Trend Analysis of Air Quality in Los Angeles

Harshal Patil, Pratik Ghawate, Vedant Ghavate

San Diego State University, Fowler College of Business

1. Introduction

Air Quality Index is an important metric to determine the intensity of hazards, that the air in that particular environment will affect us. With air pollution on the rise due to vehicular emissions and factories as well as natural calamities, the Air quality index is a metric that informs us about all aforementioned factors.

The air quality of the Los Angeles Long Beach metropolitan area has been a significant concern for several decades. This area is home to one of the largest and most complex urban environments in the world, with a population of over 13 million people. Air pollution has been linked to several adverse health effects, including respiratory illnesses, heart disease, and cancer. In this report, we will analyze the air quality index (AQI) of Los Angeles Long Beach from 2000 to 2010.

To analyze the trend of AQI, we choose Long Beach, California as the region to make this analysis. Long Beach has many factors from the timeline we choose from 2000 to 2010, during this time we see many pieces of legislation being passed to curb air pollution to lower air pollution. Also, Long Beach experienced natural calamities like increasingly frequent wildfires which leave a lasting impact on air quality.

All the unique facets that affect long leach, los angeles, and its air quality, make it an interesting choice for this report as we can observe the impact of legislation regarding air quality over the years and also where air quality in Los Angeles trends.

2. Executive Summary

Significant events related to air quality in Los Angeles Long Beach from 2000 to 2010:

The California energy crisis of 2000-2001: The energy crisis caused an increase in the use of diesel-powered generators, which contributed to higher levels of particulate matter and nitrogen oxide emissions. This led to a decline in air quality in the Los Angeles Long Beach area.

The Clean Air Act Amendments of 2003: In 2003, amendments to the Clean Air Act were passed, which established new emission standards for diesel engines and nonroad vehicles, such as construction equipment and farm machinery. These standards helped to reduce emissions of particulate matter and nitrogen oxides in the Los Angeles Long Beach area.
The South Coast Air Basin 2007 Plan: In 2007, the South Coast Air Quality Management District (SCAQMD) developed a plan to reduce air pollution in the Los Angeles Long Beach area. The plan included measures such as reducing emissions from stationary sources and implementing a vehicle scrappage program.

The Porter Ranch gas leak: In 2015, a gas leak occurred in the Porter Ranch neighborhood of Los Angeles, which released large amounts of methane into the air. The leak was one of the worst environmental disasters in the United States, and it caused significant health problems for residents in the area.
The California wildfires of 2003 and 2007: Wildfires in California burned for weeks in 2003 and 2007, releasing harmful pollutants into the air and causing significant health problems for residents in the Los Angeles Long Beach area.
The Recession of 2008-2009: The recession led to a decline in economic activity and a corresponding decrease in industrial activity. This led to an improvement in air quality in the Los Angeles Long Beach area, as emissions from stationary sources decreased.

These are just a few significant events related to air quality in Los Angeles Long Beach from 2000 to 2010. It is worth noting that air quality issues in this area are ongoing, and efforts to reduce emissions and improve air quality continue to be a priority.

The AQI in Los Angeles Long Beach has improved gradually from 2000 to 2010, although there were some fluctuations. The highest AQI was recorded in 2003, while the lowest was recorded in 2010. The improvement in air quality can be attributed to several factors, including the implementation of stricter emission standards for vehicles and industrial facilities, the use of cleaner fuels, and the expansion of public transportation. However, there is still a long way to go to achieve healthy air quality levels in this area. Continued efforts are needed to reduce emissions and protect public health.

