**Forecasting with AR and ARIMA Models**

**• Create time series data set in R using the ts() function.**

Graphical user interface, text

Description automatically generated with medium confidence

Outcome:

Text

Description automatically generated

• **Develop data partition with the validation partition of 16 periods and the rest for the training partition.**

Text

Description automatically generated

Outcome:

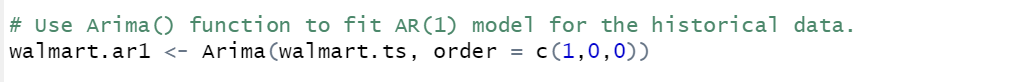
Table

Description automatically generated

**1. Identify time series predictability.**

**a. Using the AR(1) model for the historical data, Provide and explain the AR(1) model summary in your report. Explain if the Walmart revenue is predictable.**

To use the AR(1) model for the historical data:



To display the AR(1) model summary:



Model Summary:

Text, letter

Description automatically generated

The Equation of the model is shown below:

***et* = 117007.26 + 0.9269*et*-1**

The AR(1) coefficient (lag 1) is estimated to be 0.9269, indicating the strength of the autocorrelation between the current residual and the previous residuals. The mean term is estimated to be 117,007.26, representing the non-zero mean of the residuals. The standard errors for the coefficients are also provided, indicating the uncertainty in the estimates.

To check the predictability:

Graphical user interface, text, application

Description automatically generated

Outcome:

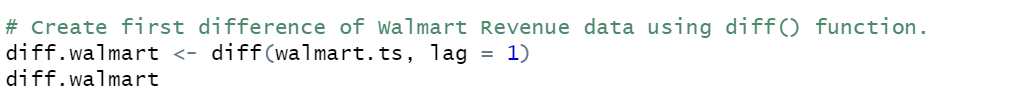
Graphical user interface, text, application

Description automatically generated

We can see the p-value is greater than 0.05, therefore we accepted our null hypothesis. This suggests that the predictability of Walmart's revenue based solely on the AR(1) model may not be strong and conclude that the Walmart revenue is not predictable.

**b. Using the first differencing (lag1) of the historical data and Acf() function, provide in the report the autocorrelation plot of the first differencing (lag1) with the maximum of 8 lags and explain if Walmart revenue is predictable.**

To calculate the first difference:



Outcome:

Table

Description automatically generated

To generate the autocorrelation plot of the first differencing (lag1) with the maximum of 8 lags, using the first differencing (lag1) of the historical data and Acf() function:

Graphical user interface, text

Description automatically generated with medium confidence

Outcome:

Chart, box and whisker chart

Description automatically generated

The autocorrelation plot shows the correlation coefficients for each lag, with confidence intervals indicating the significance of the correlations. Due to the larger autocorrelations observed at lags corresponding to multiples of the seasonal frequency (quarters), it indicates the presence of quarterly seasonality. Additionally, there is a consistently high correlation value across all lags, suggesting the existence of a trend as well. By examining the autocorrelation plot, we can say that there is significant correlation between the first differencing (lag 1) of Walmart's revenue. Therefore, it suggests that the revenue values have some predictability.

**2. Apply the two-level forecast with regression model and AR model for residuals.**

**a. For the training data set, use the tslm() function to develop a regression model with quadratic trend and seasonality. Forecast Walmart’s revenue in the validation period with the forecast() function (use the associated R code from case #2). No explanation is required in your report.**

To develop a regression model with quadratic trend and seasonality and present summary:

Text

Description automatically generated

Outcome:

Text

Description automatically generated

To Forecast Walmart’s revenue in the validation period using the forecast() function:

A picture containing text

Description automatically generated

Outcome:

Table

Description automatically generated with low confidence

**b. Identify the regression model’s residuals for the training period and use the Acf() function with the maximum of 8 lags to identify autocorrelation for these residuals. Provide the autocorrelation plot in your report and explain if it would be a good idea to add to your forecast an AR model for residuals.**

To identify autocorrelation for the regression model’s residuals for the training period, using Acf() function:

Graphical user interface, text

Description automatically generated

Running the above code will generate an autocorrelation plot of the residuals, showing the correlation coefficients for each lag up to a maximum of 8 lags.

Outcome:

Chart, box and whisker chart

Description automatically generated

The autocorrelation plot presented above shows only a few significant values, specifically at lags 1, 2, and 8. The high correlation coefficient observed at lag 1 indicates the presence of a trend. Furthermore, the significant lags at 1, 2, and 8 suggest the presence of autocorrelations among the residuals that were not accounted for in the regression model. In this case, it would be a good idea to include an AR model for the residuals to improve the forecasting accuracy.

