4/30/2022
Forecasting

Forecasting: Exam assignment

Vinay Rajagopalan

Contents

Part 1

Dataset

The data set Airpass BE contains international intra-EU air passenger transport from Belgium and EU partner countries, from January 2003 to October 2021.

Data Preprocessing

The time series is split in a training set from January 2003 up to December 2017 and a test set from January 2018 up to February 2020. The remaining data from March 2020 to October 2021 will be used as validation set for the last question.

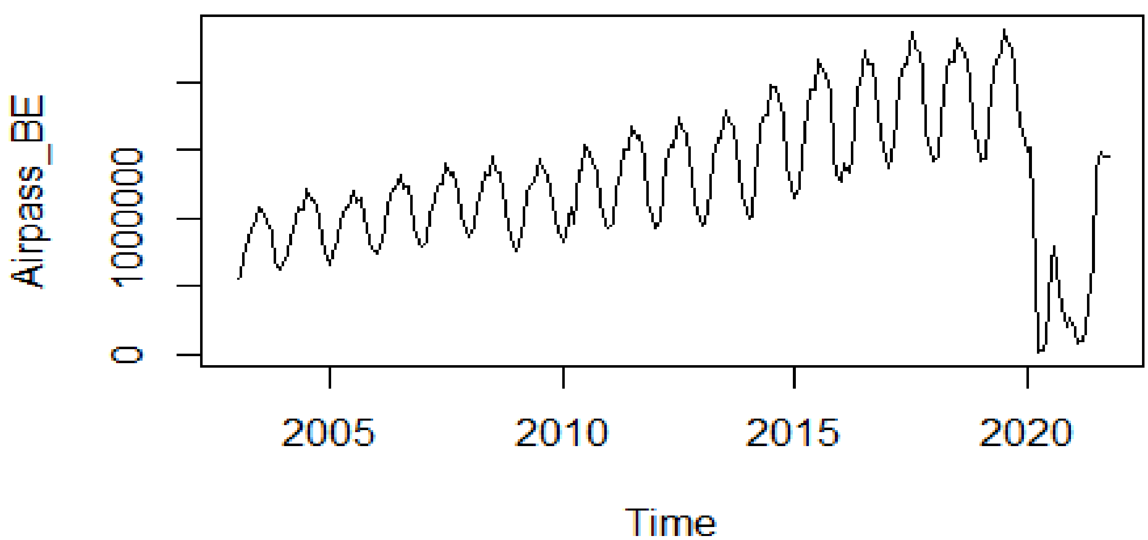
Question 1

Explore the data using relevant graphs and discuss the properties of the data. Include and discuss a time series plot, a seasonal plot, a seasonal subseries plot and a (P)ACF plot.

Ans:

The Time Series Plot shows that there is an increasing trend for most of the years until 2020 in the air passengers. There is also a strong seasonal trend that we see. In the plot we do see dip in air passengers post 2020.

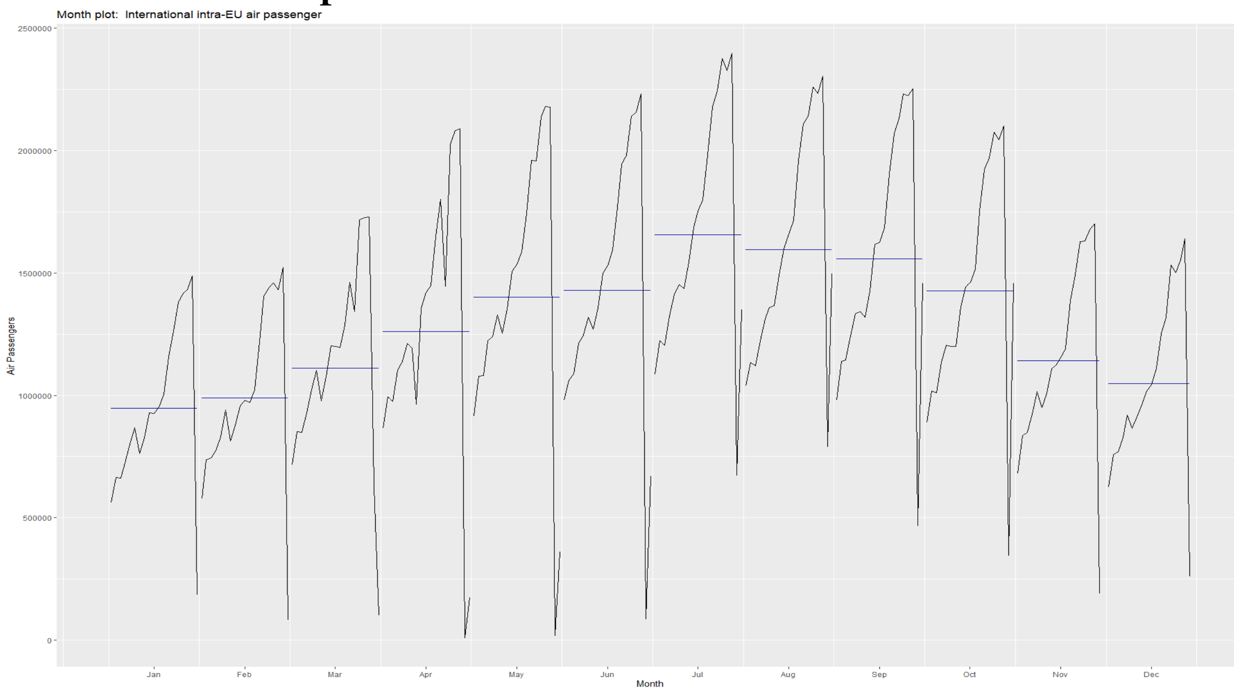
Time series plot



Seasonal and Seasonal subseries plots

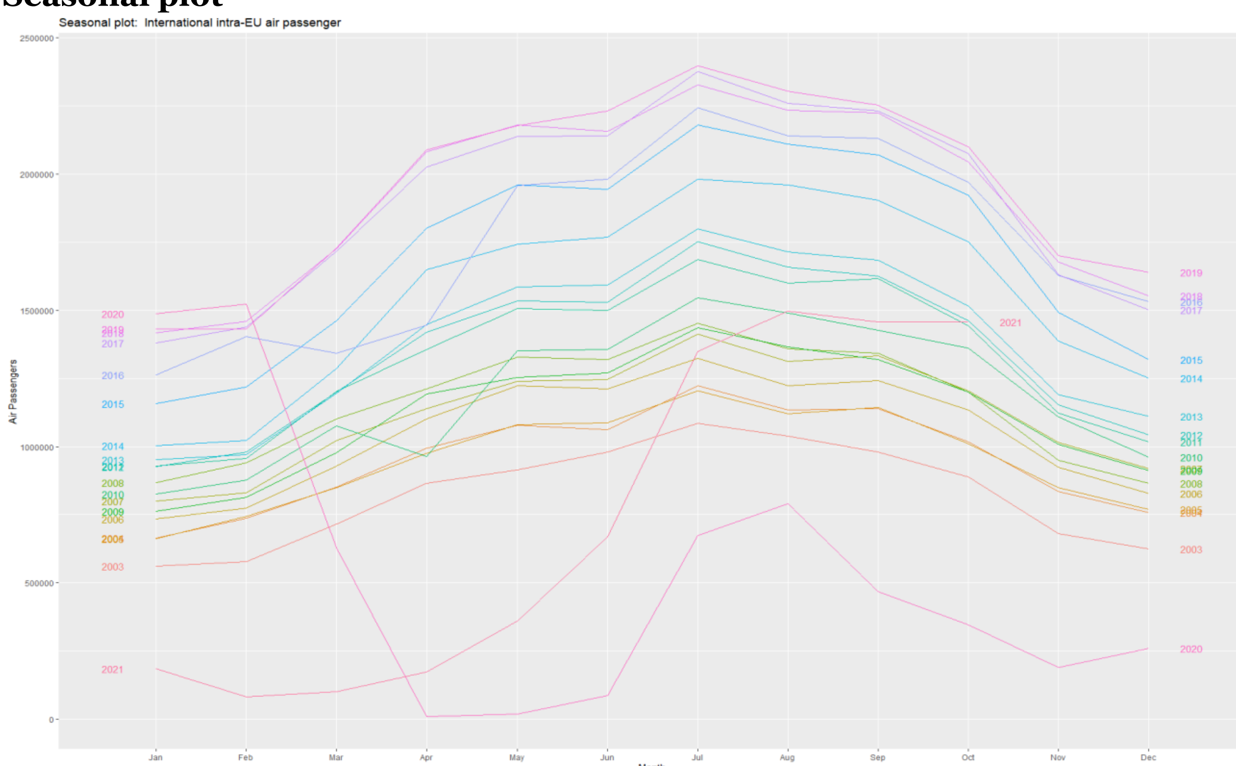
The Seasonal subseries plot shows that July has the highest air passengers this could be due to the holiday travels. We do see a dip towards the end of the month. The mean increases till July then decreases post July indicating a seasonal pattern.

Seasonal subseries plot



The seasonal plot shows an increasing trend in air passengers apart from the year of 2020 and 2021. In 2020 we see a sharp drop in customers in February which could be due to the pandemic period but there seems to be gradual increase in following months with highest air passengers in August for 2020.

Seasonal plot



We would be looking at the P(A)CF plots to check for correlations and partial correlations.

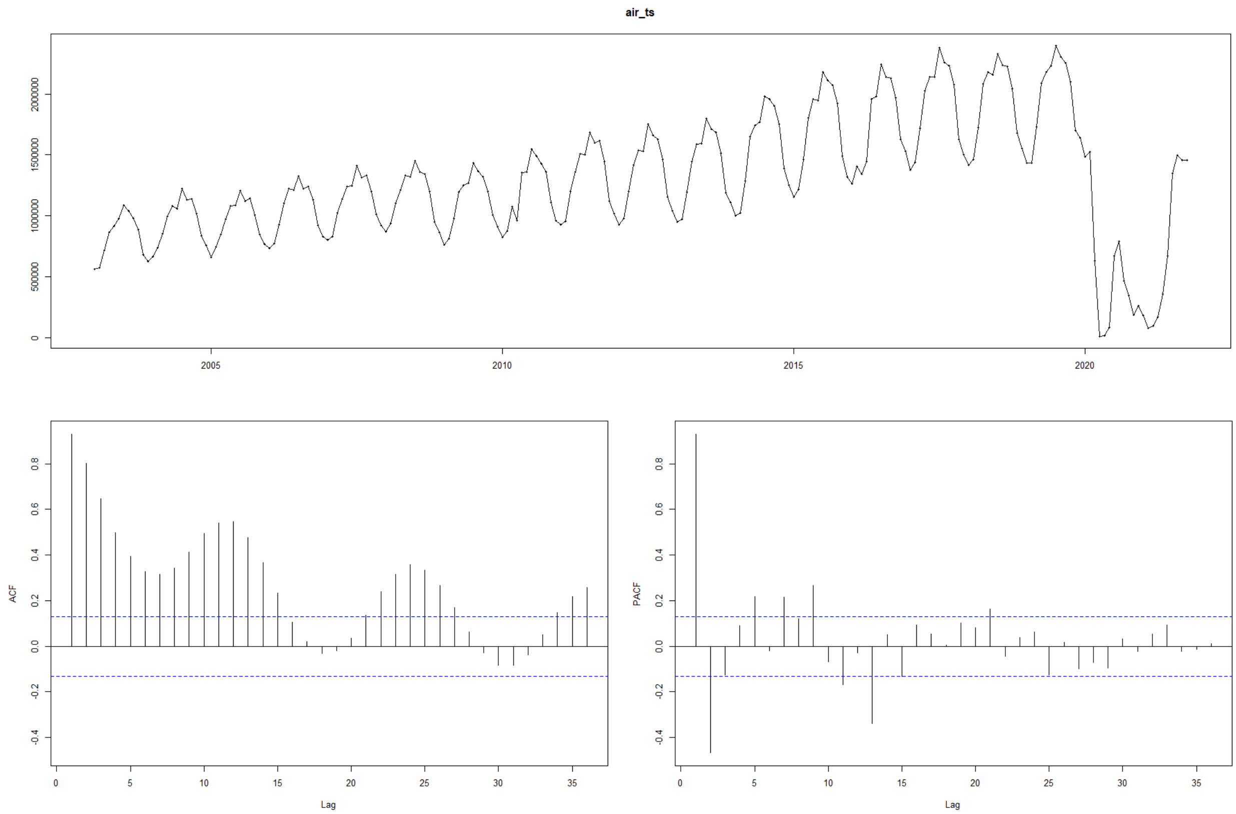
P(A)CF plots

ACF Plot

An ACF measures and plots the average correlation between data points in a time series and previous values of the series measured for different lag lengths.

PACF Plot

A PACF is like an ACF except that each partial correlation controls for any correlation between observations of a shorter lag length.



The slow decrease in the ACF as the lags increase is due to the trend, while the scalloped shape is due the seasonality. The PACF plot has the highest Partial correlation for the first and second lag.

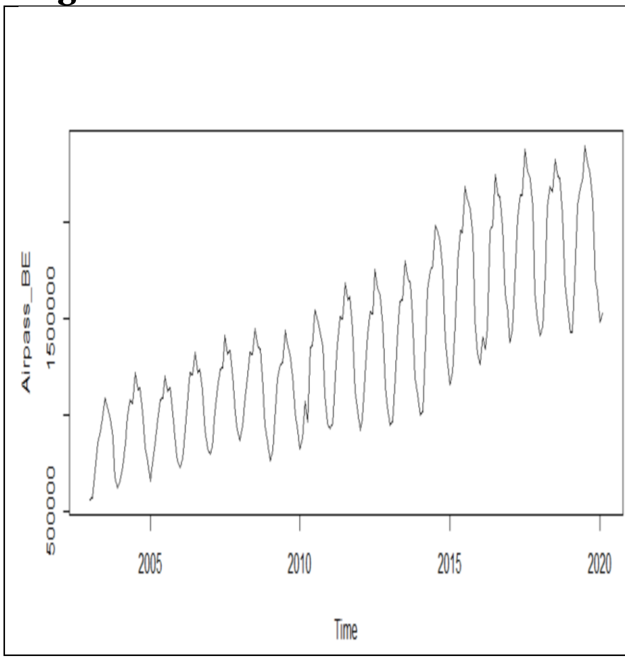
Thus, based on all the plots we could conclude that the data has an increasing trend and a seasonal trend associated with it.

