot



(b) ACF

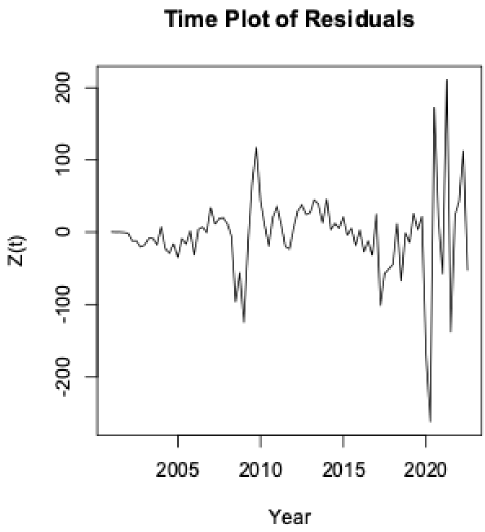


(c) P-Values

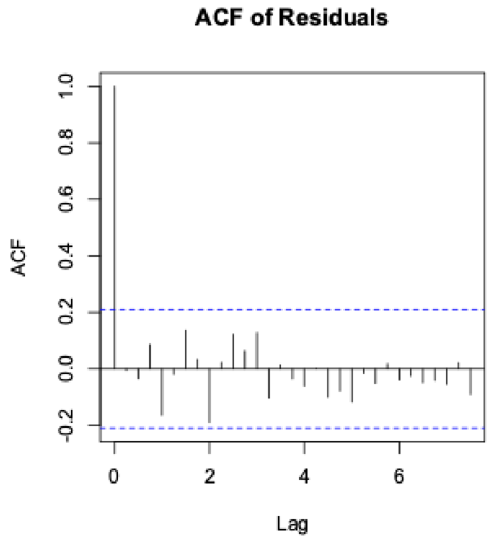
Figure 10: Time Plot, ACF and P-Values of $ARIMA(0, 1, 1)\text{x}(0, 1, 0)_4$

tion for the number of new cars registered in England. Table 4 displays information on model parameters in detail. We validated the model by checking the residuals. According to Figure 12, the residuals looked to be normally distributed, independent, and had a constant variance.

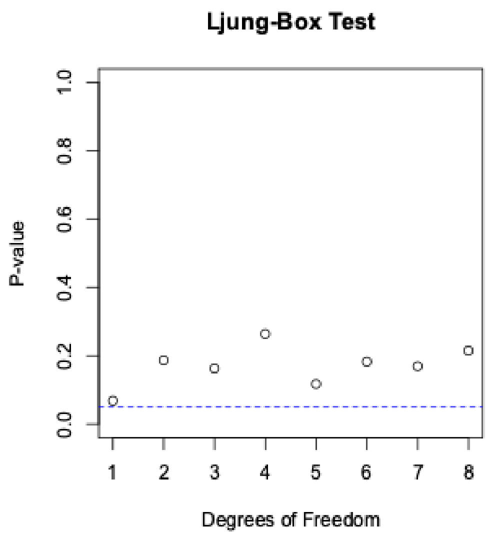
In the next section, we will present the forecasted numbers of new car registrations for Q4 2022 and Q1, Q2, and Q3 of 2023, along with associated uncertainties, using the selected model.



(a) Time plot



(b) ACF



(c) P-Values

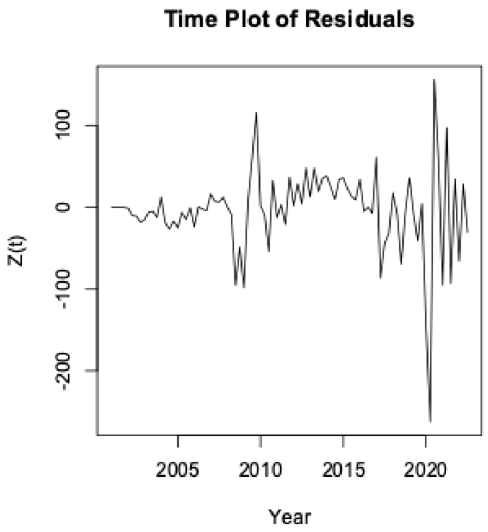
Figure 11: Time Plot, ACF and P-Values of $ARIMA(1,1,1)$x$(1,1,0)_4$

