Times Series Assignment 3

12111603 Tan Zhiheng

2023-11-10

Question 1

(a)

- $\mathbb{E}X_t = -2t + \mathbb{E}(W_t + 0.5W_{t-1}) = -2t.$
- $$\begin{aligned}\gamma(t+h,t) = Cov(X_{t+h}, X_t) &= Cov(-2t - 2h + W_{t+h} + 0.5W_{t+h-1}, -2t + W_t + 0.5W_{t-1}) \\ &= \begin{cases} 1.25\sigma^2, & h = 0 \\ 0.5\sigma^2, & h = \pm 1 \\ 0, & else \end{cases}\end{aligned}$$

The times series is non-stationary because the mean function is related to time t.

(b)

We can obtain that

$$Y_t = \nabla X_t = X_t - x_{t-1} = -2 + W_t - 0.5W_{t-1} - 0.5W_{t-2}.$$

Thus,

- $\mathbb{E}Y_t = -2 + \mathbb{E}(W_t - 0.5W_{t-1} - 0.5W_{t-2}) = -2.$
- $$\begin{aligned}\gamma(t+h,t) = Cov(Y_{t+h}, Y_t) &= Cov(-2 + W_{t+h} - 0.5W_{t+h-1} - 0.5W_{t+h-2}, -2 + W_t - 0.5W_{t-1} - 0.5W_{t-2}) \\ &= \begin{cases} 1.5\sigma^2, & h = 0 \\ -0.25\sigma^2, & h = \pm 1 \\ -0.5\sigma^2, & h = \pm 2 \\ 0, & else \end{cases}\end{aligned}$$

The times series Y_t is stationary because it is a $MA(2)$ with shift, whose characteristic function has no roots inside the unit ball on the complex plane.

Question 2

Rewrite the times series as follows:

$$\phi(B)X_t = \theta(B)W_t$$

where $\phi(z) = 1 - z + 0.25z^2$ and $\theta(z) = 1 - 0.25z$.

(a)

Since the roots of $\phi(z)$ are 2's, whose modules are greater than 1, then we can say the times series are casual.

(b)

By the formula $\psi_j = \theta_j + \phi_1\psi_{j-1} + \cdots + \phi_p\psi_{j-p}$, we can figure out that

- $\psi_1 = 0.75$
- $\psi_2 = 0.5$
- $\psi_k = \psi_{k-1} - 0.25\psi_{k-2}$ for all $k \geq 3$

Question 3

Rewrite the times series as follows:

$$\phi(B)X_t = \theta(B)W_t$$

where $\phi(z) = 1 - z + 0.25z^2$ and $\theta(z) = 1 - 0.25z$.

(a)

Since the root of $\theta(z)$ is 4, whose module is greater than 1, then we can say the times series are invertible.

(b)

Since \$(B)X_t = (B)W_t \$ and θ is invertible, then we have

$$W_t = \theta(B)^{-1}\phi(B)X_t = (\theta^{-1}\phi)(B)X_t.$$

Because \$\$

$$\begin{aligned}\theta^{-1}(z)\phi(z) &= \frac{1}{1-0.25z}(1-0.5z)^2 \\ &= (1 + \frac{1}{4}z + (\frac{1}{4}z)^2 + \cdots)(1 - z + \frac{1}{4}z^2)\end{aligned}$$

\$\$

Then we have

- $\pi_1 = -0.75$
- $\pi_2 = \frac{1}{4^2}$
- $\pi_3 = \frac{1}{4^3}$
- $\pi_k = \frac{1}{4^k}$ for all $k \geq 2$

Question 4

Since $h = 10 > p = 3$ in the $AR(p = 3)$ model, then we maintain that the best linear prediction of X_{11} becomes

$$X_{11}^{10} = Proj(X_{11}|X_1, \cdots, X_{10}) = \sum_{j=1}^{3} \phi_j X_{11-j} = \phi_1 X_{10} + \phi_2 X_9 + \phi_3 X_8.$$

Given $x_{10} = -0.74$, $x_9 = -3.5$ and $x_8 = 3.0$, then by the equation above we can make a prediction of X_{11} is $\hat{X}_{11} = 2.352$.