3. Data Preparation

Data Source

Link: https://www.kaggle.com/datasets/sogun3/uspollution

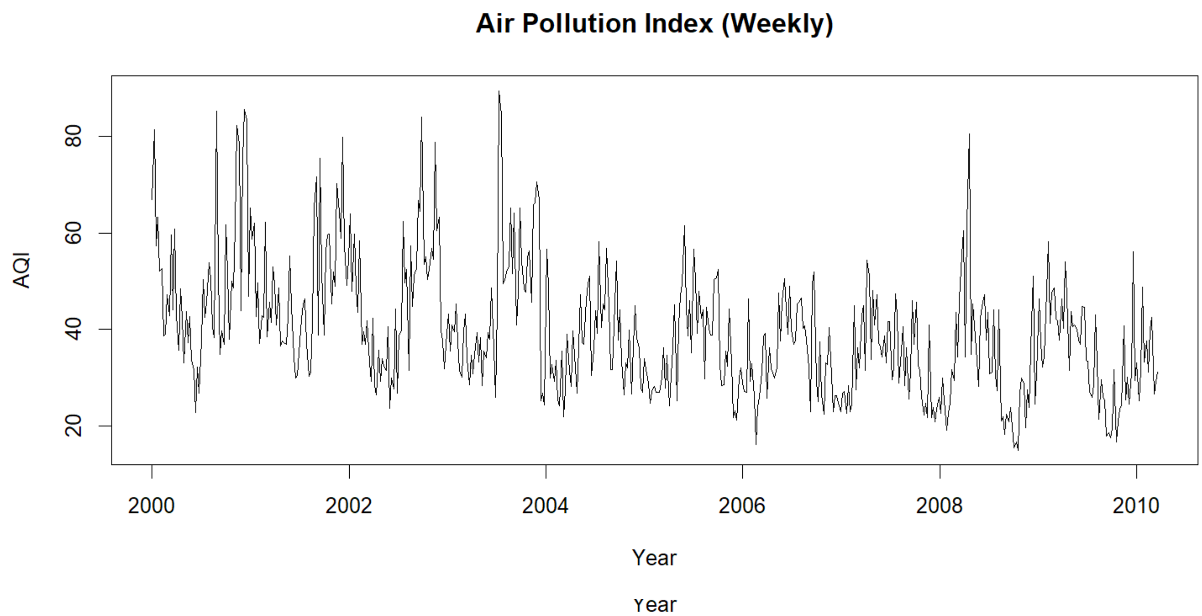
Data Filtering

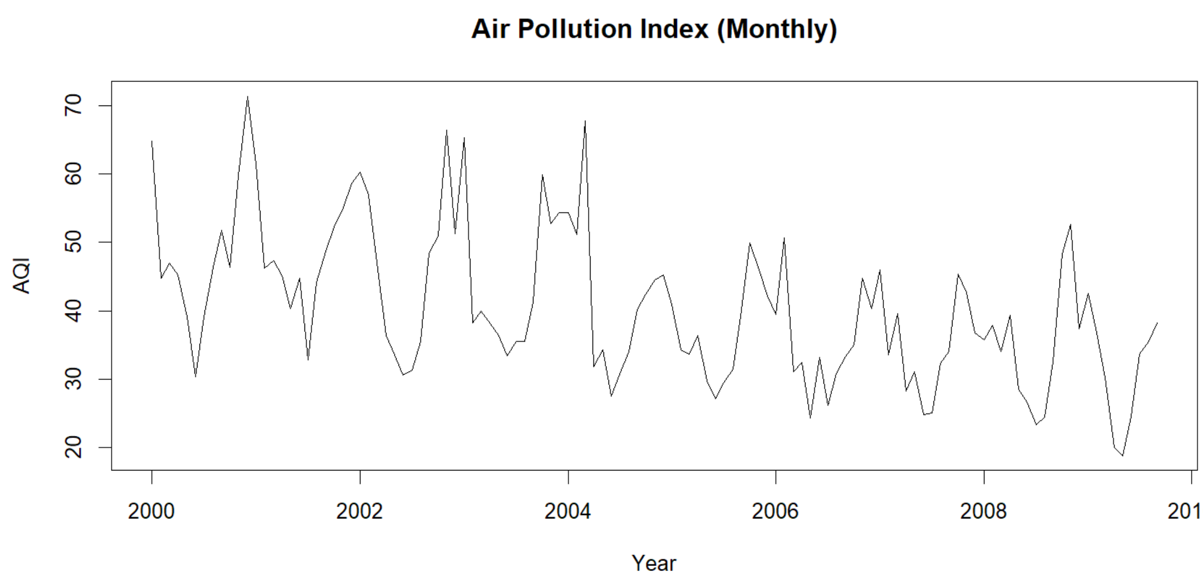
This dataset was made available by a Kaggale user @BRENDASO who has scrapped this data from the United States Environmental Protection Agency (US EPA). Firstly, The dataset contains geolocation information which includes the state, city, address, and one of the two locations in that particular zip code where air quality is measured. Secondly, it contains columns for NO2 SO2 and CO AQI for each geolocation for that particular day. Finally, it has the maximum, mean, and minimum values of AQI for that specific location on a particular day.
Through this verbose dataset, we have selected ‘3648 N. LONG BEACH BLVD. LONG BEACH’ as our specific location to conduct our trend analysis. Along with this filter, one important column we selected for our research was NO2 AQI. The most significant gas in measuring air quality is No2, hence to narrow the scope of the analysis, we have considered NO2 AQI for further research.
Finally, after selecting the required columns, we again aggregated the values into weekly averages and monthly averages. This was done by creating three different CSV files having the No2 AQI and month/ week as the column. This aggregation was necessary to explore the time series through different levels of granularity.

4. Exploring the Data

Filtering data

The dataset was aggerated and saved into different files with weekly and monthly averages of No2 AQI. This was done to explore the extent of noise that daily and weekly data carried.





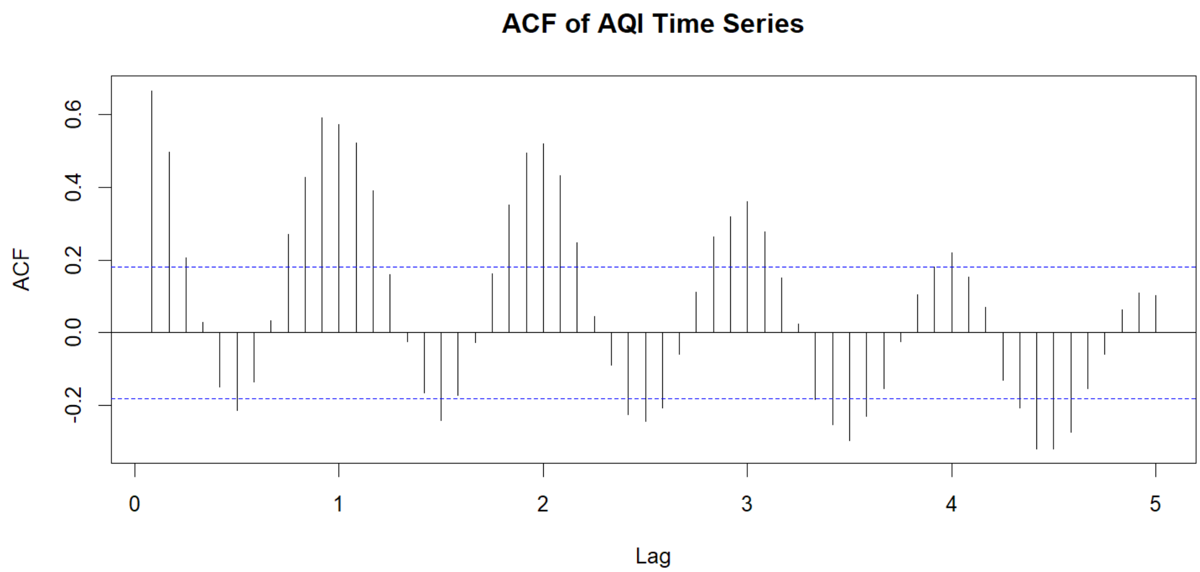
From the above charts, it is clear that the daily and weekly presented a lot of noise and AQI shows variables on a few days due to weather or very location-specific problems like windy weather needing a restaurant. The reason for noise may also include the short-term AQI changes at the location of measuring the AQI.

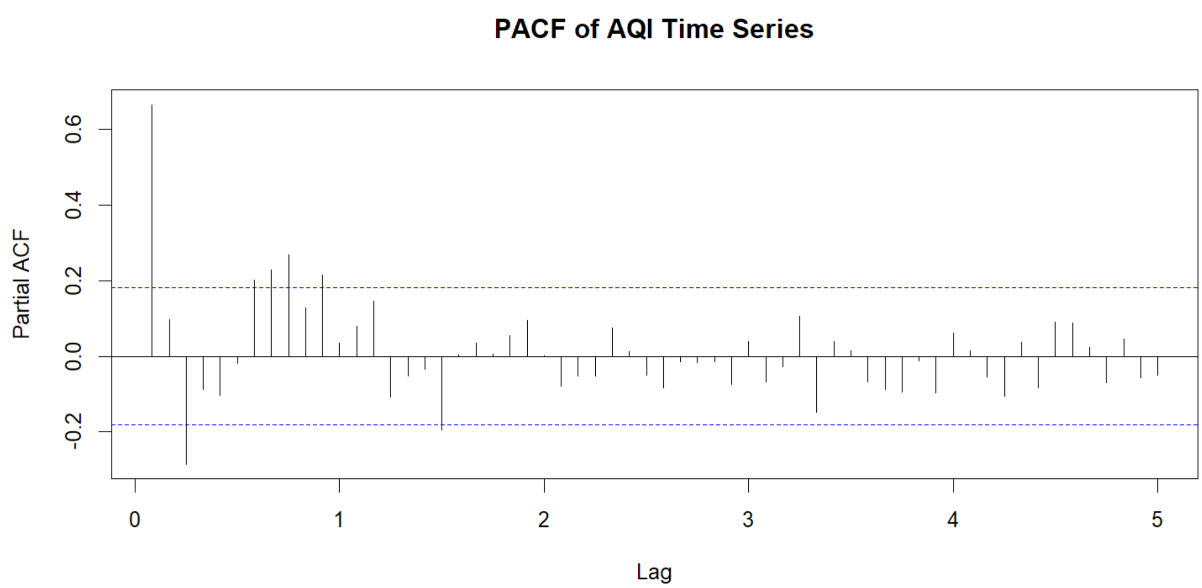
It is due to these that monthly averages of AQI presented a clear pattern and seasonality that we could explore and build a model for as well.