Question 2

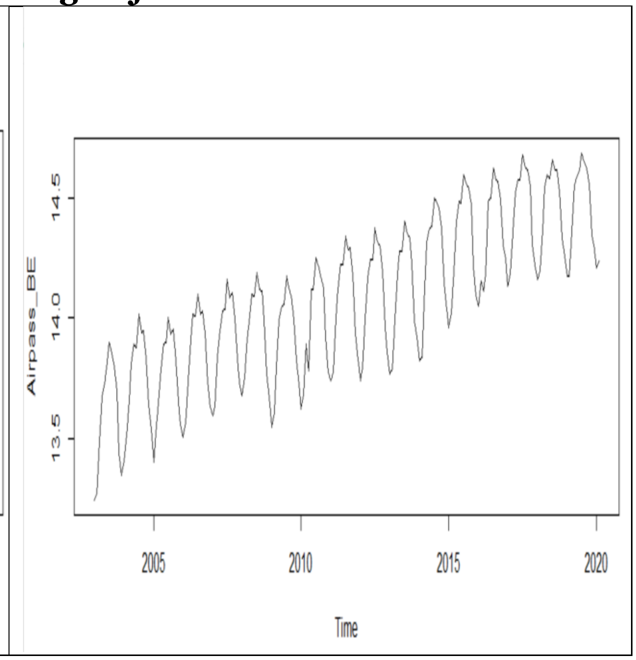
Discuss whether a transformation and/or any other adjustment of the time series would be useful. If so, apply the most appropriate transformation and/or adjustments. Also, report the optimal Box-Cox lambda value that could be used to transform the time series. Clarify how you will proceed with the transformation in the remainder of the exercise.

Ans:
A logarithmic transformation would be useful in this case. We see an increasing variation and thus it would be nice to perform logarithmic transformation to make the size of the seasonal variation about the same across the whole series.
The plot for logarithmic transformation of data.

Original Plot



Log Adjusted Plot



The optimal Box-Cox lambda value is:

```
 BoxCox.lambda(final_air_ts)
1] 0.01461759
```

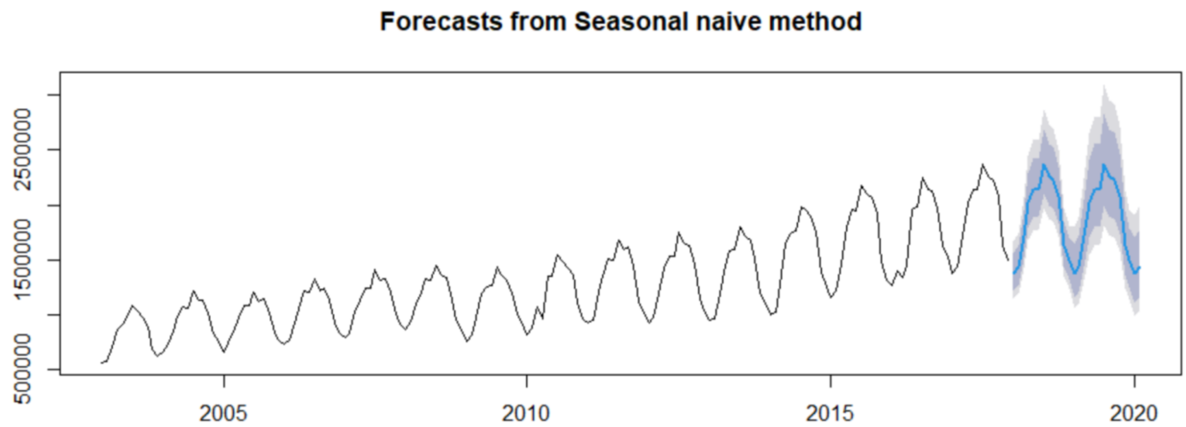
In rest of the procedures, for all models we would be performing log transformation to train and test data.

Question 3

Create forecasts using the seasonal naive method. Check the residual diagnostics (including the Ljung-Box test) and the forecast accuracy (on the test set).

Ans:

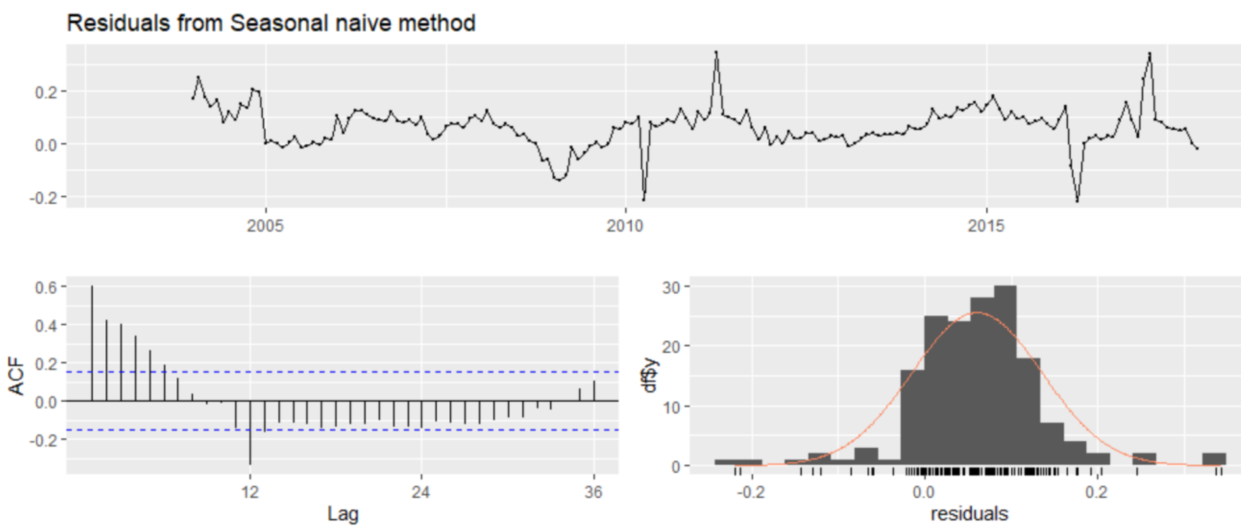
Forecasts from Seasonal Naïve Method



The Forecast Accuracy

```
> accuracy(f1,test)[,c(2,3,5,6)]
                  RMSE      MAE     MAPE      MASE
Training set 124232.13 95011.47 7.225186 1.0000000
Test set      55167.45 45156.54 2.518184 0.4752746
```

Residual Diagnostic




```
        Ljung-Box test

data:  Residuals from Seasonal naive method
Q* = 225.73, df = 24, p-value < 2.2e-16

Model df: 0.   Total lags used: 24
```

Based on the Ljung-Box Test we see that the p value is less than 0.05 thus helping us reject the null hypothesis of white noise and there are still some trends in the residuals that is not captured by the model based on the ACF and distribution plots.

Question 4

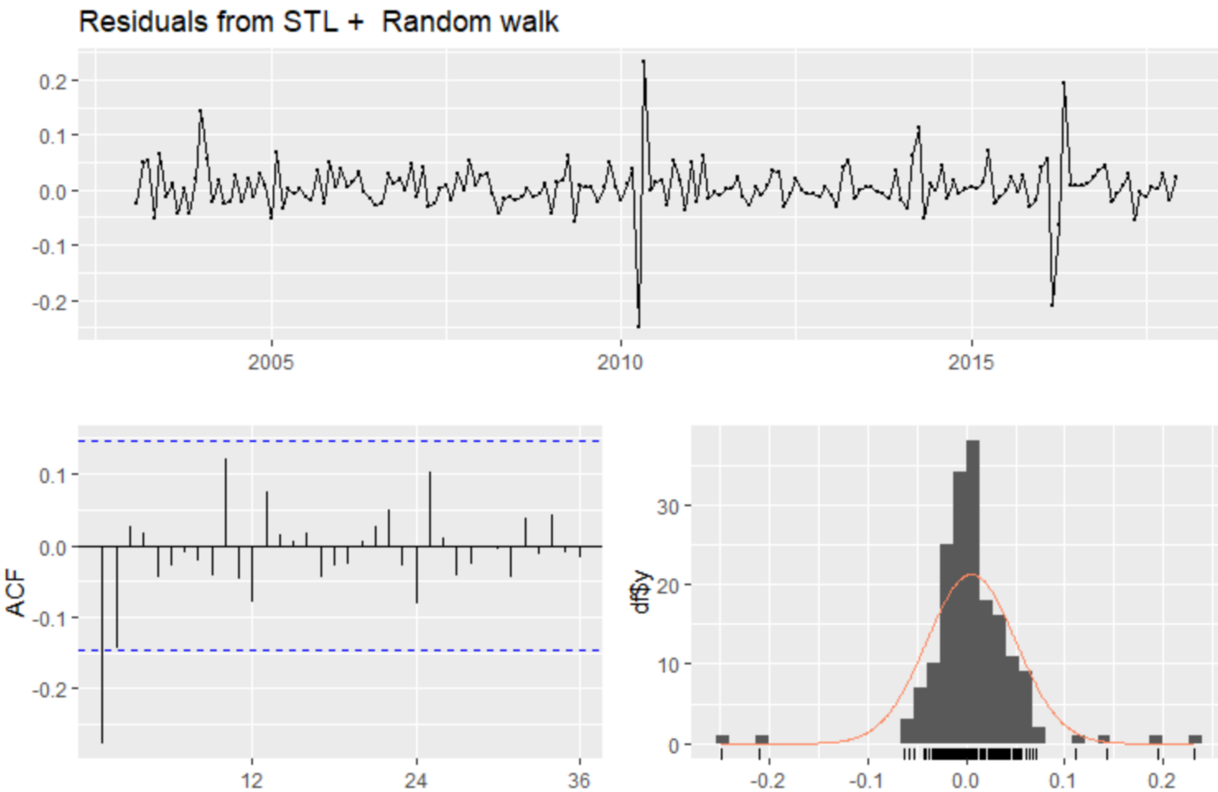
Use an STL decomposition to forecast the time series. Use the various underlying forecasting methods for the seasonally adjusted data (naive, rwdrift, ets, arima). Check the residual diagnostics and the forecast accuracy and select the best performing STL decomposition.

Ans:
Log transformation was applied to the train and test dataset.
For STL models the parameters set were s.window = 'periodic ' and t.window = 12.

STL Naïve
Accuracy

	RMSE	MAE	MAPE	MASE
Training set	0.04529692	0.02761211	0.1974604	0.3649924
Test set	0.03734065	0.02964263	0.2053523	0.3918330

Residual Diagnostic




```
        Ljung-Box test

data:  Residuals from STL +  Random walk
Q* = 27.42, df = 24, p-value = 0.2852

Model df: 0.   Total lags used: 24
```

Based on the Ljung-Box Test we see that the p value is greater than 0.05 thus helping us accept the null hypothesis of white noise

STL RWDRIFT
Accuracy

	RMSE	MAE	MAPE	MASE
Training set	0.04501769	0.02755904	0.1970480	0.364291
Test set	0.07403706	0.06468040	0.4461167	0.854982

Residual Diagnostic

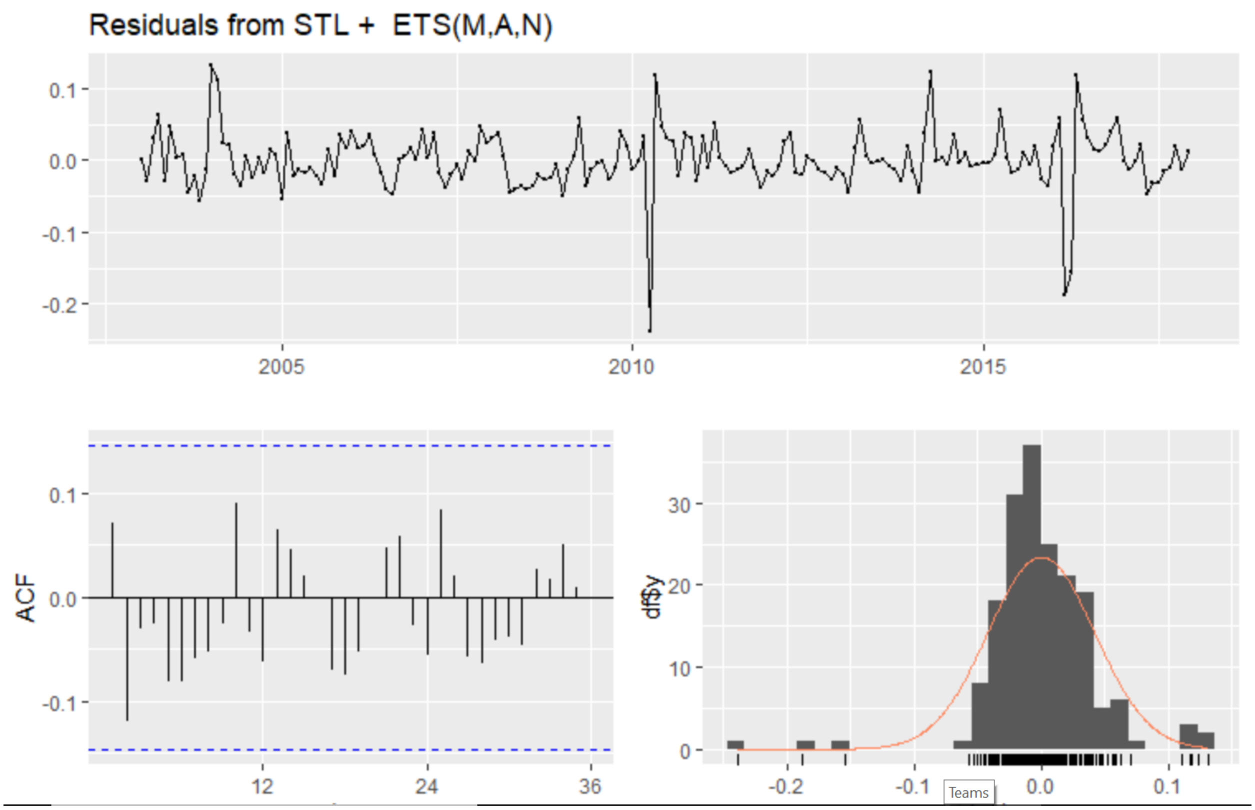

```
        Ljung-Box test

data:  Residuals from STL +  Random walk with drift
Q* = 27.42, df = 23, p-value = 0.2385

Model df: 1.   Total lags used: 24
```