Table 4: Selected Model Parameters

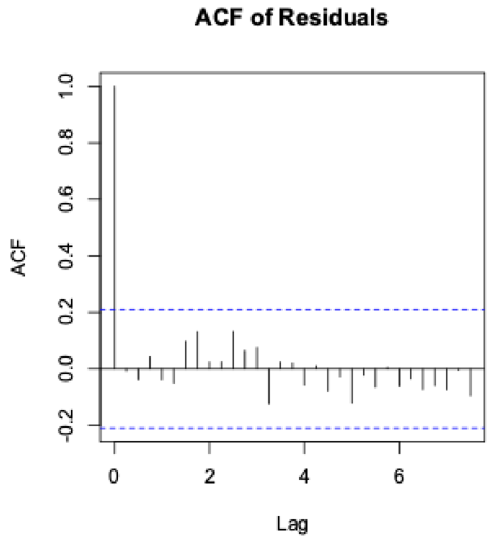
Model	**ar1**	**ma1**	**sma1**	$sigma^2$	aic
$ARIMA(1,1,1)$x$(0,1,1)_4$	-0.0088	-0.5412	-0.9413	2898	903.31

2.5 Forecasting

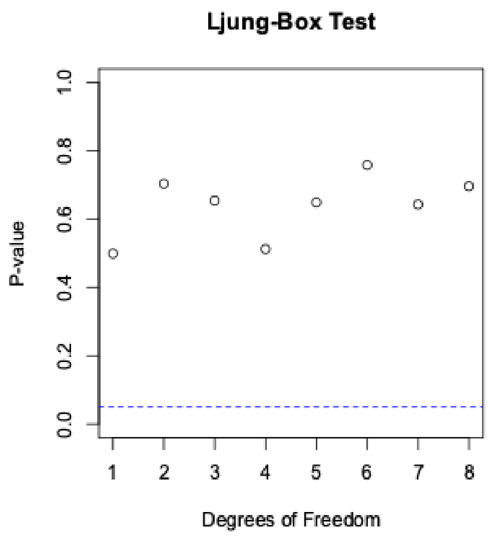
In this section, we will present the forecasted numbers of new car registrations for Q4 2022 and Q1, Q2, and Q3 of 2023, along with associated uncertainties, using the selected



(a) Time plot



(b) ACF

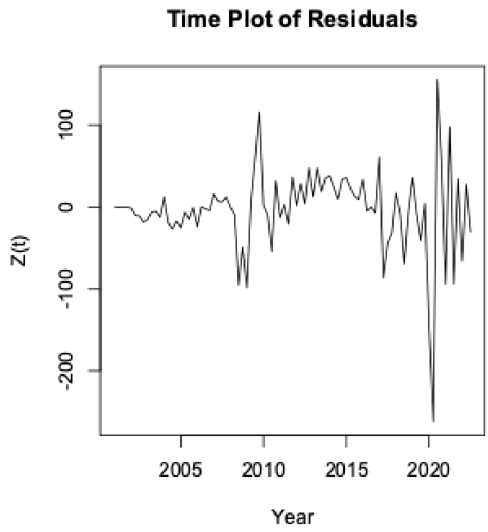


(c) P-Values

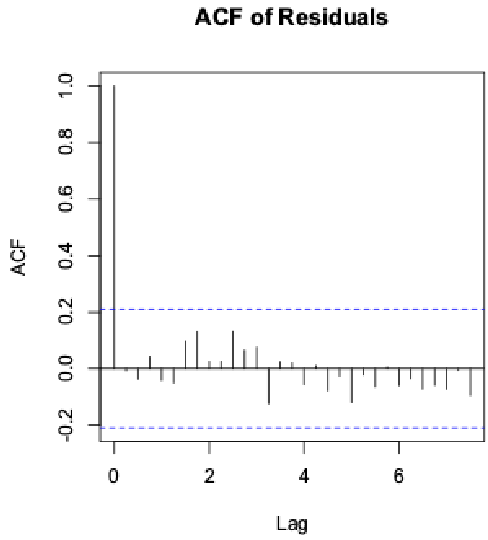
Figure 12: Time Plot, ACF and P-Values of $ARIMA(1,1,1)\text{x}(0,1,1)_4$

