# pstat274_hw05_aoxu

AO XU

2022-11-01

# 1. A glossary of R-commands for time series.

### define working directory

getwd()

### read and plot data

read.csv()

ts.plot()

### simulate and plot ARMA models

$x_1$ <- arima.sim(n = 100,model = list(ma=c($\theta_1$,$\theta_2$)))

plot(X1)

### add trend and mean line to the original time series plot

abline(lm(data~as.numeric(1:length(data))))

abline(lm(data~as.numeric(1:length(data))))

### calculate and plot theoretical acf/pacf for ARMA models

plot(ARMAacf(ar=. . . ,ma=. . . ,lag.max=. . . ,pacf=T/F))

### calculate and plot sample acf/pacf

acf or pacf(model,lag.max=. . . , plot=T/F)

### check whether a particular model is causal/invertible (R commands to find and plot roots of polynomials)

polyroot(c(. . . ))

### perform Box-Cox transforms

boxcox(data~as.numeric(1:length(data)) # library(MASS)

### perform differencing data at lags 1 and 12

diff(x,lag = 1) diff(x, lag = 12)

### perform Yule-Walker estimation and find standard deviations of the estimates

data.ywest<- arima(data, aic = TRUE, order.max = NULL, method=c(“yule-walker”))

sqrt(data.ywest$var.pred)

### perform MLE and check AICC associated with the model

arima(data, aic = TRUE, order.max = NULL, method = c(“mle”))

AICc()

# 2. Choose a dataset

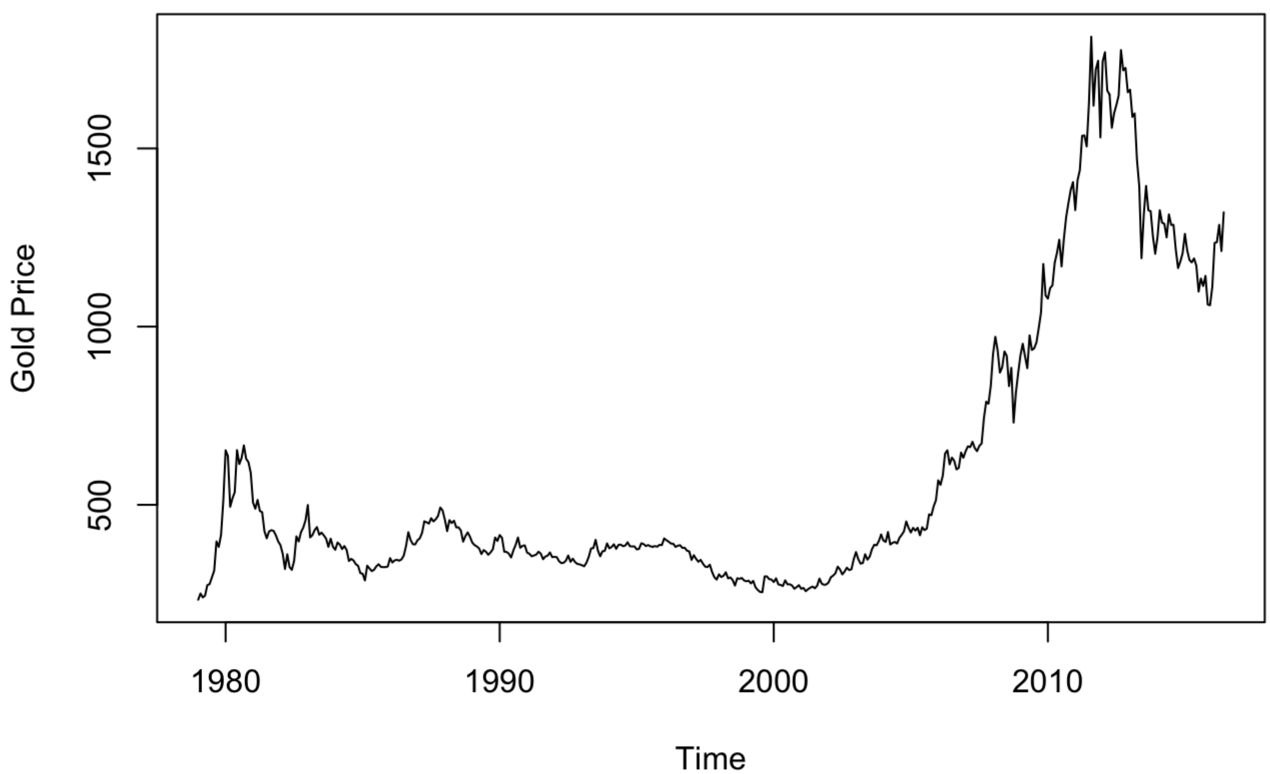
a. It’s about gold price monthly. Contains prices in respective currencies for each country monthly from January 1979 to July 2021.