**c. Develop an AR(1) model for the regression residuals, present and explain the model and its equation in your report. Use the Acf() function for the residuals of the AR(1) model (residuals of residuals), present the autocorrelation chart, and explain it in your report.**

To develop an AR(1) model for the regression residuals and identifying the parameters:

A picture containing logo

Description automatically generated

Outcome:

Text, letter

Description automatically generated

The Equation of the model is shown below:

***et* = 123.4899 + 0.7585 *et*-1**

The AR(1) coefficient (lag 1) is estimated to be 0.7585, indicating the strength of the autocorrelation between the current residual and the previous residual. The mean term is estimated to be 123.4899, representing the non-zero mean of the residuals. The standard errors for the coefficients are also provided, indicating the uncertainty in the estimates.

The autocorrelation chart:

Chart

Description automatically generated

The graph indicates a lack of significant autocorrelation among the lags, suggesting that all eight lags are already included in the model. Consequently, combining this information with the regression model is expected to enhance the accuracy of the forecast results.

**d. Create a two-level forecasting model (regression model with quadratic trend and seasonality + AR(1) model for residuals) for the validation period. Show in your report a table with the validation data, regression forecast for the validation data, AR(1) forecast for the validation data, and combined forecast for the validation period.**

To Create a two-level forecasting model for the validation period and display a table with the validation data, regression forecast for the validation data, AR(1) forecast for the validation data, and combined forecast for the validation period:

Text

Description automatically generated

Outcome:

Table

Description automatically generated

**e. Develop a two-level forecast (regression model with quadratic trend and seasonality and AR(1) model for residuals) for the entire data set. Provide in your report the autocorrelation chart for the AR(1) model’s residuals and explain it. Also, provide a data table with the models’ forecasts for Walmart revenue in Q1-Q4 of 2023 and 2024 (regression model, AR(1) for residuals, and two-level combined forecast).**

To fit the regression model with quadratic trend and seasonality for the entire dataset and forecast the Walmart revenue in Q1-Q4 of 2023 and 2024 using Regression Model:

Graphical user interface, text

Description automatically generated

Outcome:

Text

Description automatically generated with medium confidence

To fit AR(1) model for the entire dataset and forecast the Walmart revenue in Q1-Q4 of 2023 and 2024:

Text

Description automatically generated

Forecast Outcome:

Text

Description automatically generated

Summary Outcome:

Text, letter

Description automatically generated

The Equation of the model is shown below:

***et* = 296.4102 + 0.9129 *et*-1**

To generate the autocorrelation chart for the AR(1) model’s residuals:

Text

Description automatically generated with low confidence

Outcome:

Chart, box and whisker chart

Description automatically generated

The autocorrelation chart of the AR(1) model's residuals reveals that the autocorrelations appear to be random, indicating that the AR(1) model successfully accounted for any significant autocorrelations present in the residuals. The chart does not show any significant lags or residual patterns, as all the autocorrelation coefficients are within the threshold lines. This suggests that the AR(1) model adequately captures the remaining dependencies in the residuals, and no further autocorrelation needs to be considered in the forecast.

To provide a data table with the models’ forecasts for Walmart revenue in Q1-Q4 of 2023 and 2024 (regression model, AR(1) for residuals, and two-level combined forecast):

Text

Description automatically generated

Outcome:

Table

Description automatically generated

**3. Use ARIMA Model and Compare Various Methods.**

**a. Use Arima() function to fit ARIMA(1,1,1)(1,1,1) model for the training data set. Insert in your report the summary of this ARIMA model, present and briefly explain the ARIMA model and its equation in your report. Using this model, forecast revenue for the validation period and present it in your report.**

To fit ARIMA(1,1,1)(1,1,1) model using Arima() function for the training data set and to generate the summary of this ARIMA model:

Text

Description automatically generated

Outcome:

Text, letter

Description automatically generated

The ARIMA(1,1,1)(1,1,1)[4] means the following:

p=1, order 1 autoregressive model AR(1)

d=1, order 1 differencing to remove linear trend

q=1, order 1 moving average MA(1) for error lags

P=1, order 1 autoregressive model AR(1) for seasonality

D=1, order 1 differencing to remove linear trend

Q=1, order 1 moving average MA(1) for seasonal error lags

m=4, for quarterly seasonality

The equation of the ARIMA model with seasonality is shown below:

***yt* - *yt*-1 = - *0.7265* (*yt*-1 -*yt*-2) *+ 0.6765et-1* + 0.2647(*yt*-1 - *yt*-4)** ***- 0.8859rt-1***

The estimated coefficients of the ARIMA(1,1,1)(1,1,1)[4] model are as follows:

ar1 = -0.7265

ma1 = 0.6765

sar1 = 0.2647

sma1 = -0.8859

The standard errors (s.e.) of these coefficients indicate their uncertainty.

