Homework8

Hsueh-Pin Liu

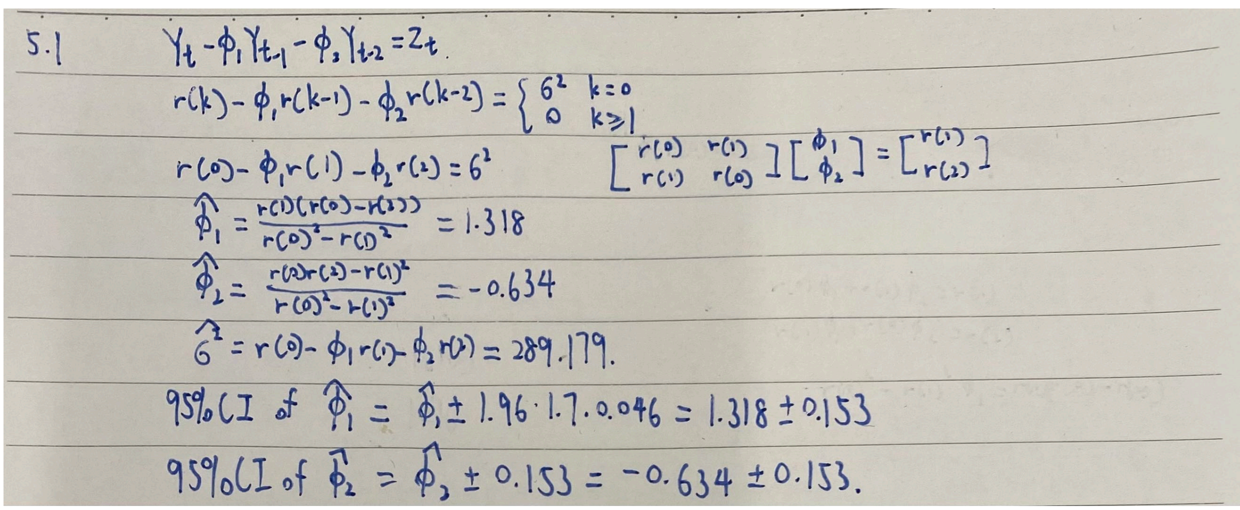
2023-04-24

Problem 1

Do the following problems: 5.1, 5.2, 5.4

5.1

The sunspot numbers $\{X_t, t = 1, \ldots, 100\}$, filed as SUNSPOTS.TSM, have sample autocovariances $\hat{\gamma}(0) = 1382.2$, $\hat{\gamma}(1) = 1114.4$, $\hat{\gamma}(2) = 591.734$, and $\hat{\gamma}(3) = 96.216$. Use these values to find the Yule–Walker estimates of ϕ_1, ϕ_2, and σ^2 in the model $Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + Z_t, Z_t \sim WN(0, \sigma^2)$, for the mean-corrected series $Y_t = X_t - 46.93, t = 1, \ldots, 100$. Assuming that the data really are a realization of an AR(2) process, find 95 % confidence intervals for ϕ_1 and ϕ_2.



5.1 $Y_t - \phi_1 Y_{t-1} - \phi_2 Y_{t-2} = Z_t$.

$r(k) - \phi_1 r(k-1) - \phi_2 r(k-2) = \begin{cases} \sigma^2 & k=0 \\ 0 & k \geq 1 \end{cases}$

$r(0) - \phi_1 r(1) - \phi_2 r(2) = \sigma^2$ $\quad \begin{bmatrix} r(0) & r(1) \\ r(1) & r(0) \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \end{bmatrix} = \begin{bmatrix} r(1) \\ r(2) \end{bmatrix}$

$\hat{\phi}_1 = \frac{r(1)(r(0)-r(2))}{r(0)^2 - r(1)^2} = 1.318$

$\hat{\phi}_2 = \frac{r(0)r(2) - r(1)^2}{r(0)^2 - r(1)^2} = -0.634$

$\hat{\sigma}^2 = r(0) - \phi_1 r(1) - \phi_2 r(2) = 289.179.$

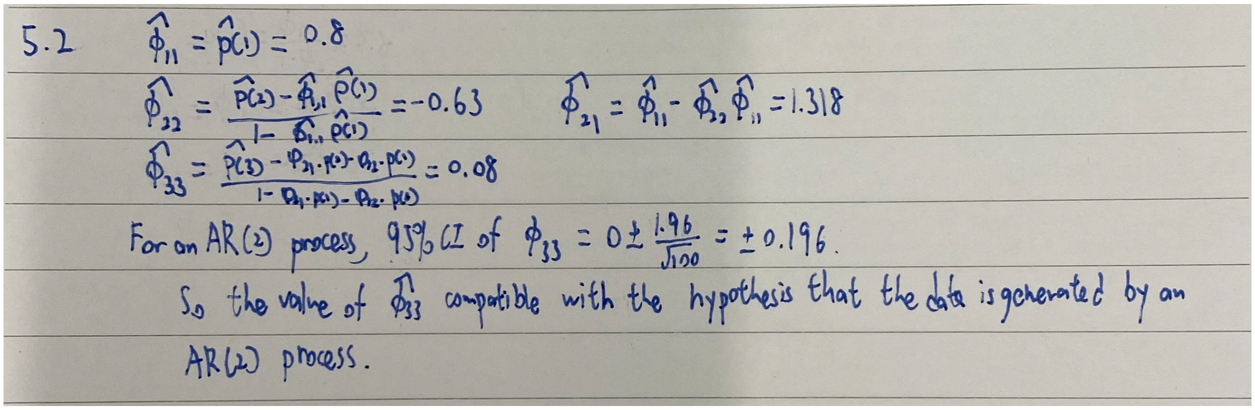
95% CI of $\hat{\phi}_1 = \hat{\phi}_1 \pm 1.96 \cdot 1.7 \cdot 0.046 = 1.318 \pm 0.153$

95% CI of $\hat{\phi}_2 = \hat{\phi}_2 \pm 0.153 = -0.634 \pm 0.153.$

Figure 1: Problem 1 5.1

5.2

From the information given in the previous problem, use the Durbin–Levinson algorithm to compute the sample partial autocorrelations $\hat{\phi}_{11}$, $\hat{\phi}_{22}$ and $\hat{\phi}_{33}$ of the sunspot series. Is the value of $\hat{\phi}_{33}$ compatible with the hypothesis that the data are generated by an AR(2) process? (Use significance level 0.05.)



