STA457_project

prajwol poudel

08/04/2023

```
library(fUnitRoots)
library(caTools)
library(tseries)
library(TSA)
library(psych)
setwd("C:\\Users\\poude\\Desktop\\time_series project")
df = read.csv("Geomagnetic_Intensity_Data.csv")
```

Including Plots

You can also embed plots, for example:

```
x = new_train$North_Geomagnetic_Pole
cummeanx <- cumsum(x)/ seq_along(x)
par(mfrow=c(2,1), mar = c(4,4,4,4))
plot(x, type="l", xlab = "Time",
     ylab="Time series of index values for north", main = "Realization of values")
plot(cummeanx, type = "l", xlab="Time", ylab= "Mean level")
acf(x,xlab= "Lag",ylab = "Sample ACF", main ="")
acf(x,type="partial",xlab = "Lag", ylab= "Sample PACF", main="")
x.adf <- adfTest(x, lags = 17)
x.adf2 <- adfTest(x, lags = 8)
x.adf3 <- adfTest(x, lags = 3)
x.kpss_level <- kpss.test(x, null = "Level")
x.kpss_trend <- kpss.test(x, null = "Trend")
x.adf
x.adf2
x.adf3
x.kpss_level
x.kpss_trend
```

```
x.diff2 = diff(diff(x))
cummeanx.diff2 <- cumsum(x.diff2)/ seq_along(x.diff2)
par(mfrow=c(2,1), mar = c(4,4,4,4))
plot(x.diff2, type="l", xlab = "Time",
     ylab="Time series of index values", main = "Realization of values")
```

```
plot(cummeanx.diff2, type = "l", xlab="Time", ylab= "Mean level")
acf(x.diff2,xlab= "Lag",ylab = "Sample ACF", main ="")
acf(x.diff2,type="partial",xlab = "Lag", ylab= "Sample PACF", main="")
x.diff2.adf <- adfTest(x.diff2, lags = 17)
x.diff2.adf2 <- adfTest(x.diff2, lags = 4)
x.diff2.adf3 <- adfTest(x.diff2, lags = 8)
x.diff2.kpss_level <- kpss.test(x.diff2, null = "Level")
x.diff2.kpss_trend <- kpss.test(x.diff2, null = "Trend")
x.diff2.adf
x.diff2.adf2
x.diff2.adf3
x.diff2.kpss_level
x.diff2.kpss_trend
```

North From the ACF diagram we can see a decay signature similar to an AR(1)

```
x.diff2.eacf <-eacf(x.diff2,ar.max=5,ma.max=5)
```

```
x2 <- x.diff2
x2.aic <- matrix(0,5,5)
x2.bic <- matrix(0,5,5)

for (i in 0:4) for (j in 0:4){
  library(gdata)
  x2.fit <- arima(x2, order=c(i,0,j), method = "ML",  include.mean = TRUE)
  x2.aic[i+1,j+1] <-x2.fit$aic
  x2.bic[i+1,j+1] <- BIC(x2.fit)
}

x2.aic_vec <- sort(unmatrix(x2.aic, byrow=FALSE))[1:13]
x2.bic_vec <- sort(unmatrix(x2.bic, byrow = FALSE))[1:13]

x2.aic_vec
x2.bic_vec
```

Checking the model with the lowest AIC and BIC: ARMA(3,5) model

```
x2.fit_1 <- arima(x2, order=c(3,0,5), method="ML",include.mean=TRUE)
tsdiag(x2.fit_1)
qqnorm(residuals(x2.fit_1))
qqline(residuals(x2.fit_1))
shapiro.test(residuals(x2.fit_1))
harmonic.mean(x2.fit_1$residuals)
```

From the residual plot the errors look to be uncorrelated but There is some significant acf values in the residual acf plot at lag 16.