b. It’s interesting since it’s well-known that people like to invest on gold and it could maintain the wealth even in such a high-inflation world. I want to forcast how gold price would change in the next five years or 10 years.

c.

```
library(ggplot2)
data <- read.csv("gold.csv")
```

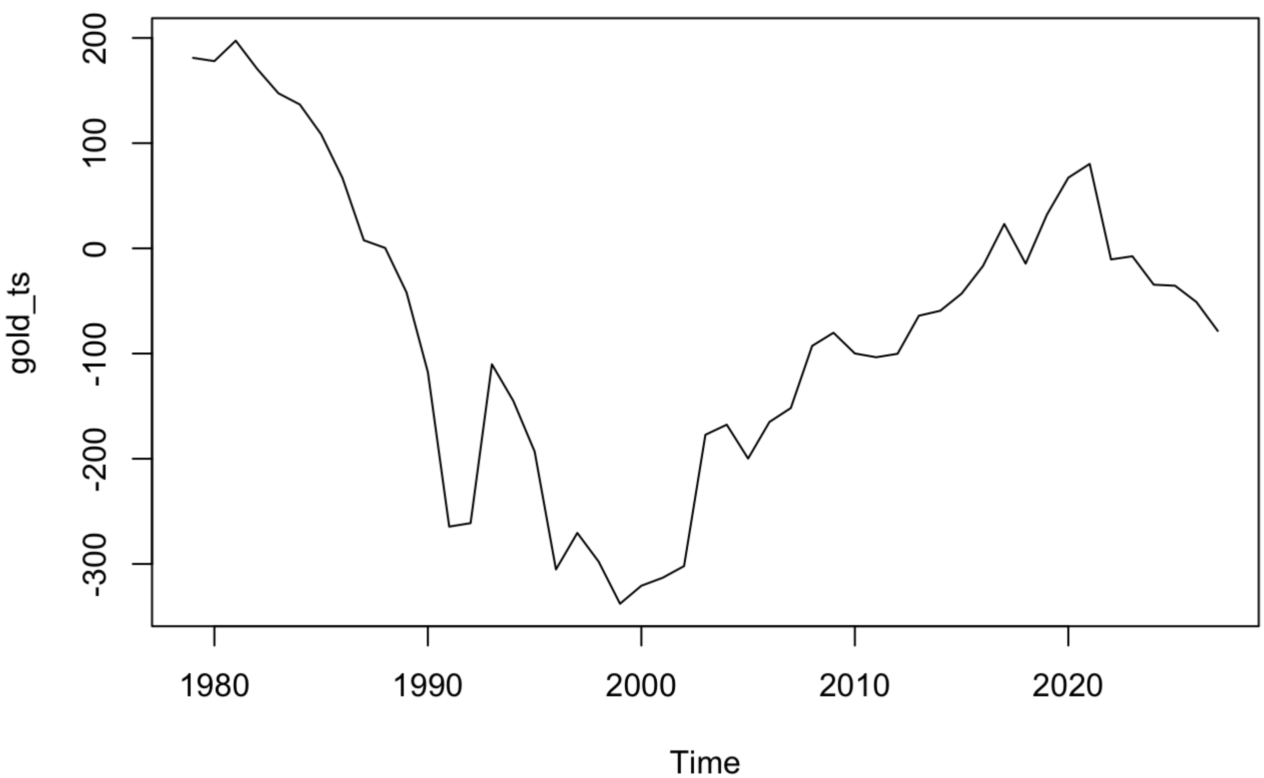
```
train <- data[1:450,2]
test <- data[451:511,2]
mydata <- ts(train,start=c(1979,1),frequency = 12)
plot(mydata,ylab="Gold Price")
```




From the graph, we could see that (i) no trend between 1980 and 2000, but a positive trend between 2000 and 2021 (ii) no seasonal component (iii) we could see that there is an apparent sharp change between 2000 and 2010.

d.

