Stats326 - Assignment3

Cameron Todd (5611230)

29 Jan 2018

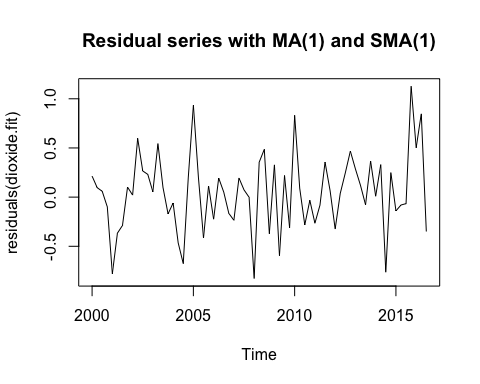
dioxide\_data = read.table("quarterly dioxide.2000.1.2017.3.txt", header = T)

Question 1

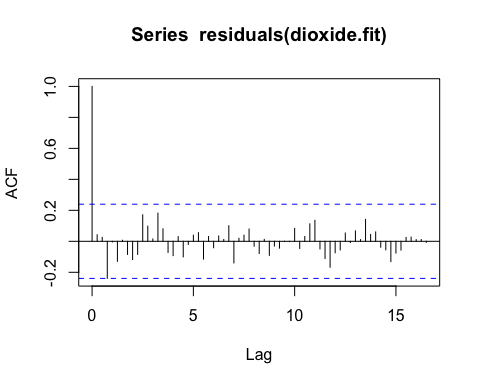
Time = 1:67  
Quarter = factor(c(rep(1:4,16),(1:3)))  
dioxide.ts = ts(dioxide\_data$dioxide[1:67],start=2000,frequency=4)  
  
dioxide.fit = arima(dioxide.ts, order = c(0,1,1), seasonal = list(order = c(0,1,1), period = 4))  
dioxide.fit

##   
## Call:  
## arima(x = dioxide.ts, order = c(0, 1, 1), seasonal = list(order = c(0, 1, 1),   
## period = 4))  
##   
## Coefficients:  
## ma1 sma1  
## -0.3444 -0.8198  
## s.e. 0.1367 0.1556  
##   
## sigma^2 estimated as 0.1551: log likelihood = -32.49, aic = 70.98

plot.ts(residuals(dioxide.fit), main= "Residual series with MA(1) and SMA(1)")



acf(residuals(dioxide.fit),lag.max = 67)



dioxide.predictions = predict(dioxide.fit, n.ahead = 4)  
dioxide.predictions

## $pred  
## Qtr1 Qtr2 Qtr3 Qtr4  
## 2016 404.4292  
## 2017 407.3150 408.7430 403.4886   
##   
## $se  
## Qtr1 Qtr2 Qtr3 Qtr4  
## 2016 0.3938820  
## 2017 0.4709804 0.5371132 0.5959518

actual.values = dioxide\_data$dioxide[68:71]  
actual.values

## [1] 404.42 407.18 408.84 403.38

RMSEP.dioxide.fit = sqrt(1/4\*sum((actual.values - dioxide.predictions$pred)^2))  
RMSEP.dioxide.fit

## [1] 0.099392

Comment on dioxide.fit –

The residual series appears to be random independent observations with reasonably constant scatter. Their appears to be slight increase in variability of the residual series towards the end. The autocorrelation plot of the residuals shows no significant lags.

Better predicting model from Ass2 -

As shown above the RMSEP for my SARIMA model is 0.099392 is slightly smaller then the RMSEP of assignment2 optimal model of 0.8533238, therefore making our SARIMA model a slightly better predicting model. I think the SARIMA model works better because it doesn’t just deal with the seasonality and trend component (like the seasonal factor model in Ass2) but it also deals with the underlying white noise in the data with an MA term.