checking if the residuals are normal by using a Q-Q plot and a shapiro-wilk test

```
x2.fit_1 <- arima(x2, order=c(3,2,5), method="ML",include.mean=TRUE)
tsdiag(x2.fit_1)
qqnorm(residuals(x2.fit_1))
```

```
qqline(residuals(x2.fit_1))
shapiro.test(residuals(x2.fit_1))
harmonic.mean(x2.fit_1$residuals)
```

```
x2.fit_1 <- arima(x2, order=c(3,1,5), method="ML",include.mean=TRUE)
tsdiag(x2.fit_1)
qqnorm(residuals(x2.fit_1))
qqline(residuals(x2.fit_1))
shapiro.test(residuals(x2.fit_1))
harmonic.mean(x2.fit_1$residuals)
BIC(x2.fit_1)
```

```
x2.fit_1 <- arima(x2, order=c(2,1,2), method="ML",include.mean=TRUE)
tsdiag(x2.fit_1)
qqnorm(residuals(x2.fit_1))
qqline(residuals(x2.fit_1))
shapiro.test(residuals(x2.fit_1))
harmonic.mean(x2.fit_1$residuals)
```

```
x2.fit_1 <- arima(x2, order=c(5,2,4), method="ML",include.mean=TRUE)
tsdiag(x2.fit_1)
qqnorm(residuals(x2.fit_1))
qqline(residuals(x2.fit_1))
shapiro.test(residuals(x2.fit_1))
harmonic.mean(x2.fit_1$residuals)
BIC(x2.fit_1)
```

```
x2.fit_1 <- arima(x2, order=c(4,1,5), method="ML",include.mean=TRUE)
tsdiag(x2.fit_1)
qqnorm(residuals(x2.fit_1))
qqline(residuals(x2.fit_1))
shapiro.test(residuals(x2.fit_1))
harmonic.mean(x2.fit_1$residuals)
BIC(x2.fit_1)
```

```
#candidate model
arima(x2, order=c(3,1,5), method="ML",include.mean=TRUE)
arima(x2, order=c(5,2,4), method="ML",include.mean=TRUE)
arima(x2, order=c(4,1,5), method="ML",include.mean=TRUE)
```

```
x = test$North_Geomagnetic_Pole
test.diff = diff(diff(x))
test.fit = arima(test.diff,order=c(4,1,5), method="ML",include.mean=TRUE)
plot(test.fit,n.ahead=5,type='b',xlab='Time',
ylab='Geomagnetic strength')
abline(h=coef(test.fit)[names(coef(test.fit))=='intercept'])
prediction<- predict(test.fit,n.ahead = 10)
pred <- prediction$pred
test
library(stats)
forecast<- diffinv(pred, xi = x[44])
forecast
```

```
RMSE <- sqrt(sum((test.fit$residuals-mean(test.fit$residuals))**2)/nrow(test))
RMSE
plot(x,type = "l",xlab='Time',ylab='Geomagnetic strength', xlim = c(0,50)) +
  lines(forecast,col="blue", type = "b")
```

```
y = train$South_Geomagnetic_Pole
cummeany <- cumsum(y)/ seq_along(y)
par(mfrow=c(2,1), mar = c(4,4,4,4))
plot(y, type="l", xlab = "Time",
     ylab="Time series of index values for south", main = "Realization of values")
plot(cummeany, type = "l", xlab="Time", ylab= "Mean level")
acf(y,xlab= "Lag",ylab = "Sample ACF", main ="")
acf(y,type="partial",xlab = "Lag", ylab= "Sample PACF", main="")
y.adf <- adfTest(y, lags = 17)
y.adf2 <- adfTest(y, lags = 4)
y.adf3 <- adfTest(y, lags = 8)
y.kpss_level <- kpss.test(y, null = "Level")
y.kpss_trend <- kpss.test(y, null = "Trend")
y.adf
y.adf2
y.adf3
y.kpss_level
y.kpss_trend
```

```
library(astsa)
y.diff2 = diff(diff(y))
#y.diff2 = detrend(y, lowess = TRUE)
cummeany.diff2 <- cumsum(y.diff2)/ seq_along(y.diff2)
par(mfrow=c(2,1), mar = c(4,4,4,4))
plot(y.diff2, type="l", xlab = "Time",
     ylab="Time series of index values", main = "Realization of values")
plot(cummeany.diff2, type = "l", xlab="Time", ylab= "Mean level")
acf(y.diff2,xlab= "Lag",ylab = "Sample ACF", main ="")
acf(y.diff2,type="partial",xlab = "Lag", ylab= "Sample PACF", main="")
y.diff2.adf <- adfTest(y.diff2, lags = 17)
y.diff2.adf2 <- adfTest(y.diff2, lags = 4)
y.diff2.adf3 <- adfTest(y.diff2, lags = 8)
y.diff2.kpss_level <- kpss.test(y.diff2, null = "Level")
y.diff2.kpss_trend <- kpss.test(y.diff2, null = "Trend")
