Multiple Linear Regression and Time Series analysis

Valentina Bernal Gomez
Student number: 23395745
Statistics & Optimisation
x23395745@student.ncirl.ie

I. Section A: Multiple Linear Regression

A. Introduction

This section will be dedicated to multiple linear regression analysis. In multiple linear regression models the independent variables can be any type of predictor and the idea is to estimate the value of the independent variable. In this case we will use three independent variables (x1, x2 and x3) to find the best model to predict the variable y.

Having said that, exploratory data analysis, data preparation, modelling, interpretation, diagnostics and evaluation will be developed during this report.

B. Exploratory data analysis

It is important to look over the variables, with this pair plot we can identify patterns or linear relationships between the variables, moreover this plot shows in its diagonal univariate distributions.

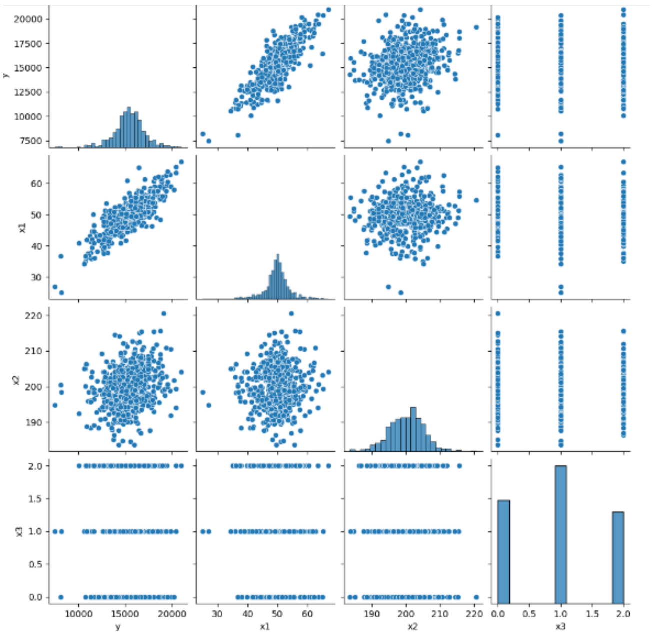
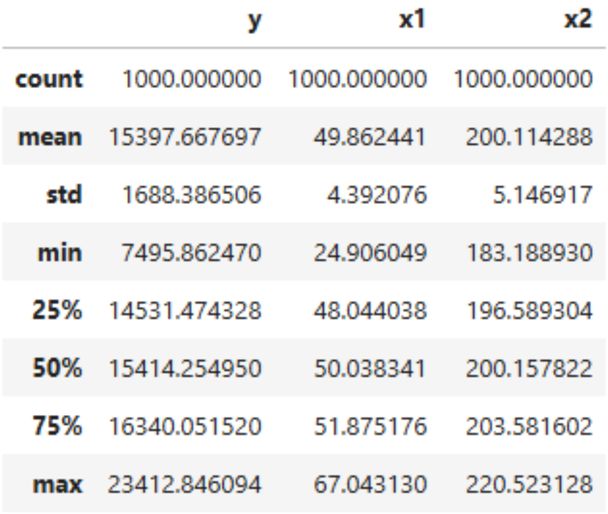


Fig. 1. Pair plot between dependent and independent variables

From figure 1, we can see that there is a linear relationship between the dependent variable y and the independent variable x1, and the diagonal shows a normal distribution for each variable.

It is important to understand the analysis for the variables, having said that we plot the statistics summary.

TABLE 1. Descriptive statistics



	y	x1	x2
count	1000.000000	1000.000000	1000.000000
mean	15397.667697	49.862441	200.114288
std	1688.386506	4.392076	5.146917
min	7495.862470	24.906049	183.188930
25%	14531.474328	48.044038	196.589304
50%	15414.254950	50.038341	200.157822
75%	16340.051520	51.875176	203.581602
max	23412.846094	67.043130	220.523128

From table 1, we can conclude that:

1. The average values for y, x1 and x2 are 15397,66, 49.86 and 200.11 respectively.
2. The standard deviation suggest data are gathered around the mean.
3. The data for the dependent variable varies between 7495.86 and 23412.84. For the variable x1 and x2 the data fluctuate between 24.9 and 67.04, 183.18 and 220.52 in that order.

As we cannot see any useful information from the descriptive statistics for the variable x3, we plot and visualize this variable.

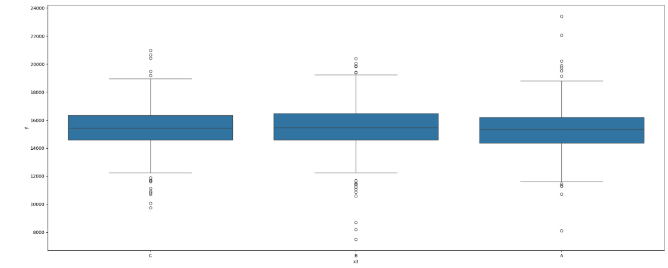


Fig. 2. Box plot for categorical variable x3

Outliers are present in this categorical variable, however, as we do not have any context for the dataset, we decide to continue with these values and do not drop any outliers, they might be important data.

After splitting the data and encoding the categorical variable, we can visualize the dependent variable.