5.2 $\hat{\phi}_{11} = \hat{\rho}(1) = 0.8$

$\hat{\phi}_{22} = \frac{\hat{\rho}(2) - \hat{\phi}_{11}\hat{\rho}(1)}{1 - \hat{\phi}_{11}\hat{\rho}(1)} = -0.63 \qquad \hat{\phi}_{21} = \hat{\phi}_{11} - \hat{\phi}_{22}\hat{\phi}_{11} = 1.318$

$\hat{\phi}_{33} = \frac{\hat{\rho}(3) - \phi_{21}\cdot\rho(2) - \phi_{22}\cdot\rho(1)}{1 - \phi_{21}\cdot\rho(1) - \phi_{22}\cdot\rho(2)} = 0.08$

For an AR(2) process, 95% CI of $\phi_{33} = 0 \pm \frac{1.96}{\sqrt{100}} = \pm 0.196$.

So the value of $\hat{\phi}_{33}$ compatible with the hypothesis that the data is generated by an AR(2) process.

Figure 2: Problem 1 5.2

5.4

Two hundred observations of a time series, $X_1, ..., X_{200}$, gave the following sample statistics:

sample mean: $\overline{X}_{200} = 3.82$;

sample variance: $\hat{\gamma}(0) = 1.15$;

sample ACF: $\hat{\rho}(1) = 0.427$; $\hat{\rho}(2) = 0.475$; $\hat{\rho}(3) = 0.169$

a. Based on these sample statistics, is it reasonable to suppose that $X_t - \mu$ is white noise?

b. Assuming that $X_t - \mu$ can be modeled as the AR(2) process $X_t - \mu - \phi 1(X_{t-1} - \mu) - \phi_2(X_{t-2} - \mu) = Z_t$, where $Z_t \sim IID(0, \sigma^2)$, find estimates of $\mu, \phi 1, \phi 2$ and σ^2.

c. Would you conclude that $\mu = 0$?

d. Construct 95 % confidence intervals for $\phi 1$ and $\phi 2$.

e. Assuming that the data were generated from an AR(2) model, derive estimates of the PACF for all lags $h \geq 1$.

5.4. (a) $\pm \frac{1.96}{\sqrt{n}} \approx \pm 0.138$ $\hat\rho(1), \hat\rho(2), \hat\rho(3)$ are outside of it

So it's not reasonable to suppose that $\{X_t-\mu\}$ is white noise.

(b) $\hat\phi_1 = \frac{\rho(1)-\rho(1)\rho(2)}{1-\rho(1)^2} = 0.274$

$\hat\phi_2 = \frac{\rho(2)-\rho(1)^2}{1-\rho(1)^2} = 0.358$

$\hat\sigma^2 = \hat r(0) - \hat\phi_1 \hat\rho(1)\hat r(0) - \hat\phi_2 \hat\rho(2)\hat r(0) = 0.82$

$\hat\mu = \overline{X_{200}} = 3.82$

(c) $\bar X_n - \mu \sim N(0, \frac{1}{n}\sum_{|h|<\infty} r(h)) \approx N(0, 3.61)$ $3.82 > 1.96\sqrt{\frac{3.61}{200}} \approx 0.26$

So we reject the hypothesis that $\mu=0$.

(d) 95% CI of $\hat\phi_1 = \hat\phi_1 \pm 1.96 \cdot \sqrt{\frac{0.82}{200}}\sqrt{\frac{1-0.427^2}{1.15}} = 0.274 \pm 0.13$

95% CI of $\hat\phi_2 = \hat\phi_2 \pm 0.13 = 0.358 \pm 0.13$

Figure 3: Problem 1 5.4

Problem 2

Derive the ACF of a SARIMA $(0, 0, 0) \times (1, 0, 0)_{12}$ model.

2. SARIMA $(0,0,0)\times(1,0,0)_{12}$

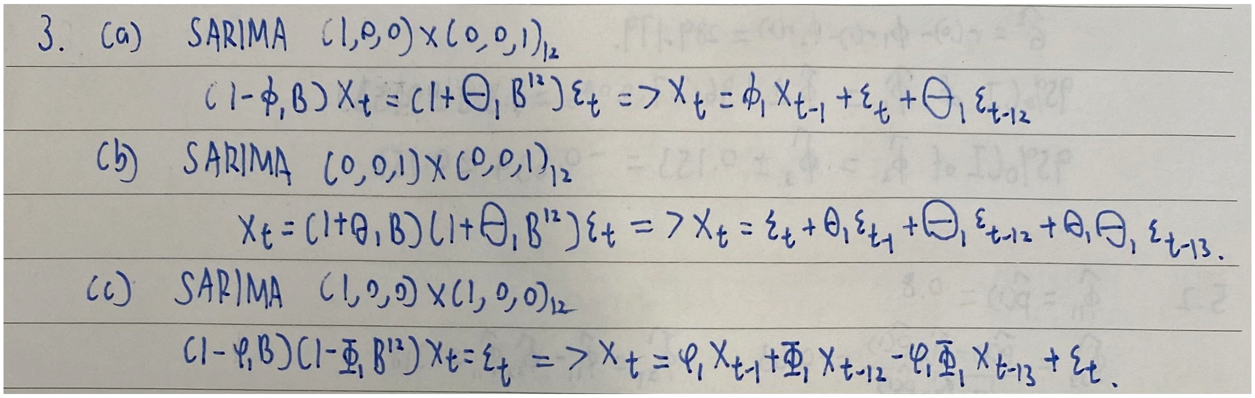
$(1-\Phi_1 B^{12})X_t = \varepsilon_t \Rightarrow X_t = \Phi_1 X_{t-12} + \varepsilon_t$

$r(h) = E(X_{t+h}X_t) = \Phi_1 r(h-12) \Rightarrow \rho(h) = \Phi_1 \rho(h-12) = \Phi_1^k \rho(t)$ $h=12k+t$ $k\in\mathbb{Z}$ $0\le t<12$.

