Stats 326 Assignment 5

Hasnain Cheena

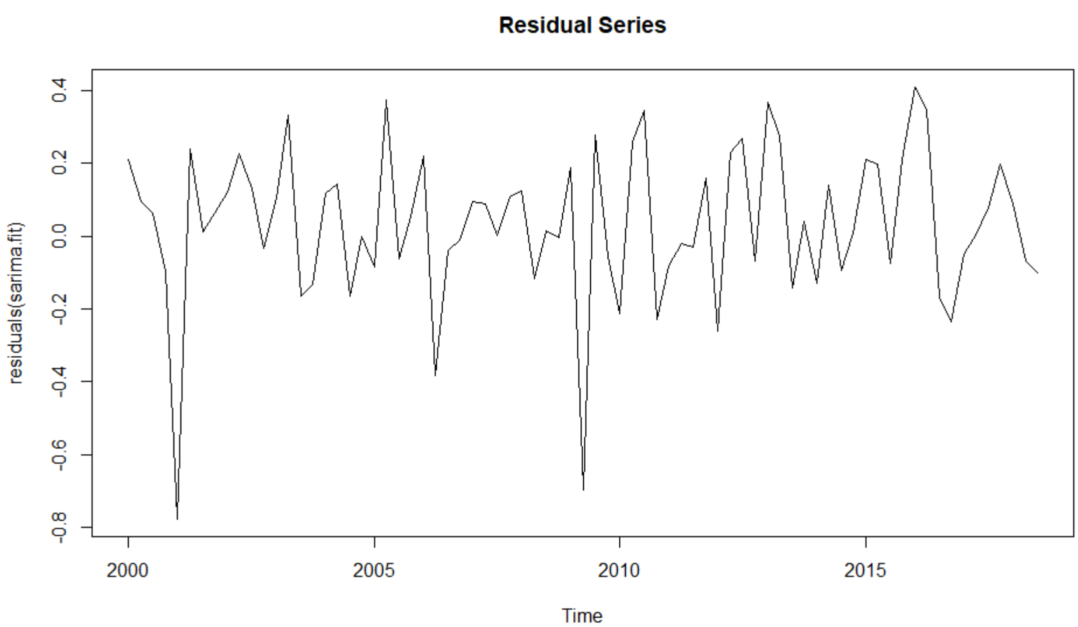
28/05/2020

Question 1

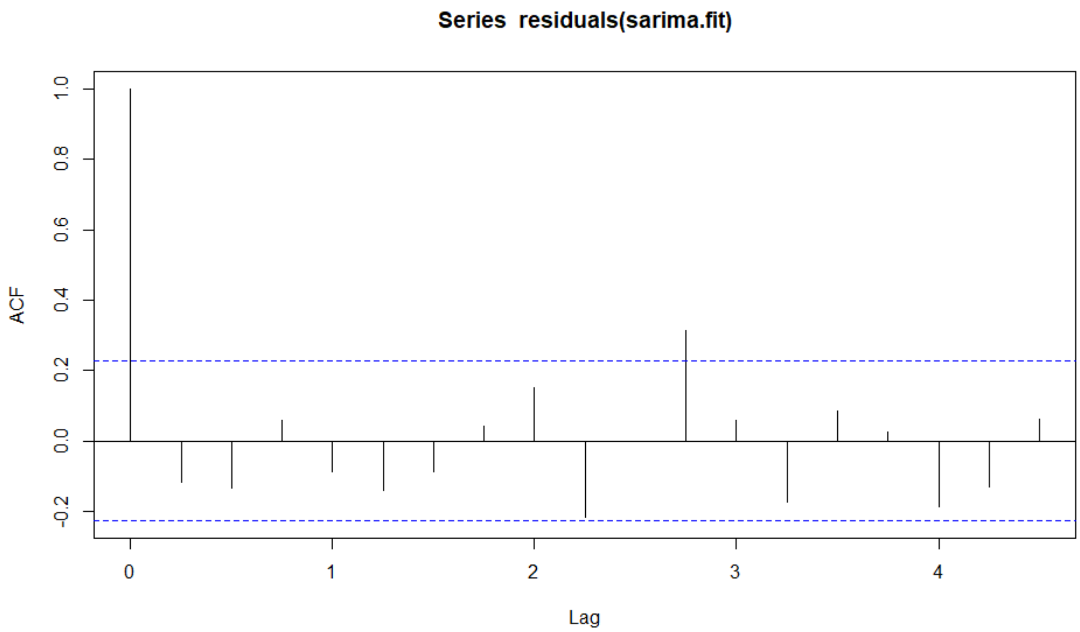
```
#best fitting SARIMA
sarima.fit = arima(red.CO2.ts,order=c(0,1,1),
seasonal=list(order=c(0,1,1),period=4))
sarima.fit

##
## Call:
## arima(x = red.CO2.ts, order = c(0, 1, 1), seasonal = list(order = c(0, 1,
1),
##     period = 4))
##
## Coefficients:
##          ma1     sma1
##       0.5764  -0.8992
## s.e.  0.1129   0.1326
##
## sigma^2 estimated as 0.03919:  log likelihood = 10.48,  aic = -14.96

plot.ts(residuals(sarima.fit),main="Residual Series")
```



```r
acf(residuals(sarima.fit))
```