Based on the Ljung-Box Test we see that the p value is greater than 0.05 thus helping us accept the null hypothesis of white noise

STL ETS
Accuracy

	RMSE	MAE	MAPE	MASE
Training set	0.04195469	0.02776864	0.1985727	0.3670616
Test set	0.07001430	0.06046777	0.4169925	0.7992971

Residual Diagnostic




```
Ljung-Box test

data:  Residuals from STL +  ETS(M,A,N)
Q* = 16.221, df = 20, p-value = 0.7028

Model df: 4.   Total lags used: 24
```

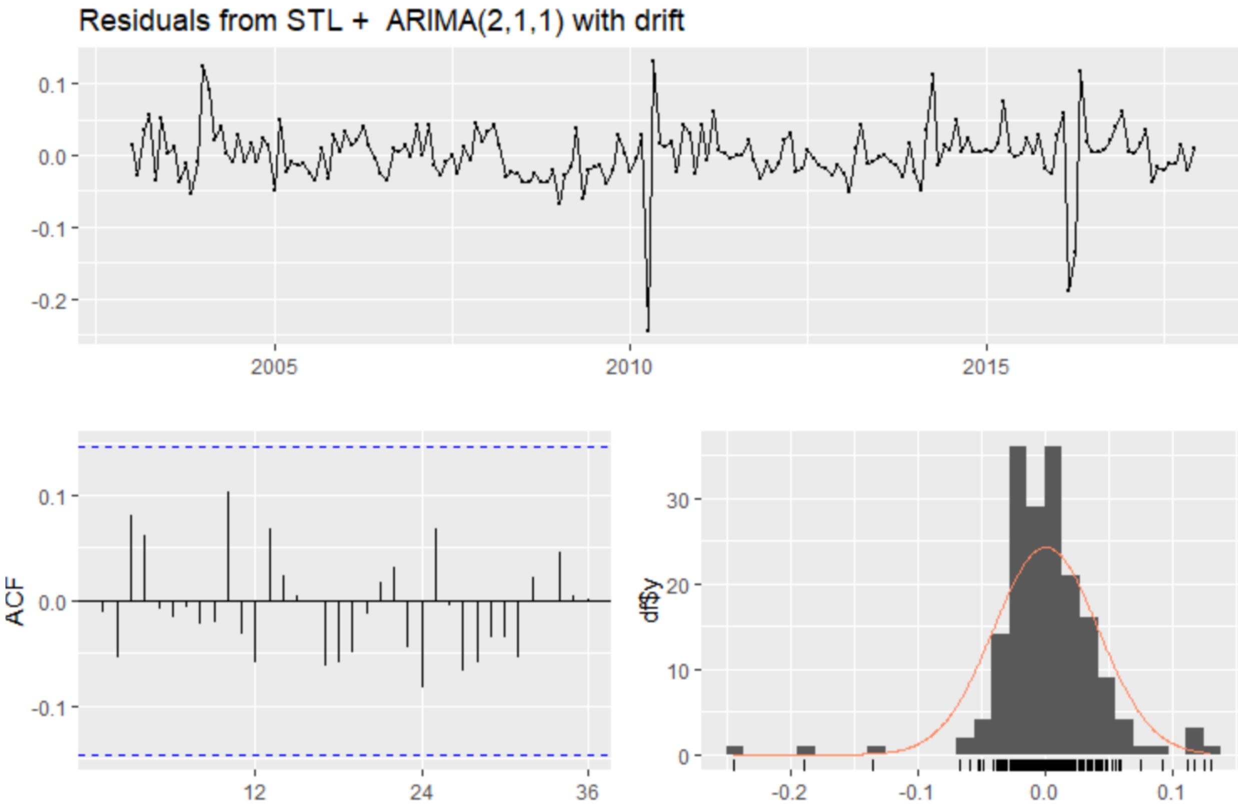
Based on the Ljung-Box Test we see that the p value is greater than 0.05 thus helping us accept the null hypothesis of white noise

STL ARIMA

Accuracy

	RMSE	MAE	MAPE	MASE
Training set	0.04112637	0.02707404	0.1937901	0.3578800
Test set	0.07321852	0.06398411	0.4412890	0.8457781

Residual Diagnostic

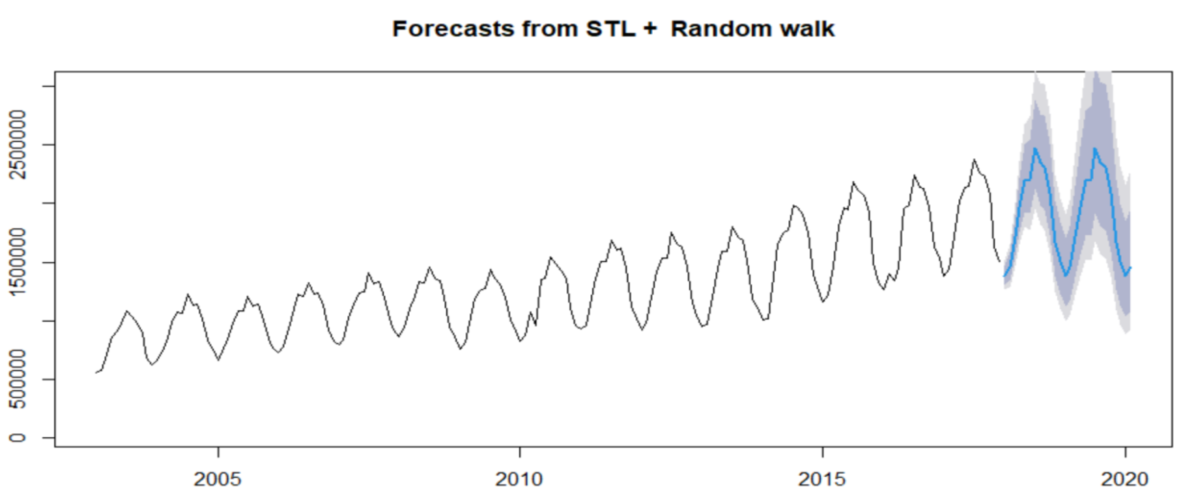



```
    Ljung-Box test

data:  Residuals from STL +  ARIMA(2,1,1) with drift
Q* = 10.866, df = 20, p-value = 0.9496

Model df: 4.   Total lags used: 24
```

Based on the Ljung-Box Test see that the p value is greater than 0.05 thus helping us accept the null hypothesis of white noise

Based on all 4 models we see that **STL NAÏVE (Random Walk)** has the best performance in terms of accuracy (**0.39 MASE**) and then on evaluating residuals and performing Ljung Box we can say that the null hypothesis of white noise is accepted and based on residuals the model is able to capture the trends in the data.

Forecast With STL NAÏVE (Random Walk)



Question 5

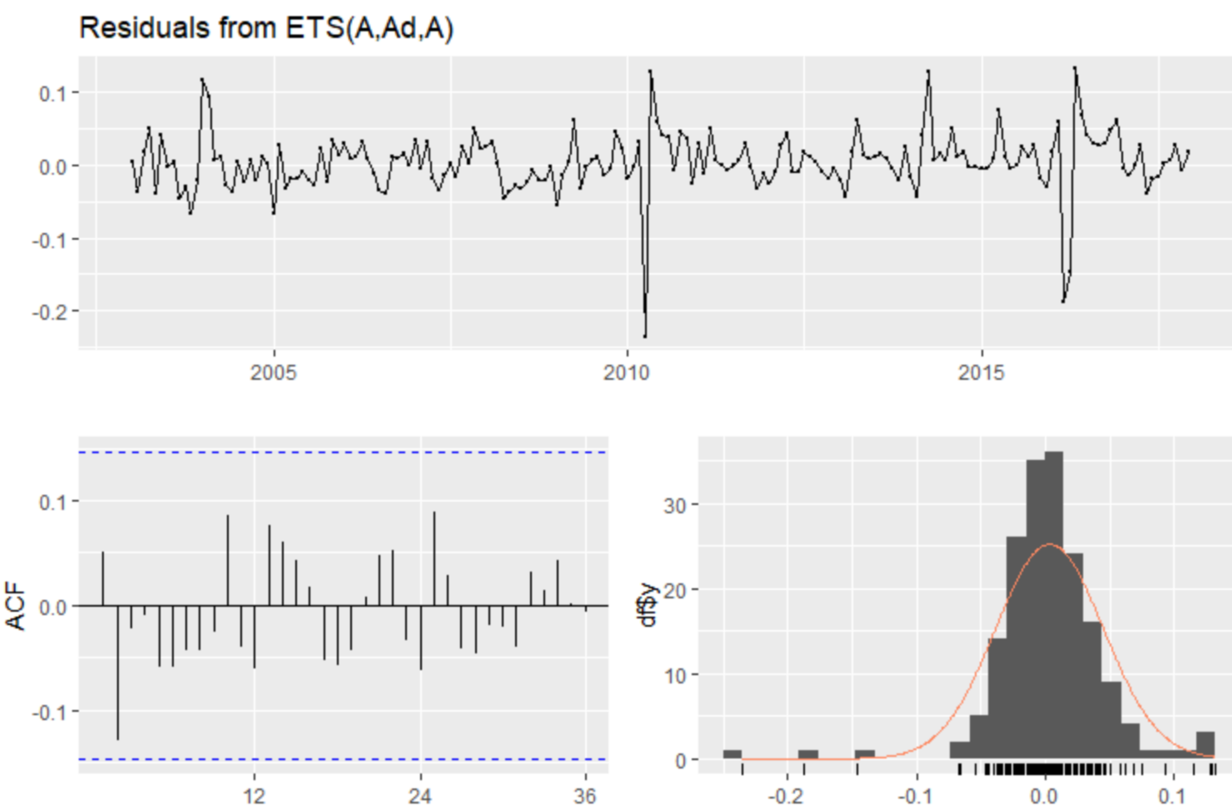
Generate forecasts using ETS. First select the appropriate models yourself and discuss their performance. Compare these models with the results of the automated ETS procedure. Check the residual diagnostics and the forecast accuracy for the various ETS models you've considered. Present the parameters of the final ETS model and show the forecasts in a graph.

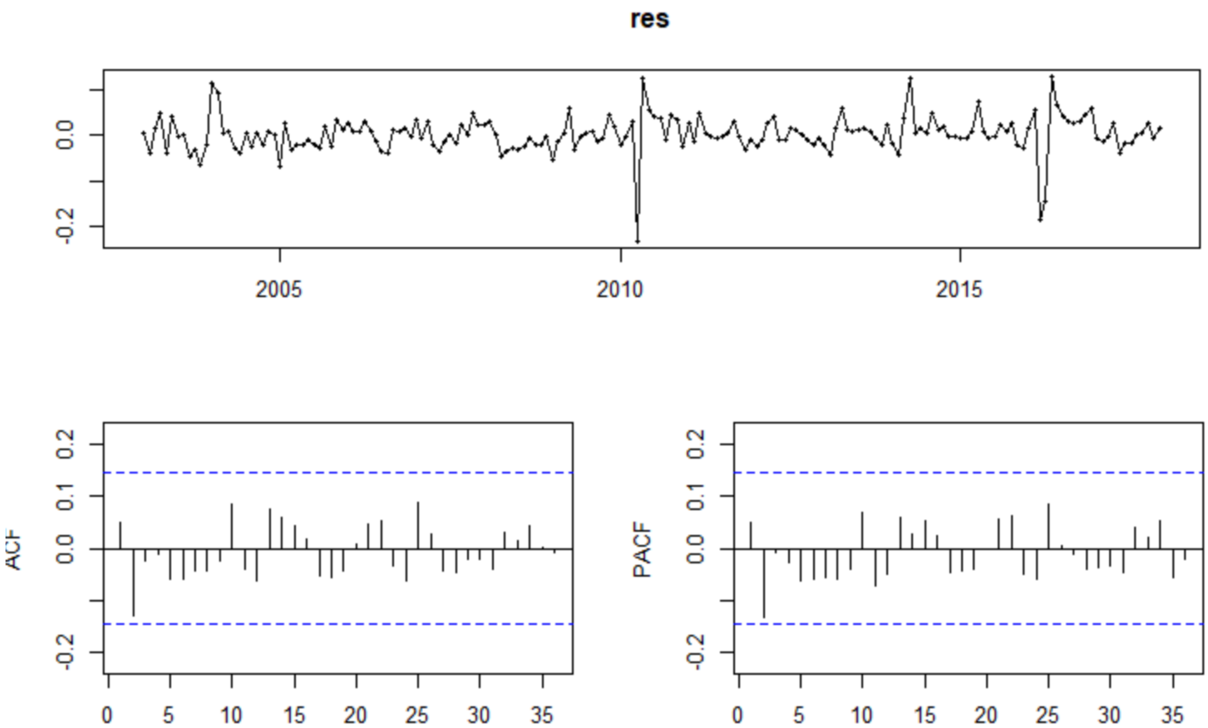
Ans:

A Log transformation has been performed for the train and test set.
The selected ETS models and their Test Accuracy with P value are:

Model	RMSE	MAE	MAPE	MASE	Ljung-Box test (P value)
AAA	**0.03369911**	**0.02735409**	**0.1893455**	**0.3615819**	**0.06587**
ANA	0.04711739	0.03886378	0.2700495	0.5137234	0.04245
MAM	0.03717391	0.03136867	0.2168918	0.4146488	0.06667
MNM	0.05096862	0.03873082	0.2695781	0.5119659	2.402e-08
MAAd	0.03479951	0.02780305	0.1926222	0.3675164	0.06825
MAA	0.03479951	0.02780305	0.1926222	0.3675164	0.06825
MNA	0.03755805	0.03050815	0.2113932	0.4032740	0.1171

Based on the test accuracy, Ljung Box test and residual diagnostic we see that the AAA model has the best performance in terms of accuracy and p value is greater than 0.05 thus null hypothesis of white noise is accepted. We will be considering AAA model for comparison with automated ETS model.
The automated ETS also provides an AAA model with best results. We will be further evaluating the AAA model for Ljung Box test and residual diagnostic






```
        Ljung-Box test

data:  Residuals from ETS(A,Ad,A)
Q* = 16.049, df = 9, p-value = 0.06587

Model df: 17.   Total lags used: 26
```

Based on the Ljung-Box Test see that the p value is greater than 0.05 thus helping us accept the null hypothesis of white noise

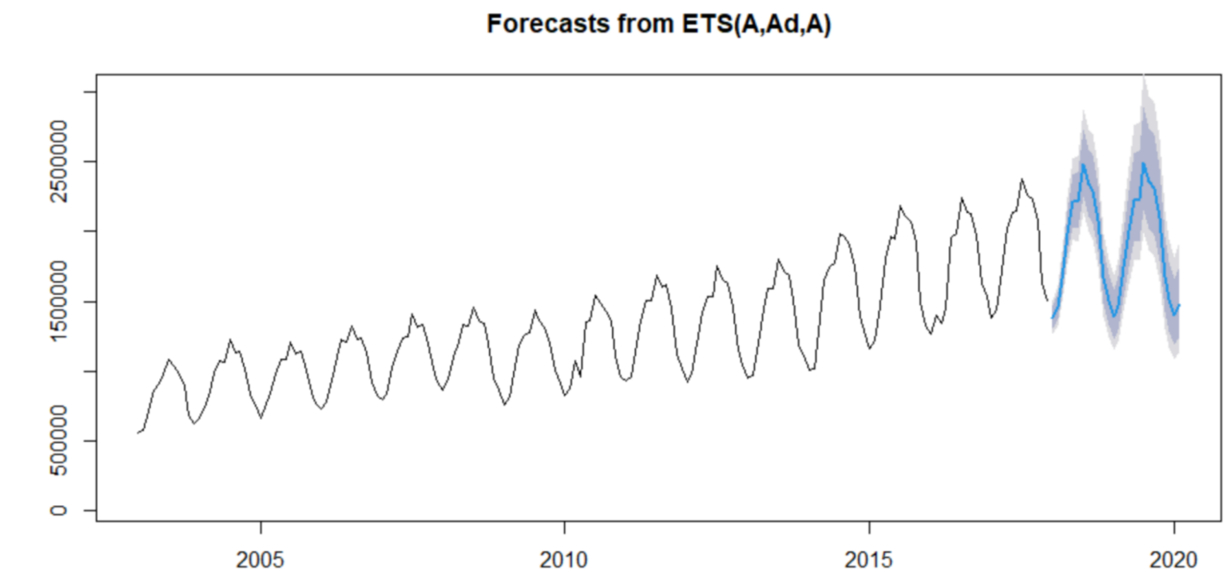
The final parameters for the model are:
Smoothing parameters:

- alpha = 0.5585
- beta = 0.0013
- gamma = 1e-04
- phi = 0.98

Initial states:

- l = 13.5418
- b = 0.0124
- s = -0.2434 -0.14 0.0836 0.1811 0.2009 0.2603 0.1476 0.1473 0.0414 -0.0954 -0.2649 -0.3185
- sigma= 0.0441

Forecast With ETS(AAA) model

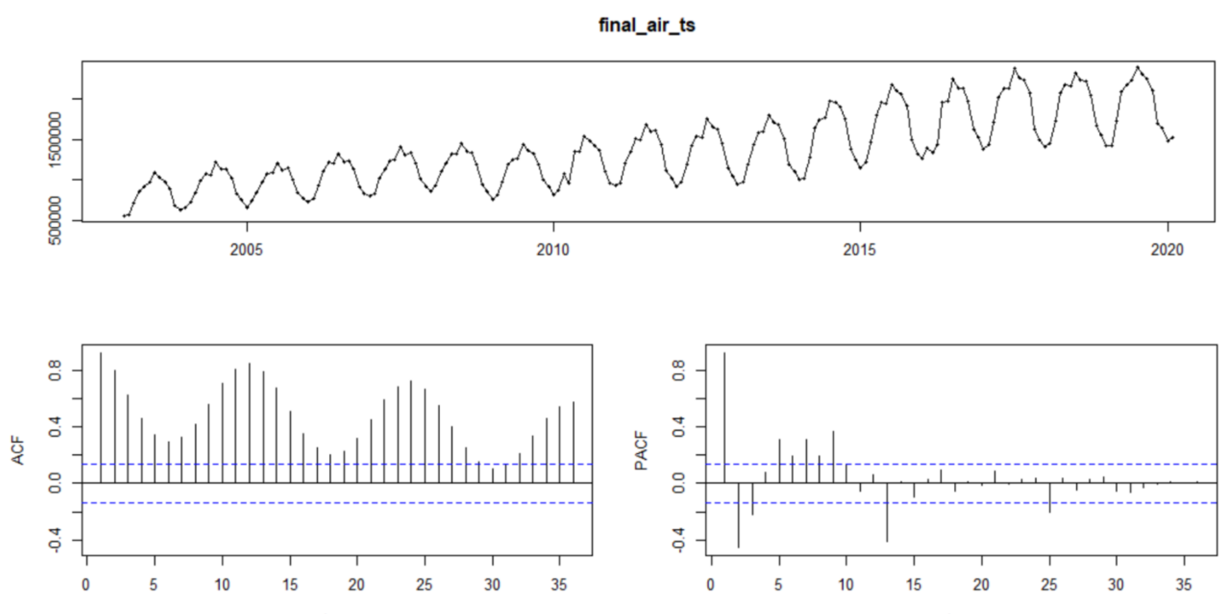


Question 6

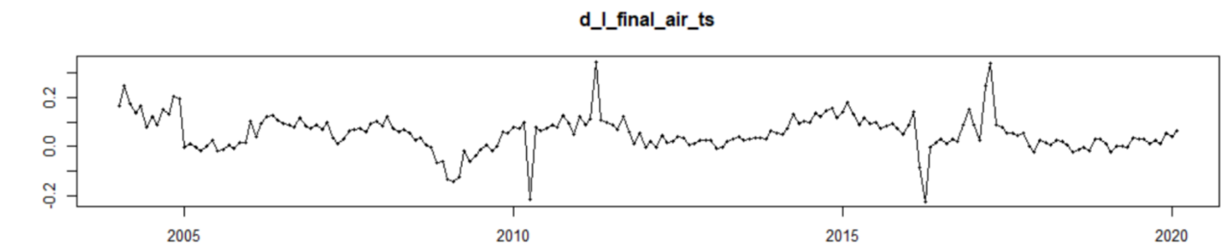
Generate forecasts using the auto.arima procedure. Present the estimated model using the backward shift operator. Include the parameter estimates. Check the residual diagnostics and the forecast accuracy. Discuss your results, and if necessary compare these with other possible ARIMA models (e.g. if small changes in the model specification improve the properties of the residuals and/or the forecast accuracy).

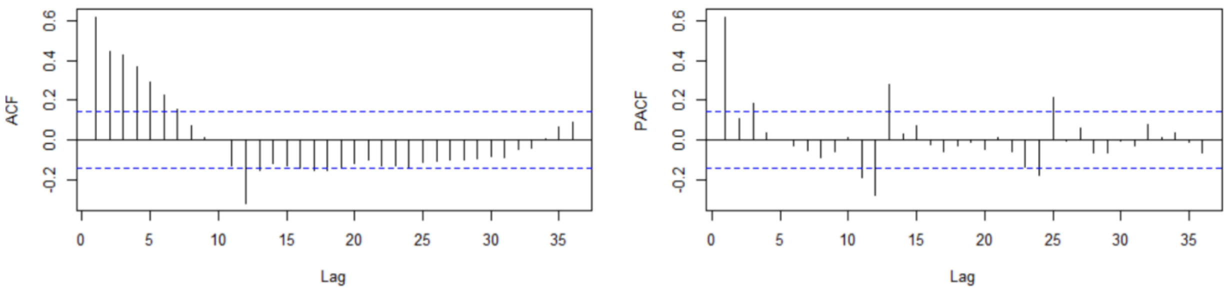
Ans:

Before Seasonality Difference



Seasonality differenced plot





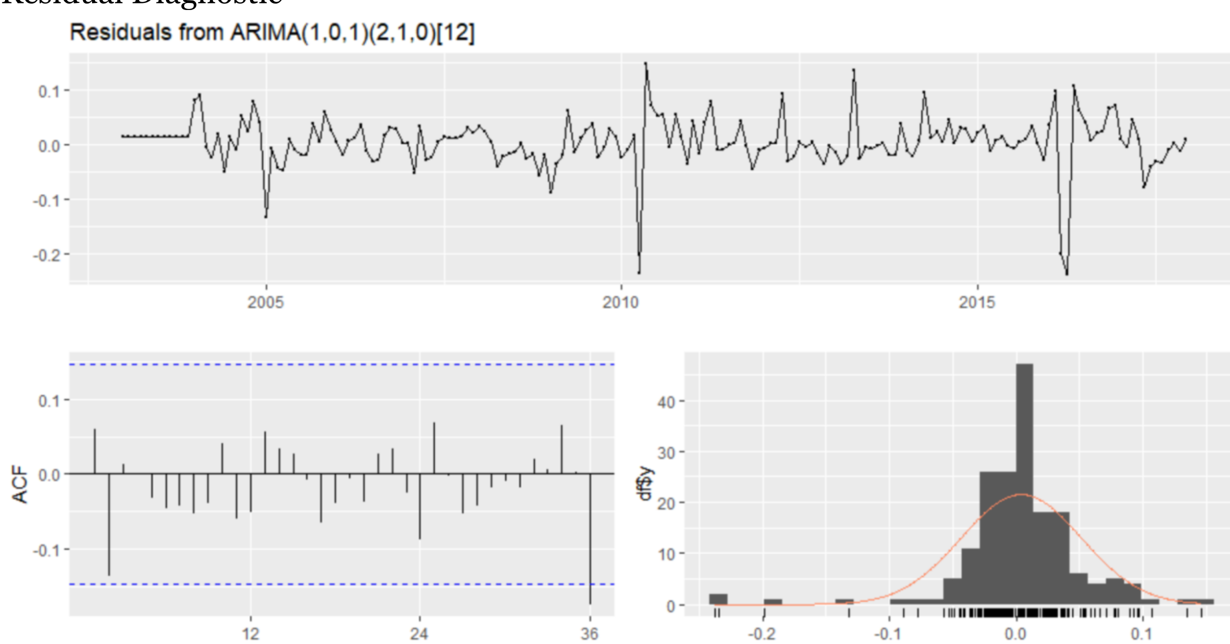
ARIMA(1,0,1)(2,1,0)[12] is the best model being selected by auto.arima
Coefficients :

	ar1	ma1	sar1	sar2
	0.9847	-0.4899	-0.6961	-0.3835
s.e.	0.0141	0.0820	0.0776	0.0835

Accuracy

	RMSE	MAE	MAPE	MASE
Training set	0.04738070	0.03030293	0.2163503	0.4005613
Test set	0.04676839	0.03907579	0.2697082	0.5165258

Residual Diagnostic



```
        Ljung-Box test

data:  Residuals from ARIMA(1,0,1)(2,1,0)[12]
Q* = 11.971, df = 20, p-value = 0.9171

Model df: 4.   Total lags used: 24
```

Based on the Ljung-Box Test see that the p value is greater than 0.05 thus helping us accept the null hypothesis of white noise

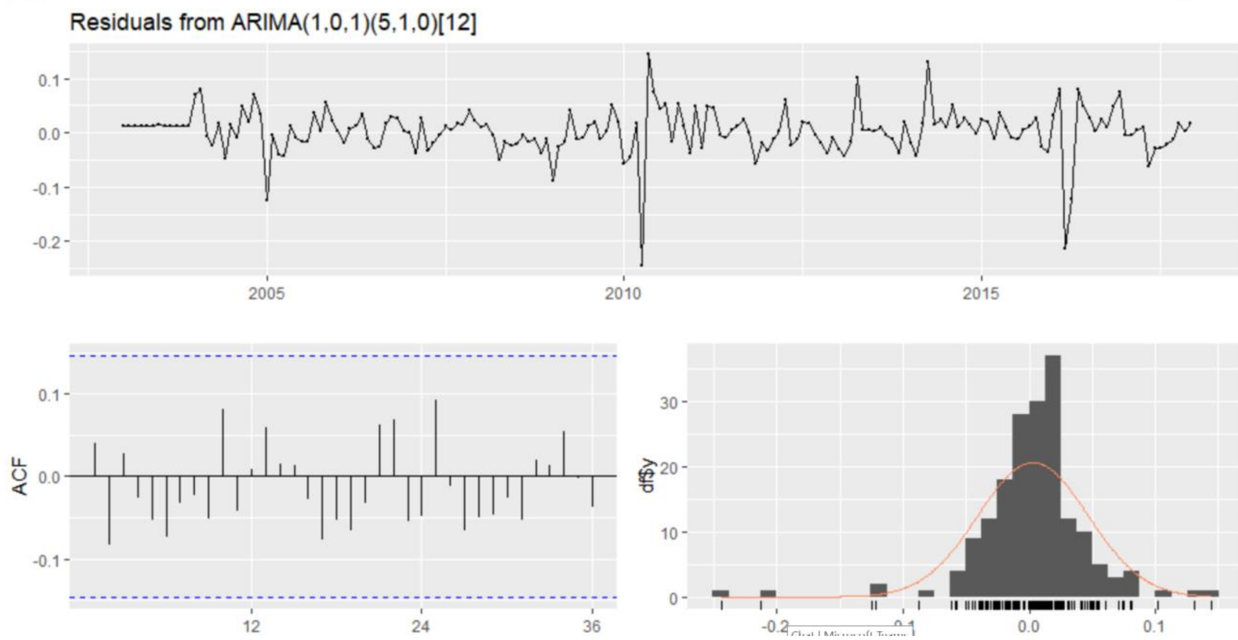
ARIMA(1,0,1)(5,1,0)[12]