Question 5

Suppose $X_t = \phi X_{t-1} + W_t$, which is an $AR(1)$ model and assume that $|\phi| \neq 1$.

Let Y be the linear prediction of X_2 given X_1 and X_3, then Y can be rewritten as the form of $Y = \alpha X_1 + \beta X_3$ where α and β are two constants to be solved.

Since the best linear prediction is essencially a projection, then we can say that the difference between the variable to be predicted and its projection is perpendicular to the plane spanned by given variables, which means

$$(X_2 - Y) \perp Span\{X_1, X_3\}.$$

Thus, we can establish our equation systems as follows:

$$\begin{cases} Cov(X_1, X_2 - Y) = \mathbb{E}[X_1(X_2 - Y)] = \gamma(1) - \alpha\gamma(0) - \beta\gamma(2) = 0 \\ Cov(X_3, X_2 - Y) = \mathbb{E}[X_3(X_2 - Y)] = \gamma(1) - \alpha\gamma(2) - \beta\gamma(0) = 0 \end{cases}$$

Invoke that $\gamma(h) = \frac{\phi^h}{1-\phi^2}$ in $AR(1)$ model, we can solve the equation systems above to obtain that $\alpha = \beta = \frac{\phi}{1+\phi^2}$.

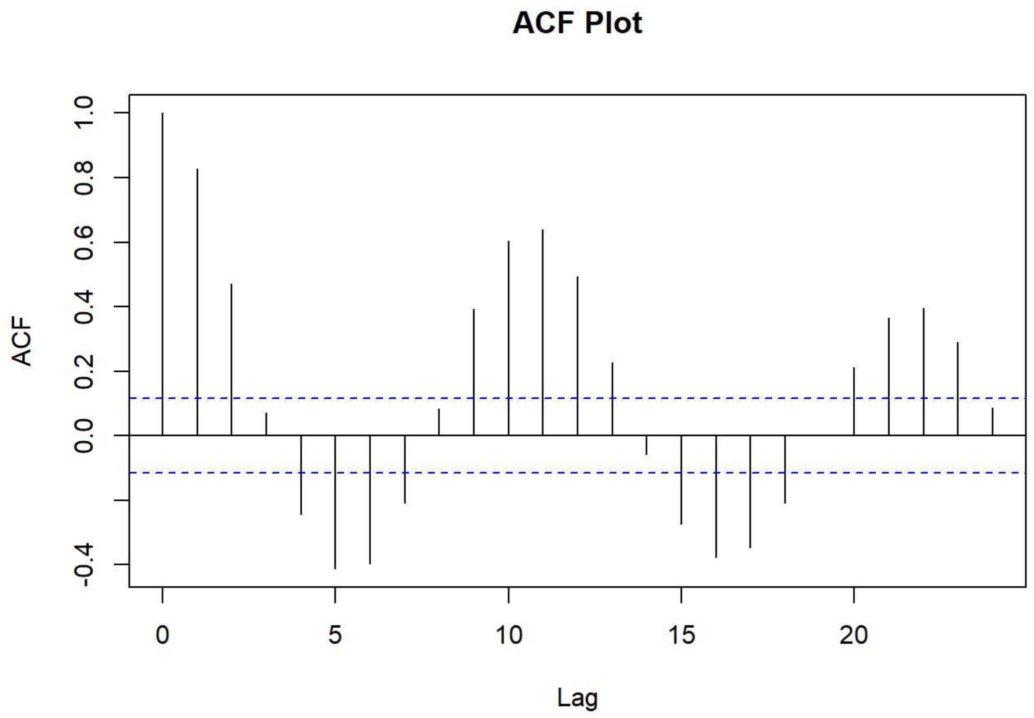
Consequently, the best linear prediction of X_2 given X_1 and X_3 is