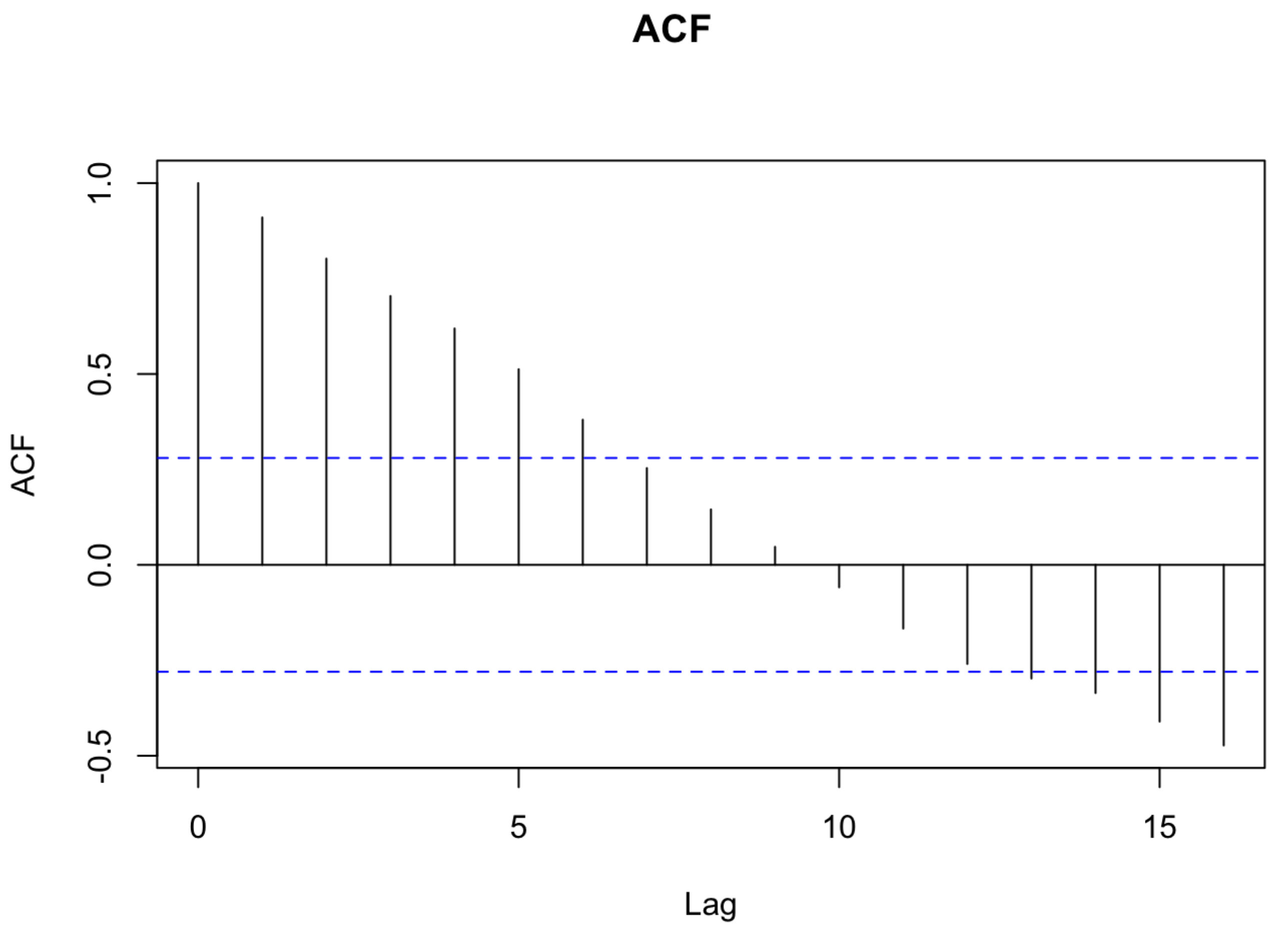
```
gold_diff <- diff(mydata,50)
gold_ts <- ts(gold_diff,start = c(1979,1), end = c(2021,7))
plot.ts(gold_ts,type="l")
```