Accuracy

```
> accuracy(forecast(fcast, h = h), log(test))[,c(2,3,5,6)]
```

	RMSE	MAE	MAPE	MASE
Training set	0.04363351	0.02881506	0.2059586	0.3808938
Test set	0.04383422	0.03626566	0.2495016	0.4793800

Residual Diagnostic




```
        Ljung-Box test

data:  Residuals from ARIMA(1,0,1)(5,1,0)[12]
Q* = 12.532, df = 17, p-value = 0.7669

Model df: 7.   Total lags used: 24
```

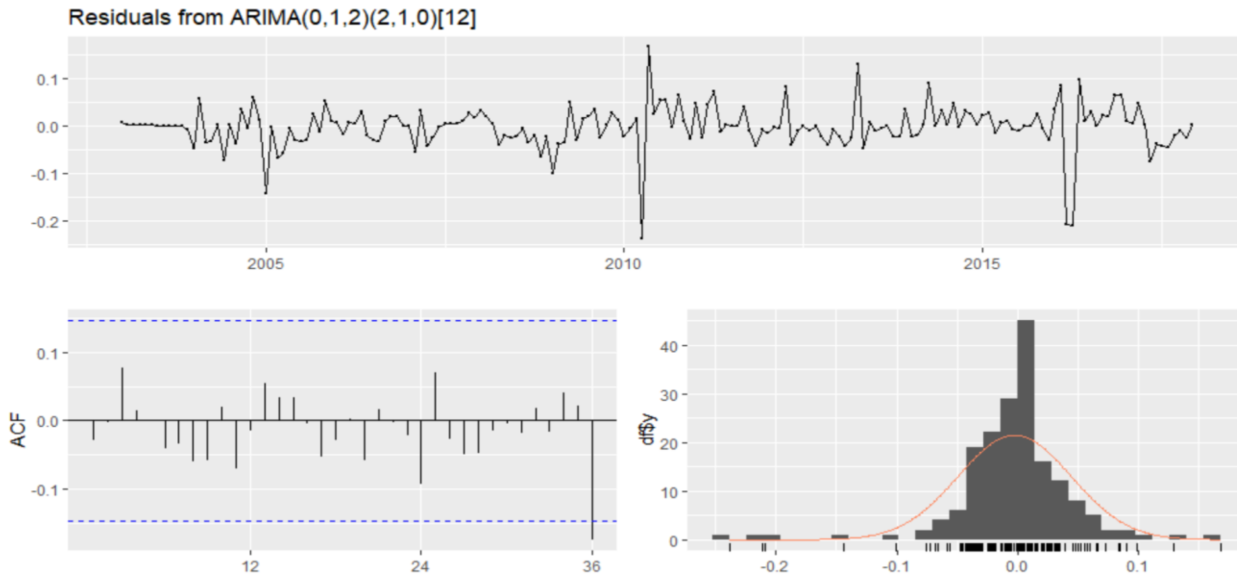
Based on the Ljung-Box Test see that the p value is greater than 0.05 thus helping us accept the null hypothesis of white noise.

With Backward shift operator ARIMA (0,1,2)(2,1,0)[12] was selected.

Accuracy

	RMSE	MAE	MAPE	MASE
Training set	0.04677045	0.02924478	0.2086746	0.3865741
Test set	0.06655270	0.06028846	0.4165622	0.7969269

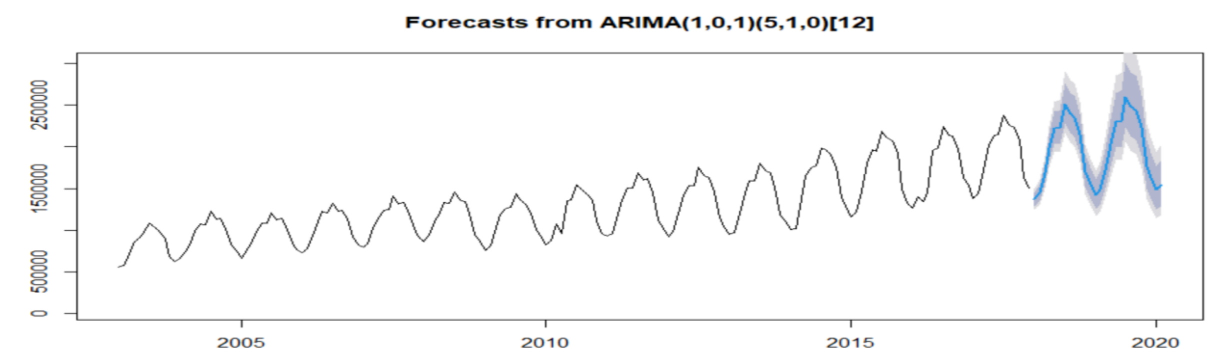
Residual Diagnostic




```
        Ljung-Box test

data:  Residuals from ARIMA(0,1,2)(2,1,0)[12]
Q* = 8.7155, df = 20, p-value = 0.986

Model df: 4.   Total lags used: 24
```

Based on the Ljung-Box Test see that the p value is greater than 0.05 thus helping us accept the null hypothesis of white noise.

We do see that on comparison between the two ARIMA models ARIMA(1,0,1)(5,1,0)[12] seems to be performing better in terms of accuracy. We would be generating forecasts for the same.



Question 7

Compare the different models (naive, STL, ETS, ARIMA) in terms of residual diagnostics and forecast accuracy. Present the results in a summary table. Analyze your results and select your final model.

Ans:

Model	RMSE	MAE	MAPE	MASE	Ljung-Box test(P value)
ETS(AAA)	**0.03369911**	**0.02735409**	**0.1893455**	**0.3615819**	**0.06587**
ARIMA(1,0,1)(5,1,0)[12]	0.04383422	0.03626566	0.2495016	0.4793800	0.7669

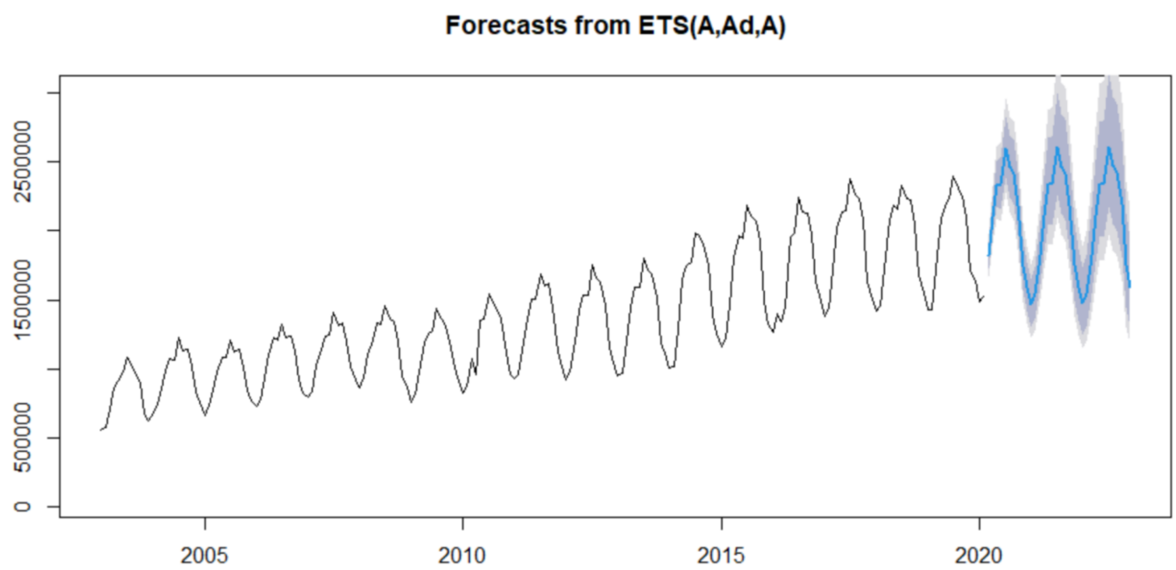
STL NAÏVE (Random Walk)	0.0373406 5	0.02964263	0.205352 3	0.391833 0	0.2852

Based on the results we can say that **ETS(AAA)** is the best performing model in terms of accuracy. The p value in Ljung Box Test is greater than 0.05 indicating that the null hypothesis of white noise is true in terms of three models.

Question 8

Generate out of sample forecasts up to December 2022, based on the complete time series (January 2003 - February 2020). Present your results.

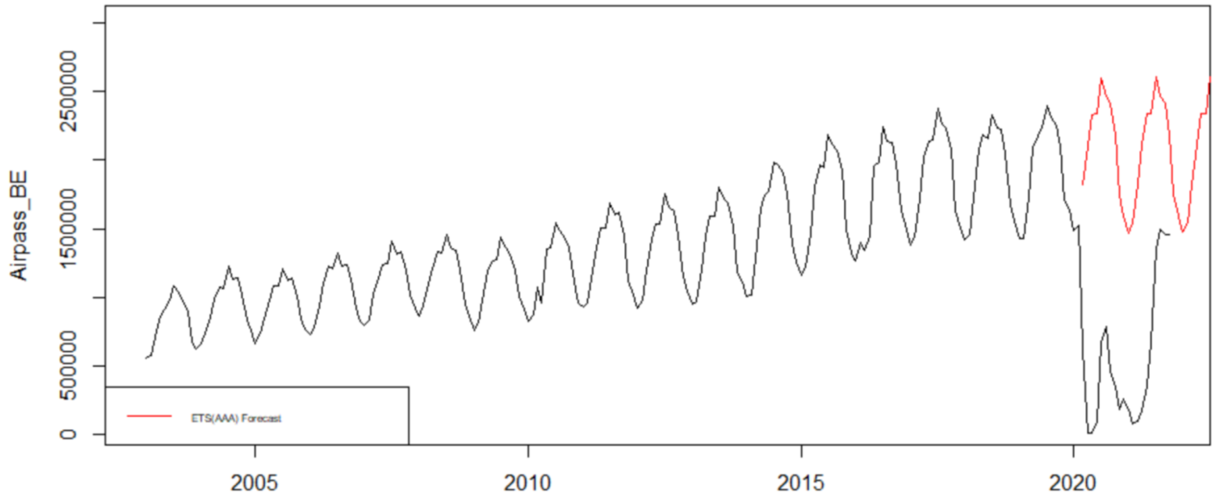
Ans:



Question 9

Now consider the last observations in the time series (March 2020 - October 2021). They correspond to the COVID pandemic times. What do you learn about the impact of the pandemic on air passenger transport between Belgium and other EU countries, based on the data and your final forecasts?

Ans:



Based on the forecast and the existing data we see that Pandemic had a major impact on the Air passengers. There was a sharp decline and the forecast for the same period do not align with the actual trends. It could be quite difficult to capture this trend in a model. The timeseries also shows that towards the end there is an increase in the air passengers.

Part 2

Dataset

The data set Real Personal Consumption Expenditures contains consumption expenditure in Billions of Dollars per quarter for US since Q1 2002 to Q4 2021.

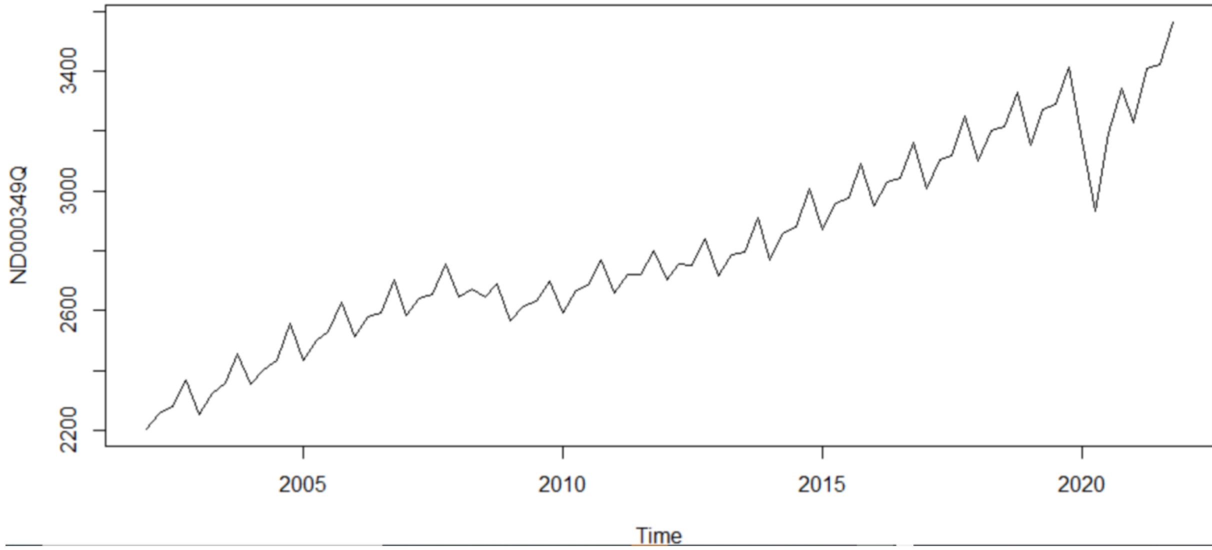
Data Preprocessing

The time series is split in a training set from Q1 2002 up to Q4 2017 and a test set from Q1 2018 up to Q4 2019. A final set would be the validation set to check the impact of pandemic on our validation set comprising of Q1 2020 to Q4 2021.