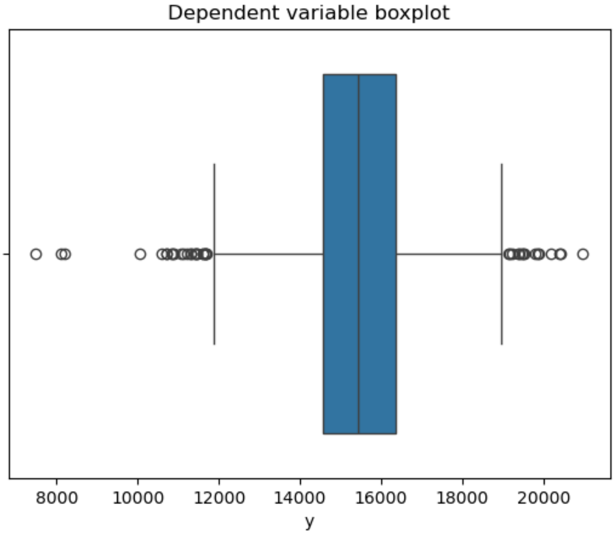


Fig. 3. Box plot for the dependent variable

The dependent variable has outliers, those can significantly influence the model, but considering elimination is not an option in this case, as we do not have any context for the data.

Heatmaps are an excellent tool when understanding the correlation between the variables:

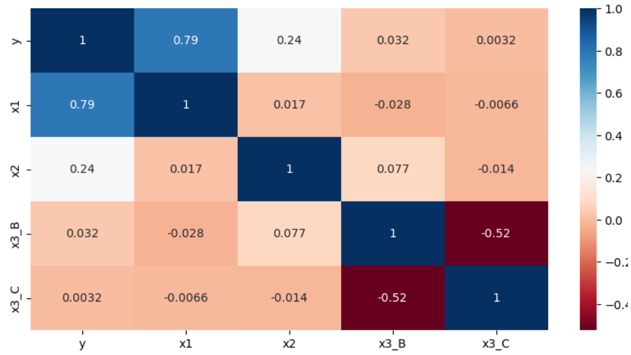


Fig. 4. Heatmap for all the variables

In figure 4 we can see that the dependent variable (y) and the independent variable (x1) are highly correlated (0.79), this is a good signal for the model and indicates that (x1) has a strongly effect on the dependent variable (y) and this might help with the model performance.

Likewise, from the heatmap, we can see that it is not necessary to eliminate any independent variable as they are not closely related to another (collinearity).

C. Data Preparation

To ensure that the data is prepared to be trained to the model, we must do some data pre-processing, this includes reviewing the missing values, encoding categorical data, and normalizing numerical features.

As the dataset has a categorical variable (x3), dummy variables have been used in the model as numerical variables, to allow their use in the OLS model.

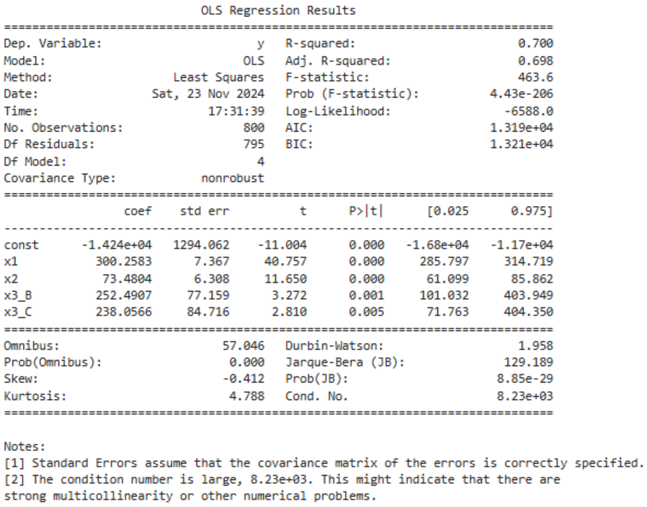
Splitting the dataset is important to ensure that the model is accurate and performs well to unseen data. For this report we have split the data into train (80%) and test (20%) partitions, using random seed, this random seed is defined with the student number (23395745)

In this case, the data does not have any missing values, and it is not necessary to scale the data as they have comparable ranges.

D. Modelling

Model building starts deciding which variables to include in the model. Using the heatmap we can conclude that the three variables have a positive correlation with the dependent variable, this means that when the independent variables (x1, x2 and x3) increase, the dependent variable (y) rises as well.

Now, we can fit an OLS model and print a summary of the fitted model.



