Heathrow Monthly Cargo Volume (2005–2024) time series model Report

Shubhankar Shahade

University of Nottingham

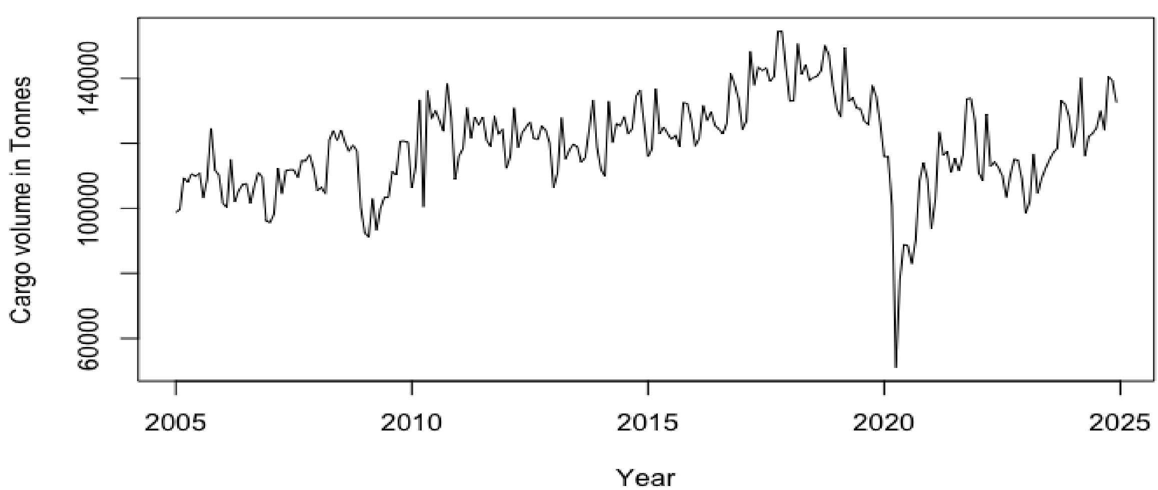
I. Introduction

The following report analyses the monthly volumes of cargo (in metric tonnes) handled at London's Heathrow airport from January 2005 to Decem 2024 (inclusive). The data was sourced from www.heathrow.com/company/investor-centre . The objective was to fit a suitable time series model to describe the data and report the identified fit for the time series model. Seasonal behaviors and trends which were visible in the series has been identified and validated using appropriate methods including time plots, autocorrelation analysis, differencing, model fitting, and residual checks.

II. Exploratory Data Analysis

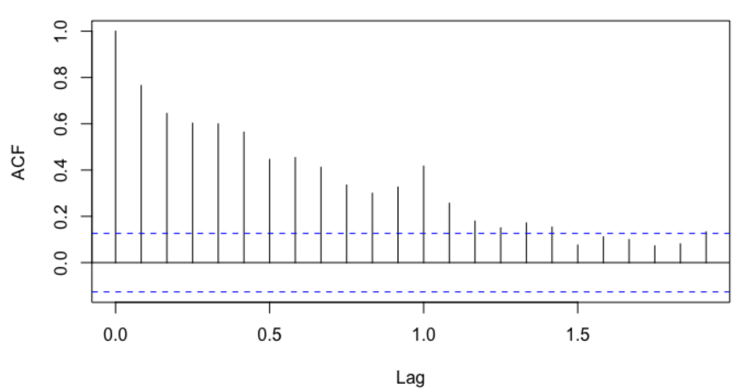
A time series plot (Figure 1) was created to visualize the time series data. The plot below shows the monthly volume of cargo handled in tonnes from 2005 to 2024. We observe an upward trend over the period along with an obvious seasonal pattern where the volume rises for certain months of the year. There is a dip observed around 2008 – 2009 and then a sharp dip again around 2020. The one around 2008-09 can likely be attributed to the 2008 financial crisis and the one in 2020 due to the COVID- 19 pandemic.

Figure 1: Heathrow Cargo Tonnage (2005-2024)



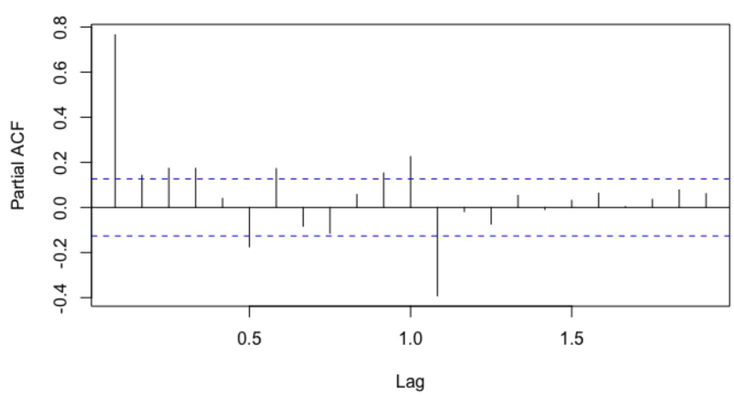
To help identify a suitable model, A sample autocorrelation function (ACF) plot (Figure 2) and a sample partial autocorrelation function (PACF) plot (Figure 3) of the dataset was examined. The ACF plot shows how observations are correlated with their past values and can suggest appropriate MA terms, while the PACF plot shows the direct effect of past observations after accounting for intermediate lags, helping suggest AR terms.

Figure 2: ACF plot for Heathrow Cargo volume



The ACF plot displays a slow decay, suggesting non-stationarity and the need for differencing.