$ARIMA(1,1,1)\text{x}(0,1,1)_4$ model. We used the forecast function in R to generate the forecasts with a 95% confidence level. The forecasted values and their associated uncertainties are presented in table 5.

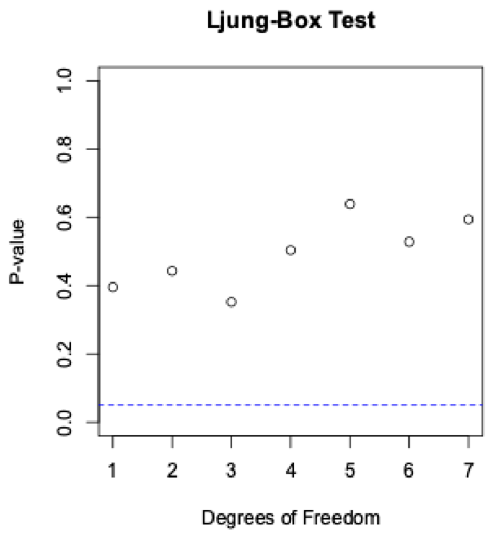
As seen in Figure 14, we also presented the forecasted values alongside the historical data. The graphic demonstrates that the model captures the seasonality and trend in the data and offers accurate estimates for the following few quarters.



(a) Time plot



(b) ACF



(c) P-Values

Figure 13: Time Plot, ACF and P-Values of $ARIMA(1,1,1)\text{x}(1,1,1)_4$

2.6 Conclusion

In conclusion, the $ARIMA(1,1,1)\text{x}(0,1,1)_4$ model offers a decent fit to the historical data as well as precise projections for the next quarters. For stakeholders in the sector, these forecasts can be helpful in making knowledgeable choices about manufacturing, marketing, and sales strategies. It is crucial to remember that forecasts are unreliable and that actual values may differ from predictions as a result of a variety of reasons, including

Table 5: Forecasting of Q4 2022 and Q1, Q2, and Q3 of 2023

Point	**Forecast**	**Lo 95**	**Hi 95**
2022 Q4	232.4415	126.4075	338.4756
2023 Q1	381.7304	265.4602	498.0007
2023 Q2	292.0404	166.1776	417.9033
2023 Q3	357.4431	222.6702	492.2159