To forecast the revenue for the validation period:

Graphical user interface, text

Description automatically generated

Outcome:

Table

Description automatically generated

**b. Use the auto.arima() function to develop an ARIMA model using the training data set. Insert in your report the summary of this ARIMA model, present and explain the ARIMA model and its equation in your report. Use this model to forecast revenue in the validation period and present this forecast in your report.**

To develop an ARIMA model for the training data set using the auto.arima() function and generate the summary of the ARIMA model:

A picture containing text

Description automatically generated

Outcome:

Text

Description automatically generated

The ARIMA(0,1,0)(1,1,1)[4] means the following:

p=0, no autoregressive model

d=1, order 1 differencing to remove linear trend

q=0, no moving average for error lags

P=1, order 1 autoregressive model AR(1) for seasonality

D=1, order 1 differencing for seasonality

Q=1, order 1 moving average

m=4, for quarterly seasonality

The equation of the ARIMA model with seasonality is shown below:

***yt* - *yt*-1 = 0.2992(*yt*-1 - *yt*-5)** ***- 0.8859rt-1***

The estimated coefficients of the ARIMA(0,1,0)(1,1,1)[4] model are as follows:

sar1 = 0.2992

sma1 = -0.8859

The standard errors (s.e.) of these coefficients indicate their uncertainty. The training set error measures provide information about the accuracy of the model in predicting the training data. These measures include mean error (ME), root mean squared error (RMSE), mean absolute error (MAE), mean percentage error (MPE), mean absolute percentage error (MAPE), mean absolute scaled error (MASE), and autocorrelation of residuals (ACF1). These measures help evaluate the performance of the model in capturing the patterns and variability in the training data.

To forecast the revenue for the validation period:

Graphical user interface, text, application

Description automatically generated

Outcome:

Table

Description automatically generated

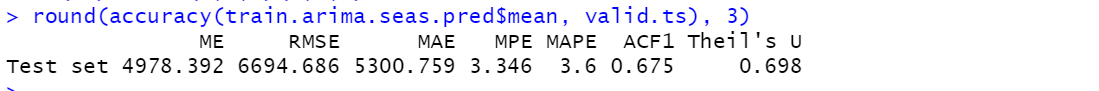
**c. Apply the accuracy() function to compare performance measures of the two ARIMA models in 3a and 3b. Present the accuracy measures in your report, compare them and identify, using MAPE and RMSE, the best ARIMA model to apply.**

To find the performance measures of ARIMA(1,1,1)(1,1,1) and Auto ARIMA model:

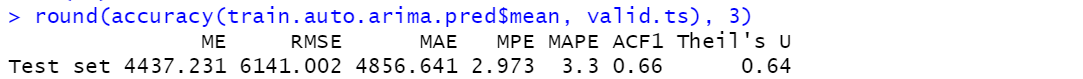
Graphical user interface, text

Description automatically generated

Performance measures of ARIMA(1,1,1)(1,1,1) model:



Performance measures of Auto ARIMA model:



|  |  |  |
| --- | --- | --- |
|  | **RMSE** | **MAPE** |
| **ARIMA(1,1,1)(1,1,1)** | 6694.686 | 3.6 |
| **Auto ARIMA** | 6141.002 | 3.3 |

Based on the accuracy measures shown above for the two models, we can compare them to identify the best ARIMA model to apply based on MAPE and RMSE.

Comparing the MAPE and RMSE values, we can see that the Auto ARIMA model (3b) has a lower MAPE (3.3%) and RMSE (6141.002) compared to the ARIMA(1,1,1)(1,1,1) model. This indicates that the Auto ARIMA model provides a more accurate forecast compared to the other model. Therefore, based on the MAPE and RMSE values, the best ARIMA model to apply would be the Auto ARIMA model (3b).