Figure 3: PACF plot for Heathrow Cargo volume

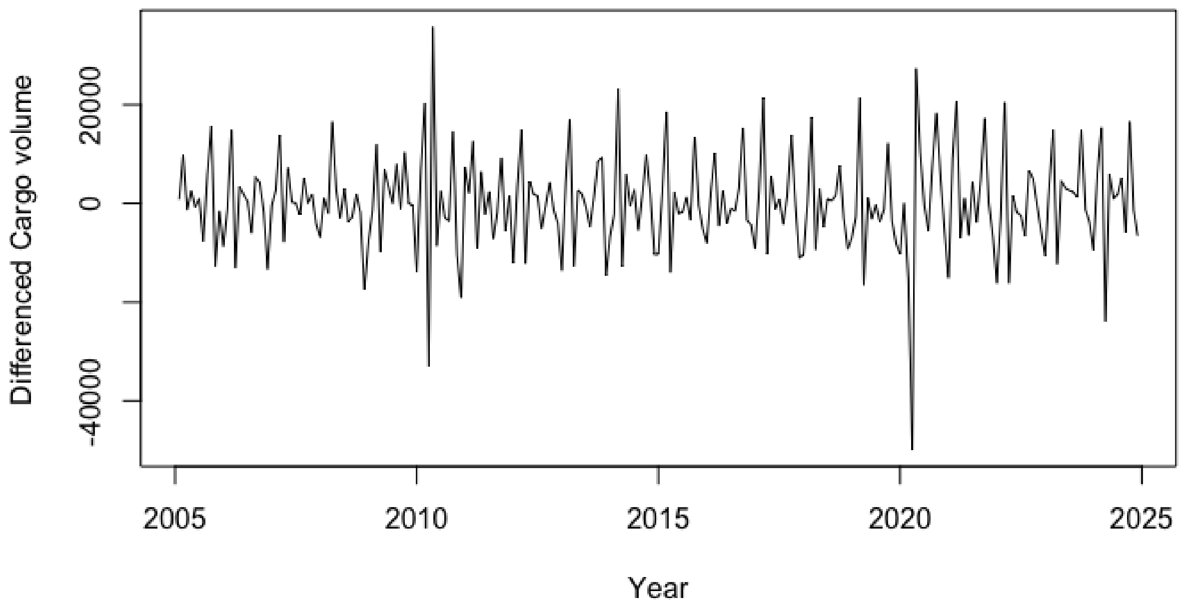


The PACF plot shows a significant spike at lag 1, indicating possible short-term autoregressive structure. Given the presence of trend and strong autocorrelation, first differencing was applied to stabilize the mean.

III. Data transformation: Differencing

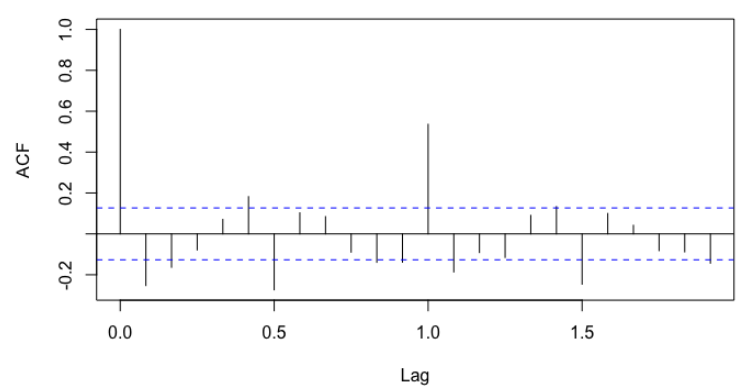
Differencing removes trend and stabilizes the mean in a non-stationary time series, making it suitable for ARIMA modeling. The differenced series (Figure 4) below appears stationary, fluctuating around a constant mean.

Figure 4: Time Plot of First-Differenced time series



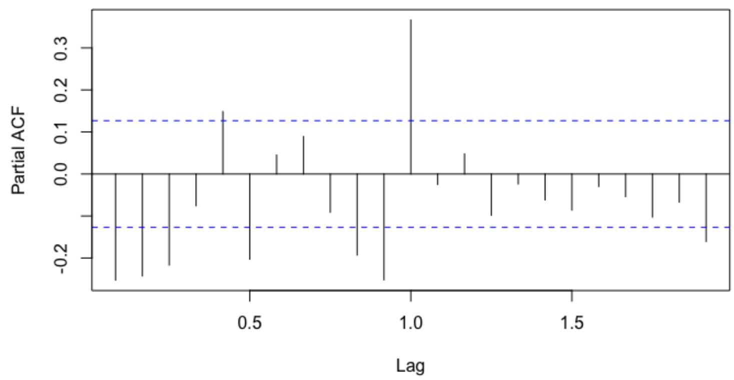
The ACF (Figure 5) plot of the differenced series below shows a sharp cutoff after lag 1, suggesting moving average (MA) behavior.

Figure 5: ACF plot of first-differenced cargo



The PACF (Figure 6) plot below also indicates short-term autoregressive (AR) structure at lag 1.

Figure 6: PACF plot of first-differenced cargo



IV. Model Identification and Fitting

Based on the ACF and PACF plots, observations and given annual seasonality in original series, various ARIMA models were considered. The models are given below as follows:

- ARIMA(0,1,1) — MA(1) model
- ARIMA(1,1,0) — AR(1) model
- ARIMA(1,1,1) — ARMA(1,1) model
- ARIMA(0,1,1)(2,0,0)[12] — seasonal ARIMA model with seasonal AR(2)

A summary output of the chosen models was done. The summary output of ARIMA(0,1,1)(2,0,0)[12] is shown below in Table 1.