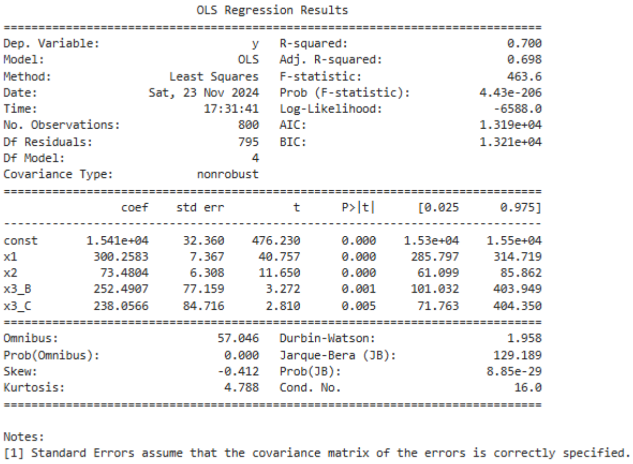
```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.700
Model:                            OLS   Adj. R-squared:                  0.698
Method:                 Least Squares   F-statistic:                     463.6
Date:                Sat, 23 Nov 2024   Prob (F-statistic):          4.43e-206
Time:                        17:31:39   Log-Likelihood:                -6588.0
No. Observations:                 800   AIC:                         1.319e+04
Df Residuals:                     795   BIC:                         1.321e+04
Df Model:                           4
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const      -1.424e+04   1294.062    -11.004      0.000   -1.68e+04   -1.17e+04
x1           300.2583      7.367     40.757      0.000     285.797     314.719
x2            73.4804      6.308     11.650      0.000      61.099      85.862
x3_B         252.4907     77.159      3.272      0.001     101.032     403.949
x3_C         238.0566     84.716      2.810      0.005      71.763     404.350
==============================================================================
Omnibus:                       57.046   Durbin-Watson:                   1.958
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              129.189
Skew:                          -0.412   Prob(JB):                     8.85e-29
Kurtosis:                       4.788   Cond. No.                     8.23e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 8.23e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

Fig. 5. OLS regression results – first model

After fitting the model, there is a warning at the end of the report that suggests multicollinearity may be present.

This can be solved by centering the independent variables, which subtract the mean of each variable from its original values and will help with multicollinearity and reduce VIF. After centering the independent variables, we fit the model, and the results are shown below:



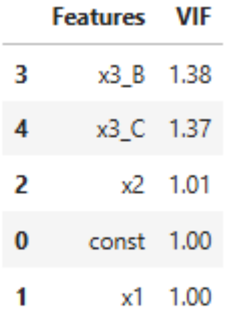
```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.700
Model:                            OLS   Adj. R-squared:                  0.698
Method:                 Least Squares   F-statistic:                     463.6
Date:                Sat, 23 Nov 2024   Prob (F-statistic):          4.43e-206
Time:                        17:31:41   Log-Likelihood:                -6588.0
No. Observations:                 800   AIC:                         1.319e+04
Df Residuals:                     795   BIC:                         1.321e+04
Df Model:                           4
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const       1.541e+04     32.360    476.230      0.000    1.53e+04    1.55e+04
x1           300.2583      7.367     40.757      0.000     285.797     314.719
x2            73.4804      6.308     11.650      0.000      61.099      85.862
x3_B         252.4907     77.159      3.272      0.001     101.032     403.949
x3_C         238.0566     84.716      2.810      0.005      71.763     404.350
==============================================================================
Omnibus:                       57.046   Durbin-Watson:                   1.958
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              129.189
Skew:                          -0.412   Prob(JB):                     8.85e-29
Kurtosis:                       4.788   Cond. No.                         16.0
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

Fig. 6. OLS regression results – Second model

The condition number is not present anymore. We will check for multicollinearity by examining the Variance Inflation Factor of the independent variables.

TABLE 2. Variance Inflation Error



	Features	VIF
3	x3_B	1.38
4	x3_C	1.37
2	x2	1.01
0	const	1.00
1	x1	1.00

There is not a VIF over 5, so we can conclude there is no multicollinearity.

E. Interpretation

The coef column shows the intercept value and the slope for the independent variables. These are interpreted as follows:

- The const coefficient value or intercept indicates that the value of y will be 15410 when all independent variables are at 0.
- The value for the x1 coefficient indicates that the value of y will change by 300.25 when the value of x1 changes by 1 and all other x variables are unchanged.
- The value for the x2 coefficient indicates that the value of y will change by 73.48 when the value of x2 changes by 1 and all other x variables are unchanged.
- The value for the x3_B coefficient indicates that the value of y will change by 252.49 compared to the variable x3_A when all other x variables are unchanged.
- The value for the x3_C coefficient indicates that the value of y will change by 238.05 compared to the variable x3_A when all other x variables are unchanged.

The P>|t| column shows the p value for the intercept and the slopes. This tells us how important each is to the model. The p value for all the variables is less than 0.5 (assuming alpha 0f 0.05), this would mean they are likely to make a significant contribution to the model. This also means their coefficients are not likely to be more than 0 by random chance.

Prob (F-statistic): This p-value for the overall F-statistics of the model. As the value is infinitesimally small at 4.43e-206, so the model is more likely to fit the data better than the one with just the intercept.

F. Diagnostics

To continue with the regression diagnostics, we must extract required measures, first we must store some measures that we will need in one or more plots (standardised and studentised residuals). Both of these allow us to determine the magnitude of the residuals in standard deviation units, permitting the easy identification of outliers.

1. Residuals vs fitted plot

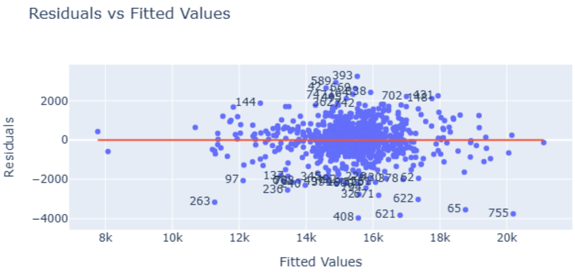


Fig. 6. Residual vs fitted plot

With this plot we can check for linearity by observing if the line is flat and close to zero. In this instance, the line is nearly flat and is close to zero, showing that the mean value of the residuals is also close to zero.