Q2  
dioxide\_full.ts = ts(dioxide\_data$dioxide, start = 2000, frequency = 4)  
  
dioxide\_full.fit = arima(dioxide\_full.ts, order = c(0,1,1), seasonal = list(order = c(0,1,1), period = 4))  
dioxide\_full.fit

##   
## Call:  
## arima(x = dioxide\_full.ts, order = c(0, 1, 1), seasonal = list(order = c(0,   
## 1, 1), period = 4))  
##   
## Coefficients:  
## ma1 sma1  
## -0.3496 -0.8070  
## s.e. 0.1346 0.1173  
##   
## sigma^2 estimated as 0.1472: log likelihood = -32.62, aic = 71.23

Model Workings

*ARIMA(0,1,1)\*(0,1,1)4*

*⇒ (1 – B)(1 – B4)yt = (1 + α1B)(1 + A1B4) εt*

*⇒ yt – yt-1 - yt-4 + yt-5 = (1 + α1B + A1B4 + α1A1B5) εt*

*⇔ Yt = yt-1 + yt-2 – yt-3 + εt + α1εt-1 + A1εt-4 + α1 A1εt-5*

Predictions

*Yt+1 = yt + yt-3 – yt-4 + εt+1 + α1εt + A1εt-3 + α1 A1εt-4  where t = 71*

*⇒ Yt+1 = 403.38 + 404.42 – 401.03 + 0 + (-0.3496 \* -0.1314232706) + (-0.807 \* -0.0298380118) + (-0.3496 \* -0.807 \* -0.3382097126)*

*Y72 = 406.74*

*Yt+2 = yt+1 + yt-2 – yt-3 + εt+2 + α1εt+1 + A1εt-2 + α1 A1εt-3*

*⇒ Yt+2 = 406.74 + 407.18 – 404.42 + 0 + (-0.3496 \* 0) + (-0.807 \* -0.1349186971) + (-0.3496 \* -0.807 \* -0.0298380118)*

*Y73 = 409.6*

*Yt+3 = yt+2 + yt-1 – yt-2 + εt+3 + α1εt+2 + A1εt-1 + α1 A1εt-2*

*⇒ Yt+2 = 409.6 + 408.84 – 407.18 + 0 + (-0.3496 \* 0) + (-0.807 \* 0.1736324853) + (-0.3496 \* -0.807 \* -0.1349186971)*

*Y74 = 411.08*

*Yt+4 = yt+3 + yt – yt-1 + εt+4 + α1εt+3 + A1εt + α1 A1εt-1*

*⇒ Yt+2 = 411.08 + 403.38 – 408.84+ 0 + (-0.3496 \* 0) + (-0.807 \* -0.1314232706) + (-0.3496 \* -0.807 \* 0.1736324853)*

*Y75 = 405.78*

Executive Summary –

We were interested in finding the best predicting SARIMA model and compare its predictions with those from the seasonal factor model produced in assignment 2 and to also use the SARIMA model to generate the predictions of carbon dioxide (ppm) for the next year.

I found the best predicting SARIMA model by already knowing the data we knew their was a strong trend and seasonal component so I differenced it twice to count for the seasonal trend component i.e. SARIMA(0,1,0)\*(0,1,0). I then added a AR term and MA term separately and checked to see if the AIC was reduced and if the AR and MA terms were significant. I also tried adding all AR and MA terms i.e. SARIMA(1,1,1)\*(1,1,1) this model was suprisingly worse with a larger AIC then my final model of SARIMA(0,1,1)\*(0,1,1) with an AIC of 70.98

From the below table you can see how I have compared RMSEP values for our SARIMA model and seasonal factor model from assignment 2, calculated by comparing our predicted values for 2016.4 – 2017.3 vs the actual values. You can see that the RMSEP for the SARIMA model is slightly lower therefore making it the better model to use for predictions. You can also see in the below table we have made predictions of quarterly carbon dioxide in parts per million (ppm) for the next 4 quarters from 2017.4 – 2018.3, using the SARIMA model.

|  |  |  |
| --- | --- | --- |
| Model Summary | SARIMA(0,1,1)\*(0,1,1) | SFM (Ass2) |
| 2017.4 | 406.74 | 406.30 |
| 2018.1 | 409.6 | 409.06 |
| 2018.2 | 411.08 | 410.30 |
| 2018.3 | 405.78 | 405.09 |
| RMSEP (calc from 2016.4 - 2017.3 actual vs pred) | 0.099392 | 0.8533238 |