Explanatory Data Analysis

The Time Series Plot shows that there is an increasing trend for most of the years until 2020 in the consumption expenditure. There is also a strong seasonal trend that we see. In the plot we do see dip in consumption expenditure post 2020 following an increase since Q1 of 2020.

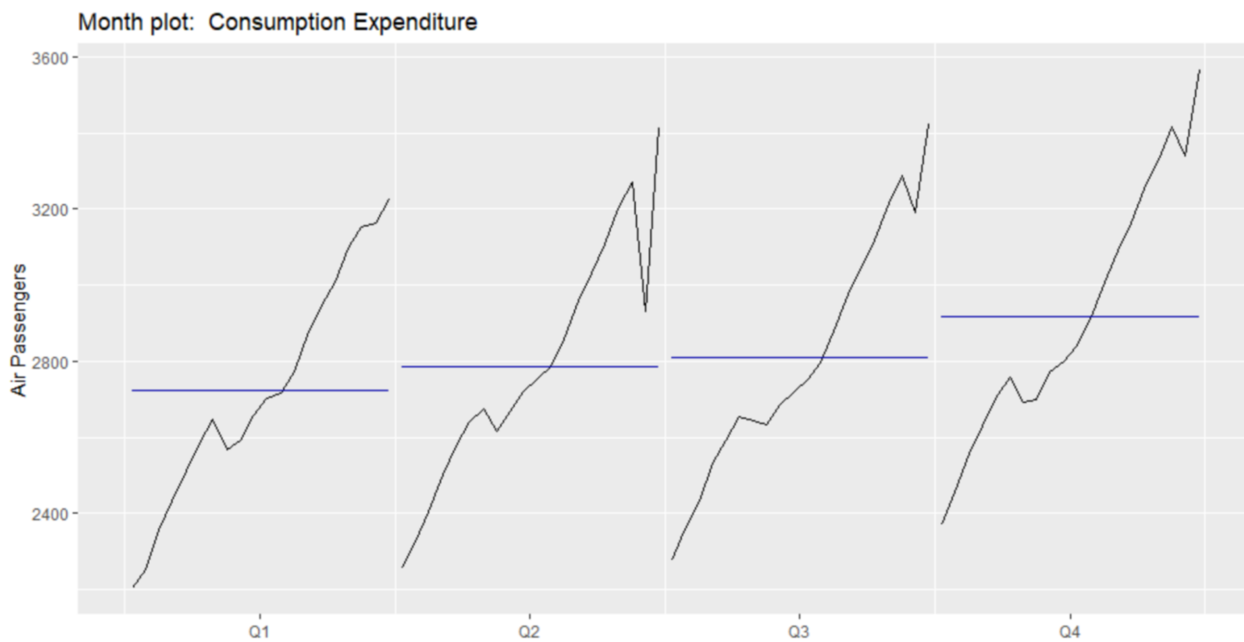
Time series plot



Seasonal and Seasonal subseries plots

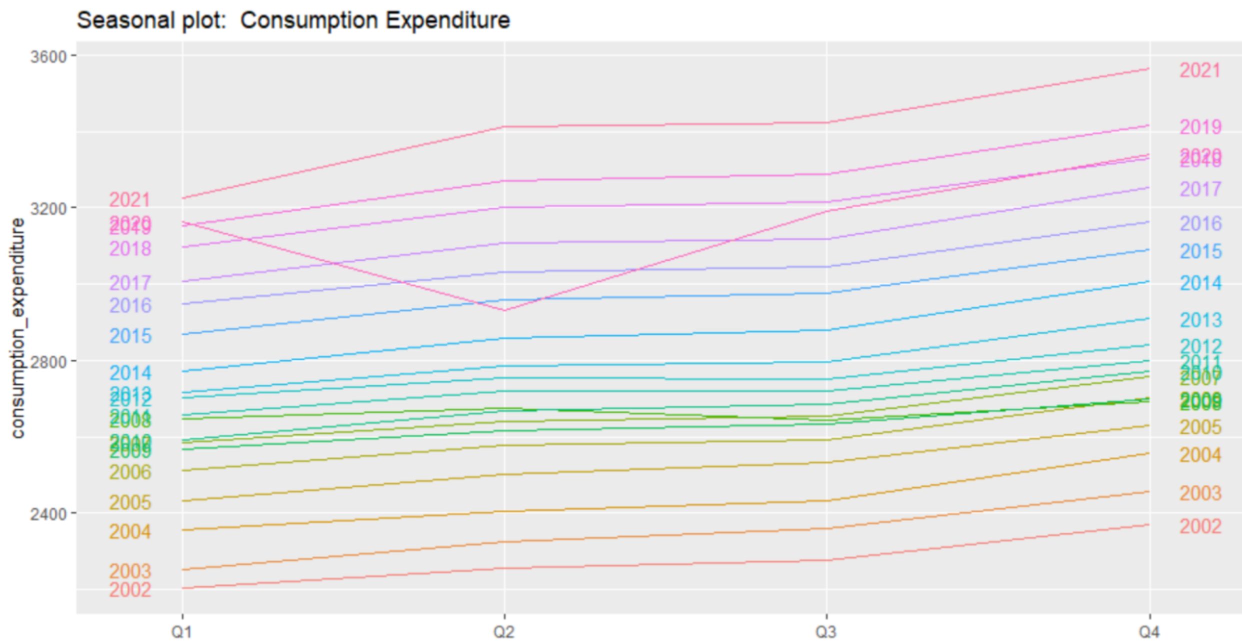
The Seasonal subseries plot shows that there is an increasing trend per Quarter for consumption expenditure. We see that Q4 has the highest consumption expenditure in terms of mean.

Seasonal subseries plot



The seasonal plot shows an increasing trend in consumption expenditure apart from the year of 2020 and 2021. In 2020 we see a sharp drop in consumption expenditure from Q1 to Q2 which could be due to the pandemic period but there seems to be gradual increase in following months.

Seasonal plot



We would be looking at the P(A)CF plots to check for correlations and partial correlations.

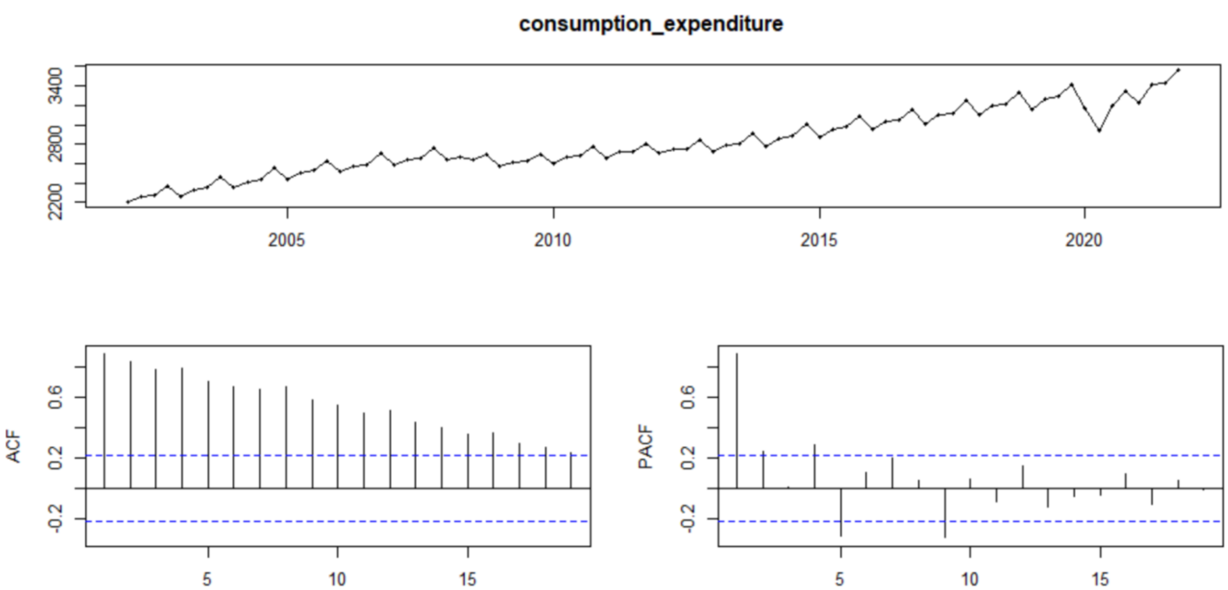
P(A)CF plots

ACF Plot

An ACF measures and plots the average correlation between data points in a time series and previous values of the series measured for different lag lengths.

PACF Plot

A PACF is like an ACF except that each partial correlation controls for any correlation between observations of a shorter lag length.



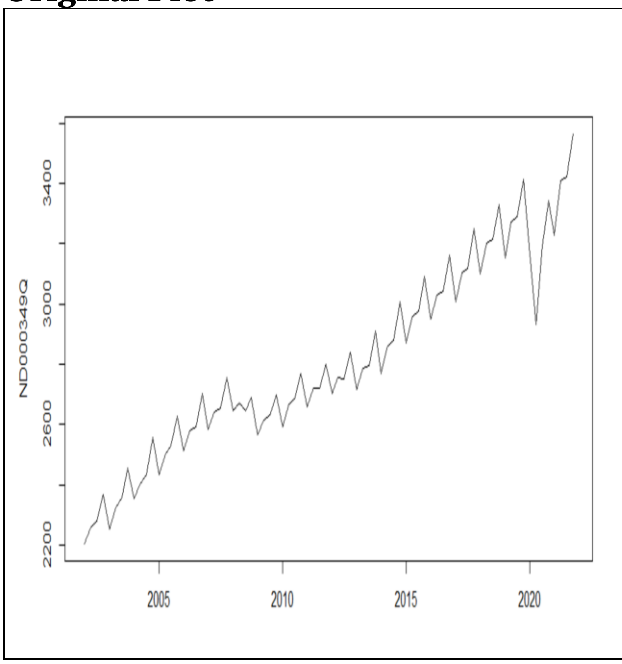
The slow decrease in the ACF as the lags increase is due to the trend, while the scalloped shape is due the seasonality. The PACF plot has the highest Partial correlation for the first second and forth last lag.

Thus, based on all the plots we could conclude that the data has an increasing trend and a seasonal trend associated with it.

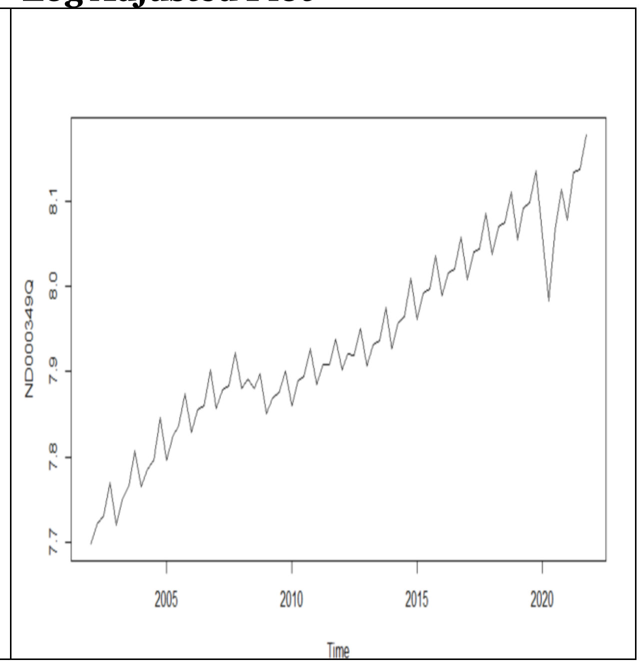
Data Transformation

A logarithmic transformation would be useful in this case. We see an increasing variation and thus it would be nice to perform logarithmic transformation to make the size of the seasonal variation about the same across the whole series.
The plot for logarithmic transformation of data.

Original Plot



Log Adjusted Plot



The optimal Box-Cox lambda value is:

```
> BoxCox.lambda(final_consumption_expenditure)
[1] 0.09499272
```

In rest of the procedures, for all models we would be performing log transformation to train and test data.

Model Evaluation

STL Decomposition

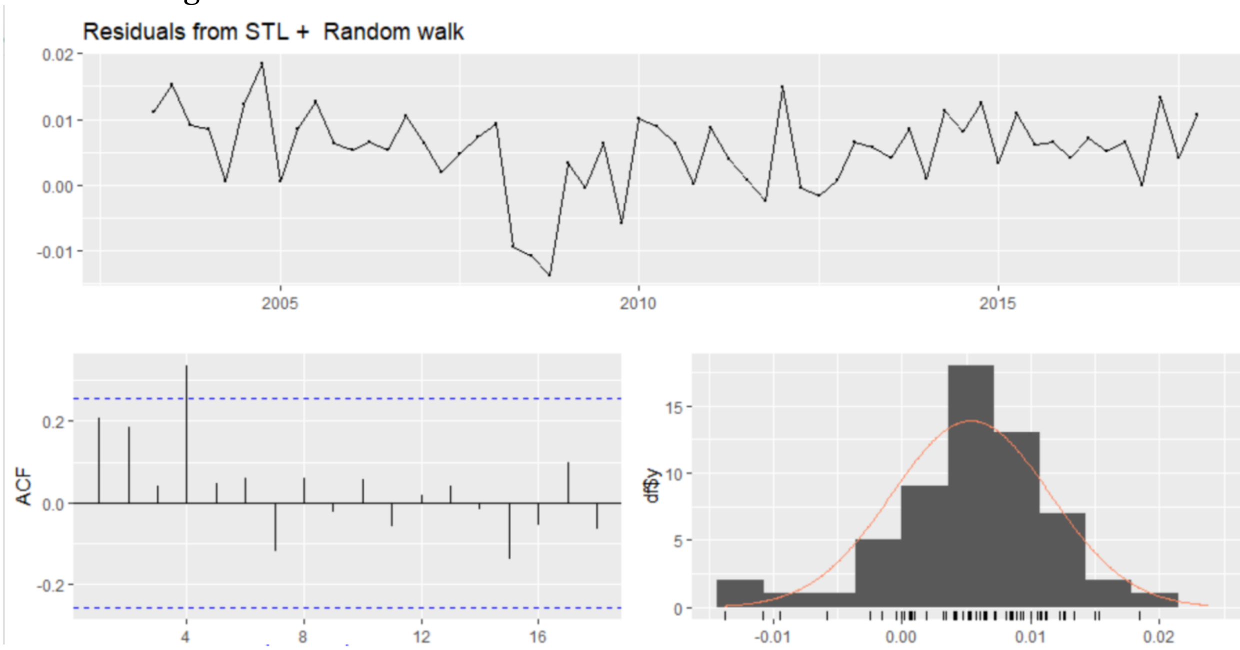
The data shows a seasonal and an increasing trend. We would therefore first be using STL decomposition with naïve, RWDRIFT, ets and arima model and evaluate their performance.