Table 1

```
Series: cargo_ts
ARIMA(0,1,1)(2,0,0)[12]

Coefficients:
          ma1     sar1     sar2
      -0.3159   0.3578   0.3176
s.e.   0.0680   0.0607   0.0612

sigma^2 = 55422625:  log likelihood = -2471.62
AIC=4951.24   AICc=4951.41   BIC=4965.14

Training set error measures:
                   ME     RMSE      MAE        MPE     MAPE      MASE       ACF1
Training set 129.8099 7382.338 4883.683 -0.2548229 4.430568 0.4530837 0.03106527
[1] 4951.237
```

The output consists of estimated coefficients, standard errors of those coefficients, log likelihood for model fitting, AIC values and error measures. All these metrics help us choose the model that fits the data well. The (AIC) is a measure used to compare different statistical models, balancing model fit and complexity. A lower AIC value indicates a model that better explains the data without unnecessary complexity.

The Akaike Information Criterion (AIC) values for these models were compared and as seen from Table 2 below, the seasonal ARIMA(0,1,1)(2,0,0)[12] model achieved the lowest AIC (4951.24), suggesting it was the best fit and was therefore, selected for further analysis.

Table 2

Model	AIC
ARIMA(0,1,1)	5041.87
ARIMA(1,1,0)	5056.84
ARIMA(1,1,1)	5036.58
ARIMA(0,1,1)(2,0,0)[12]	**4951.24**

V. Residual analysis & model validation

After fitting the seasonal ARIMA(0,1,1)(2,0,0)[12] model, residual analysis was performed to assess model adequacy. The residual time plot below (Figure 7) shows random fluctuations around zero without any visible pattern. This suggests that the model has captured the main structure of the data.

The ACF plot of the residuals below (Figure 8) shows no significant autocorrelation at any lag, supporting the randomness of the residuals. Additionally, the Ljung-Box test was performed at lag 20, returning a p-value of 0.253 (> 0.05), which fails to reject the null hypothesis of no autocorrelation. This indicates that the residuals are consistent with white noise and that the fitted model is adequate.

Figure 7: Residuals of ARIMA(0,1,1)(2,0,0)[12] Model

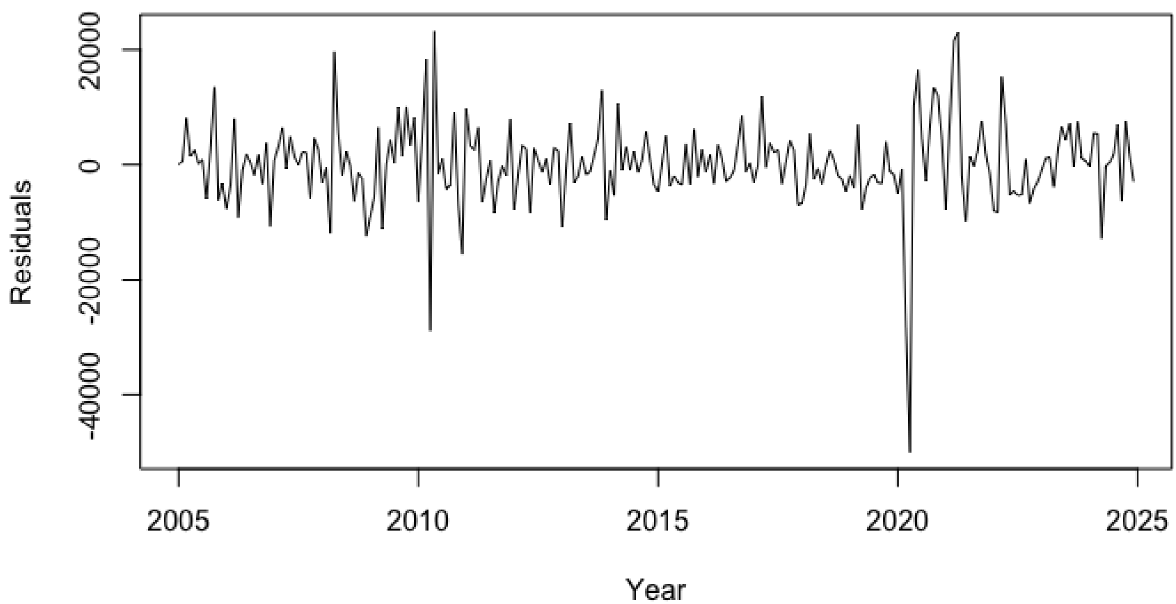
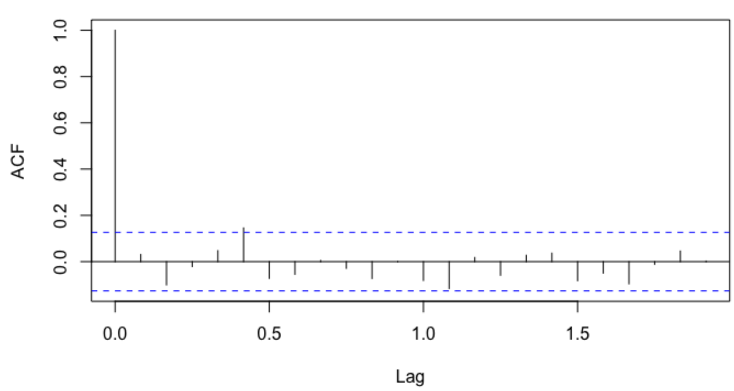


Figure 8: ACF of Residuals



VI. Final Model Equation

The equation of the final fitted model is :

$$(1 - 0.3578\, B^{12} - 0.3176\, B^{24})\,(1 - B)Y_t = (1 - 0.3159\, B)\, \epsilon_t$$

where:

- Y_t represents the monthly cargo tonnage,

- B is the backshift operator,
- ϵ_t is the error term at time t.

After fitting the seasonal ARIMA(0,1,1)(2,0,0)[12] model to the data, the estimated coefficients were found. The model includes a short-term moving average (MA(1)) term with value –0.3159, and two seasonal autoregressive terms: one at lag 12 months (coefficient 0.3578) and one at lag 24 months (coefficient 0.3176). These coefficients appear in the final model equation, where the seasonal structure is captured through the terms involving B^{12} and B^{24}, which is one and two years ago respectively. The short-term structure is captured through the first-order differencing and MA(1) component.