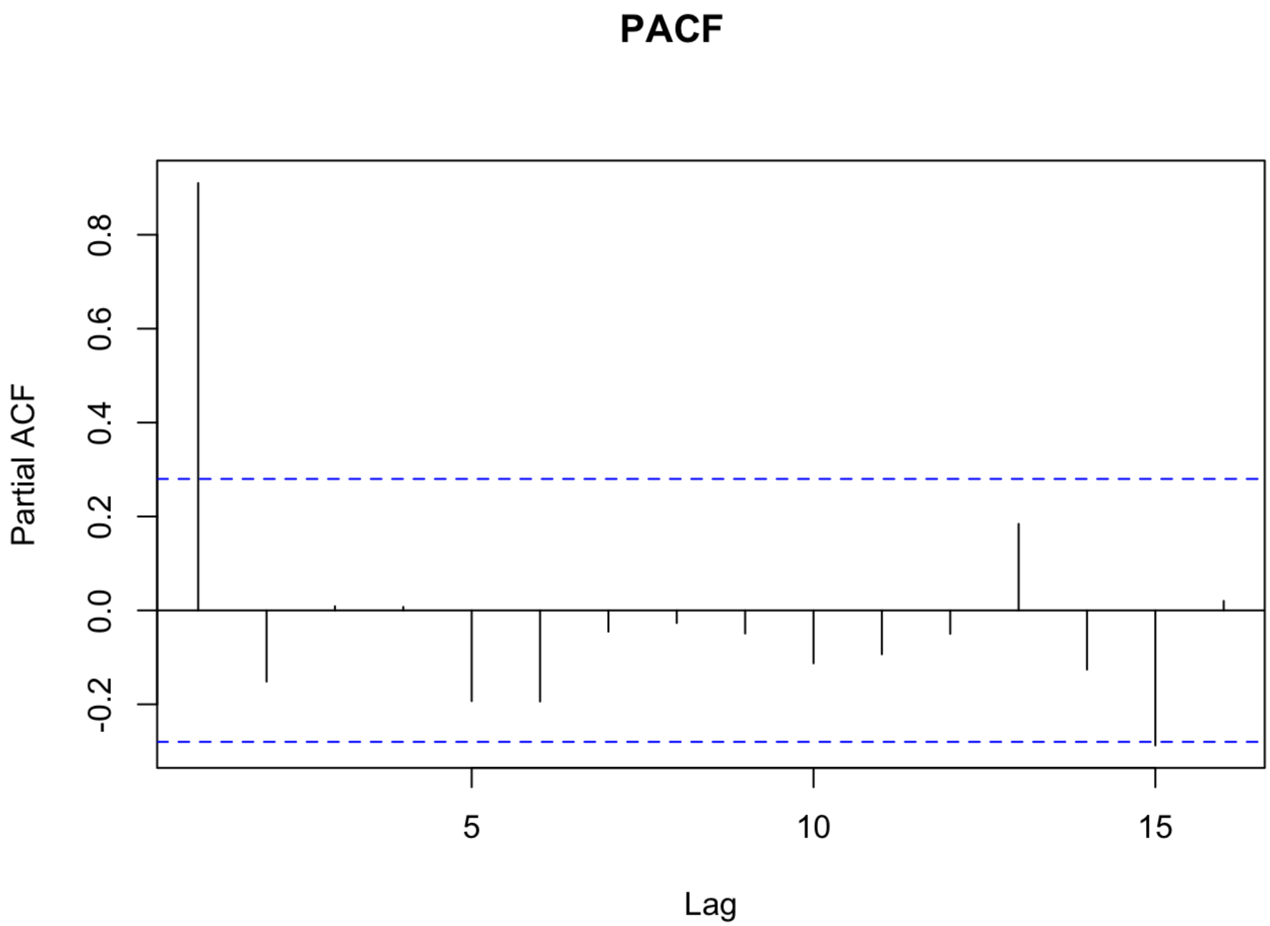

used lag = 50 to difference the data to get stationary series. The transformed data is staionary now since it follows the stationary distribution.

e.

```
acf(gold_ts,main="")
title("ACF",line=-1,outer=TRUE)
```




```
pacf(gold_ts,main="")
title("PACF",line=-1,outer=TRUE)
```




From ACF, we could get that it do not follow pattern of MA(q) and should be AR model.