We can check for homoscedasticity by observing the spread of residuals around the line. The residuals appear to be gathered around the center. This might indicate homoscedasticity in not hold.

We can determine if there are outliers. There are some outliers in the plot, extreme values that are much higher than the rest.

2. Normal Quantile-Quantile plot

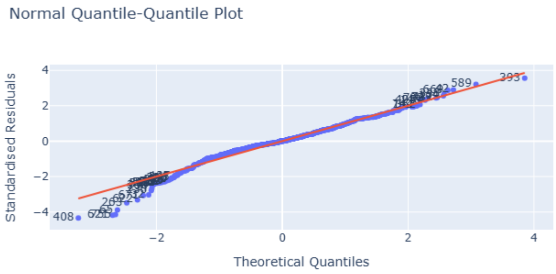


Fig. 7. Normal Quantile-Quantile plot

The normal QQ is used to determine if the residuals are normally distributed. In this instance, as the points are for the most part located near the line, it is likely that the residuals are close to normally distributed.

3. Scale-Location plot

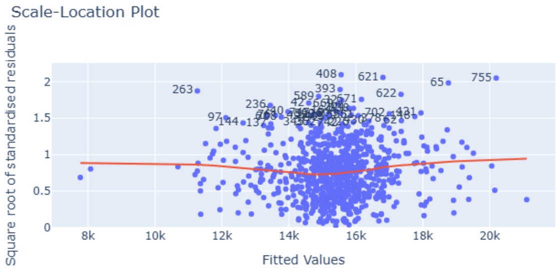


Fig. 8. Scale-Location plot

This plot shows the estimated (fitted) values the regression model on the x-axis and the squared root of the standardised residual on the y-axis.
There is a particular pattern to the residuals, but the line is roughly horizontal, indicating that variance of the residuals is approximately the same for all fitted values(homoscedasticity); for some steps above we confirm there is no homoscedasticity.

4. Residuals vs Leverage plot

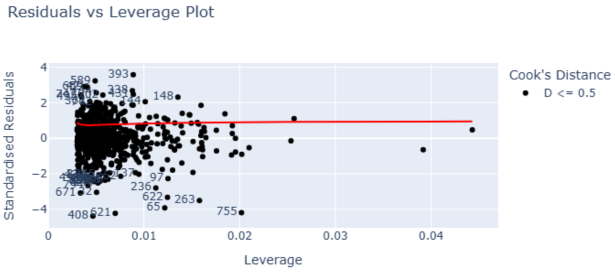


Fig. 9. Residuals vs Leverage plot

This plot is used to identify influential data points on the model.
From this plot we can see that no outliers have a Cook's distance over 5, so there are not influential data points in the model.

5. Normality – The Shapiro-Wilk test

Ho: The sample is drawn from a normal distribution
Hi: The sample is drawn from a non-normal distribution.

We will set alpha to 0.05 for this test. The result for this test is: "The test statistic is 0.976. The p-value is 0.000."

As the p-value is lower than α, we reject the null hypothesis and conclude that the residuals are not normally distributed.

6. Homoscedasticity – The Breusch-Pagan test

Ho: The residuals exhibit homoscedasticity (the error variance is constant)
Hi: The residuals exhibit heteroscedasticity (the error variance is non-constant)

We will set α to 0.05 for this test. The result for this test is: "The test statistic is 44.104. The p-value is 0.000."

As the returned p-value is less than α we reject the null hypothesis and conclude that the residuals do not exhibit constant variance.

7. Serial correlation – The Durbin-Watson test

Ho: There is no autocorrelation in the residuals
Hi: There is autocorrelation in the residuals.

The result for this test is: 1.958

The test statistic ranges from 0 to 4, with a value around 2 indicating that there is no autocorrelation, a value closer to 0 indicating the presence of positive autocorrelation and a value closer to 4 indicating the presence of negative autocorrelation.
As the test statistic is close to 2, we can conclude that there is no dependence among the residuals.

Since we have the results of the diagnostics test and their respective plots, we can conclude that the model does not meet the Gauss-Markov assumptions. This can impact the interpretation of the model, and the results cannot be reliable. Further analysis should be carried out and it is important to keep in mind outliers were held during the model and OLS model is sensitive to outliers.

G. *Evaluation*

We will use the model to predict the values of y for all values of the independent variables in the test data.
We produce the Mean Absolute Error metric for the model, the result is 29766.002 and the Root Mean Squared Error metric for the model is 172.55, this means the predictions are 172.55 units far from the real values for y.

R2 and adjusted R2:

- R2 = 0.70: this means that 70% of the variation in the dependent variable is explained by the independent variables included in the model.
- Adjusted R2 = 0.698, adjusted R2 adjusts the R2 accounting for the number of variables.
- The outcome for both R2 and adjusted R2 are so close, this means that the independent variables included in the model are meaningfully contributing to the fit.

The model before and after centering the variables have the same values for AIC and BIC. If more models are produced, for instance with no outliers, the model with an AIC and BIC lower is better.

II. Section B: Time Series Analysis

A. *Introduction*

This section will be focused on time series analysis, in this analysis the independent variables include past observations or exogenous variables, but in this case we will be focused on past observations.

The main idea is to analyze and forecast the time series, finding the best ARIMA model.

B. *Exploratory data analysis*

Before creating a model, it is good practice to plot the series to identify what patterns are present.