VII. Conclusion

This report analyzed monthly cargo tonnage data at Heathrow Airport from January 2005 to December 2024, with the aim of identifying a suitable time series model to describe and capture the underlying patterns in the data. Exploratory analysis of the original series revealed a clear upward trend and annual seasonality, which led to the application of first differencing to stabilize the mean and achieve stationarity. After examining the ACF and PACF plots of the differenced series, several ARIMA models were considered. Model selection was based on the Akaike Information Criterion (AIC) with the best fit being the model with the lowest AIC values. Thus, the seasonal ARIMA(0,1,1)(2,0,0)[12] model was found to provide the best fit.

Furthermore, Residual analysis confirmed the adequacy of the chosen model. The residuals showed no significant autocorrelation and appeared to behave like white noise, as indicated by the Ljung-Box test. The model successfully captured both the short-term dependencies and the seasonal structure present in the cargo data.

Overall, the seasonal ARIMA(0,1,1)(2,0,0)[12] model provides an appropriate and reliable description of the Heathrow cargo tonnage series, and it can be used for future forecasting or analysis if needed.

Appendix: R Code

```
```{r}
# Load necessary libraries
library(ggplot2)
library(forecast)
library(tseries)

# Read dataset and display first few rows
heathrow_cargo <- read.csv("heathrow_cargo.csv")
head(heathrow_cargo)
```

```{r}
#Create time series from dataset
cargo_ts <- ts(heathrow_cargo$cargo_tonnes,start=c(2005,1),frequency=12)

# Plot time series object
plot(cargo_ts, main="Figure 1: Heathrow Cargo Tonnage (2005-2024)",
      ylab="Cargo volume in Tonnes", xlab="Year")
```

```{r}
# Plot ACF and PACF for the original data
acf(cargo_ts, main="Figure 2: ACF plot for Heathrow Cargo volume")
pacf(cargo_ts, main="Figure 3: PACF plot for Heathrow Cargo volume")
```

```{r}
# First difference for stationarity
cargo_ts_diff <- diff(cargo_ts)

# Plot of first-differenced time series
plot(cargo_ts_diff, main="Figure 4: Time Plot of First-Differenced time series",
      ylab="Differenced Cargo volume", xlab="Year")

# ACF and PACF for the differenced series
acf(cargo_ts_diff, main="Figure 5: ACF plot of first-differenced cargo")
pacf(cargo_ts_diff, main="Figure 6: PACF plot of first-differenced cargo")
```

```{r}
# Model 1: ARIMA(0,1,1) Moving average (MA) model
model1 <- Arima(cargo_ts, order = c(0,1,1), seasonal = list(order = c(0,0,0), period=12))
summary(model1)
AIC(model1)

# Model 2: ARIMA(1,1,0) Autoregressive (AR) model
model2 <- Arima(cargo_ts, order = c(1,1,0), seasonal = list(order = c(0,0,0), period=12))
summary(model2)
AIC(model2)

# Model 3: ARIMA(1,1,1) Combined AR and MA model
model3 <- Arima(cargo_ts, order = c(1,1,1), seasonal = list(order = c(0,0,0), period=12))
summary(model3)
AIC(model3)

# Model 4: ARIMA(0,1,1)(2,0,0)[12] Seasonal ARIMA model with seasonal AR(2)
model4 <- Arima(cargo_ts, order = c(0,1,1), seasonal = list(order = c(2,0,0), period=12))
summary(model4)
AIC(model4)
```
```

```{r}
# Plot residuals over time to check for any patterns
plot(residuals(model4),
     main = "Figure 7: Residuals of ARIMA(0,1,1)(2,0,0)[12] Model",
     ylab = "Residuals", xlab = "Year")

# ACF residuals plot
acf(residuals(model4),
    main = "Figure 8: ACF of Residuals")

# Ljung-Box test (20 lags) for randomness
Box.test(residuals(model4), lag = 20, type = "Ljung-Box")

```

Heathrow Monthly Cargo Volume (2005–2024) time series model Report

Shubhankar Shahade

University of Nottingham

Executive Summary

This report analyses quarterly data on new car registrations in Scotland from early 2001 to mid-2024, with the aim of predicting future registrations for the next year, Q3, Q4 of 2024 and Q1, Q2 of 2025 to be specific. Understanding these trends is important to assist business planning and decision-making in the car industry.

Upon conducting exploratory analysis and visualizing the plot, the data showed clear seasonal patterns, with more cars registered in certain parts of the year, and long-term changes over time. A couple of years served as outliers or anomalies, but they can be attributed to the 2008 financial crisis and the COVID- 19 pandemic.

All features were adjusted for during the analysis to better capture the key drivers of future patterns. Past patterns were carefully studied to build a model that reflects both seasonal effects and short-term changes.

Using standard forecasting methods, a model was built that captures both the seasonal patterns and underlying trends in the data. The model was checked carefully against past data using statistical tests and showed a good fit, meaning the forecasts are considered reliable for short-term planning.

The forecasts suggest that registrations will likely increase during Q3 2024 to around 60,000 cars, fall in Q4 2024 as typically expected during winter, and then recover somewhat in early 2025.Uncertainty ranges have been provided to give a best-case and worst-case estimate for planning purposes.

Overall, the results offer helpful insights to support sales, inventory management, and marketing strategies for the year ahead.

I. Aim

The main aim of this project was to forecast the number of new car registrations in Scotland for the next four quarters which are, Q3-Q4 of 2024 and Q1-Q2 of 2025, based on historical quarterly data spanning from Q1 2001 to Q2 2024.