From PACF, we could that it should be AR(1) model.

# 3.

Option A.

```
X1 = 2.91
X2 = 0.98
X3 = 3.53
Xt = 2.637 + 0.252*(X1 - 2.637) + 0.061*(X2 - 2.637) - 0.202*(X3 - 2.637) + 0
Xt
```

```
## [1] 2.424333
```

Since 2.424333 < 3.0, then we choose A.

# 4.

AR(1) model with mean 0: $\rho(2)$ = 0.215, $\rho(3)$ = –0.100, $X_T$ = –0.431

Solution:

Since $\rho(3)$ = –0.100, then $\phi < 0$.

```
phi = -sqrt(0.215)
XT = -0.431
XTplus1 = XT*phi
XTplus1
```

```
## [1] 0.1998465
```

# 5.

The five models, AR(1), ARMA(1, 1), ARMA(1, 2), ARMA(2, 3), and ARMA(4, 3) are fitted to the same time series. The models are ranked using Akaike Information Criterion (AIC): AIC = −2× log-likelihood + 2 ×(p + q + 2).

```
AR1 = -650
ARMA11 = -641
ARMA12 = -636
ARMA23 = -630
ARMA43 = -629
AIC_AR1 = -2 * AR1 + 2*(1+0+2)
AIC_ARMA11 = -2 * ARMA11 + 2*(1+1+2)
AIC_ARMA12 = -2 * ARMA12 + 2*(1+2+2)
AIC_ARMA23 = -2 * ARMA23 + 2*(2+3+2)
AIC_ARMA43 = -2 * ARMA43 + 2*(4+3+2)
AIC_AR1
```

```
## [1] 1306
```

```
AIC_ARMA11
```

```
## [1] 1290
```

```
AIC_ARMA12
```

```
## [1] 1282
```

```
AIC_ARMA23
```

```
## [1] 1274
```

```
AIC_ARMA43
```

```
## [1] 1276
```

Therefore, ARMA(2,3) is the best model, which has the lowest AIC 1274.

# G1

$w_{11}$ = 1 − 3$\rho(1)^2$ + 4$\rho(1)^4$

$w_{ii}$ = 1 + 2$\rho(1)^2$ for i > 1

$\hat{\rho}(1)$ = 0.438

$\hat{\rho}(2)$ = 0.145

95% CI for $\rho(1)$: $\frac{|\hat{\rho}(1)-\rho(1)|}{\sqrt{\frac{1-3\rho(1)^2+4\rho(1)^4}{100}}} < 1.96$

$\theta = 0.6, \rho(1) = \frac{\theta}{1+\theta^2}$ = $\frac{15}{34}$

$|\hat{\rho}(1) - \frac{15}{34}|$ < 1.96 * $\sqrt{\frac{1-3(\frac{15}{34})^2+4(\frac{15}{34})^4}{100}}$

$\Rightarrow 0.294 < \hat{\rho}(1) < 0.589$

95% CI for $\rho(2)$: $\frac{|(\hat{\rho}2)-\rho(2)|}{\sqrt{\frac{1+2\rho(1)^2}{100}}} < 1.96$

$|\hat{\rho}(1) - 0|$ < 1.96 * $\sqrt{\frac{1+2(\frac{15}{34})^2}{100}}$

$\Rightarrow -0.23 < \hat{\rho}(2) < 0.23$

Since 0.294 < $\hat{\rho}(1)$ = 0.438 < 0.589, 0.23 < $\hat{\rho}(2)$ = 0.145 < 0.23,

then we could get that the data are consistent with an MA(1) model with θ = 0.6