$$\frac{\phi}{1+\phi^2}(X_1 + X_3).$$

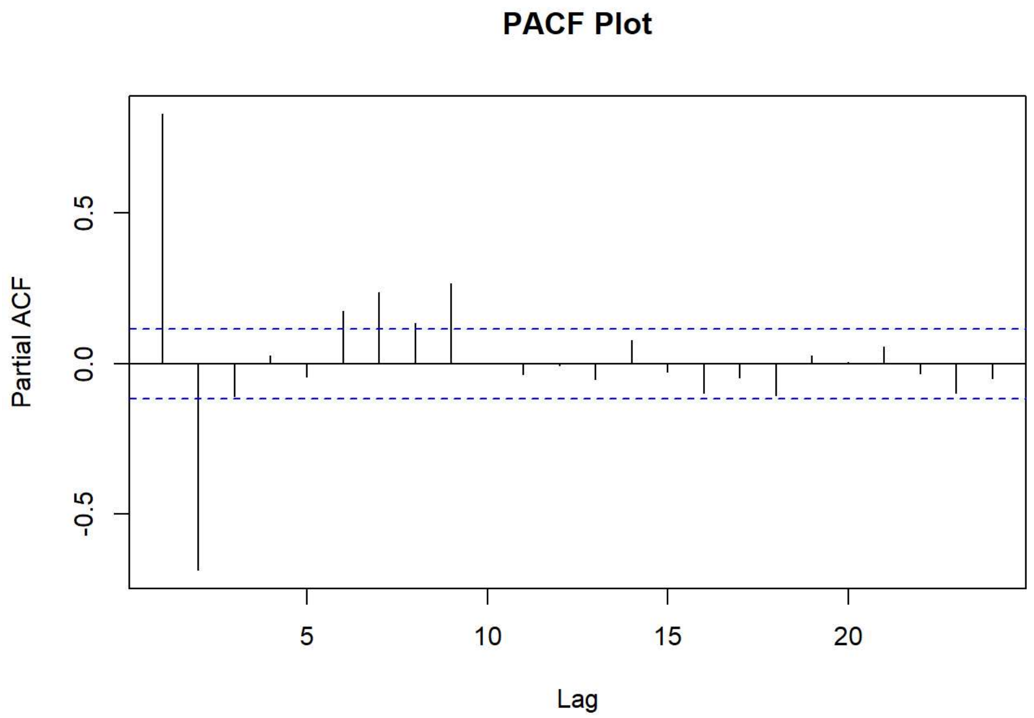
Question 5

(a)

```
data <- read.csv("C:/Users/Lenovo/Desktop/sunspot(2).dat")

data$X5.00 <- sqrt(data$X5.00)
ts_data <- ts(data)
acf(ts_data, main = "ACF Plot")
```




```
pacf(ts_data, main = "PACF Plot")
```



(b)

There does not exist any truncation in the ACF plot, which means it cannot be an $MA(q)$ model. However, we see a cut off at $h = 2$ in the PACF plot, which indicates that we can use an $AR(2)$ model to fit it.

```
library(forecast)
```

```
## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo
```

```
model <- arima(ts_data, order = c(2,0,0))
summary(model)
```

```
##
## Call:
## arima(x = ts_data, order = c(2, 0, 0))
##
## Coefficients:
##          ar1      ar2  intercept
##       1.4029  -0.6916     6.3471
## s.e.  0.0425   0.0425     0.2390
##
## sigma^2 estimated as 1.353:  log likelihood = -447.12,  aic = 902.25
##
## Training set error measures:
##                      ME     RMSE       MAE  MPE MAPE      MASE        ACF1
## Training set 0.00195595 1.163115 0.9109023 -Inf  Inf 0.6511489 -0.07022899
```

Thus, we maintain that the $AR(2)$ model is

$$X_t = 1.40X_{t-1} - 0.69X_{t-2} + 6.35 + W_t,$$

where $W_t \sim WN(0, 1.353)$.