5. Model Planning

We use the Unit root test, ACF, and PACF to explore the time series data.

```
Augmented Dickey-Fuller,Test

data:  datamts
Dickey-Fuller = -7.7102,
Lag order = 4, p-value = 0.01
alternative hypothesis: stationary
```

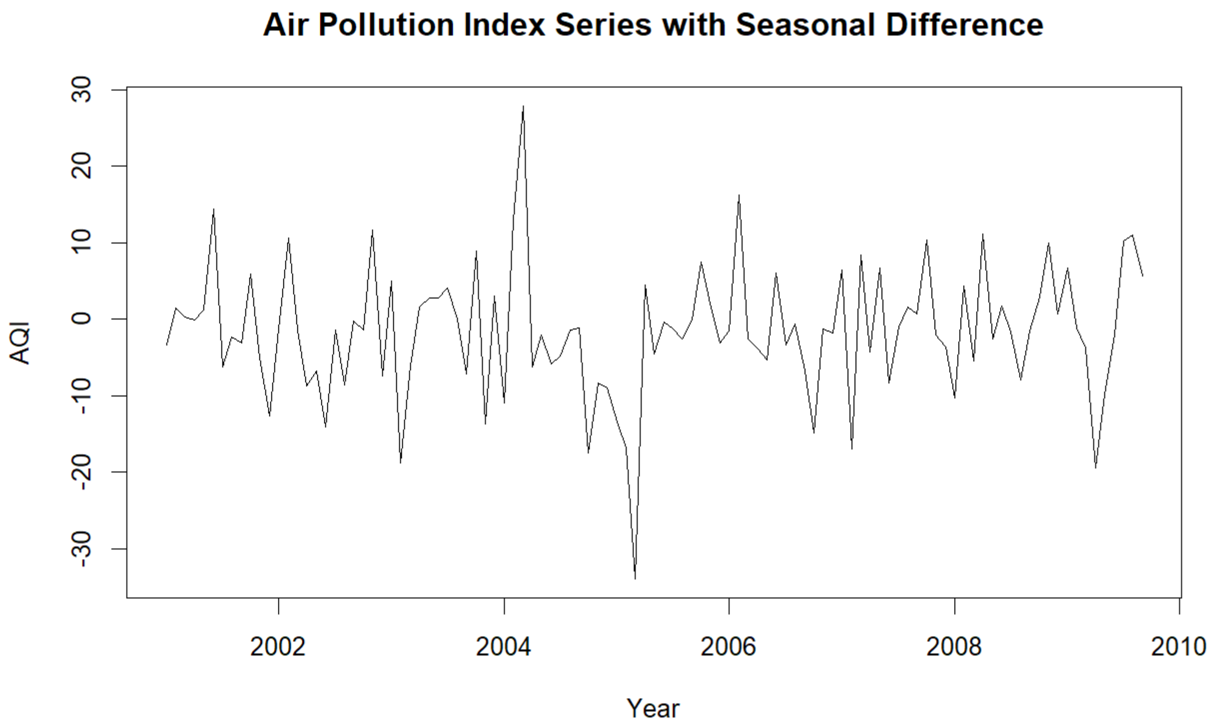




The unit root test confirms that the series is stationary and does not require a log transformation and difference to be considered. Hence, we proceed without any transformation. The ACF and PACF suggest seasonality in periods 12, 24, 36, and so on.

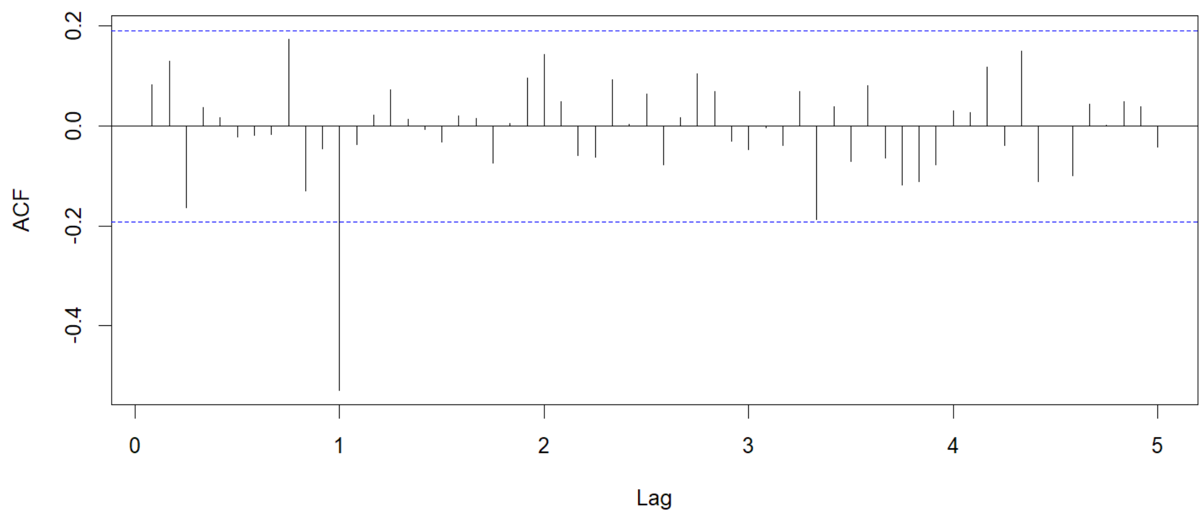
Model Selection - Seasonality Effect

It is visible from the time series graph that there is seasonality in the series with period 12. Hence we take seasonal differences with period 12.

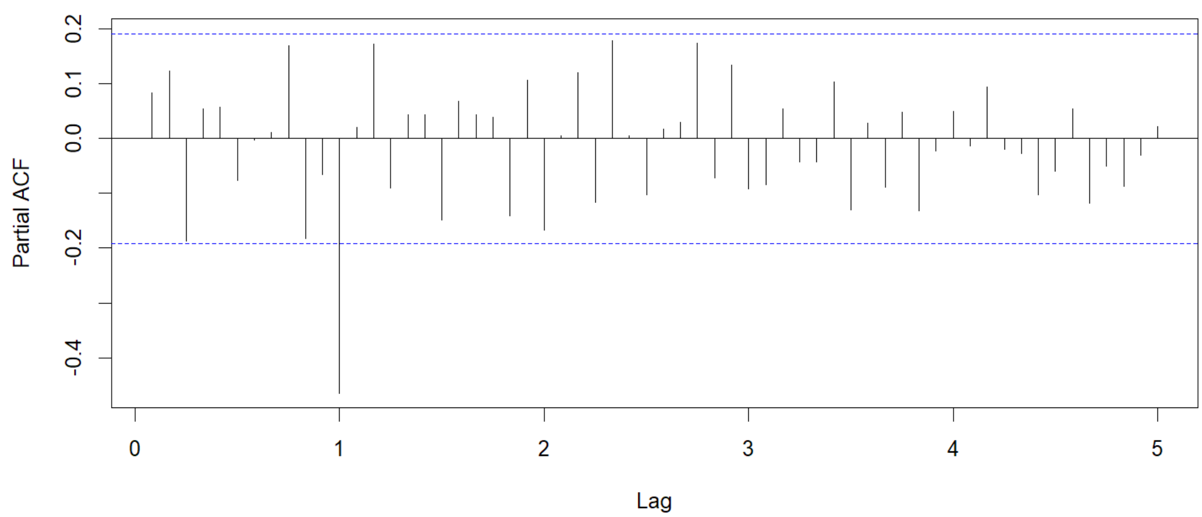


We can observe that the seasonality is no longer visible in the graph above after taking the seasonal difference.