STL Naïve

Accuracy

	RMSE	MAE	MAPE	MASE
Training set	0.008122138	0.006874482	0.08698412	0.2944801
Test set	0.029751465	0.025980476	0.32079393	1.1129176

Residual Diagnostic




```
        Ljung-Box test

data:  Residuals from STL +  Random walk
Q* = 13.943, df = 8, p-value = 0.08326

Model df: 0.   Total lags used: 8
```

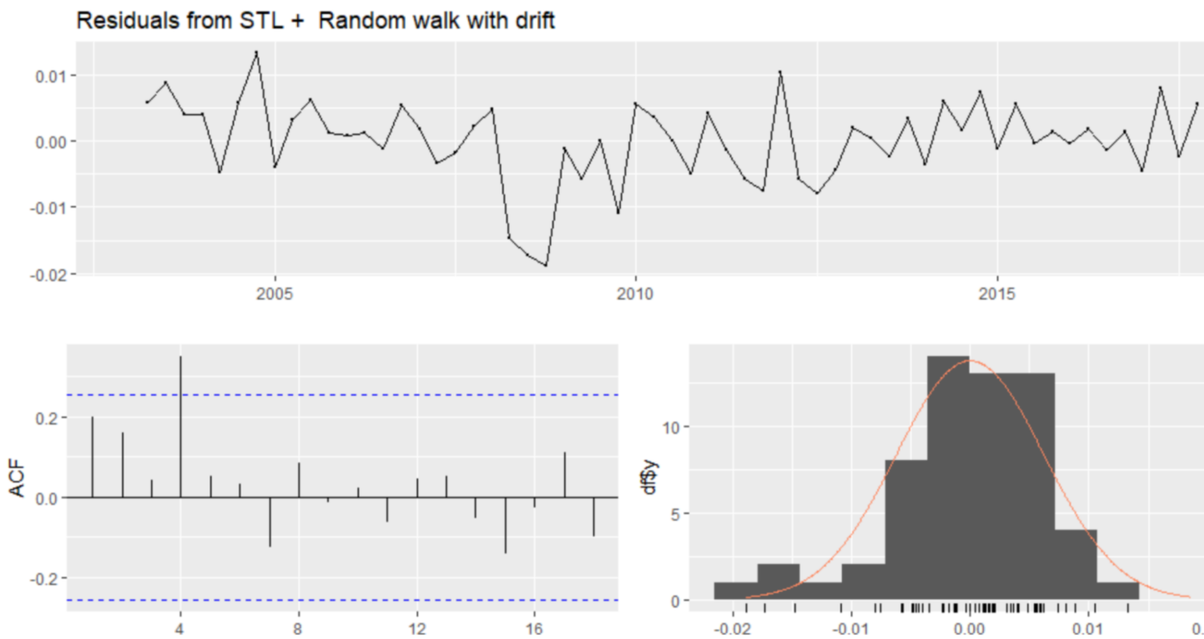
Based on the Ljung-Box Test we see that the p value is greater than 0.05 thus helping us accept the null hypothesis of white noise. We do see an ACF with significant value at lag 4.

STL RWDRIFT

Accuracy

	RMSE	MAE	MAPE	MASE
Training set	0.006161171	0.004650652	0.05887844	0.1992186
Test set	0.004892134	0.004599182	0.05687578	0.1970137

Residual Diagnostic



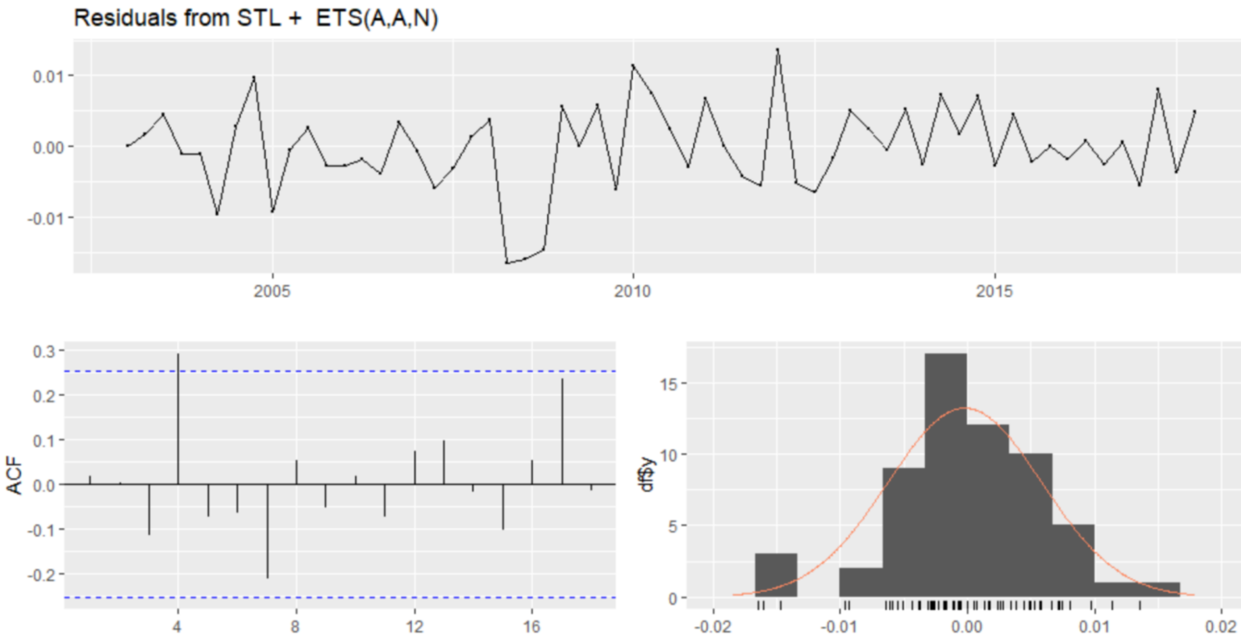

```
     Ljung-Box test

data:  Residuals from STL +  Random walk with drift
Q* = 14.007, df = 7, p-value = 0.05106

Model df: 1.   Total lags used: 8
```

Based on the Ljung-Box Test we see that the p value is greater than 0.05 thus helping us accept the null hypothesis of white noise. We do see an ACF with significant value at lag 4.

STL ETS
Accuracy

	RMSE	MAE	MAPE	MASE
Training set	0.006005729	0.004567187	0.05779737	0.1956432
Test set	0.006768444	0.005525037	0.06835582	0.2366743

Residual Diagnostic




```
        Ljung-Box test

data:  Residuals from STL +  ETS(A,A,N)
Q* = 10.558, df = 4, p-value = 0.03201

Model df: 4.   Total lags used: 8
```

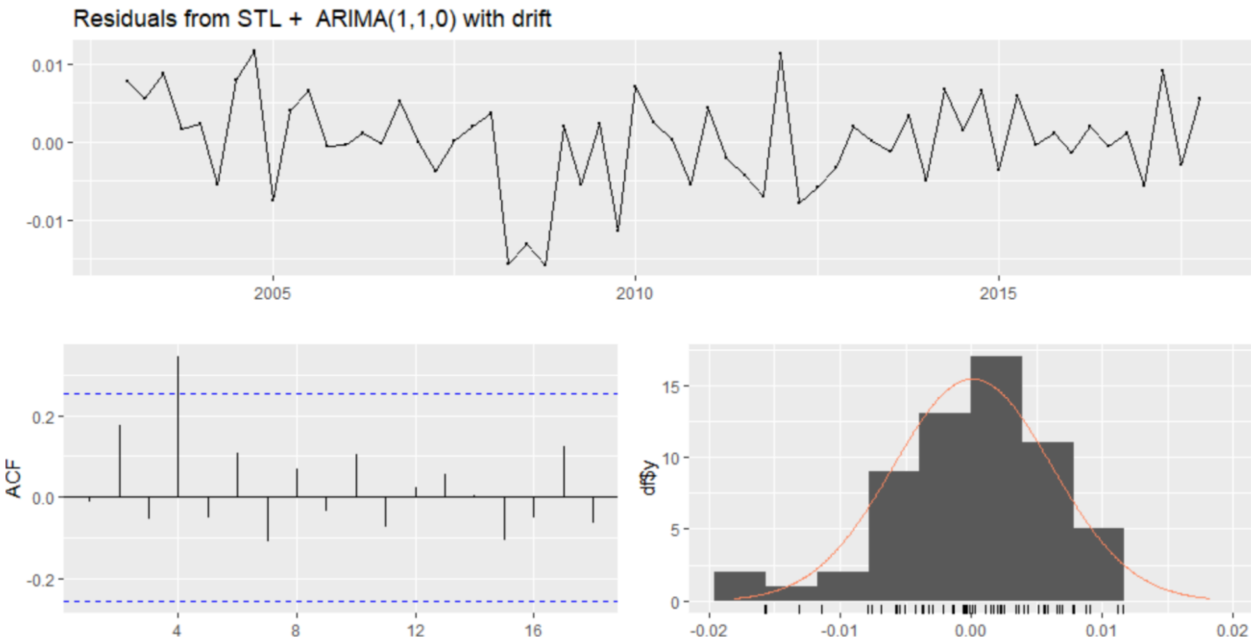
Based on the Ljung-Box Test we see that the p value is lesser than 0.05 thus helping us reject the null hypothesis of white noise

STL ARIMA

Accuracy

	RMSE	MAE	MAPE	MASE
Training set	0.006066921	0.004781162	0.06054864	0.2048092
Test set	0.004314008	0.003730538	0.04617689	0.1598039

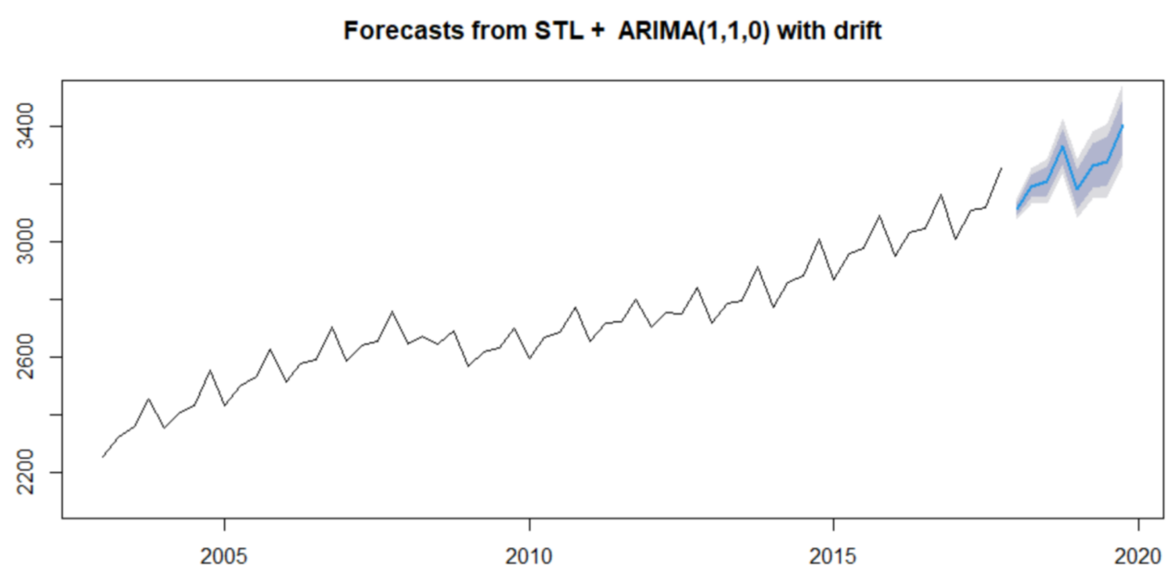
Residual Diagnostic




```
        Ljung-Box test

data:  Residuals from STL +  ARIMA(1,1,0) with drift
Q* = 12.109, df = 6, p-value = 0.05958

Model df: 2.   Total lags used: 8
```

Based on the Ljung-Box Test see that the p value is greater than 0.05 thus helping us accept the null hypothesis of white noise