Figure 4: Problem 2

Problem 3

For the following SARIMA models, write the models in their standard forms, and find the ACF and PACF using R (choose your own parameters). Describe the ACF and PACF behavior in words. a. $\text{SARIMA}(1,0,0) \times (0,0,1)_{12}$ b. $\text{SARIMA}(0,0,1) \times (0,0,1)_{12}$ c. $\text{SARIMA}(1,0,0) \times (1,0,0)_{12}$



3. (a) SARIMA $(1,0,0)\times(0,0,1)_{12}$

$(1-\phi_1 B)X_t = (1+\Theta_1 B^{12})\varepsilon_t \Rightarrow X_t = \phi_1 X_{t-1} + \varepsilon_t + \Theta_1 \varepsilon_{t-12}$

(b) SARIMA $(0,0,1)\times(0,0,1)_{12}$

$X_t = (1+\theta_1 B)(1+\Theta_1 B^{12})\varepsilon_t \Rightarrow X_t = \varepsilon_t + \theta_1\varepsilon_{t-1} + \Theta_1\varepsilon_{t-12} + \theta_1\Theta_1\varepsilon_{t-13}.$

(c) SARIMA $(1,0,0)\times(1,0,0)_{12}$

$(1-\varphi_1 B)(1-\Phi_1 B^{12})X_t = \varepsilon_t \Rightarrow X_t = \varphi_1 X_{t-1} + \Phi_1 X_{t-12} - \varphi_1\Phi_1 X_{t-13} + \varepsilon_t.$