ACF of AQI Time Series with Seasonal Difference



PACF of AQI Time Series with Seasonal Difference



Through the ACF, we observe that only lag 1 is significant, this indicates that we can use the seasonal MA(1) model. The PACF however, indicates that Lag 1 is significant, and we can also consider the seasonal AR(1) model.

Model Building - Comparing Models

After eliminating seasonality from the series, we can use seasonal MA(1) and seasonal AR(1) to build the model. Since we have taken seasonal differences, the seasonal model i.e. (p,s,q) will consider s = 1.

Therefore we consider the seasonal model (1,1,1) where we have both AR and MA components, seasonal (0,1,1) weight only the MA component, and lastly, seasonal (1,1,0) with only the AR component. The difference component is taken for all.

	SARIMA(1,1,1)	SARIMA(0,1,1)	SARIMA(1,1,0)
AIC	701.07	701.88	712.87

We compare the AIC for the three seasonal models and we can conclude that seasonal arma (1,1,1) has the lowest AIC, hence we consider this seasonal model for further processing.

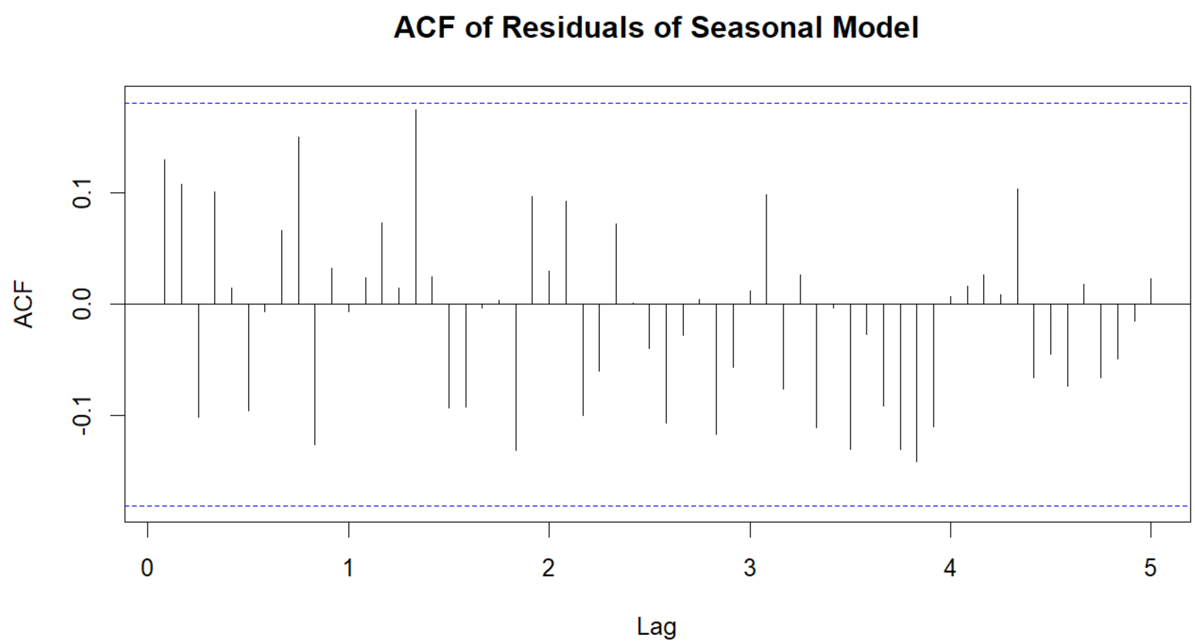
```
      Estimate Std. Error z value  Pr(>|z|)    
sar1 -0.238901   0.133602 -1.7882   0.07375 .  
sma1 -0.683435   0.142371 -4.8004 1.584e-06 ***
xreg -0.160567   0.018602 -8.6319 < 2.2e-16 ***
```

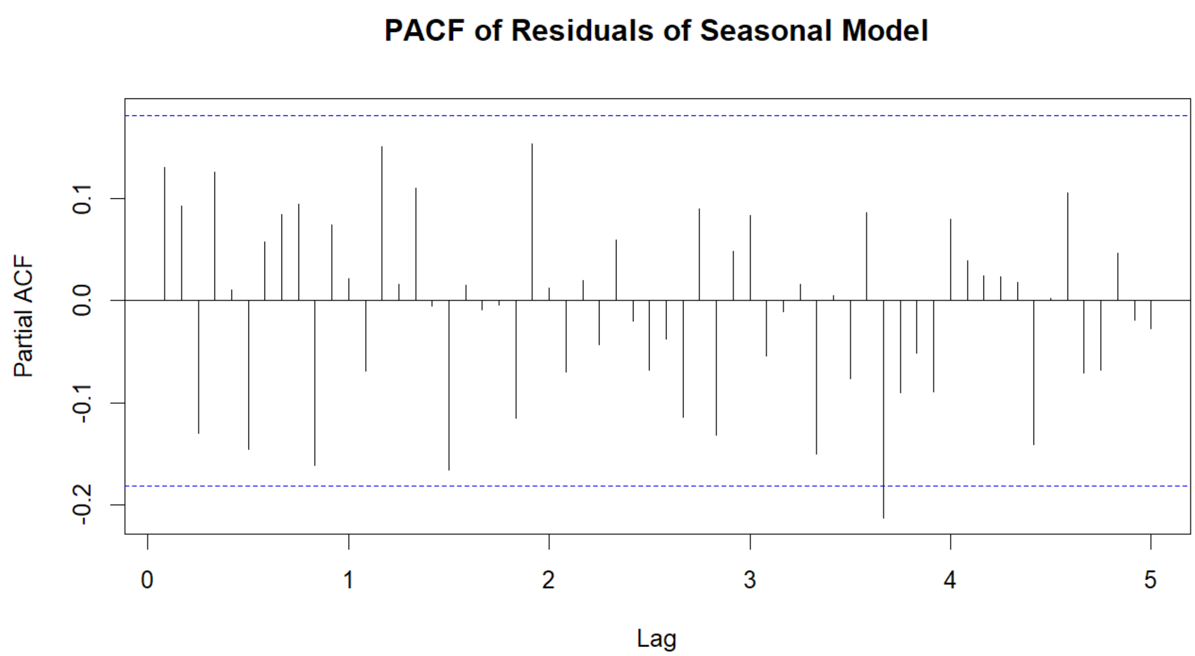
Coefficient test for seasonal SARIMA(1,1,1) Model

We need to observe the residuals of the seasonal model for further analysis, hence we evaluate the residuals on the Box-Ljung test and plot the ACF and PACF for the residuals. We finally create the EACF to get the best model.

```
Box-Ljung test

data:  out$residuals
X-squared = 60.075, df = 60, p-value = 0.473
```

From the Box-Ljung test, we can conclude that there is no serial dependence in the results and we have extracted all relevant rend from the residuals. They also conclude that the residuals are white noise and do not have more insights left.