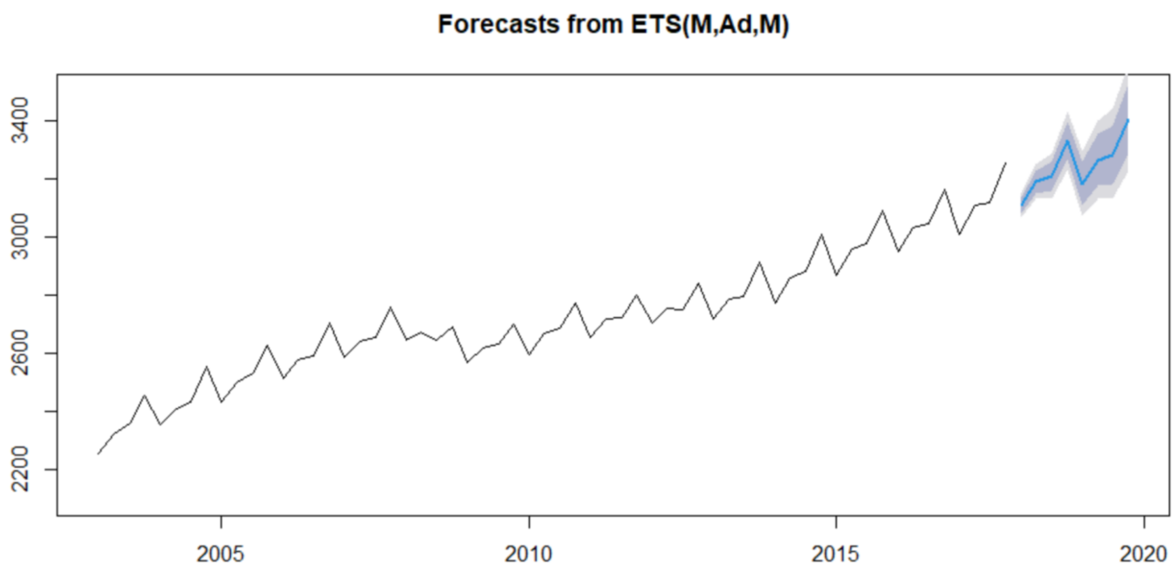
Based on all 4 models we see that **STL ARIMA (1,1,0) with drift** has the best performance in terms of accuracy (**0.16 MASE**) and then on evaluating residuals and performing Ljung Box we can say that the null hypothesis of white noise is accepted and based on residuals the model is able to capture the trends in the data.
Forecast using **STL ARIMA (1,1,0)**



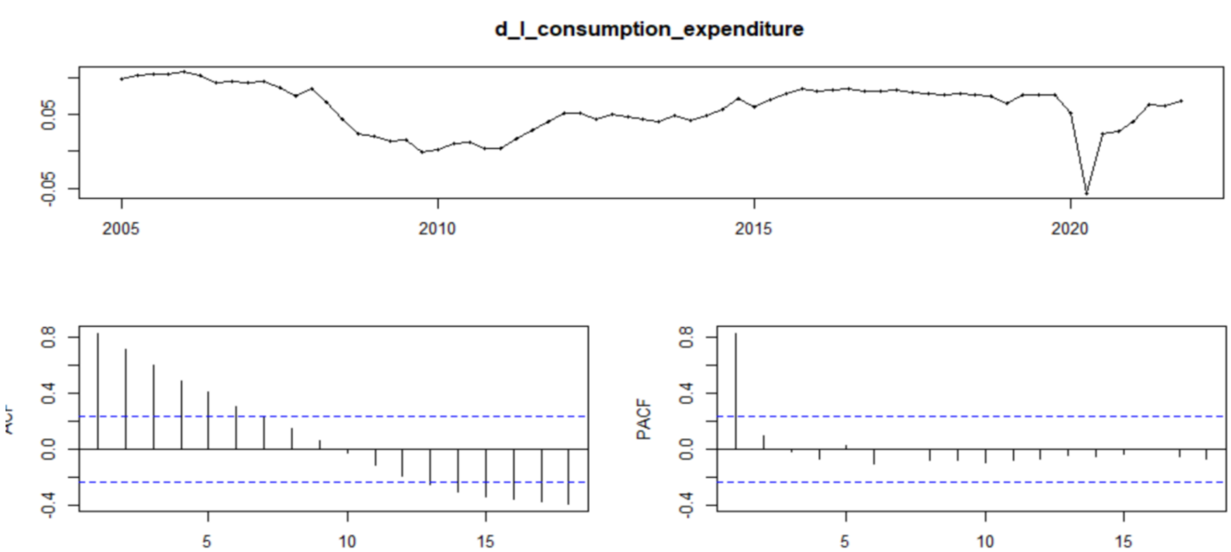
In the next part we will be using ETS Model with Multiplicative and Additive seasonal component and Additive trend component to account for the seasonal and trend we see in the time series.

Model	RMSE	MAE	MAPE	MASE	Ljung-Box test (P value)
AAA	0.007992627	0.006517575	0.08061865	0.2791913	0.07112
MAM	**0.004290688**	**0.003485435**	**0.04315847**	**0.1493045**	**0.09021**
MAA	0.007051173	0.005565148	0.06886978	0.2383925	0.08902

Based on the test accuracy and Ljung Box and residual diagnostic we see that the MAM model has the best performance in terms of accuracy and p value is greater than 0.05 thus null hypothesis of white noise is accepted.
The forecasts using the ETS MAM model will be



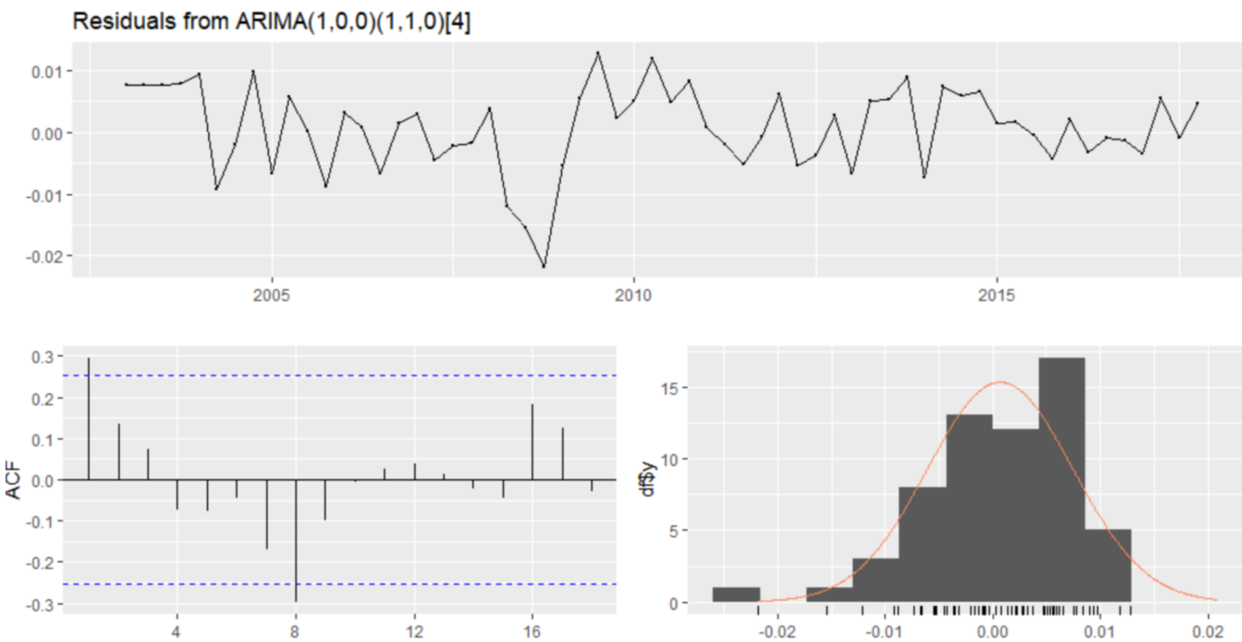
We will now be using the ARIMA model
Seasonality differenced plot



Based on the plot we will be running a ARIMA model (1, 0, 0) (1,1,0)[4]. We would be trying various p and q values to check if we get better performance.
Various ARIMA model was tried ARIMA model (1, 0, 1) (1,1,0) [4], ARIMA model (2, 0, 0) (1,1,0)[4], ARIMA model (1, 0, 1) (1,1,1)[4], ARIMA model (2, 0, 0) (1,1,1)[4]
On evaluation of residual diagnostic and accuracy we see that ARIMA model (1, 0, 0) (1,1,0)[4] has the performance.

Based on the plot we will be running a ARIMA model (1, 0, 0) (1,1,0) [4]

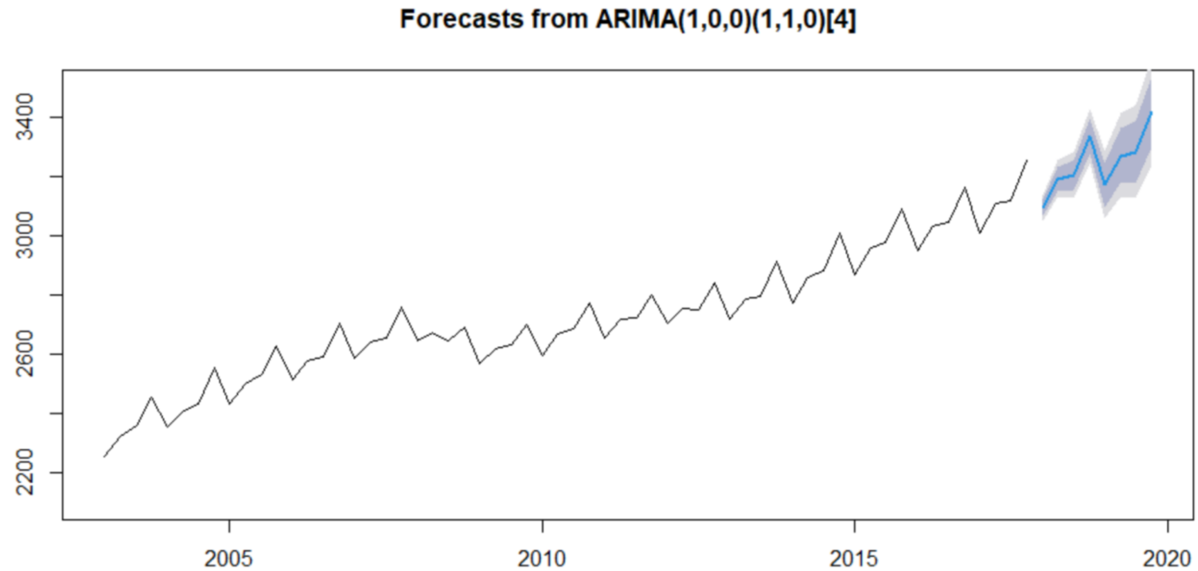
```
                     RMSE         MAE        MAPE      MASE
Training set 0.006723233 0.005413345 0.06862954 0.2318898
Test set     0.003054403 0.002505895 0.03104084 0.1073442
```




```
        Ljung-Box test

data:  Residuals from ARIMA(1,0,0)(1,1,0)[4]
Q* = 20.102, df = 14, p-value = 0.127

Model df: 2.   Total lags used: 16
```

Based on the Ljung-Box Test see that the p value is greater than 0.05 thus helping us accept the null hypothesis of white noise

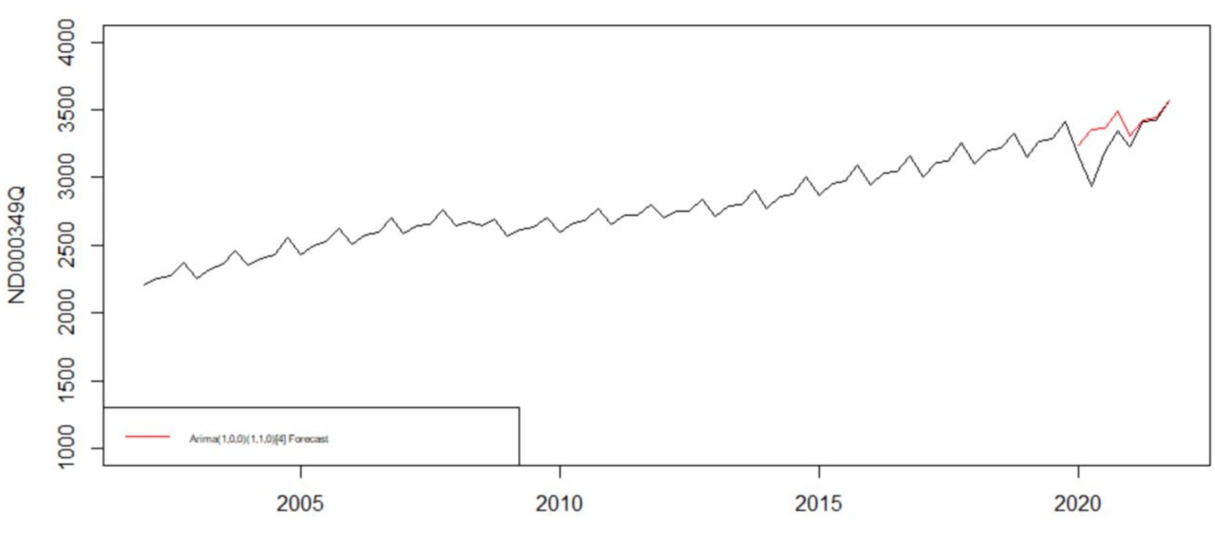
Forecast using the A ARIMA model (1, 0, 0) (1,1,0) [4]



Comparison for all models we see that Arima Model has the best performance in terms of accuracy and P value.

Model	RMSE	MAE	MAPE	MASE	Ljung-Box test(P value)
MAM	0.004290688	0.003485435	0.04315847	0.1493045	0.09021
ARIMA(1,0,1)(1,1,0)[4]	**0.003054403**	**0.002505895**	**0.03104084**	**0.1073442**	**0.1425**
STL ARIMA (1,1,0) with drift	0.004314008	0.003730538	0.04617689	0.1598039	0.05958

We will be using the ARIMA (1,0,1) (1,1,0)[4] for forecasting on the validation and comparing our results.
Forecasting results for the validation dataset



Based on the results we see our model was not able to capture the sudden dip in 2020 due to the pandemic and this would be an expected behavior.

References

- https://fred.stlouisfed.org/series/ND000349Q
- https://otexts.com/fpp2/AR.html
- https://otexts.com/fpp2/residuals.html
- https://otexts.com/fpp2/transformations.html
- https://www.rampfesthudson.com/can-box-cox-lambda-be-negative/#Can_Box_Cox_lambda_be_negative
- https://towardsdatascience.com/identifying-ar-and-ma-terms-using-acf-and-pacf-plots-in-time-series-forecasting-ccb9fd073db8
- https://www.rdocumentation.org/packages/forecast/versions/8.16/topics/Acf
- Lecture Slides and Code for forecasting LABS