y.diff2.adf
y.diff2.kpss_level
y.diff2.kpss_trend
```

```
y.diff2.eacf <-eacf(y.diff2,ar.max=5,ma.max=5)
```

```r
y2 <- y.diff2
y2.aic <- matrix(0,5,5)
y2.bic <- matrix(0,5,5)

for (i in 0:4) for (j in 0:4){
  library(gdata)
  y2.fit <- arima(y2, order=c(i,0,j), method = "ML",  include.mean = TRUE)
  y2.aic[i+1,j+1] <-y2.fit$aic
  y2.bic[i+1,j+1] <- BIC(y2.fit)
}

y2.aic_vec <- sort(unmatrix(y2.aic, byrow=FALSE))[1:13]
y2.bic_vec <- sort(unmatrix(y2.bic, byrow = FALSE))[1:13]

y2.aic_vec
y2.bic_vec
```

```r
y.fit_1 <- arima(y2, order=c(5,0,5), method="ML",include.mean=TRUE)
tsdiag(y.fit_1)
qqnorm(residuals(y.fit_1))
qqline(residuals(y.fit_1))
shapiro.test(residuals(y.fit_1))
harmonic.mean(y.fit_1$residuals)
```

```r
y.fit_1 <- arima(y2, order=c(5,1,5), method="ML",include.mean=TRUE)
tsdiag(y.fit_1)
qqnorm(residuals(y.fit_1))
qqline(residuals(y.fit_1))
shapiro.test(residuals(y.fit_1))
harmonic.mean(y.fit_1$residuals)
```

```r
y.fit_1 <- arima(y2, order=c(5,2,5), method="ML",include.mean=TRUE)
tsdiag(y.fit_1)
qqnorm(residuals(y.fit_1))
qqline(residuals(y.fit_1))
shapiro.test(residuals(y.fit_1))
harmonic.mean(y.fit_1$residuals)
BIC(y.fit_1)
```

```r
y.fit_1 <- arima(y2, order=c(3,0,5), method="ML",include.mean=TRUE)
tsdiag(y.fit_1)
qqnorm(residuals(y.fit_1))
qqline(residuals(y.fit_1))
shapiro.test(residuals(y.fit_1))
harmonic.mean(y.fit_1$residuals)
```

```
y.fit_1 <- arima(y2, order=c(3,1,5), method="ML",include.mean=TRUE)
tsdiag(y.fit_1)
qqnorm(residuals(y.fit_1))
qqline(residuals(y.fit_1))
shapiro.test(residuals(y.fit_1))
harmonic.mean(y.fit_1$residuals)
```

```
y.fit_1 <- arima(y2, order=c(5,0,4), method="ML",include.mean=TRUE)
tsdiag(y.fit_1)
qqnorm(residuals(y.fit_1))
qqline(residuals(y.fit_1))
shapiro.test(residuals(y.fit_1))
harmonic.mean(y.fit_1$residuals)
```

```
y.fit_1 <- arima(y2, order=c(5,1,4), method="ML",include.mean=TRUE)
tsdiag(y.fit_1)
qqnorm(residuals(y.fit_1))
qqline(residuals(y.fit_1))
shapiro.test(residuals(y.fit_1))
harmonic.mean(y.fit_1$residuals)
```

```
y.fit_1 <- arima(y2, order=c(4,0,5), method="ML",include.mean=TRUE)
tsdiag(y.fit_1)
qqnorm(residuals(y.fit_1))
qqline(residuals(y.fit_1))
shapiro.test(residuals(y.fit_1))
harmonic.mean(y.fit_1$residuals)
```

```
y.fit_1 <- arima(y2, order=c(4,1,5), method="ML",include.mean=TRUE)
tsdiag(y.fit_1)
qqnorm(residuals(y.fit_1))
qqline(residuals(y.fit_1))
shapiro.test(residuals(y.fit_1))
harmonic.mean(y.fit_1$residuals)
```

```
library(psych)
arima(y2, order=c(5,0,5), method="ML",include.mean=TRUE) #candidate model
arima(y2, order=c(5,0,4), method="ML",include.mean=TRUE)
arima(y2, order=c(5,2,5), method="ML",include.mean=TRUE)
arima(y2, order=c(5,1,5), method="ML",include.mean=TRUE)
```

```r
y = test$South_Geomagnetic_Pole
test.diff2 = diff(diff(y))
test.fit2 = arima(test.diff2,order=c(5,0,4), method="ML",include.mean=TRUE)
plot(test.fit2,n.ahead=10,type='b',xlab='Time',
ylab='Geomagnetic strength')
abline(h=coef(test.fit2)[names(coef(test.fit2))=='intercept'])
prediction<- predict(test.fit2,n.ahead = 10)
pred <- prediction$pred
test
library(stats)
forecast <- diffinv(pred, xi = y[44])
forecast
```

```r
RMSE <- sqrt(sum((test.fit2$residuals-mean(test.fit2$residuals))**2)/nrow(test))
RMSE
plot(y,type = "l",xlab='Time',ylab='Geomagnetic strength', xlim= c(0,50))
+ lines(forecast,col="blue", type = "b")
```

Southern data