(c)

```
forecast_result <- forecast(model, h = 4)

print(forecast_result)
```

```
##     Point Forecast    Lo 80    Hi 80     Lo 95     Hi 95
## 285       5.692866 4.202274 7.183459 3.4132023  7.972531
## 286       5.133391 2.565413 7.701369 1.2060073  9.060774
## 287       5.096921 1.900942 8.292901 0.2090931  9.984750
## 288       5.432706 2.010790 8.854621 0.1993380 10.666073
```

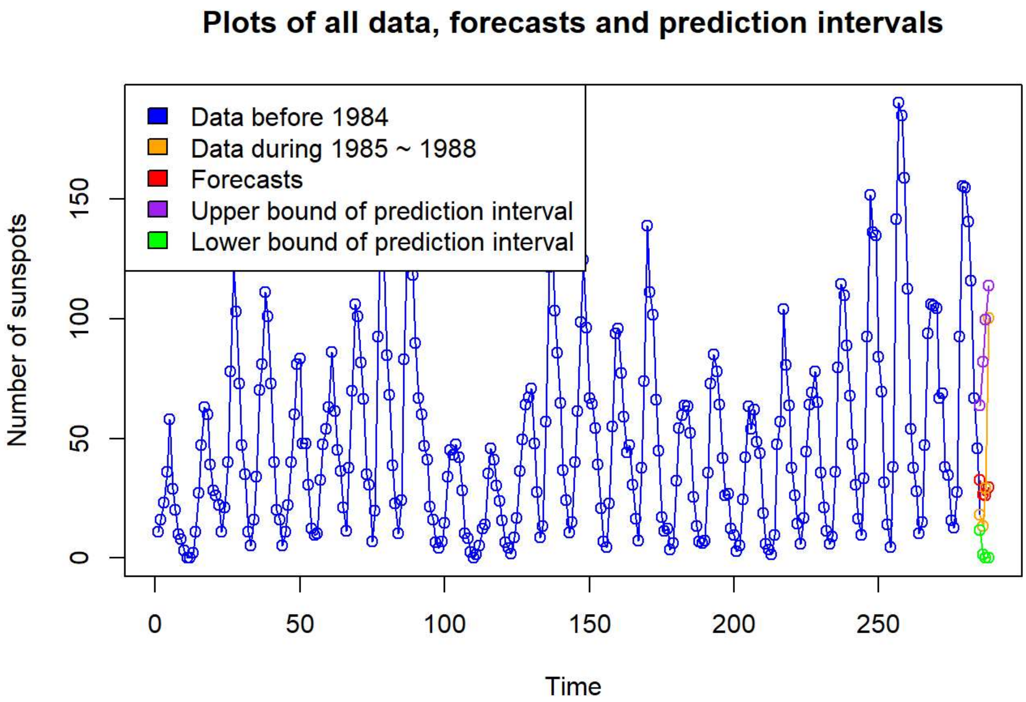
- $X_{285}^{284} \approx 5.69$, whose 95% CI is approximately $[3.41, 7.97]$
- $X_{286}^{284} \approx 5.13$, whose 95% CI is approximately $[1.21, 9.06]$
- $X_{287}^{284} \approx 5.10$, whose 95% CI is approximately $[0.21, 9.98]$
- $X_{288}^{284} \approx 5.43$, whose 95% CI is approximately $[0.20, 10.67]$

(d)

```
data2 <- read.csv("C:/Users/Lenovo/Desktop/sunspot2(2).dat")
ts_data2 <- ts(data2)

forecast_result$mean <- forecast_result$mean^2
forecast_result$lower <- forecast_result$lower^2
forecast_result$upper <- forecast_result$upper^2

data2[285:288,]
```