The ACF and PACF also affirm our conclusion the residuals are white noise, as no log is significant in both ACF and PACF.

```
AR/MA
  0 1 2 3 4 5 6 7 8 9 10 11 12 13
0 o o o o o o o o o o o  o  o  o
1 x o o o o o o o o o o  o  o  o
2 x o o o o o o o o o o  o  o  o
3 x o x o o o o o o o o  o  o  o
4 o x o x o o o o o o o  o  o  o
5 o x x o o o o o o o o  o  o  o
6 x x o o o o o o o o o  o  o  o
7 x o x x o x x o o o o  o  o  o
```

The EACF indicates that the MA and AR should be 0, this may be because seasonality was the only relevant trend in this time series data. However, we consider ARIMA(1,0,1) and ARIMA(0,0,0) for comparison to further affirm our analysis. The delta component of the model will remain zero as we have not taken any difference.

ARIMA(0,0,0) x SARIMA(1,1,1)

```
      Estimate Std. Error z value  Pr(>|z|)
sar1 -0.238901   0.133602 -1.7882   0.07375 .
sma1 -0.683435   0.142371 -4.8004 1.584e-06 ***
xreg -0.160567   0.018602 -8.6319 < 2.2e-16 ***
```

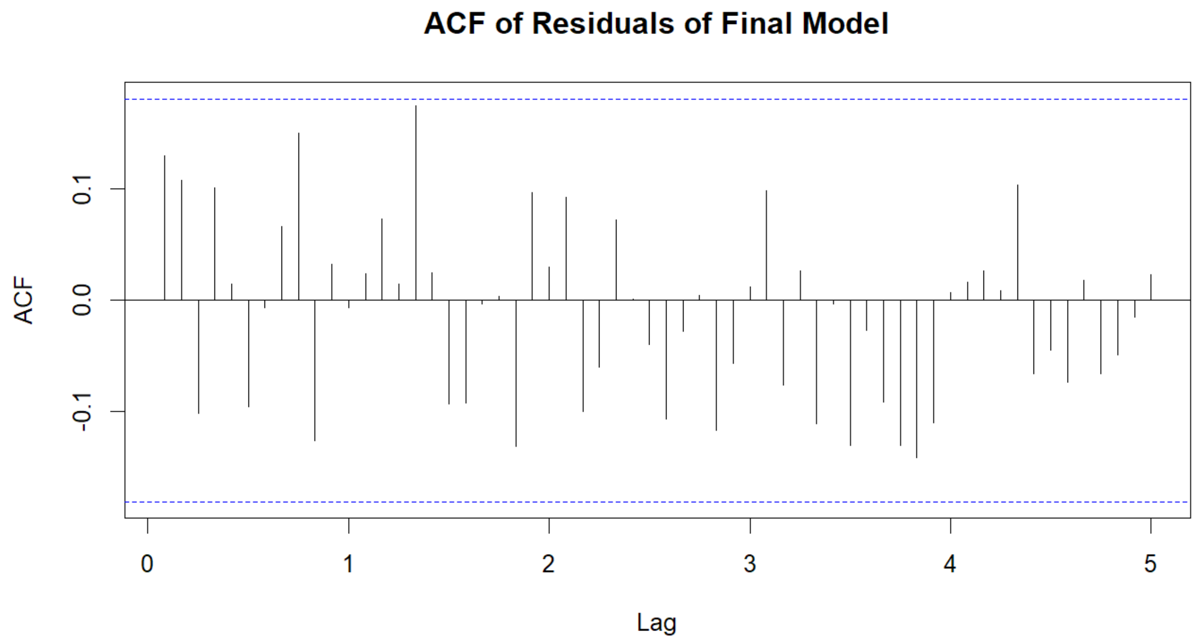
ARIMA(1,0,1) x SARIMA(1,1,1)

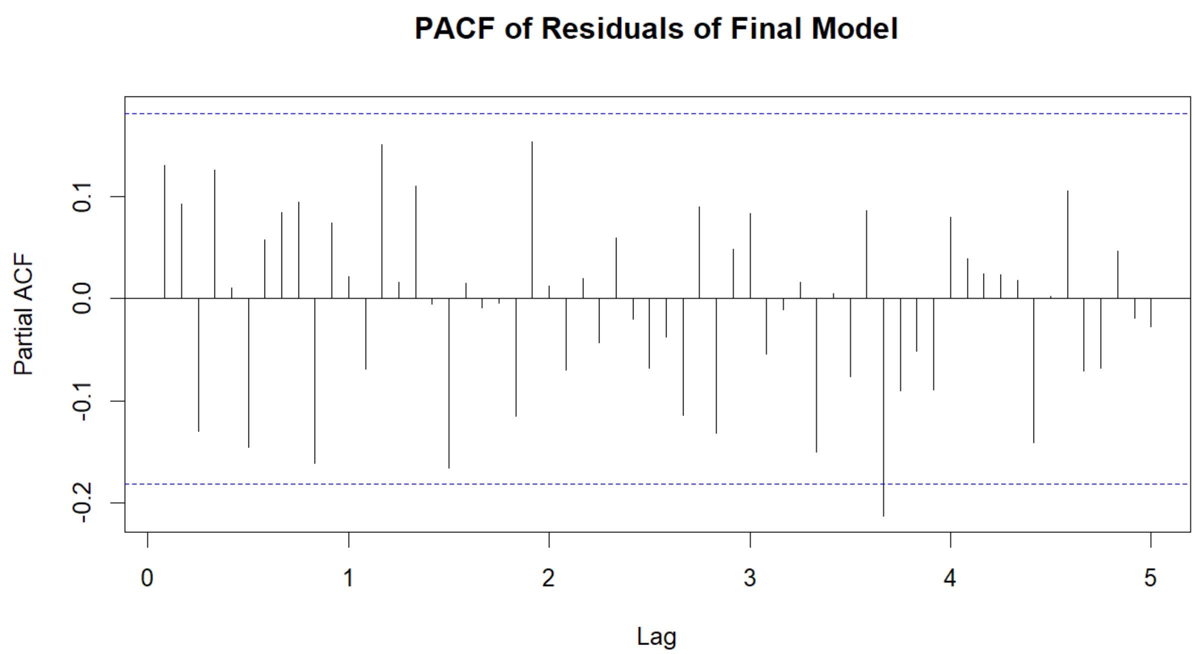
```
      Estimate Std. Error z value  Pr(>|z|)
ar1   0.496597   0.451718  1.0994   0.27161
ma1  -0.362687   0.479339 -0.7566   0.44927
sar1 -0.238723   0.130540 -1.8287   0.06744 .
sma1 -0.702092   0.138716 -5.0614 4.163e-07 ***
xreg -0.159566   0.022535 -7.0807 1.435e-12 ***
```

	ARIMA(0,0,0) x SARIMA(1,1,1)	ARIMA(1,0,1) x SARIMA(1,1,1)
AIC	701.07	702.7
Rolling Forecasting	508.4677 928.0959 1654.7196	509.0062 928.0828 1654.9629

It is clear from the AIC and 3 Step Rolling Forecast that the ARIMA(0,0,0) x SARIMA(1,1,1) is the better model.