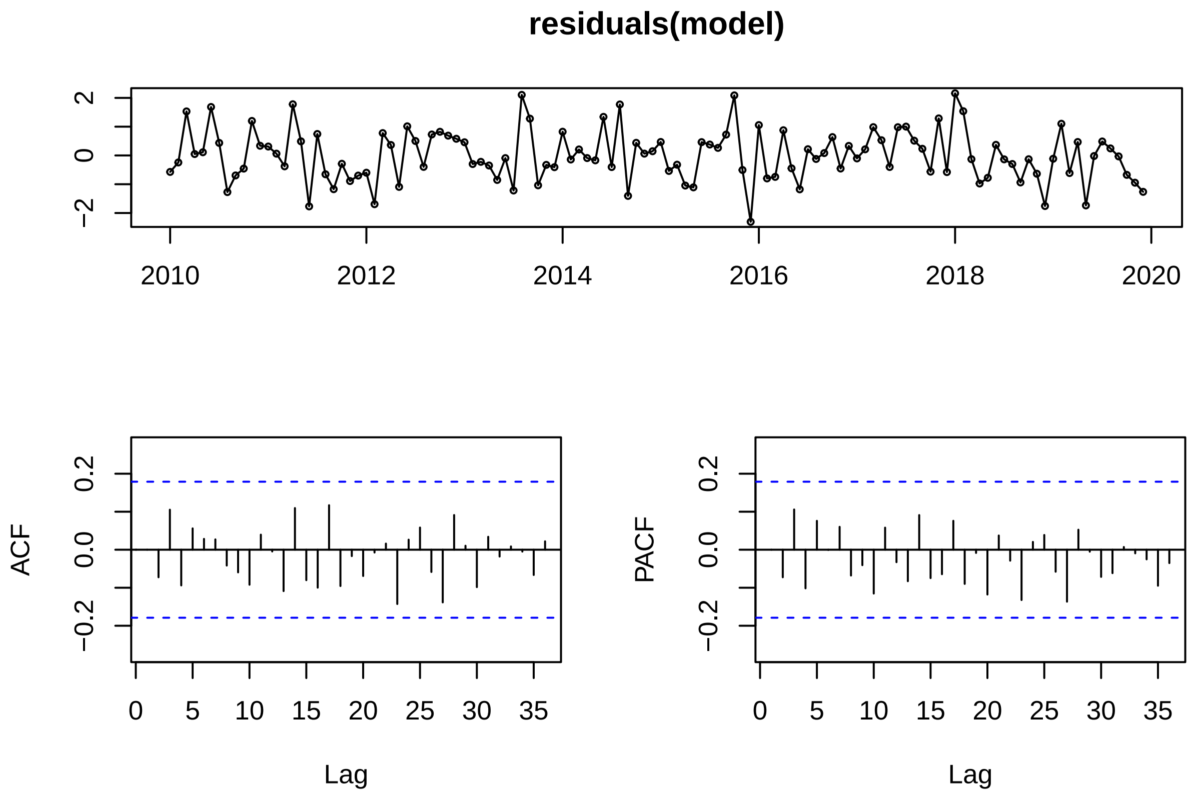
Figure 5: Problem 3

(a)

```r
library(forecast)
```

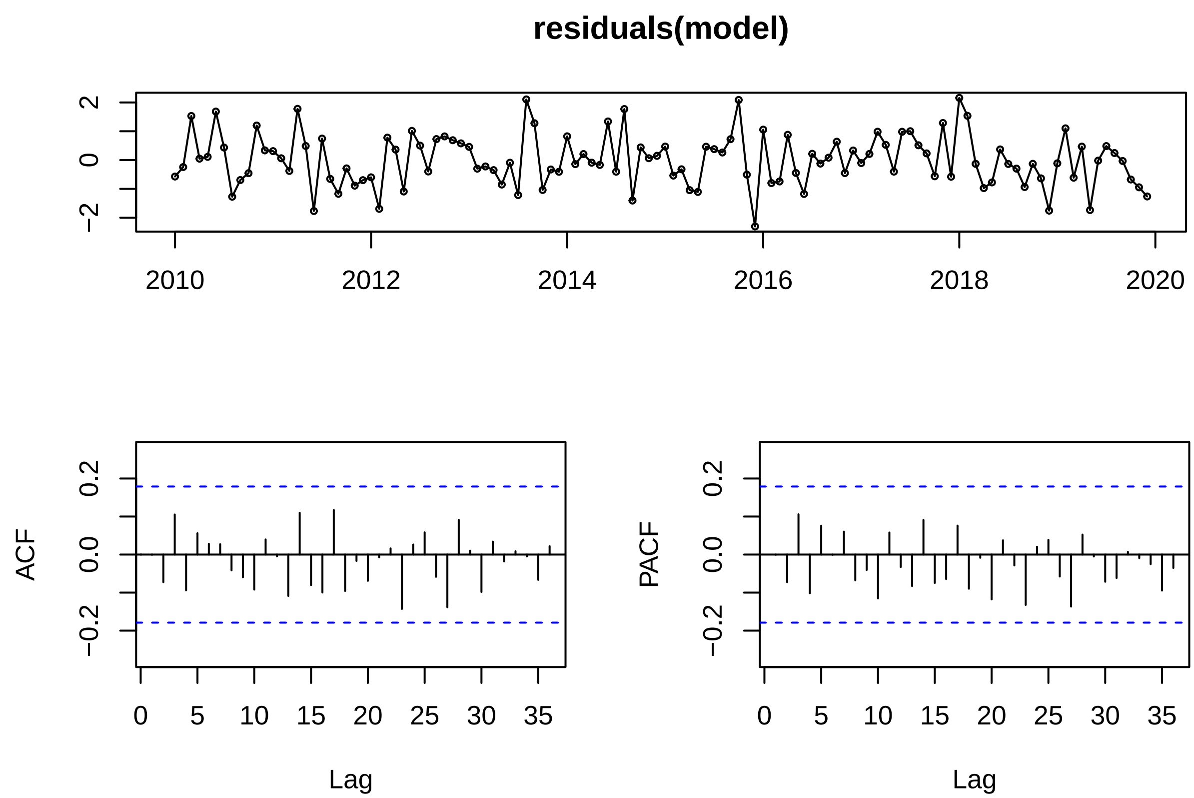
```
## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo
```

```r
set.seed(123)
data <- ts(rnorm(120), frequency = 12, start = c(2010, 1))
model <- arima(data, order = c(1,0,0), seasonal = list(order = c(0,0,1), period = 12))
tsdisplay(residuals(model), lag.max = 36)
```



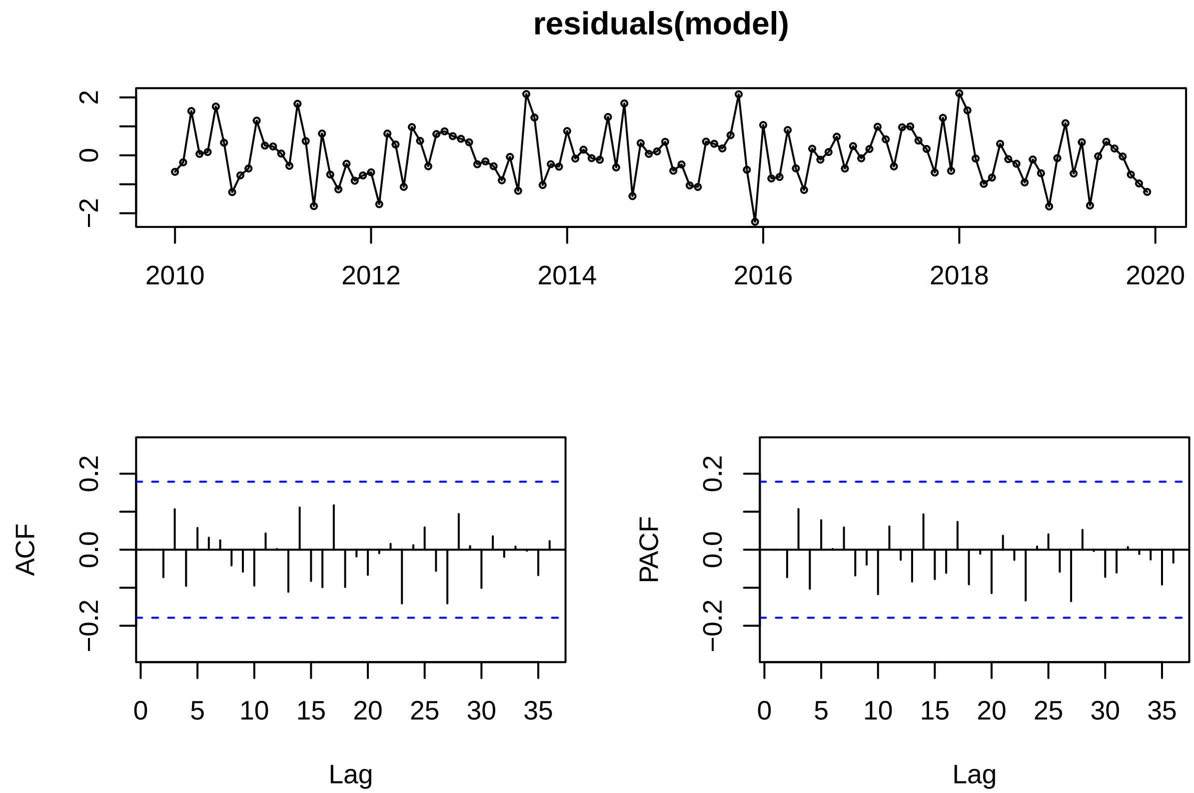
(b)

```
set.seed(123)
data <- ts(rnorm(120), frequency = 12, start = c(2010, 1))
model <- arima(data, order = c(0,0,1), seasonal = list(order = c(0,0,1), period = 12))
tsdisplay(residuals(model), lag.max = 36)
```



(c)

```
set.seed(123)
data <- ts(rnorm(120), frequency = 12, start = c(2010, 1))
model <- arima(data, order = c(1,0,0), seasonal = list(order = c(1,0,0), period = 12))
tsdisplay(residuals(model), lag.max = 36)
```


residuals(model)
2
0
−2
2010
2012
2014
2016
2018
2020
ACF
0.2
0.0
−0.2
0
5
10
15
20
25
30
35
Lag
PACF
0.2
0.0
−0.2
0
5
10
15
20
25
30
35
Lag

Problem 4

Consider the co2 data in the dataset pacakage in R, which is Mauna Loa atmospheric CO2 Concentration. Set aside the last 24 observations as the test data and the rest as the training data.

a. Plot the data and apply Box-Cox transformation, if necessary.

b. Forecast 1: Use subset selection method to fit an ARIMA model to the data. Verify if the model is adequate. Forecast the 24 values along with the forecast intervals.