The forecasts can provide useful insights for planning and decision-making in the near future for concerned stakeholders in the car industry. This report aims to convey the important details and findings of the models fitted by this project.

II. Exploratory Data Analysis

Exploratory analysis of the data was conducted. A time plot (Figure 1) was drawn to visualize the new car registrations from Q1 2001 – Q2 2024. The time plot below shows notable fluctuations throughout the year. While there was a steady rise from 2011 to 2018, there was a sharp decline in two periods, 2008-09 and 2020 which can be attributed to the 2008 financial crisis and the COVID-19 pandemic. The data exhibited a clear trend and seasonal pattern, as seen in the raw time series plot.

Figure 1: Quarterly New Car Registrations (2001-Q1 to 2024-Q2)

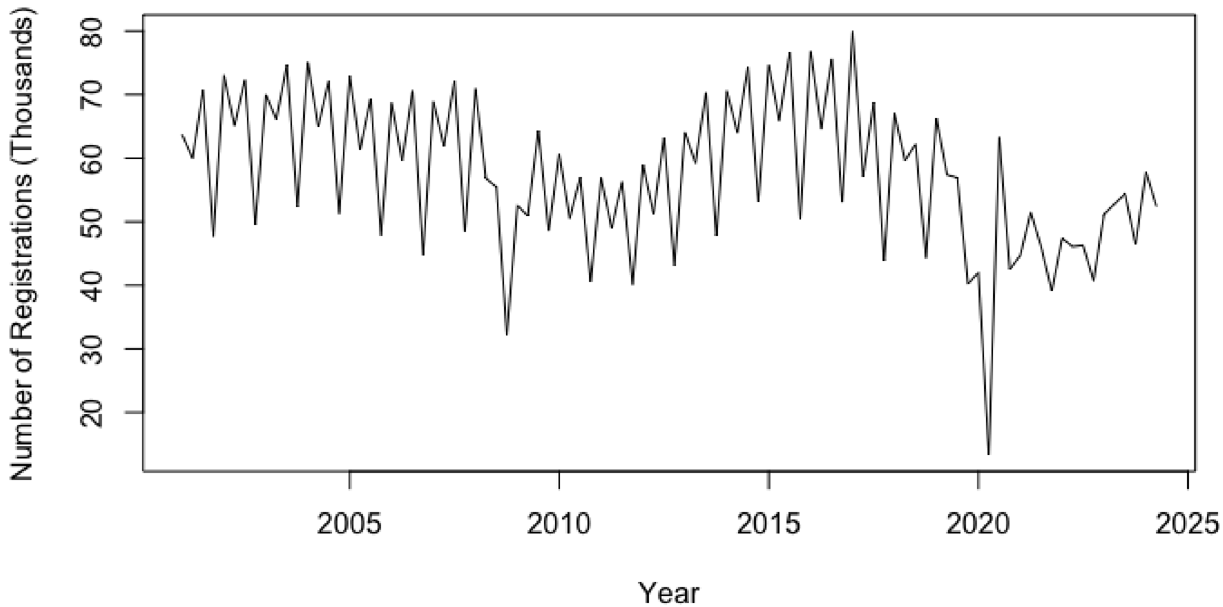
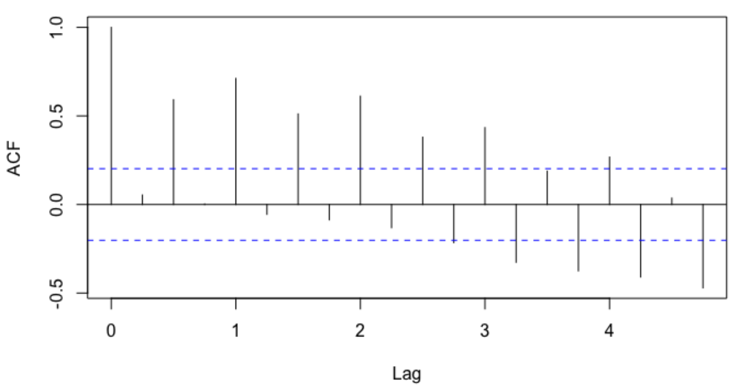


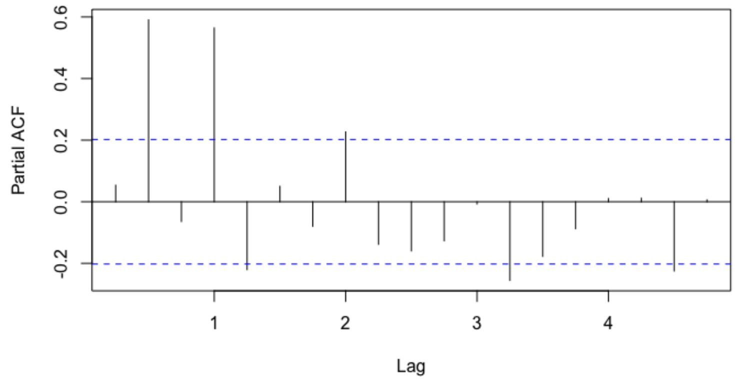
Figure 2: ACF of Car Registrations Time Series



To help identify a suitable model, A sample autocorrelation function (ACF) plot (Figure 2) above was examined. At multiples of lag 4, the ACF plot shows strong positive autocorrelations. This suggests that the seasonality is quarterly for this particular time series. Also, the slow decay of autocorrelations being consistent with a non-stationary process suggests the need for differencing

Similarly, and a sample partial autocorrelation function (PACF) plot (Figure 3) given below shows a significant spike at lag 1, followed by small insignificant spikes at higher lags. Thus it might potentially be an autoregressive structure of order 1.