**d. Use two ARIMA models from 3a and 3b for the entire data set. Present models’ summaries in your report. Use these ARIMA models to forecast Walmart revenue in Q1- Q4 of 2023 and 2024 and present these forecasts in your report.**

To fit ARIMA(1,1,1)(1,1,1) model for the entire data set and generate model summary:

Text

Description automatically generated with medium confidence

Model summary of ARIMA(1,1,1)(1,1,1) model:

Text, letter

Description automatically generated

The ARIMA(1,1,1)(1,1,1)[4] means the following:

p=1, order 1 autoregressive model AR(1)

d=1, order 1 differencing to remove linear trend

q=1, order 1 moving average MA(1) for error lags

P=1, order 1 autoregressive model AR(1) for seasonality

D=1, order 1 differencing to remove linear trend

Q=1, order 1 moving average MA(1) for seasonal error lags

m=4, for quarterly seasonality

The equation of the ARIMA model with seasonality is shown below:

***yt* - *yt*-1 = 0.3224 (yt-1 -yt-2) - 0.3978 et-1 + 0.0788 (yt-1 - yt-4) -1.0000 rt-1**

The estimated coefficients of the ARIMA(1,1,1)(1,1,1)[4] model are as follows:

ar1 = 0.3224

ma1 = -0.3978

sar1 = 0.0788

sma1 = -1.0000

The standard errors (s.e.) of these coefficients indicate their uncertainty.

To forecast the Walmart revenue in Q1- Q4 of 2023 and 2024:

Text

Description automatically generated with medium confidence

Outcome:

Text

Description automatically generated with low confidence

To fit Auto ARIMA model for the entire data set and generate model summary:

A picture containing logo

Description automatically generated

Model summary of Auto ARIMA model:

Text, letter

Description automatically generated

The ARIMA(1,0,0)(2,1,0)[4] means the following:

p=1, order 1 autoregressive model AR(1)

d=0, no differencing to remove linear trend

q=0, no moving average for error lags

P=2, order 2 autoregressive model AR(2) for seasonality

D=1, order 1 differencing to remove linear trend

Q=0, no moving average for seasonal error lags

m=4, for quarterly seasonality

The equation of the ARIMA model with seasonality is shown below:

***yt = 1196.3907 + 0.8771 yt-1 -*** ***0.5464 (yt-1 -yt-5) - 0.2607 (yt-2 -yt-6)***

The estimated coefficients of the ARIMA(1,1,1)(1,1,1)[4] model are as follows:

ar1 = 0.8771

sar1 = - 0.5464

sar2 = - 0.2607

drift = 1196.3907

The standard errors (s.e.) of these coefficients indicate their uncertainty.

To forecast the Walmart revenue in Q1- Q4 of 2023 and 2024:

Text

Description automatically generated

Outcome:

Text

Description automatically generated with low confidence

**e. Apply the accuracy() function to compare performance measures of the following forecasting models for the entire data set: (1) regression model with quadratic trend and seasonality; (2) two-level model (with AR(1) model for residuals); (3) ARIMA(1,1,1)(1,1,1) model; (4) auto ARIMA model; and (5) seasonal naïve forecast for the entire data set. Present the accuracy measures in your report, compare them, and identify, using MAPE and RMSE, the best model to use for forecasting Walmart’s revenue in in Q1-Q4 of 2023 and 2024.**

To find the performance measures of the following forecasting models for the entire data set: (1) regression model with quadratic trend and seasonality; (2) two-level model (with AR(1) model for residuals); (3) ARIMA(1,1,1)(1,1,1) model; (4) auto ARIMA model; and (5) seasonal naïve forecast for the entire data set, using accuracy() function:

Graphical user interface, text, application

Description automatically generated

Outcome:

Text

Description automatically generated

|  |  |  |
| --- | --- | --- |
|  | **RMSE** | **MAPE** |
| **Regression Model** | 4050.66 | 2.935 |
| **Two-level Model** | 1868.012 | 1.169 |
| **ARIMA(1,1,1)(1,1,1) Model** | 1839.248 | 1.068 |
| **Auto ARIMA Model** | 2091.467 | 1.313 |
| **Seasonal Naïve Forecast** | 5599.183 | 3.985 |

Based on the MAPE and RMSE values, the best model to use for forecasting Walmart's revenue in Q1-Q4 of 2023 and 2024 is the ARIMA (1,1,1)(1,1,1) model. This model has the lowest MAPE value of 1.068 and the lowest RMSE value of 1839.248. Therefore, we can say that the ARIMA model seems to be a better choice for this dataset.