Forecast 2: Now identify potential SARIMA models for ACF and PACF plots. Fit the candidate models and compare AICC to choose your final model. Use the model to Forecast the 24 values along with the forecast intervals.

Forecast 3: Use Holt-Winters seasonal forecasting method to predict the 24 values along with the forecast intervals.

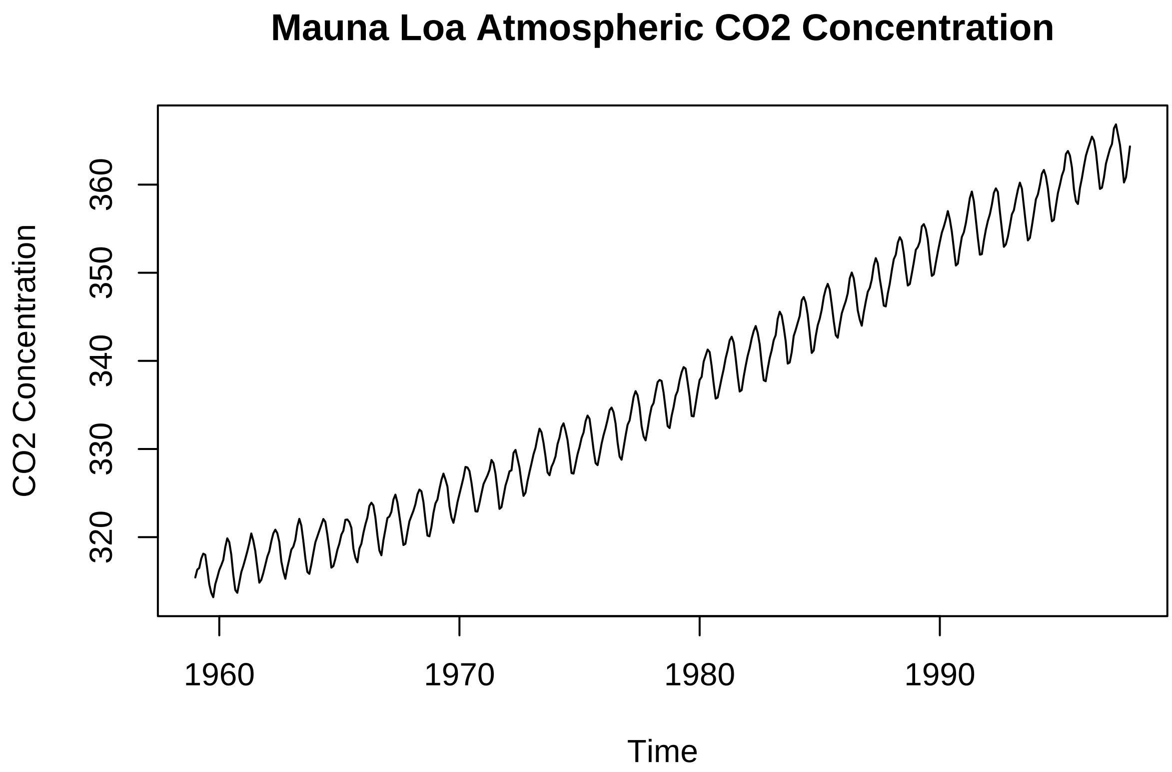
c. Now complete the following table to compare between the forecasts:

Criteria	Forecast 1	Forecast 2	Forecast 3
RMSE (Root Mean Squared Error)			
MAPE (Mean Average Percent Error)			

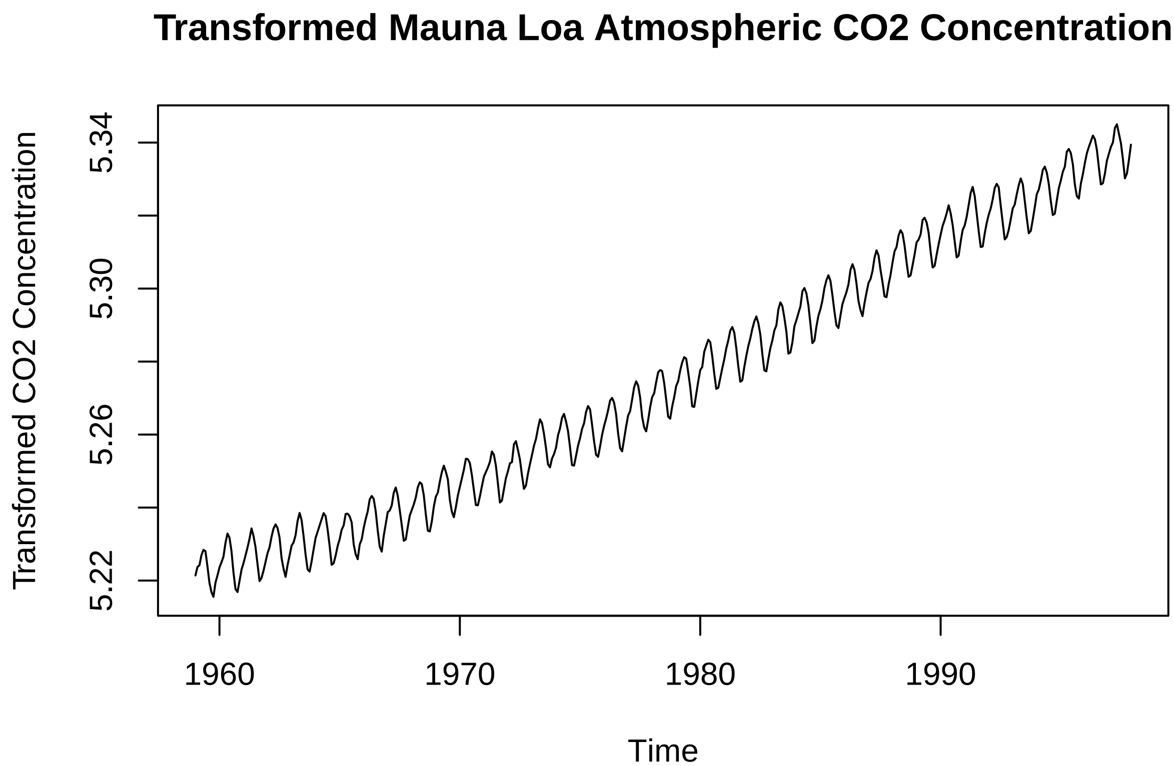
What is your conclusion?