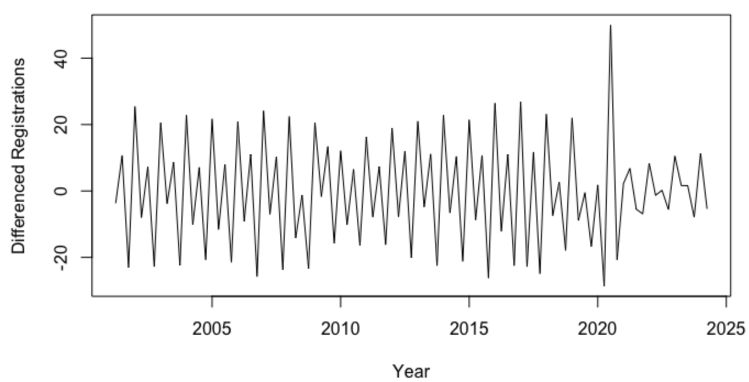
Figure 3: PACF of Car Registrations Time Series

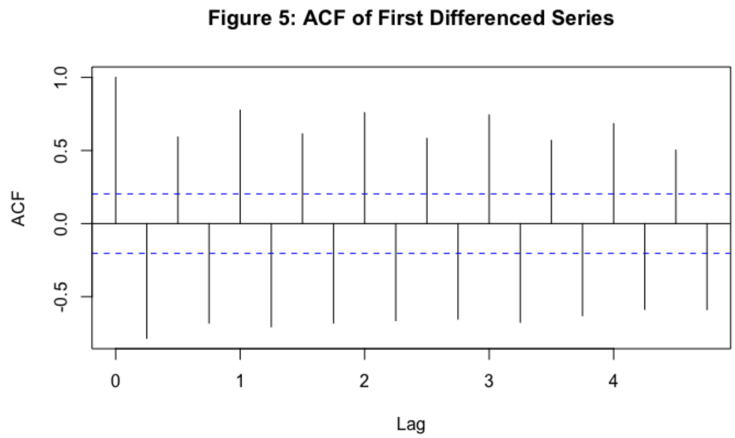


III. Data transformation: differencing

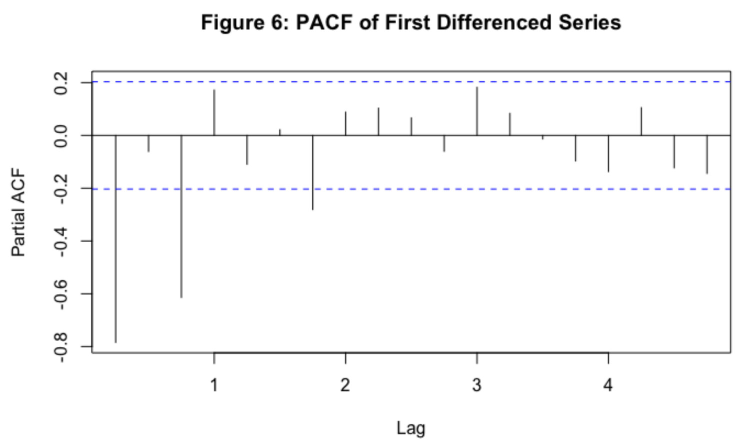
Given that the original series is of non-stationary nature, a first difference was applied to remove trends and stabilize the mean. The time plot of the first differenced series (Figure 4) shows a series fluctuating around a constant mean, with much of the original trend removed. However, some seasonal patterns still appear to remain.

Figure 4: First Differenced Car Registrations Time Series





The ACF plot of the differenced series above (Figure 5) exhibits significant spikes at lags that are multiples of 4, indicating that seasonal autocorrelation persists even after first differencing. This suggested that we might still need additional seasonal difference at lag 4 to remove the seasonal effects.



The PACF plot of the differenced series above (Figure 6) shows a strong negative spike at lag 1, consistent with an autoregressive component of order 1 (AR(1)). There's no other significant spikes at higher lags. This serves as evidence for the use of a simple AR(1) structure in the non-seasonal part of the model.

Based on the findings from reading the plot, a Seasonal ARIMA model of the form SARIMA(1,1,0)(0,1,1)[4] was selected for fitting:

- The non-seasonal AR(1) term accounts for short-term dependence,
- The non-seasonal differencing (d=1) removes trend,
- The seasonal differencing (D=1) accounts for quarterly seasonality,
- The seasonal MA(1) term captures seasonal dependence.

Further validation was required to confirm that SARIMA(1,1,0)(0,1,1)[4] is the appropriate fit for the data.

IV. Model Validation

To evaluate the performance of the selected SARIMA(1,1,0)(0,1,1)[4] model, key model summary statistics were examined. Table 1 given below summarizes the estimated model coefficients, their standard errors, and various model fit and forecast accuracy measures.

The fitted AR(1) coefficient is -0.4393 (standard error = 0.0966), and the seasonal MA(1) coefficient is -0.8839 (standard error = 0.0787), both of which are relatively large in magnitude and statistically significant. The standard errors are reasonably small compared to the coefficients, which means that the parameter estimation was done reliably.

The residual variance ($sigma^2$) is estimated as 52.41, and the model's AIC and BIC values are 615.11 and 622.58, respectively. These information criteria indicate a reasonably good model fit, with lower values being desirable in model selection. Additionally, The model's in-sample accuracy measures show a Root Mean Squared Error (RMSE) of 6.96 and a Mean Absolute Error (MAE) of 4.14. Both values suggest that the fitted SARIMA(1,1,0)(0,1,1)[4] model is a reasonable fit to the data, with relatively small forecasting errors.

Table 1

```
Series: car_ts
ARIMA(1,1,0)(0,1,1)[4]

Coefficients:
          ar1     sma1
      -0.4393  -0.8839
s.e.   0.0966   0.0787

sigma^2 = 52.41:  log likelihood = -304.56
AIC=615.11   AICc=615.4   BIC=622.58

Training set error measures:
                     ME     RMSE      MAE       MPE     MAPE      MASE        ACF1
Training set -0.3138063 6.964417 4.138162 -2.333017 9.530991 0.7357459 -0.09005943
```

Residual check of the model was done. Figure 7 given below consists of the time plot, the ACF plot of residuals and a histogram of the residuals. All these figures helped us in validating our model.