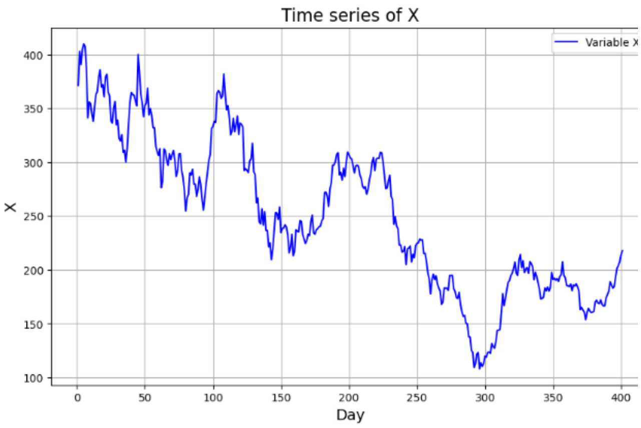


Fig. 10. Time series of X

From this figure we can see the variance appears to be steady over time, however, the mean of the series changes over time. We can check for stationarity using the Augmented Dickey-Fuller test. The hypotheses for this are:

Ho: The time is non-stationary. There is a change in either or both of the mean and variance over time

Hi: The time series is stationary. There is no change in the mean.

The result for this test is:

(-1.6286681325611188, 0.4681879425986012, 9, 310, {'1%': -3.451621854687657, '5%': -2.870908950689806, '10%': -2.571761810613944}, 2306.0264548811965)

From this result and the plot, we can conclude the time series is non-stationary, as the p-value (0.46818) is less than α (0.05), we can not reject the null hypothesis (the time series is non-stationary).

C. *Data preparation*

After reviewing and confirming that the time series is non stationary, it is important and necessary to apply differencing in order to transform the non-stationary series into a stationary one.

The plot after differencing is shown as:

```
<Axes: xlabel='Day'>
```

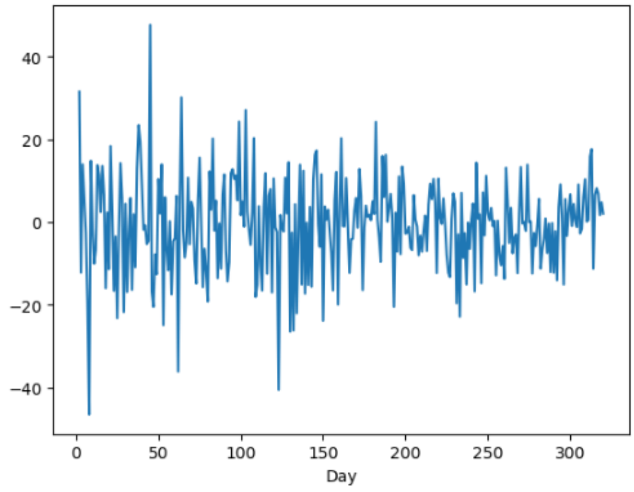


Fig. 11. First differencing plot

The plot can suggest that the time series is still non-stationary, but it is necessary to confirm this with the Augmented Dickey-Fuller test.

The result for this test after differencing is:

0.0

As the p-value is less than 0.001, we reject the null hypothesis and conclude that the first differenced series is stationary.

The ARIMA model includes 3 parameters, and one of them is already known as differencing was done to determine the order of the other two parameters (p and q), so we can produce the autocorrelation function and partial autocorrelation function.