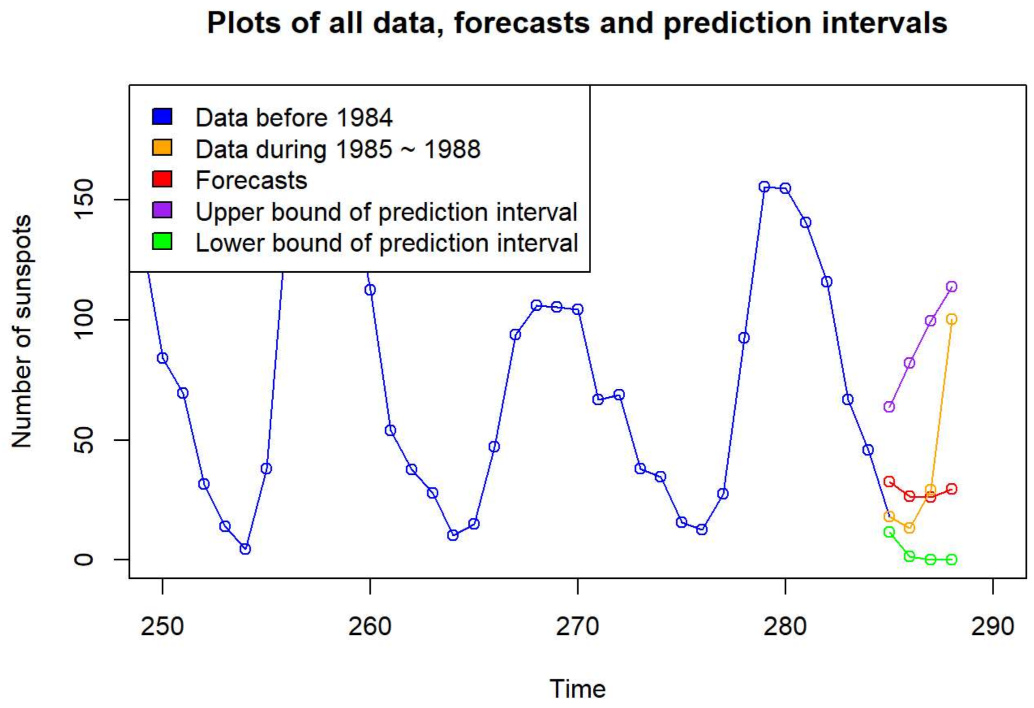
```
## [1]  17.9  13.4  29.2 100.2
```

```
x1 <- 285:288
y1 <- c(17.9,13.4,29.2,100.2)



plot(ts_data2, col = "blue",type = "o",
     main = "Plots of all data, forecasts and prediction intervals", xlab = "Time",
     ylab = "Number of sunspots")
legend("topleft",  legend = c("Data before 1984", "Data during 1985 ~ 1988",
                              "Forecasts", "Upper bound of prediction interval"
                              ,"Lower bound of prediction interval"),
       fill = c("blue","orange","red", "purple","green"), cex = 1
)
lines(forecast_result$mean, col = "red",type  = "o")
lines(x1,y1, col = "orange", type = "o")
lines(x1,forecast_result$lower[,2],col = "green",type = "o")
lines(x1,forecast_result$upper[,2],col = "purple",type = "o")
```



To have an explicit insight of the effect of our prediction, we filter out the figure during the time from 1950 to 1988.

```
plot(ts_data2, col = "blue",type = "o",xlim = c(250,290),
     main = "Plots of all data, forecasts and prediction intervals", xlab = "Time",
     ylab = "Number of sunspots")
legend("topleft",  legend = c("Data before 1984", "Data during 1985 ~ 1988",
                                  "Forecasts", "Upper bound of prediction interval"
                                  ,"Lower bound of prediction interval"),
       fill = c("blue","orange","red", "purple","green"), cex = 1
)
lines(forecast_result$mean, col = "red",type  = "o")
lines(x1,y1, col = "orange", type = "o")
lines(x1,forecast_result$lower[,2],col = "green",type = "o")
lines(x1,forecast_result$upper[,2],col = "purple",type = "o")
```



Plots of all data, forecasts and prediction intervals
Data before 1984
Data during 1985 ~ 1988
Forecasts
Upper bound of prediction interval
Lower bound of prediction interval
Number of sunspots
0
50
100
150
250
260
270
280
290
Time