(a)

```
data(co2)
train <- window(co2, end = c(1995, 12))
test <- window(co2, start = c(1996, 1))
plot(co2, main = "Mauna Loa Atmospheric CO2 Concentration", ylab = "CO2 Concentration")
```




```
lambda <- BoxCox.lambda(co2)
co2_bc <- BoxCox(co2,lambda=lambda)
plot(co2_bc, main = "Transformed Mauna Loa Atmospheric CO2 Concentration", ylab = "Transformed CO2 Conc
```



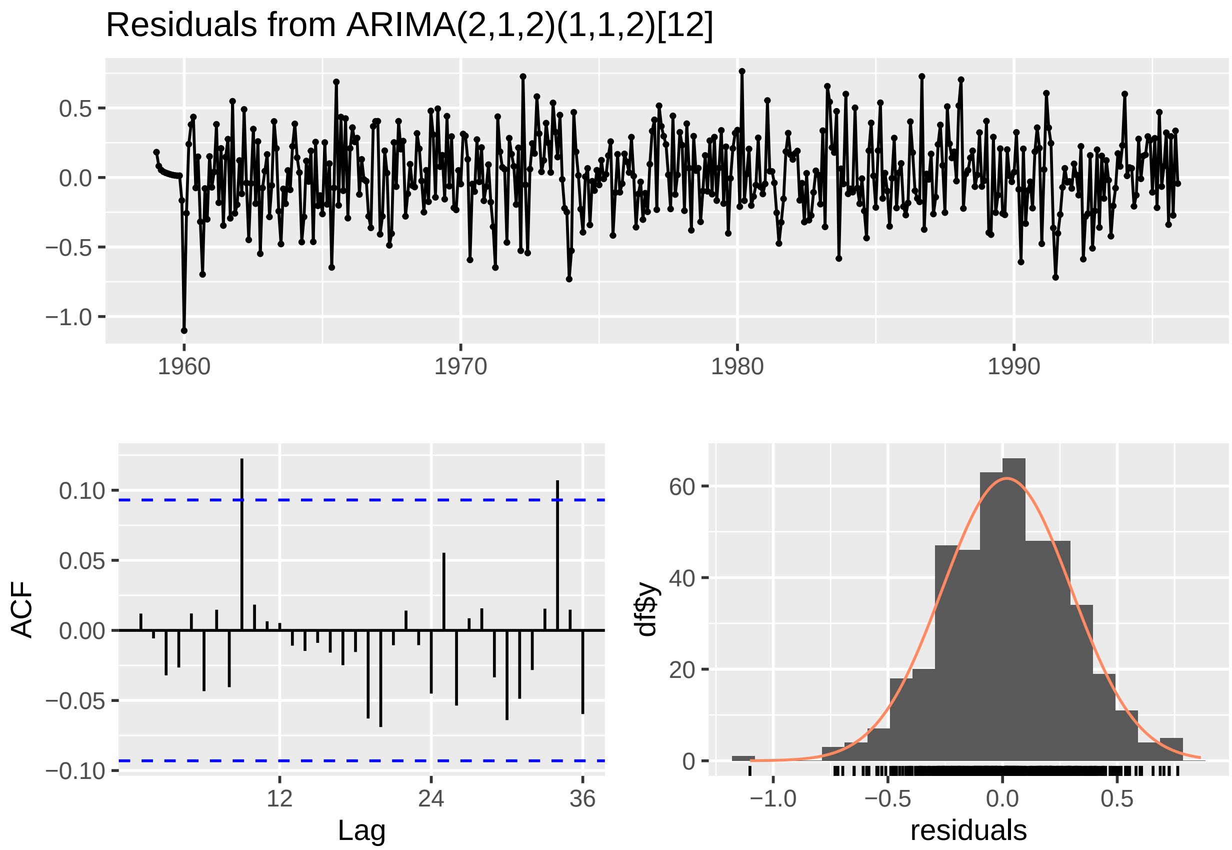
(b)