- ACF plot

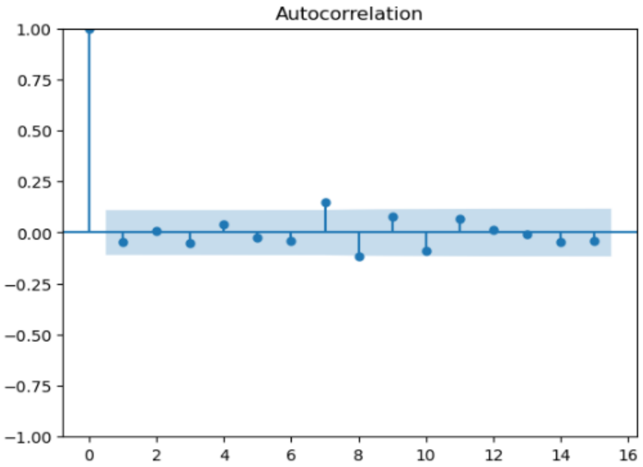


Fig. 12. ACF plot

- PACF plot

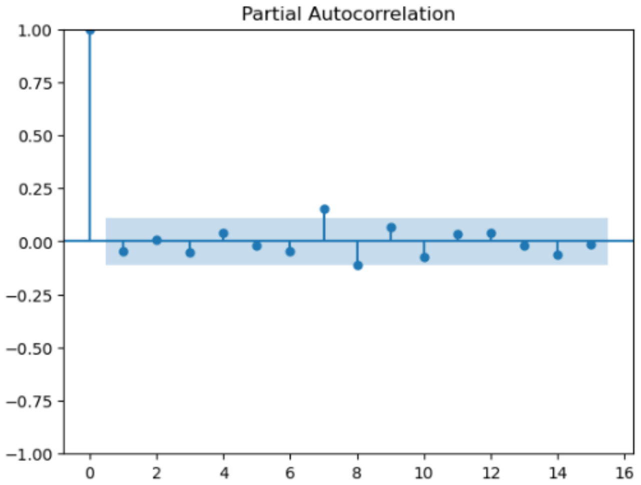


Fig. 13. PACF plot

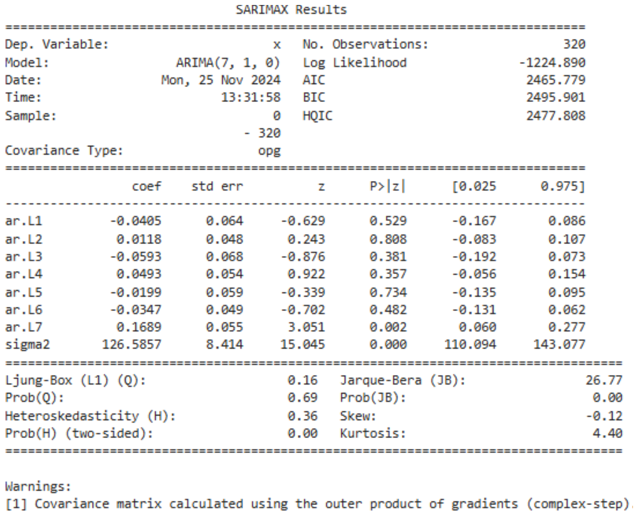
The PACF plot shows a cut off in the lag 7, this indicates the most suitable model is AR(7).

D. *Modelling*

We will first fit an ARIMA (7,1,0) and also we will fit some variations of this model, such as ARIMA(6,1,0), ARIMA(8,1,0) and ARIMA(7,1,1) to compare models and determine which one fits better and produces the best predictions.

The result for the different variations of ARIMA are:

- ARIMA (7,1,0)

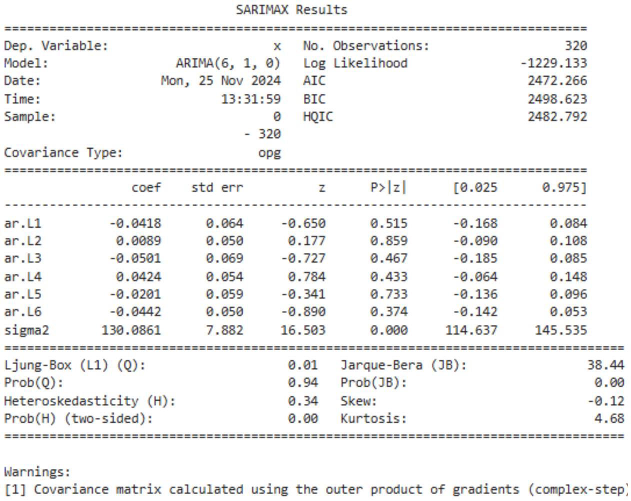


```
                               SARIMAX Results
==============================================================================
Dep. Variable:                      x   No. Observations:                  320
Model:                ARIMA(7, 1, 0)   Log Likelihood               -1224.890
Date:               Mon, 25 Nov 2024   AIC                           2465.779
Time:                       13:31:58   BIC                           2495.901
Sample:                            0   HQIC                          2477.808
                               - 320
Covariance Type:                 opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -0.0405      0.064     -0.629      0.529      -0.167       0.086
ar.L2          0.0118      0.048      0.243      0.808      -0.083       0.107
ar.L3         -0.0593      0.068     -0.876      0.381      -0.192       0.073
ar.L4          0.0493      0.054      0.922      0.357      -0.056       0.154
ar.L5         -0.0199      0.059     -0.339      0.734      -0.135       0.095
ar.L6         -0.0347      0.049     -0.702      0.482      -0.131       0.062
ar.L7          0.1689      0.055      3.051      0.002       0.060       0.277
sigma2       126.5857      8.414     15.045      0.000     110.094     143.077
===================================================================================
Ljung-Box (L1) (Q):                   0.16   Jarque-Bera (JB):                26.77
Prob(Q):                              0.69   Prob(JB):                         0.00
Heteroskedasticity (H):               0.36   Skew:                            -0.12
Prob(H) (two-sided):                  0.00   Kurtosis:                         4.40
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```

Fig. 14. ARIMA (7,1,0)

- ARIMA(6,1,0)

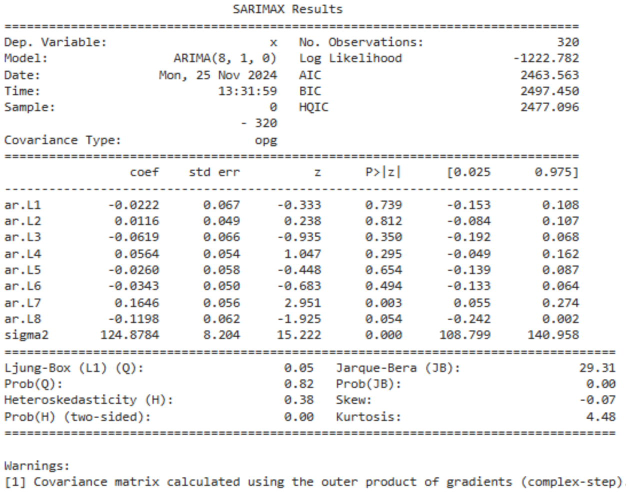