We proceed to analyze the residuals from this model to confirm if all insights are extracted.





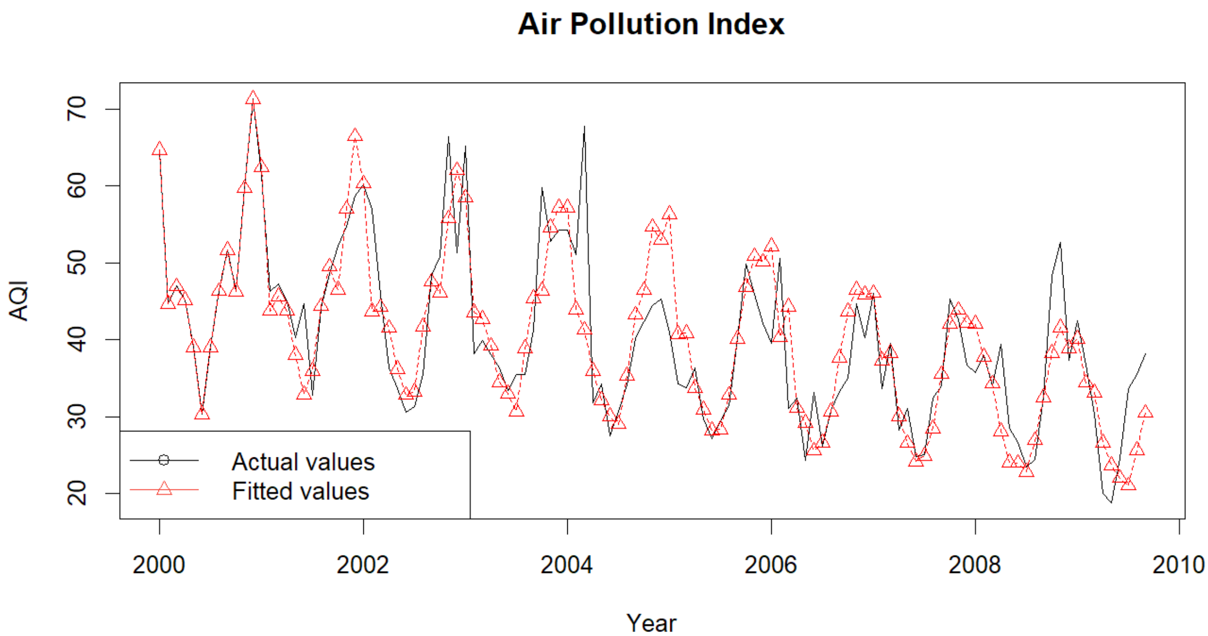

```
Box-Ljung test

data:  out$residuals
X-squared = 60.075, df = 60, p-value = 0.473
```

We can conclude that the same results after extracting seasonality, and that the residuals are white noise.

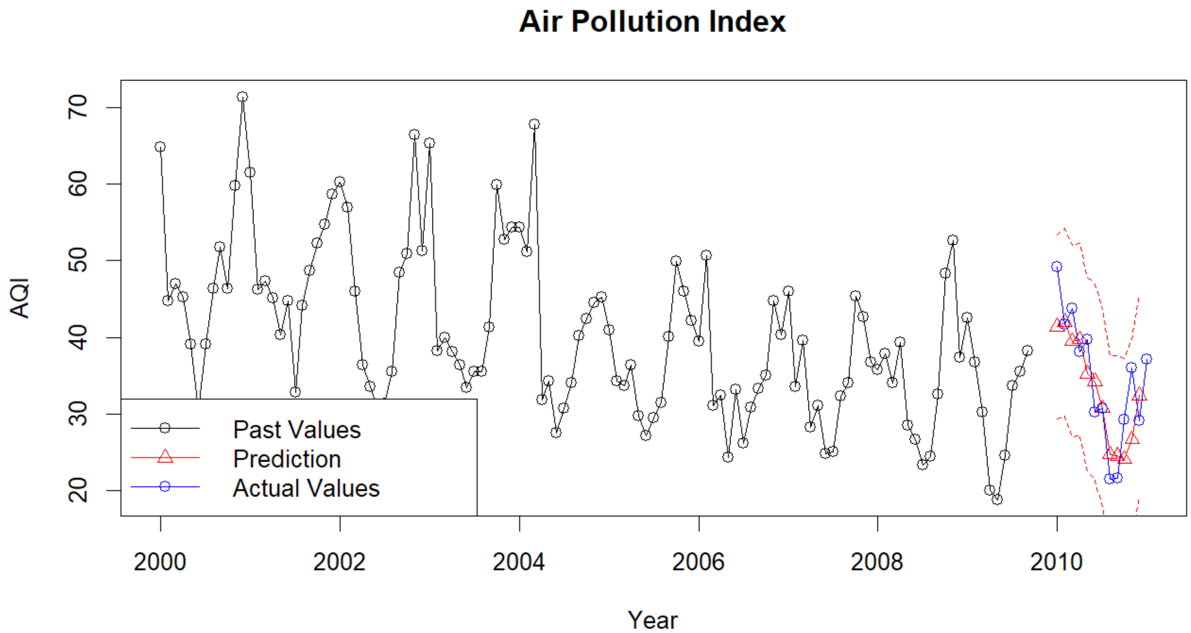
6. Forecasting

Fitted Values vs Actuals Values

We see a slightly improper fit in 2003, 2005, and 2009, this may be due to wildfires in 2003, air quality-related legislation being passed in 2005, and wildfires again in 2009.



Predicting values



We had already split the dataset to exclude the data from January 2010 to December 2010, the model that we built will now predict the AQI in 2010 as visible in the chart above.
The prediction appears to be quite close to the actual values and very well within the upper and lower bound of prediction. The model also shows the gradual decrease in AQI in Long Beach, Los Angeles.

7. Conclusion

The model complied captures the improvement in AQI over the years. We predicted the AQI from January 2010 to December 2010. Wildfires remain a concern as they can cause a temporary spike in pollution, however, our model has managed to capture them through the years 2003 and 2007. Incidents related to Air Pollution would remain inventions and have occurred in 2015, however, this is outside the data selected for this analysis.

Findings and Application of our data analysis

We found strong evidence that the main trend of the air quality index is seasonality with peaks in the summer months, apart from seasonality there was no conclusive evidence to support any other trend or insight into the pattern of AQI.

Another important finding was that AQI shows a consistent decrease through the years, the model we built also reflects the same.

The model, however, does capture the decline in AQI, which is caused to legislation being passed to improve air quality, the dramatic drop in 2005 is not captured by the model.

Temporary spikes in AQI due to wildfires have been somewhat captured in the model, hoever additional data or a bigger timeline would be able to capture the AQI changes due to wildfires as well.

Our analysis confirms the AQI is on a slow decline and that seasonal spikes will be observed in the future, however, there appears to be no other trend in the AQI time series.