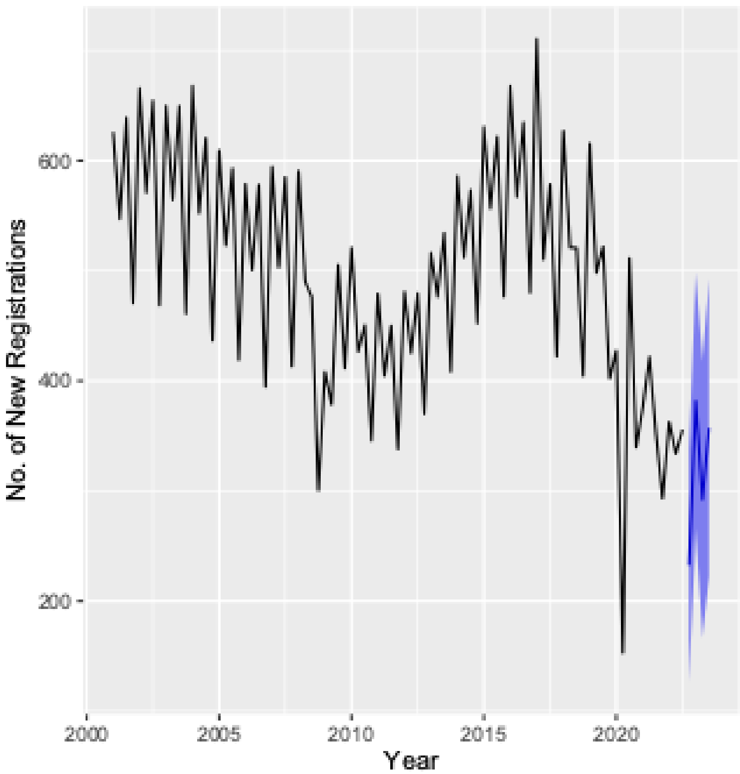


Figure 14: Forecasted New Car Registrations in England

shifting economic conditions, shifting consumer preferences, and shifting governmental regulations.

References

[1] Basic information about no2. https://www.epa.gov/no2-pollution/basic-information-about-no2.

[2] the contribution of the automobile industry to technology and value ... https://www.es.kearney.com/automotive/article/-/insights/the-contribution-of-the-automobile-industry-to-technology+-and-value-creation.

[3] Hui Liu, Guangxi Yan, Zhu Duan, and Chao Chen. Intelligent modeling strategies for forecasting air quality time series: A review. *Applied Soft Computing*, 102:106957, 2021.

[4] Douglas C Montgomery, Cheryl L Jennings, and Murat Kulahci. *Introduction to time series analysis and forecasting*. John Wiley & Sons, 2015.