The Residual Series show reasonably random scatter about 0, although there are two large negative residuals at Quarter 1 2001 and Quarter 2 2009. The plot of the autocorrelation function of the Residual Series shows a significant lag at lag 11. This is an unusual lag to be significant in quarterly data and thus not of concern. Therefore, all model assumptions are satisfied.

```r
#predictions
sarima.pred = predict(sarima.fit,n.ahead=4)
sarima.pred

## $pred
##          Qtr1     Qtr2     Qtr3     Qtr4
## 2018                            405.8136
## 2019 405.7907 406.3947 407.6694
##
## $se
##           Qtr1      Qtr2      Qtr3      Qtr4
## 2018                               0.1983784
## 2019 0.3702966 0.4845881 0.5766555

#actual
pred.CO2.ts

##        Qtr1   Qtr2   Qtr3   Qtr4
## 2018                       405.83
## 2019 405.73 406.71 408.25
```

```r
#RMSEP
RMSEP.sarima = sqrt(1/4*sum((pred.CO2.ts-sarima.pred$pred)^2))
RMSEP.sarima

## [1] 0.3318586
```

The predictions of the SARIMA model are relatively close to the actual values for Quarter 4 2019 and Quarter 1 2020. The predictions then drift drastically apart from the actual values for Quarter 3 2020 and Quarter 4 2020.

The $SARIMA(0,1,1) \times (0,1,1)_4$ model had an RMSEP of 0.33 ppm. The best predicting model from previous assignments was the seasonal-trend-lowess (STL) seasonally adjusted model. The STL model had a low RMSEP of 0.2 ppm. Therefore, the SARIMA model is not better predicting than the STL model as SARIMA has a higher RMSEP.

Question 2

```
#best fitting SARIMA with all data
sarima.fit.full = arima(full.CO2.ts,order=c(0,1,1),
seasonal=list(order=c(0,1,1),period=4))
sarima.fit.full

##
## Call:
## arima(x = full.CO2.ts, order = c(0, 1, 1), seasonal = list(order = c(0, 1,
1),
##     period = 4))
##
## Coefficients:
##          ma1     sma1
##       0.5614  -0.8524
## s.e.  0.1150   0.1008
##
## sigma^2 estimated as 0.04027:  log likelihood = 10.98,  aic = -15.97
```