The time plot of residuals (topmost in figure 7) displayed no obvious patterns, trends, or seasonal structure, meaning that the model had successfully captured the structure of the data. Outside the 95% confidence interval bounds (blue highlight), the ACF plot (bottom left of figure 7) showed no significant autocorrelations. This supported the argument that the residuals behaved like a white noise process.

A histogram of residuals (Bottom right of Figure 8) is approximately symmetric and bell-shaped, although some slight departures from normality can be observed. However, for forecasting purposes, the residuals' approximate randomness was more critical than perfect normality.

Lastly, a Ljung-Box test was conducted at lag 8. Lag 8 felt like the correct parameter because it was significantly larger than the number of fitted parameters and captured two full seasonal cycles for quarterly data. This made it an appropriate choice to detect any remaining autocorrelation. The degrees of freedom in the Ljung-Box test adjusted for the number of model parameters fitted. Since we two fitted parameters and had lag 8, 6 degrees of freedom were used to fairly assess whether residual autocorrelations remain. The test returned a p-value of 0.1845 which was greater that 0.05. Therefore, we failed to reject the null hypothesis of no residual autocorrelation.

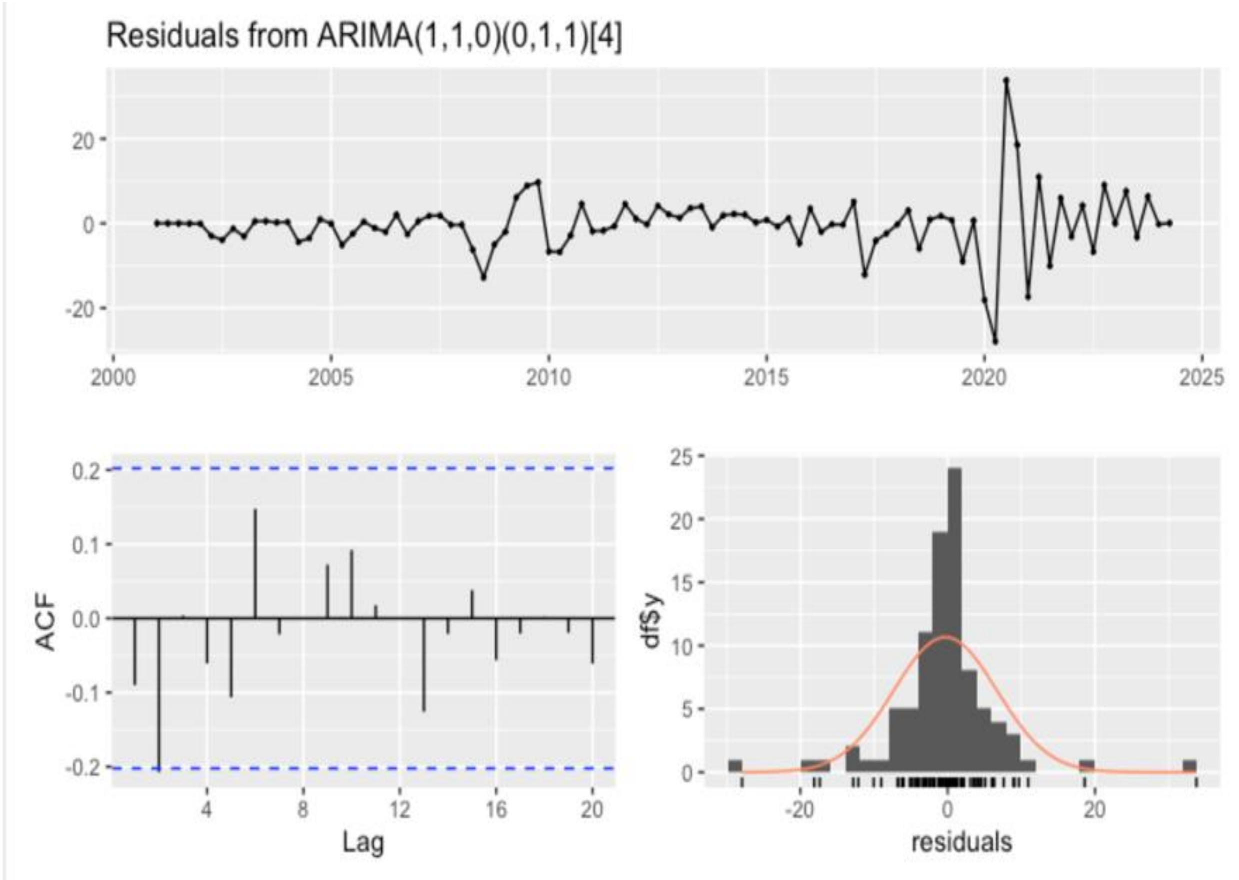


Figure 7

The tests conducted, the observations from the plot and results of the summary statistics together indicate that the SARIMA(1,1,0)(0,1,1)[4] model provides was a satisfactory fit to the data and deemed appropriate for forecasting future car registration values.

V. Forecasting Results

Forecasts for the number of new car registrations for Q3 and Q4 of 2024, and Q1 and Q2 of 2025, were produced along with 95% prediction intervals. The point forecasts, along with associated 95% prediction intervals, are summarized in Table 2 given below.