A Appendix - R Codes

A.1 Challenge 1

```
# Analysing daily mean measurements of the nitrous oxides level

# Load the data
a23_nox <- read.csv("./a23_nox.csv", header = TRUE)

# time series
daily_mean_nox <- ts(a23_nox$daily_mean_nox, start = 1,
end = 242)

# Plot the time series
par(mfrow = c(1, 1))
ts.plot(daily_mean_nox, xlab="Day", ylab="NOx (mic_g/m3)",
main="NOx Concentration at A23")

# Plot the ACF and PACF
par(mfrow = c(1, 2))
acf(daily_mean_nox, lag.max = 30, xlab="Lag", ylab="ACF",
main="ACF of NOx")
pacf(daily_mean_nox, lag.max = 30, xlab="Lag", ylab="PACF",
main="PACF of NOx")

# Apply Seasonal Differencing at lag 7
daily_mean_nox_diff <- diff(daily_mean_nox, differences = 1,
lag = 7)

par(mfrow = c(1, 1))
ts.plot(daily_mean_nox_diff, xlab="Day", ylab="NOx (mic_g/
```

```
m3)", main="7 days diff NOx Concentration at A23")

# Plot the ACF and PACF
par(mfrow = c(1, 2))
acf(daily_mean_nox_diff, lag.max = 30, xlab="Lag", ylab="ACF",
main="ACF")
pacf(daily_mean_nox_diff, lag.max = 30, xlab="Lag",
ylab="PACF", main="PACF")

# Apply Differencing at lag 1
daily_mean_nox_diff <- diff(daily_mean_nox_diff, differences =
1)

par(mfrow = c(1, 1))
ts.plot(daily_mean_nox_diff, xlab="Day", ylab="NOx (mic_g/
m3)", main="NOx Concentration at A23")

# Plot the ACF and PACF
par(mfrow = c(1, 2))
acf(daily_mean_nox_diff, lag.max = 30, xlab="Lag", ylab="ACF",
main="ACF")
pacf(daily_mean_nox_diff, lag.max = 30, xlab="Lag",
ylab="PACF", main="PACF")

# Modeling

# MA(1) Model
model.1 <- arima(daily_mean_nox, order = c(0, 1, 1), method =
"ML", seasonal = list(order = c(0, 1, 0), period = 7))
model.1
```

```
# Plot the residuals
resid.model.1 <- residuals(model.1)
par(mfrow = c(1, 1))
ts.plot(resid.model.1, xlab="Day", ylab="Residuals", main="Time
Plot or Residuals")

par(mfrow = c(1, 1))
acf(resid.model.1, lag.max = 30, xlab="Lag", ylab="ACF",
main="ACF of Residuals")

# Ljung-Box test
LB.1 <- LB_test_SARIMA(resid.model.1, max.k = 11, p = 0, q = 1,
P=0, Q=0)
plot(LB.1$deg_freedom, LB.1$LB_p_value, xlab="Degrees of
Freedom", ylab="P-value", main="Ljung-Box Test", ylim=c(0,1))
abline(h=0.05, col="blue", lty=2)

# AR(1) Model
model.2 <- arima(daily_mean_nox, order = c(1, 1, 0), method =
"ML", seasonal = list(order = c(0, 1, 0), period = 7))
model.2

# Plot the residuals
resid.model.2 <- residuals(model.2)
par(mfrow = c(1, 1))
ts.plot(resid.model.2, xlab="Day", ylab="Residuals", main="Time
Plot or Residuals")

par(mfrow = c(1, 1))
```

```
acf(resid.model.2, lag.max = 30, xlab="Lag", ylab="ACF",
main="ACF of Residuals")

# Ljung-Box test
LB.2 <- LB_test_SARIMA(resid.model.2, max.k = 11, p = 1, q = 0,
P=0, Q=0)
plot(LB.2$deg_freedom, LB.2$LB_p_value, xlab="Degrees of
Freedom", ylab="P-value", main="Ljung-Box Test", ylim=c(0,1))
abline(h=0.05, col="blue", lty=2)


# ARIMA(1, 1, 1) Model
model.3 <- arima(daily_mean_nox, order = c(1, 1, 1), method =
"ML", seasonal = list(order = c(0, 1, 0), period = 7))
model.3

# Plot the residuals
resid.model.3 <- residuals(model.3)
par(mfrow = c(1, 1))
ts.plot(resid.model.3, xlab="Day", ylab="Residuals", main="Time
Plot or Residuals")

par(mfrow = c(1, 1))
acf(resid.model.3, lag.max = 30, xlab="Lag", ylab="ACF",
main="ACF of Residuals")

# Ljung-Box test
LB.3 <- LB_test_SARIMA(resid.model.3, max.k = 11, p = 1, q = 1,
P=0, Q=0)
plot(LB.3$deg_freedom, LB.3$LB_p_value, xlab="Degrees of
```

```r
Freedom", ylab="P-value", main="Ljung-Box Test", ylim=c(0,1))
abline(h=0.05, col="blue", lty=2)


# ARIMA(1, 1, 1)(1,1,1) Model
model.4 <- arima(daily_mean_nox, order = c(1, 1, 1), method =
"ML", seasonal = list(order = c(1, 1, 1), period = 7))
model.4

# Plot the residuals
resid.model.4 <- residuals(model.4)
par(mfrow = c(1, 1))
ts.plot(resid.model.4, xlab="Day", ylab="Residuals", main="Time
Plot or Residuals")

par(mfrow = c(1, 1))
acf(resid.model.4, lag.max = 30, xlab="Lag", ylab="ACF",
main="ACF of Residuals")

# Ljung-Box test
LB.4 <- LB_test_SARIMA(resid.model.4, max.k = 11, p = 1, q = 1,
P=1, Q=1)
plot(LB.4$deg_freedom, LB.4$LB_p_value, xlab="Degrees of
Freedom", ylab="P-value", main="Ljung-Box Test", ylim=c(0,1))
abline(h=0.05, col="blue", lty=2)
```