8. References

1. 0-California Energy Commission. (2001). The Energy Crisis of 2000-2001: Causes, Consequences, and Policy Lessons. Retrieved from https://www.energy.ca.gov/2001publications/CEC-500-2001-082/CEC-500-2001-082.PDF
2. United States Environmental Protection Agency. (2004). Clean Air Act Amendments of 2003 Summary. Retrieved from https://www.epa.gov/sites/production/files/2016-09/documents/caaa_2003_summary.pdf
3. South Coast Air Quality Management District. (2007). 2007 Air Quality Management Plan. Retrieved from https://www.aqmd.gov/docs/default-source/clean-air-plans/2007-air-quality-management-plan.pdf
4. The New York Times. (2016). Porter Ranch Gas Leak Was California's Biggest, and the Worlds. Retrieved from https://www.nytimes.com/2016/07/15/us/porter-ranch-gas-leak-california-biggest-worlds.HTML
5. California Air Resources Board. (2018). Wildfires and Air Quality. Retrieved from https://ww3.arb.ca.gov/ei/wildfire/wildfire.htm
6. California Environmental Protection Agency. (2010). Air Quality in California 2010. Retrieved from https://www.arb.ca.gov/research/apr/past/apr2010/04-2010.pdf

Appendix

R Console

```
library(stringr)
Warning message:
package ‘stringr’ was built under R version 4.2.3
> library(xts)
Loading required package: zoo

Attaching package: ‘zoo’

The following objects are masked from ‘package:base’:

    as.Date, as.Date.numeric

Warning messages:
1: package ‘xts’ was built under R version 4.2.3
2: package ‘zoo’ was built under R version 4.2.3
> library(tseries)
Registered S3 method overwritten by 'quantmod':
  method            from
  as.zoo.data.frame zoo

    ‘tseries’ version: 0.10-53

    ‘tseries’ is a package for time series analysis and computational
    finance.

    See ‘library(help="tseries")’ for details.

Warning message:
package ‘tseries’ was built under R version 4.2.2
> library(TSA)

Attaching package: ‘TSA’

The following objects are masked from ‘package:stats’:

    acf, arima
```

```
The following object is masked from ‘package:utils’:

    tar

Warning message:
package ‘TSA’ was built under R version 4.2.2
> library(lmtest)
Warning message:
package ‘lmtest’ was built under R version 4.2.3
> set.seed(121)
> data = read.csv('LA_pollution.csv')
> head(data)
      X State.Code County.Code Site.Num                        Address
1 13355          6          37     4002 3648 N. LONG BEACH BLVD., LONG BEACH
2 13359          6          37     4002 3648 N. LONG BEACH BLVD., LONG BEACH
3 13363          6          37     4002 3648 N. LONG BEACH BLVD., LONG BEACH
4 13367          6          37     4002 3648 N. LONG BEACH BLVD., LONG BEACH
5 13371          6          37     4002 3648 N. LONG BEACH BLVD., LONG BEACH
6 13375          6          37     4002 3648 N. LONG BEACH BLVD., LONG BEACH
       State      County       City Date.Local        NO2.Units NO2.Mean
1 California Los Angeles Long Beach   1/1/2000 Parts per billion 33.39130
2 California Los Angeles Long Beach   1/2/2000 Parts per billion 19.43478
3 California Los Angeles Long Beach   1/3/2000 Parts per billion 49.69565
4 California Los Angeles Long Beach   1/4/2000 Parts per billion 50.21739
5 California Los Angeles Long Beach   1/5/2000 Parts per billion 60.73913
6 California Los Angeles Long Beach   1/6/2000 Parts per billion 58.52174
  NO2.1st.Max.Value NO2.1st.Max.Hour NO2.AQI         O3.Units  O3.Mean
1                44               21      42 Parts per million 0.007750
2                46               23      43 Parts per million 0.018375
3                63               17      61 Parts per million 0.003875
4                70                0      68 Parts per million 0.003750
5               101               13     101 Parts per million 0.003500
6                73                8      71 Parts per million 0.004417
  O3.1st.Max.Value O3.1st.Max.Hour O3.AQI         SO2.Units SO2.Mean
1            0.015              23     13 Parts per billion 0.739130
2            0.033               6     28 Parts per billion 0.391304
3            0.009               8      8 Parts per billion 1.652174
4            0.009               9      8 Parts per billion 1.521739
5            0.006              21      5 Parts per billion 4.173913
6            0.011               9      9 Parts per billion 3.913043
  SO2.1st.Max.Value SO2.1st.Max.Hour SO2.AQI          CO.Units  CO.Mean
1                 3               21       4 Parts per million 0.811111
2                 2               15       3 Parts per million 0.237500
3                 4               14       6 Parts per million 1.500000
```

```
4              4              8      6 Parts per million 1.916667
5             27             13     39 Parts per million 2.487500
6             11             12     16 Parts per million 1.325000
  CO.1st.Max.Value CO.1st.Max.Hour CO.AQI
1            2.0              6    23
2            0.6              3     7
3            2.3             10    26
4            2.9              4    33
5            3.9              8    44
6            2.7             10    31
> datats=ts(data[14],start=c(2000,1,1), frequency=365.25)
> plot(datats, xlab = 'Year', ylab = 'AQI', main = 'Air Pollution Index (Daily)')
> dataw = read.csv('pollution_week.csv')
> head(dataw)
      WEEK
1 67.00000
2 81.42857
3 57.42857
4 63.28571
5 52.14286
6 52.57143
> datats=ts(dataw,start=c(2000,1,1), frequency=365.25/7)
> plot(datats, xlab = 'Year', ylab = 'AQI', main = 'Air Pollution Index (Weekly)')
> datam = read.csv('pollution_monthly.csv')
> head(datam)
  X     X0       X1 X2
1 0 2/2000 64.74194 31
2 1 3/2000 44.75000 28
3 2 4/2000 47.00000 30
4 3 5/2000 45.20690 29
5 4 6/2000 39.06667 30
6 5 7/2000 30.31034 29
> datamts=ts(datam$X1[1:117],start=c(2000,1), frequency=12)
> datamts.new=ts(datam$X1[118:131],start=c(2010,1), frequency=12)
> plot(datamts, xlab = 'Year', ylab = 'AQI', main = 'Air Pollution Index (Monthly)')
> adf.test(datamts)

        Augmented Dickey-Fuller Test

data:  datamts
Dickey-Fuller = -7.7102, Lag order = 4, p-value = 0.01
alternative hypothesis: stationary

Warning message:
```

```
In adf.test(datamts) : p-value smaller than printed p-value
> acf(datamts,lag.max=60,main="ACF of AQI Time Series")
> pacf(datamts,lag.max=60,main="PACF of AQI Time Series")