Forecast 1

```
fit <- auto.arima(train)
summary(fit)
```

```
## Series: train
## ARIMA(2,1,2)(1,1,2)[12]
##
## Coefficients:
##          ar1     ar2      ma1      ma2     sar1    sma1     sma2
##       0.0598  0.2445  -0.3966  -0.2297  -0.8265  0.0117  -0.7494
## s.e.  0.3532  0.1259   0.3536   0.2056   0.2378  0.2174   0.1796
##
## sigma^2 = 0.0835:  log likelihood = -73.02
## AIC=162.05   AICc=162.39   BIC=194.58
##
## Training set error measures:
##                      ME      RMSE       MAE         MPE       MAPE      MASE
## Training set 0.01843857 0.2823781 0.2247573 0.005393578 0.06710961 0.1789641
##                    ACF1
## Training set 0.01200236
```

```
checkresiduals(fit)
```




```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(2,1,2)(1,1,2)[12]
## Q* = 15.563, df = 17, p-value = 0.555
##
## Model df: 7.   Total lags used: 24
```

```
forecast1 <- forecast(fit, h = 24)
forecast1
```

```
##          Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
## Jan 1996       361.8415 361.4712 362.2119 361.2752 362.4079
## Feb 1996       362.5952 362.1509 363.0396 361.9156 363.2748
## Mar 1996       363.5517 363.0450 364.0585 362.7767 364.3267
## Apr 1996       364.9484 364.3987 365.4981 364.1077 365.7891
## May 1996       365.5328 364.9441 366.1215 364.6324 366.4331
## Jun 1996       364.8966 364.2739 365.5194 363.9442 365.8491
## Jul 1996       363.2812 362.6265 363.9360 362.2799 364.2825
## Aug 1996       361.2174 360.5328 361.9021 360.1704 362.2645
## Sep 1996       359.4932 358.7800 360.2063 358.4025 360.5839
## Oct 1996       359.4981 358.7576 360.2386 358.3656 360.6306
```

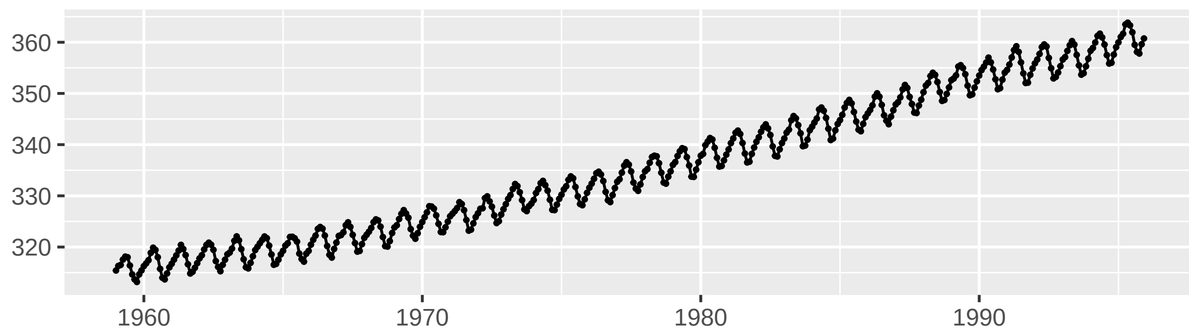
```
## Nov 1996       360.9512 360.1844 361.7180 359.7785 362.1239
## Dec 1996       362.2394 361.4472 363.0316 361.0278 363.4510
## Jan 1997       363.3589 362.5227 364.1951 362.0801 364.6378
## Feb 1997       364.1421 363.2709 365.0133 362.8097 365.4745
## Mar 1997       365.0836 364.1788 365.9883 363.6999 366.4673
## Apr 1997       366.4248 365.4891 367.3604 364.9938 367.8557
## May 1997       366.9963 366.0309 367.9618 365.5199 368.4728
## Jun 1997       366.3688 365.3749 367.3628 364.8487 367.8890
## Jul 1997       364.7772 363.7555 365.7989 363.2147 366.3398
## Aug 1997       362.7338 361.6852 363.7824 361.1301 364.3375
## Sep 1997       360.9931 359.9183 362.0679 359.3493 362.6369
## Oct 1997       361.0349 359.9346 362.1353 359.3520 362.7179
## Nov 1997       362.4477 361.3224 363.5731 360.7266 364.1689
## Dec 1997       363.7434 362.5936 364.8933 361.9849 365.5019
```

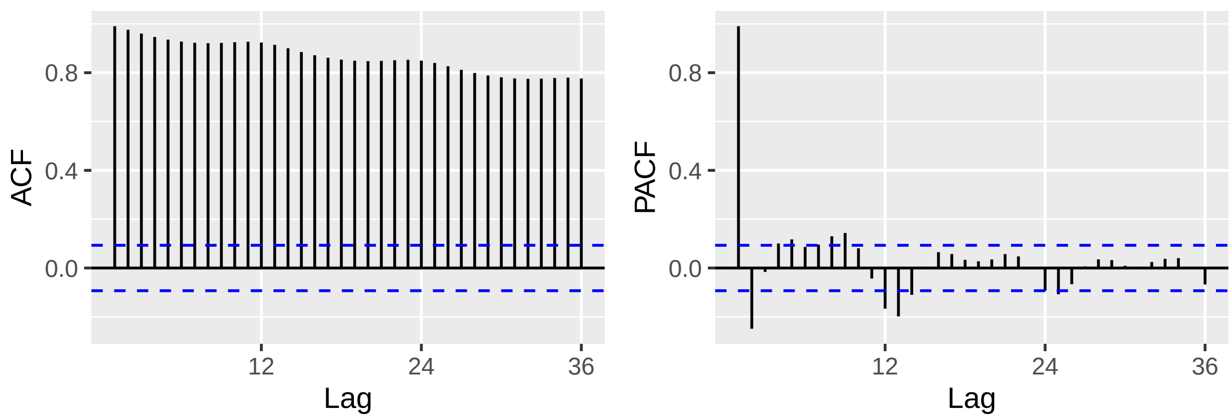
Forecast 2

```r
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.4.0      v purrr   0.3.5
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.1      v stringr 1.4.1
## v readr   2.1.3      v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
library(MuMIn)
ggtsdisplay(train)
```






```
fit1 <- Arima(train, order = c(0,1,1), seasonal = list(order = c(0,1,1), period = 12))
fit2 <- Arima(train, order = c(0,1,1), seasonal = list(order = c(1,1,1), period = 12))
fit3 <- Arima(train, order = c(1,1,1), seasonal = list(order = c(0,1,1), period = 12))
fit4 <- Arima(train, order = c(1,1,1), seasonal = list(order = c(1,1,1), period = 12))
AICc(fit1, fit2, fit3, fit4)
```

```
##      df     AICc
## fit1  3 159.6367
## fit2  4 161.0766
## fit3  4 160.5724
## fit4  5 161.9439
```

```
final_fit <- fit1
forecast2 <- forecast(final_fit, h = 24)
forecast2
```

```
##          Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
## Jan 1996       361.8635 361.4920 362.2351 361.2953 362.4317
## Feb 1996       362.6516 362.2076 363.0955 361.9726 363.3305
## Mar 1996       363.5960 363.0899 364.1021 362.8220 364.3701
## Apr 1996       364.9701 364.4087 365.5315 364.1115 365.8288
## May 1996       365.5442 364.9324 366.1559 364.6085 366.4798
## Jun 1996       364.9105 364.2523 365.5688 363.9038 365.9173
## Jul 1996       363.3029 362.6012 364.0046 362.2297 364.3760
## Aug 1996       361.2434 360.5008 361.9859 360.1077 362.3790
```

```
## Sep 1996       359.5119 358.7305 360.2932 358.3169 360.7068
## Oct 1996       359.5426 358.7243 360.3609 358.2912 360.7940
## Nov 1996       360.9788 360.1252 361.8324 359.6734 362.2842
## Dec 1996       362.2736 361.3861 363.1611 360.9163 363.6309
## Jan 1997       363.3742 362.4377 364.3107 361.9420 364.8065
## Feb 1997       364.1623 363.1849 365.1396 362.6676 365.6570
## Mar 1997       365.1067 364.0902 366.1233 363.5521 366.6614
## Apr 1997       366.4808 365.4265 367.5351 364.8684 368.0932
## May 1997       367.0549 365.9641 368.1456 365.3867 368.7230
## Jun 1997       366.4212 365.2952 367.5472 364.6992 368.1433
## Jul 1997       364.8136 363.6534 365.9738 363.0392 366.5879
## Aug 1997       362.7541 361.5607 363.9474 360.9289 364.5792
## Sep 1997       361.0226 359.7969 362.2483 359.1480 362.8971
## Oct 1997       361.0533 359.7961 362.3105 359.1306 362.9760
## Nov 1997       362.4895 361.2016 363.7774 360.5198 364.4592
## Dec 1997       363.7843 362.4664 365.1022 361.7687 365.7998
```