Model in backshift notation:

$$(1-B)(1-B^4)y_t = (1+\alpha_1 B)(1+A_1 B^4)\varepsilon_t$$

$$(1-B-B^4+B^5)y_t = (1+\alpha_1 B + A_1 B^4 + \alpha_1 A_1 B^5)\varepsilon_t$$

$$y_t - y_{t-1} - y_{t-4} + y_{t-5} = \varepsilon_t + \alpha_1\varepsilon_{t-1} + A_1\varepsilon_{t-4} + \alpha_1 A_1\varepsilon_{t-5}$$

$$y_t = y_{t-1} + y_{t-4} - y_{t-5} + \varepsilon_t + \alpha_1\varepsilon_{t-1} + A_1\varepsilon_{t-4} + \alpha_1 A_1\varepsilon_{t-5}$$

$$y_t = y_{t-1} + y_{t-4} - y_{t-5} + \varepsilon_t + 0.5614\varepsilon_{t-1} - 0.8524\varepsilon_{t-4} - 0.4785\varepsilon_{t-5}$$

Predictions

Prediction 2019 Q4:

$$y_{t+1} = y_t + y_{t-3} - y_{t-4} + \varepsilon_{t+1} + 0.5614\varepsilon_t - 0.8524\varepsilon_{t-3} - 0.4785\varepsilon_{t-4}$$

```
#2019 Q4
pred.2019Q4 = 408.25 + 405.83 - 405.56 + (0.5614 * 0.031927800) - (0.8524 * 0
.009630730) - (0.4785 * -0.101088224)
```

Prediction 2020 Q1:

$$y_{t+2} = y_{t+1} + y_{t-2} - y_{t-3} + \varepsilon_{t+2} + 0.5614\varepsilon_{t+1} - 0.8524\varepsilon_{t-2} - 0.4785\varepsilon_{t-3}$$

```
#2020Q1
pred.2020Q1 = pred.2019Q4 + 405.73 - 405.83 - (0.8524 * -0.106133913) - (0.47
85 * 0.009630730)
pred.2020Q1
```

Prediction 2020 Q2:

$$y_{t+3} = y_{t+2} + y_{t-1} - y_{t-2} + \varepsilon_{t+3} + 0.5614\varepsilon_{t+2} - 0.8524\varepsilon_{t-1} - 0.4785\varepsilon_{t-2}$$

```
#2020Q2
pred.2020Q2 = pred.2020Q1 + 406.71 - 405.73 - (0.8524 * 0.402908803) - (0.478
5 * -0.106133913)
```

Prediction 2020 Q3:

$$y_{t+4} = y_{t+3} + y_{t} - y_{t-1} + \varepsilon_{t+4} + 0.5614\varepsilon_{t+3} - 0.8524\varepsilon_{t} - 0.4785\varepsilon_{t-1}$$

```
#2020Q3
pred.2020Q3 = pred.2020Q2 + 408.25 - 406.71 - (0.8524 * 0.031927800) - (0.478
5 * 0.402908803)
```

Prediction Results:

```
#results
results.df = data.frame(Time=c("2019.4", "2020.1", "2020.2", "2020.3"),
                           Predictions=c(round(pred.2019Q4,2),round(pred.2020
Q1,2),round(pred.2020Q2,2),round(pred.2020Q3,2)))

results.df

##     Time Predictions
## 1 2019.4      408.58
## 2 2020.1      408.56
## 3 2020.2      409.25
## 4 2020.3      410.57
```

Question 3 - Executive Summary

The task was to predict the atmospheric concentration of carbon dioxide (in parts per million) at Cape Grim, in Tasmania, Australia between 2019 Quarter 4 and 2020 Quarter 3.

Several different models were built using observations between 2000 Quarter 1 and 2018 Quarter 3. These models were then used to predict 2018 Quarter 4 to 2019 Quarter 3. Each model's predictions were then compared to the actual values to find the model that produced the most accurate predictions. The best predicting model found from this method was then re-run on all the available data. After re-fitting the model on all the data, predictions for 2019 Quarter 4 to 2020 Quarter 3 were produced.

Note that we need to be wary of our predictions as we have a time series with only 79 observations. However, the best predicting model that is used is a good model and therefore the predictions should be reasonably reliable.

We predict the carbon dioxide concentration in the atmosphere above Cape Grim in Tasmania, Australia will be:

2019 Quarter 4: 408.60 ppm
2020 Quarter 1: 408.61 ppm
2020 Quarter 2: 409.34 ppm
2020 Quarter 3: 410.76 ppm