> # obvuois seasonality at period 12
> sd_datamts=diff(datamts,12)
> plot(sd_datamts,xlab = 'Year', ylab = 'AQI', main = 'Air Pollution Index Series with Seasonal Difference')
> acf(sd_datamts,lag.max=60,main="ACF of AQI Time Series with Seasonal Difference")
> pacf(sd_datamts,lag.max=60,main="PACF of AQI Time Series with Seasonal Difference")
> n=length(datamts)
> out = arima(datamts,c(0,0,0),xreg =1:n,seasonal = list(order = c(1, 1, 1), period = 12))
> out

Call:
arima(x = datamts, order = c(0, 0, 0), seasonal = list(order = c(1, 1, 1), period = 12),
   xreg = 1:n)

Coefficients:
         sar1     sma1     xreg
      -0.2389  -0.6834  -0.1606
s.e.   0.1336   0.1424   0.0186

sigma^2 estimated as 39.2:  log likelihood = -347.53,  aic = 701.07
> coeftest(out)

z test of coefficients:

      Estimate Std. Error z value  Pr(>|z|)
sar1 -0.238901   0.133602 -1.7882   0.07375 .
sma1 -0.683435   0.142371 -4.8004 1.584e-06 ***
xreg -0.160567   0.018602 -8.6319 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> out1 = arima(datamts,c(0,0,0),xreg =1:n,seasonal = list(order = c(0, 1, 1), period = 12))
> out1

Call:
arima(x = datamts, order = c(0, 0, 0), seasonal = list(order = c(0, 1, 1), period = 12),
   xreg = 1:n)

Coefficients:
         sma1     xreg
      -0.8530  -0.1591
```

```
s.e.   0.1511   0.0177

sigma^2 estimated as 39.07:  log likelihood = -348.94,  aic = 701.88
> coeftest(out1)

z test of coefficients:

      Estimate Std. Error z value  Pr(>|z|)
sma1 -0.853039   0.151058 -5.6471 1.632e-08 ***
xreg -0.159149   0.017721 -8.9808 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> out2 = arima(datamts,c(0,0,0),xreg =1:n,seasonal = list(order = c(1, 1, 0), period = 12))
> out2

Call:
arima(x = datamts, order = c(0, 0, 0), seasonal = list(order = c(1, 1, 0), period = 12),
    xreg = 1:n)

Coefficients:
         sar1     xreg
      -0.5772  -0.1454
s.e.   0.0785   0.0373

sigma^2 estimated as 47.8:  log likelihood = -354.43,  aic = 712.87
> coeftest(out2)

z test of coefficients:

      Estimate Std. Error z value  Pr(>|z|)
sar1 -0.577242   0.078467 -7.3565 1.888e-13 ***
xreg -0.145408   0.037258 -3.9027 9.511e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> Box.test(out$residuals, type = 'Ljung',lag = 60)

             Box-Ljung test

data:  out$residuals
X-squared = 60.075, df = 60, p-value = 0.473

> acf(out$residuals,lag.max=60,main="ACF of Residuals of Seasonal Model")
```

```
> pacf(out$residuals,lag.max=60,main="PACF of Residuals of Seasonal Model")
> eacf(out$residuals)
AR/MA
  0 1 2 3 4 5 6 7 8 9 10 11 12 13
0 o o o o o o o o o o o  o  o  o
1 x o o o o o o o o o o  o  o  o
2 x o o o o o o o o o o  o  o  o
3 x o x o o o o o o o o  o  o  o
4 o x o x o o o o o o o  o  o  o
5 o x x o o o o o o o o  o  o  o
6 x x o o o o o o o o o  o  o  o
7 x o x x o x x o o o o  o  o  o
> outf = arima(datamts,c(0,0,0),xreg =1:n,seasonal = list(order = c(1, 1, 1), period = 12))
> outf

Call:
arima(x = datamts, order = c(0, 0, 0), seasonal = list(order = c(1, 1, 1), period = 12),
    xreg = 1:n)

Coefficients:
         sar1     sma1     xreg
      -0.2389  -0.6834  -0.1606
s.e.   0.1336   0.1424   0.0186

sigma^2 estimated as 39.2:  log likelihood = -347.53,  aic = 701.07
> coeftest(outf)

z test of coefficients:

      Estimate Std. Error z value  Pr(>|z|)
sar1 -0.238901   0.133602 -1.7882   0.07375 .
sma1 -0.683435   0.142371 -4.8004 1.584e-06 ***
xreg -0.160567   0.018602 -8.6319 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> source("rolling.forecast.R")
> rolling.forecast(datamts, 3, length(datamts)-10, c(0,0,0),seasonal = list(order = c(1, 1, 1)))
[1]  508.4677  928.0959 1654.7196
> outf1 = arima(datamts,c(1,0,1),xreg =1:n,seasonal = list(order = c(1, 1, 1), period = 12))
> outf1

Call:
arima(x = datamts, order = c(1, 0, 1), seasonal = list(order = c(1, 1, 1), period = 12),
```

```
    xreg = 1:n)

Coefficients:
         ar1      ma1     sar1     sma1     xreg
      0.4966  -0.3627  -0.2387  -0.7021  -0.1596
s.e.  0.4517   0.4793   0.1305   0.1387   0.0225

sigma^2 estimated as 38.07:  log likelihood = -346.35,  aic = 702.7
> coeftest(outf1)

z test of coefficients:

      Estimate Std. Error z value  Pr(>|z|)
ar1   0.496597   0.451718  1.0994   0.27161
ma1  -0.362687   0.479339 -0.7566   0.44927
sar1 -0.238723   0.130540 -1.8287   0.06744 .
sma1 -0.702092   0.138716 -5.0614 4.163e-07 ***
xreg -0.159566   0.022535 -7.0807 1.435e-12 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> rolling.forecast(datamts, 3, length(datamts)-10,c(1,0,1),seasonal = list(order = c(1, 1, 1)))
[1]  509.0062  928.0828 1654.9629
> Box.test(outf$residuals, type = 'Ljung',lag = 60)

            Box-Ljung test

data:  outf$residuals
X-squared = 60.075, df = 60, p-value = 0.473

> acf(outf$residuals,lag.max=60,main = "ACF of Residuals of Final Model")
> pacf(outf$residuals,lag.max=60, main = "PACF of Residuals of Final Model")
> fit = ts(fitted.values(outf), start = 2000, frequency = 12)
> #pdf('fit_0417.pdf',width=8,heigh t=4)
> par(mfrow = c(1,1))
> plot(datamts, ylim=c(min(datamts), max(fit)), xlab = 'Year', ylab = 'AQI', main = 'Air Pollution Index')
> points(fit, col = 'red', pch = 2)
> lines(fit, col = 'red', lty = 2)
> legend.text=c("Actual values", "Fitted values")
> legend("bottomleft", legend.text, lty = rep(1,2), col = 1:2, pch = 1:2)
> #pdf('pred_0417.pdf',height=5,width=8)
> time.pred = time(datamts)[118:length(datamts)]
> pp = predict(outf, 12)
Error in predict.Arima(outf, 12) :
```

```
  'xreg' and 'newxreg' have different numbers of columns
> pred = ts(pp$pred, start =2010 , frequency = 12)
> pred.upp = ts(pp$pred+2*pp$se, start = 2010, frequency = 12)
> pred.low = ts(pp$pred-2*pp$se, start = 2010, frequency = 12)
> plot(datamts, type = 'o',xlim = c(2000,2011), xlab = 'Year', ylab = 'AQI', main = 'Air Pollution Index')
> lines(pred, col = 'red')
> points(pred, col = 'red', pch = 2)
> lines(pred.low, col = 'red', lty = 2)
> lines(pred.upp, col = 'red', lty = 2)
> lines(datamts.new, col = 'blue', type = 'o')
> legend.text=c("Past Values", "Prediction","Actual Values")
> legend("bottomleft", legend.text, lty = rep(1,2,3), col = c(1,'red','blue'), pch = c(1,2,1))
>
> #dev.off()
>
```