Table 2

Quarter	Point Forecast (Thousands)	95% Prediction Interval (Thousands)
2024 Q3	59.96	(45.76, 74.15)
2024 Q4	43.85	(27.57, 60.12)
2025 Q1	58.93	(39.45, 78.40)
2025 Q2	51.81	(30.14, 73.47)

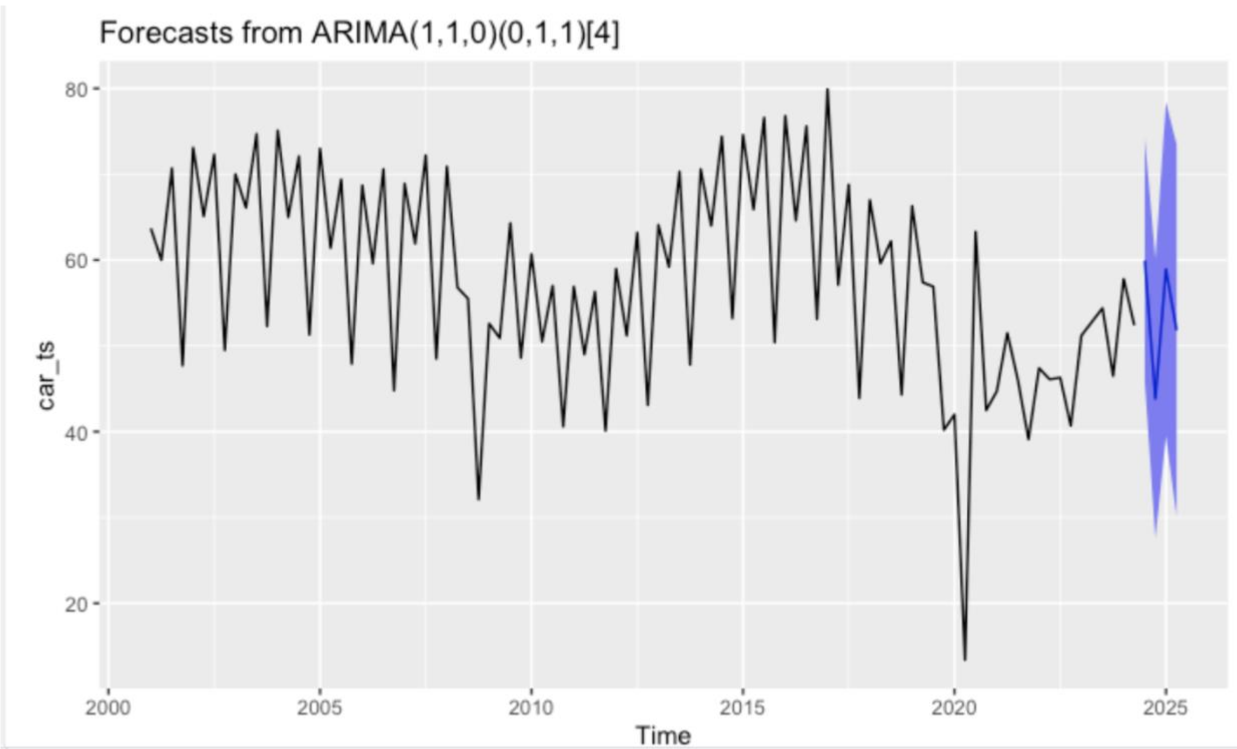


Figure 8

From the plot, we forecast that new car registrations will increase the during Q3 of 24 before declining during Q4 of 2024 and then rising again in 2025. Although the prediction intervals widen slightly further into the future, reflecting increasing uncertainty, the forecasts imply that new car registrations are expected to in line with the recent historical patterns.

VI. Conclusion

This project successfully analysed quarterly new car registration data for Scotland from 2001 to 2024, using time series modelling techniques. Following exploratory data analysis, the patterns observed in the ACF/PACF plots and model validation criteria, a SARIMA(1,1,0)(0,1,1)[4] model was selected. Residual diagnostics, including ACF of residuals and the Ljung-Box test, indicated that the model provided a good fit with no significant autocorrelation remaining. Point forecasts and associated 95% prediction intervals were produced successfully for Q3–Q4 2024 and Q1–Q2 2025. The model performance suggests it is suitable for short-term forecasting applications in the Scottish car industry.

Appendix: R code

```
```{r}
# Read dataset and display first few rows
scot_car_data <- read.csv("scot_car_data.csv")
head(scot_car_data)

#Create a time series
car_ts <- ts(scot_car_data$no_new_reg, start = c(2001,1), frequency = 4)

#Plot the time series
plot(car_ts,
     main = "Figure 1: Quarterly New Car Registrations (2001-Q1 to 2024-Q2)",
     ylab = "Number of Registrations (Thousands)",
     xlab = "Year")
```

```{r}
#Plot the ACF and PACF
acf(car_ts, main = "Figure 2: ACF of Car Registrations Time Series")
pacf(car_ts, main = "Figure 3: PACF of Car Registrations Time Series")
```

```{r}
#First differencing to get stationarity
diff_car_ts <- diff(car_ts)

#Plot the differenced series
plot(diff_car_ts,
     main = "Figure 4: First Differenced Car Registrations Time Series",
     ylab = "Differenced Registrations",
     xlab = "Year")

#Plot ACF and PACF of differenced data
acf(diff_car_ts, main = "Figure 5: ACF of First Differenced Series")
pacf(diff_car_ts, main = "Figure 6: PACF of First Differenced Series")
```

```{r}
#Manually fit model based on ACF and PACF plots
manual_model <- Arima(car_ts, order=c(1,1,0), seasonal=list(order=c(0,1,1), period=4))
summary(manual_model)
```

```{r}
#Model validation. Check residuals of fitted model
checkresiduals(manual_model)
```

```{r}
# Forecast for next 4 quarters
future_forecast <- forecast(manual_model, h=4, level=95)

# Simple plot of forecast along with intervals
autoplot(future_forecast)
print(future_forecast)
```
```