```
                               SARIMAX Results
==============================================================================
Dep. Variable:                      x   No. Observations:                  320
Model:                ARIMA(6, 1, 0)   Log Likelihood               -1229.133
Date:               Mon, 25 Nov 2024   AIC                           2472.266
Time:                       13:31:59   BIC                           2498.623
Sample:                            0   HQIC                          2482.792
                               - 320
Covariance Type:                 opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -0.0418      0.064     -0.650      0.515      -0.168       0.084
ar.L2          0.0089      0.050      0.177      0.859      -0.090       0.108
ar.L3         -0.0501      0.069     -0.727      0.467      -0.185       0.085
ar.L4          0.0424      0.054      0.784      0.433      -0.064       0.148
ar.L5         -0.0201      0.059     -0.341      0.733      -0.136       0.096
ar.L6         -0.0442      0.050     -0.890      0.374      -0.142       0.053
sigma2       130.0861      7.882     16.503      0.000     114.637     145.535
===================================================================================
Ljung-Box (L1) (Q):                   0.01   Jarque-Bera (JB):                38.44
Prob(Q):                              0.94   Prob(JB):                         0.00
Heteroskedasticity (H):               0.34   Skew:                            -0.12
Prob(H) (two-sided):                  0.00   Kurtosis:                         4.68
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step)
```

Fig. 15. ARIMA (6,1,0)

- ARIMA (8,1,0)

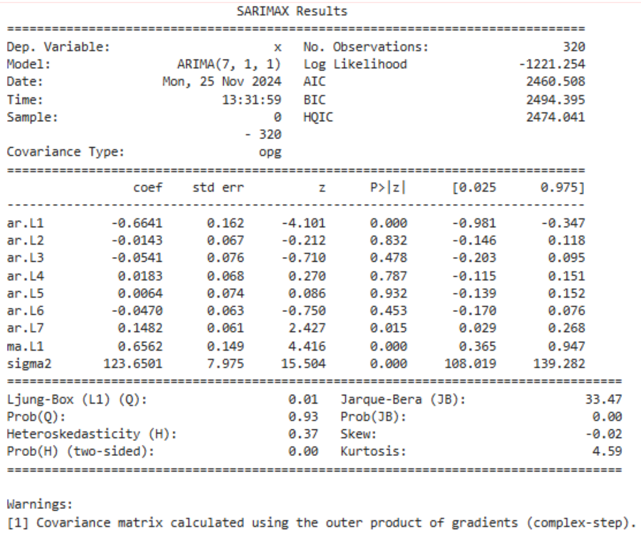


```
                               SARIMAX Results
==============================================================================
Dep. Variable:                      x   No. Observations:                  320
Model:                ARIMA(8, 1, 0)   Log Likelihood               -1222.782
Date:               Mon, 25 Nov 2024   AIC                           2463.563
Time:                       13:31:59   BIC                           2497.450
Sample:                            0   HQIC                          2477.096
                               - 320
Covariance Type:                 opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -0.0222      0.067     -0.333      0.739      -0.153       0.108
ar.L2          0.0116      0.049      0.238      0.812      -0.084       0.107
ar.L3         -0.0619      0.066     -0.935      0.350      -0.192       0.068
ar.L4          0.0564      0.054      1.047      0.295      -0.049       0.162
ar.L5         -0.0260      0.058     -0.448      0.654      -0.139       0.087
ar.L6         -0.0343      0.050     -0.683      0.494      -0.133       0.064
ar.L7          0.1646      0.056      2.951      0.003       0.055       0.274
ar.L8         -0.1198      0.062     -1.925      0.054      -0.242       0.002
sigma2       124.8784      8.204     15.222      0.000     108.799     140.958
===================================================================================
Ljung-Box (L1) (Q):                   0.05   Jarque-Bera (JB):                29.31
Prob(Q):                              0.82   Prob(JB):                         0.00
Heteroskedasticity (H):               0.38   Skew:                            -0.07
Prob(H) (two-sided):                  0.00   Kurtosis:                         4.48
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```

Fig. 16. ARIMA (8,1,0)