A.2 Challenge 2

```r
# Load the Data
eng_car_reg <- read.csv("./eng_car_reg.csv", header = TRUE)

# time series
no_new_reg <- ts(eng_car_reg$no_new_reg,start=c(2001,1),frequency=4)

# Plot the time series
par(mfrow = c(1, 1))
ts.plot(no_new_reg, xlab="Year", ylab="Number of New
Registrations (in thousands)", main="Number of New
Registrations of Cars in England")

# Plot the ACF and PACF
par(mfrow = c(1, 1))
acf(no_new_reg, lag.max = 30, xlab="Lag", ylab="ACF", main="ACF
of Number of New Registrations of Cars")
pacf(no_new_reg, lag.max = 30, xlab="Lag", ylab="PACF",
main="PACF of Number of New Registrations of Cars")

# difference seasonally at lag 4
no_new_reg_diff <- diff(no_new_reg, differences = 1, lag = 4)

# Produce the time plot
par(mfrow = c(1, 1))
ts.plot(no_new_reg_diff, xlab="Year", ylab="New Registrations
(in thousands)", main="New Registrations of Cars - Difference")

# ACF of difference
par(mfrow = c(1, 1))
```

```
acf(no_new_reg_diff, lag.max = 30, xlab="Lag", ylab="ACF",
main="New Registrations of Cars ACF - Difference")

# take the first difference of the seasonally differenced data.
no_new_reg_diff2 <- diff(no_new_reg_diff, differences = 1)

# Produce the time plot
par(mfrow = c(1, 1))
ts.plot(no_new_reg_diff2, xlab="Year", ylab="New Registrations
(in thousands)", main="New Registrations of Cars - 1st
Difference")

# ACF of difference
par(mfrow = c(1, 1))
acf(no_new_reg_diff2, lag.max = 30, xlab="Lag", ylab="ACF",
main=paste("New Registrations of Cars ACF", "\n- 1st
Difference"))
pacf(no_new_reg_diff2, lag.max = 30, xlab="Lag", ylab="PACF",
main=paste("New Registrations of Cars PACF", "\n- 1st
Difference"))



# ARIMA(0,1,1)(0,1,0)4
model.1 <- arima(no_new_reg, order = c(0, 1, 1), seasonal =
list(order = c(0, 1, 0), period = 4))
model.1

resid.model.1 <- model.1$residuals
par(mfrow = c(1, 1))
ts.plot(resid.model.1, xlab="Year", ylab="Z(t)", main="Time
```

```
Plot of Residuals")
acf(resid.model.1, lag.max = 30, xlab="Lag", ylab="ACF",
main="ACF of Residuals")

# Ljung-Box test
LB.1 <- LB_test_SARIMA(resid.model.1, max.k = 11, p = 0, q = 1,
P=0, Q=0)
plot(LB.1$deg_freedom, LB.1$LB_p_value, xlab="Degrees of
Freedom", ylab="P-value", main="Ljung-Box Test", ylim=c(0,1))
abline(h=0.05, col="blue", lty=2)

# ARIMA(1,1,0)(0,1,0)4
model.2 <- arima(no_new_reg, order = c(1, 1, 0), seasonal =
list(order = c(0, 1, 0), period = 4))
model.2

resid.model.2 <- model.2$residuals
par(mfrow = c(1, 1))
ts.plot(resid.model.2, xlab="Year", ylab="Z(t)", main="Time
Plot of Residuals")
acf(resid.model.2, lag.max = 30, xlab="Lag", ylab="ACF",
main="ACF of Residuals")

# Ljung-Box test
LB.2 <- LB_test_SARIMA(resid.model.2, max.k = 11, p = 1, q = 0,
P=0, Q=0)
plot(LB.2$deg_freedom, LB.2$LB_p_value, xlab="Degrees of
Freedom", ylab="P-value", main="Ljung-Box Test", ylim=c(0,1))
abline(h=0.05, col="blue", lty=2)
```

```