Forecast3

```
fit <- HoltWinters(train, seasonal = "multiplicative")
forecast3 <- forecast(fit, h = 24)
forecast3
```

```
##          Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
## Jan 1996       361.8718 361.5901 362.1535 361.4410 362.3026
## Feb 1996       362.6343 362.2861 362.9825 362.1018 363.1668
## Mar 1996       363.5799 363.1747 363.9851 362.9602 364.1996
## Apr 1996       365.0235 364.5669 365.4801 364.3252 365.7218
## May 1996       365.5662 365.0631 366.0692 364.7968 366.3355
## Jun 1996       364.9264 364.3813 365.4715 364.0927 365.7601
## Jul 1996       363.2623 362.6789 363.8458 362.3700 364.1546
## Aug 1996       361.1123 360.4933 361.7313 360.1656 362.0590
## Sep 1996       359.4310 358.7775 360.0846 358.4316 360.4305
## Oct 1996       359.4594 358.7702 360.1487 358.4053 360.5136
## Nov 1996       360.9567 360.2304 361.6830 359.8460 362.0675
## Dec 1996       362.2259 359.8247 364.6270 358.5536 365.8981
## Jan 1997       363.3423 361.4933 365.1914 360.5144 366.1703
## Feb 1997       364.1075 362.2272 365.9878 361.2318 366.9831
## Mar 1997       365.0564 363.1440 366.9689 362.1316 367.9812
## Apr 1997       366.5054 364.5582 368.4526 363.5274 369.4834
## May 1997       367.0497 365.0724 369.0271 364.0257 370.0738
## Jun 1997       366.4068 364.4057 368.4080 363.3463 369.4674
## Jul 1997       364.7356 362.7162 366.7550 361.6472 367.8240
## Aug 1997       362.5763 360.5416 364.6111 359.4644 365.6882
## Sep 1997       360.8877 358.8353 362.9402 357.7488 364.0267
## Oct 1997       360.9158 358.8362 362.9954 357.7353 364.0962
## Nov 1997       362.4186 360.3036 364.5337 359.1839 365.6533
## Dec 1997       363.6924 360.1489 367.2359 358.2731 369.1117
```

(c)

```r
m1 <- forecast1$mean
m2 <- forecast2$mean
m3 <- forecast3$mean
rmse_1 <- sqrt(mean((test - m1)^2))
mape_1 <- mean(abs(test-m1)/test) * 100
rmse_2 <- sqrt(mean((test - m2)^2))
mape_2 <- mean(abs(test-m2)/test) * 100
rmse_3 <- sqrt(mean((test - m3)^2))
mape_3 <- mean(abs(test-m3)/test) * 100
```

So the table should look like this

Criteria	Forecast 1	Forecast 2	Forecast 3
RMSE (Root Mean Squared Error)	0.354169	0.354071	0.351115
MAPE (Mean Average Percent Error)	0.07581	0.07614	0.07550

In conclusion, these 3 methods have similar RMSE and MAPE, and Forecast 3 using the Holt-Winters method has slightly lower RMSE and MAPE, so it's the most accurate forecast among the three.

Problem 5

Suppose X_t satisfies an AR(1) model with parameter ϕ. How long a series do you need to estimate ϕ so that with 95% confidence the estimation error within ± 0.1 of the true value?

5. 95% CI for φ is $\hat{\phi} \pm 1.96\sqrt{\frac{1-\hat{\phi}^2}{n}}$

So $1.96\sqrt{\frac{1-\hat{\phi}^2}{n}} \leq 0.1 \Rightarrow n \geq 19.6^2(1-\hat{\phi}^2) = 384.16(1-\hat{\phi}^2)$

Mostly the AR(1) is causal, so $n \geq 384.16$, the series should be 385 long.

Figure 6: Problem 5