- ARIMA(7,1,1)



```
                               SARIMAX Results
==============================================================================
Dep. Variable:                      x   No. Observations:                  320
Model:                ARIMA(7, 1, 1)   Log Likelihood               -1221.254
Date:               Mon, 25 Nov 2024   AIC                           2460.508
Time:                       13:31:59   BIC                           2494.395
Sample:                            0   HQIC                          2474.041
                               - 320
Covariance Type:                 opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -0.6641      0.162     -4.101      0.000      -0.981      -0.347
ar.L2         -0.0143      0.067     -0.212      0.832      -0.146       0.118
ar.L3         -0.0541      0.076     -0.710      0.478      -0.203       0.095
ar.L4          0.0183      0.068      0.270      0.787      -0.115       0.151
ar.L5          0.0064      0.074      0.086      0.932      -0.139       0.152
ar.L6         -0.0470      0.063     -0.750      0.453      -0.170       0.076
ar.L7          0.1482      0.061      2.427      0.015       0.029       0.268
ma.L1          0.6562      0.149      4.416      0.000       0.365       0.947
sigma2       123.6501      7.975     15.504      0.000     108.019     139.282
===================================================================================
Ljung-Box (L1) (Q):                   0.01   Jarque-Bera (JB):                33.47
Prob(Q):                              0.93   Prob(JB):                         0.00
Heteroskedasticity (H):               0.37   Skew:                            -0.02
Prob(H) (two-sided):                  0.00   Kurtosis:                         4.59
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```

Fig. 17. ARIMA (7,1,1)

To generate and check auto_arima it is good practice to know the best model and compare it with the previous ones. Auto_arima function suggests that the best model is (0,1,0), this might be because the time series has no dependence on previous values or errors, we will fit ARIMA (0,1,0) to check the results.

E. Interpretation

The model with the lowest AIC (2460) is ARIMA(7,1,1), this is the model that we will use for predictions.

From this model we can conclude that:

- The coefficient ar.L7 has a significant p-value, this means the lag 7 influences the model.
- The value for the ar.L7 coefficient indicates that the value of the time series will change by 0.1482 for a unit change in the value of the series at lag 7.
- The value for the ma.L1 coefficient indicates that the current value will increase by 0.6562 for a unit prediction error in the previous period.

F. Diagnostics

Checking the residuals is essential to ensure some assumptions for time series analysis.

We will plot the residuals to check if they differ from white noise.

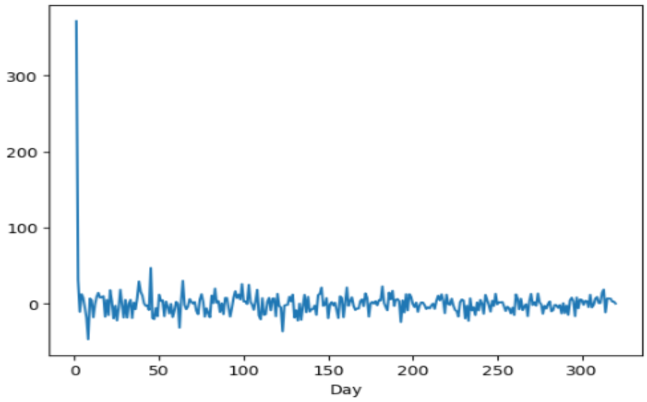


Fig. 18. Residuals plot

The residuals appear to be close to white noise, as they fluctuate around zero and appear to have constant variance. Within the assumptions we should check for correlation in the residuals. We can do this using the Ljung-Box test, this test has the following hypothesis:

Ho: The series is independently distributed
Hi: The series is not independently distributed – it exhibits serial correlation

The result for this test is:

TABLE 3. Ljung-Box test

	lb_stat	lb_pvalue
1	1.511062	0.218977
2	1.676945	0.432371
3	1.956527	0.581478
4	2.056899	0.725294
5	2.097762	0.835459
6	2.568397	0.860736
7	5.771714	0.566644
8	5.836009	0.665595
9	5.886213	0.751250
10	6.388167	0.781665

As it all lags up to 10, the p-value is not significant as they are greater than our α (0.05), indicating that no serial correlation is present.

G. Evaluation

Having the fitted model, we can produce a forecast, we will forecast 81 periods into the future, these periods correspond to the length of the test series.

Having produced the forecast, we can plot the original series, the fitted values and the forecast.

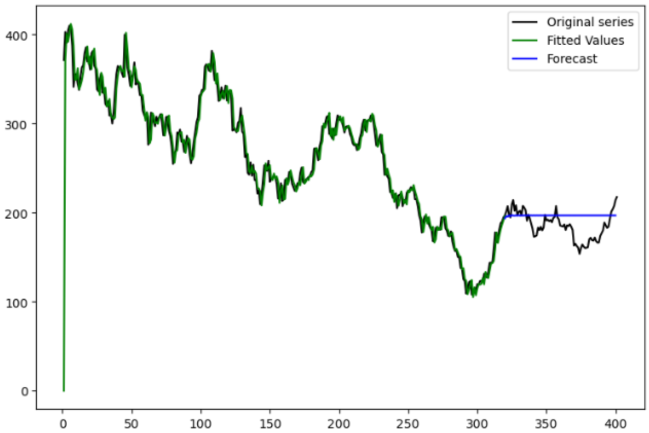


Fig. 19. ARIMA model for original, fitted and forecast values

From this plot we can see that the ARIMA model produced a flat forecast, this happened due to the ARIMA forecast mean reverting, as the greater the forecast horizon, the more they tend towards the mean.

Now, we can calculate the MAE, RMSE and MAPE for the fitted model and the forecast (measures of in sample error and out of sample error).

1. Measure of in sample error:

```
ARIMA(7,1,1) - MAE:9.538, RMSE:23.555, MAPE:0.0
36
```

1.1 MAE 9.538. On average the model makes an error of 9.538 units.
1.2 RMSE 23.555. This value is greater than MAE, this could mean the presence of some higher errors.
1.3 MAPE: 3.6%. On average the predictions are off by 3.6% from the original values.

2. Measure of out of sample error:

```
ARIMA(7,1,1) - MAE:15.328, RMSE:18.934, MAPE:0.
087
```

2.1 MAE 15.328: This measure increases for the forecast; this might be normal as there is uncertainty to predict future values than the observed ones.
2.2 RMSE 19.934: This decreases for the forecast; this might suggest that forecast errors are less variable than errors in the training phase.
2.3 MAPE 8.7%. This value increases for the forecast, suggesting that the model is having difficulty with some patterns in future data.

To sum up, the initial model performs better for training data (MAPE: 3.6%), however, for the forecast values, the model is dealing with uncertainty.