# ARIMA(1,1,1)(0,1,0)4
model.3 <- arima(no_new_reg, order = c(1, 1, 1), seasonal =
list(order = c(0, 1, 0), period = 4))
model.3

resid.model.3 <- model.3$residuals
par(mfrow = c(1, 1))
ts.plot(resid.model.3, xlab="Year", ylab="Z(t)", main="Time
Plot of Residuals")
acf(resid.model.3, lag.max = 30, xlab="Lag", ylab="ACF",
main="ACF of Residuals")

# Ljung-Box test
LB.3 <- LB_test_SARIMA(resid.model.3, max.k = 11, p = 1, q = 1,
P=0, Q=0)
plot(LB.3$deg_freedom, LB.3$LB_p_value, xlab="Degrees of
Freedom", ylab="P-value", main="Ljung-Box Test", ylim=c(0,1))
abline(h=0.05, col="blue", lty=2)

# ARIMA(1,1,1)(1,1,0)4
model.4 <- arima(no_new_reg, order = c(1, 1, 1), seasonal =
list(order = c(1, 1, 0), period = 4))
model.4

resid.model.4 <- model.4$residuals
par(mfrow = c(1, 1))
ts.plot(resid.model.4, xlab="Year", ylab="Z(t)", main="Time
Plot of Residuals")
acf(resid.model.4, lag.max = 30, xlab="Lag", ylab="ACF",
main="ACF of Residuals")
```

```
# Ljung-Box test
LB.4 <- LB_test_SARIMA(resid.model.4, max.k = 11, p = 1, q = 1,
P=1, Q=0)
plot(LB.4$deg_freedom, LB.4$LB_p_value, xlab="Degrees of
Freedom", ylab="P-value", main="Ljung-Box Test", ylim=c(0,1))
abline(h=0.05, col="blue", lty=2)

# ARIMA(1,1,1)(0,1,1)4
model.5 <- arima(no_new_reg, order = c(1, 1, 1), seasonal =
list(order = c(0, 1, 1), period = 4))
model.5

resid.model.5 <- model.5$residuals
par(mfrow = c(1, 1))
ts.plot(resid.model.5, xlab="Year", ylab="Z(t)", main="Time
Plot of Residuals")
acf(resid.model.5, lag.max = 30, xlab="Lag", ylab="ACF",
main="ACF of Residuals")

# Ljung-Box test
LB.5 <- LB_test_SARIMA(resid.model.5, max.k = 11, p = 1, q = 1,
P=0, Q=1)
plot(LB.5$deg_freedom, LB.5$LB_p_value, xlab="Degrees of
Freedom", ylab="P-value", main="Ljung-Box Test", ylim=c(0,1))
abline(h=0.05, col="blue", lty=2)

# ARIMA(1,1,1)(1,1,1)4
model.6 <- arima(no_new_reg, order = c(1, 1, 1), seasonal =
list(order = c(1, 1, 1), period = 4))
```

```
model.6

resid.model.6 <- model.6$residuals
par(mfrow = c(1, 1))
ts.plot(resid.model.6, xlab="Year", ylab="Z(t)", main="Time
Plot of Residuals")
acf(resid.model.6, lag.max = 30, xlab="Lag", ylab="ACF",
main="ACF of Residuals")

# Ljung-Box test
LB.6 <- LB_test_SARIMA(resid.model.6, max.k = 11, p = 1, q = 1,
P=1, Q=1)
plot(LB.6$deg_freedom, LB.6$LB_p_value, xlab="Degrees of
Freedom", ylab="P-value", main="Ljung-Box Test", ylim=c(0,1))
abline(h=0.05, col="blue", lty=2)

# Forecasting

# Fit the model
model <- arima(no_new_reg, order = c(1, 1, 1), seasonal =
list(order = c(0, 1, 1), period = 4))

# Forecast for the next 1 quarter
library(forecast)
fc_1q <- forecast(no_new_reg,h=1,model=model,level=95)
autoplot(fc_1q, main="Forecast Q4 2022", xlab="Year", ylab="No.
of New Registrations")

# Forecast for the next 4 quarters with 95% and 80% confidence
intervals
```

```
fc_4q <- forecast(no_new_reg,h=4,model=model, level=95)
autoplot(fc_4q, main=" ", xlab="Year", ylab="No. of New
Registrations")
fc_4q
```