Statistics 5350/7110 22 March 2023
P. Shaman

Assignment 3 Solutions

1. This problem involves examination of monthly Lydia Pinkham data for the period 1940 through 1952. It addresses model construction and estimation of the 90 per cent duration interval. The data are in Lydiamonthly4052.txt.

(a) Construct a model relating sales and advertising with sales as the response. Explore the inclusion of lagged sales and lagged advertising variables. Include estimation of seasonal structure, dummies for outliers, and necessary calendar variables. Describe your fitted model.

```
> model1a<-lm(msales ~ madv + fmonth + madvl1 + madvl2 + madvl3 + madvl4
+ msalesl1 + msalesl2 + msalesl3 + msalesl4 + c220 + s220 + c348 + s348 +
c432 + s432 + feb44 + dec44 + jan45 + sep45)
> summary(model1a)

Call:
lm(formula = msales ~ madv + fmonth + madvl1 + madvl2 + madvl3 +
    madvl4 + msalesl1 + msalesl2 + msalesl3 + msalesl4 + c220 +
    s220 + c348 + s348 + c432 + s432 + feb44 + dec44 + jan45 +
    sep45)

Residuals:
   Min     1Q Median     3Q    Max
-29771  -8251    -29   6479  38058

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.446e+04  9.630e+03   2.540 0.012315 *
madv         2.129e-01  4.720e-02   4.510 1.47e-05 ***
fmonth2     -1.870e+04  9.583e+03  -1.951 0.053235 .
fmonth3     -1.204e+04  9.822e+03  -1.226 0.222438
fmonth4     -2.855e+04  9.249e+03  -3.087 0.002492 **
fmonth5     -3.026e+04  7.968e+03  -3.798 0.000227 ***
fmonth6     -2.745e+04  6.878e+03  -3.991 0.000112 ***
fmonth7     -1.002e+04  6.844e+03  -1.463 0.145865
fmonth8     -1.231e+04  7.623e+03  -1.615 0.108774
fmonth9     -8.944e+03  8.584e+03  -1.042 0.299435
fmonth10    -3.883e+03  8.728e+03  -0.445 0.657209
fmonth11    -5.000e+04  9.278e+03  -5.388 3.40e-07 ***
fmonth12    -5.291e+04  7.931e+03  -6.671 7.37e-10 ***
madvl1       1.179e-01  5.278e-02   2.233 0.027327 *
madvl2      -6.808e-03  5.464e-02  -0.125 0.901042
madvl3      -3.101e-02  5.206e-02  -0.596 0.552483
madvl4       9.280e-03  4.963e-02   0.187 0.851967
msalesl1     3.954e-01  7.991e-02   4.948 2.37e-06 ***
msalesl2     2.903e-01  8.245e-02   3.521 0.000601 ***
msalesl3     1.053e-01  8.189e-02   1.286 0.200917
msalesl4     5.409e-02  7.168e-02   0.755 0.451903
c220         1.633e+02  1.690e+03   0.097 0.923159
s220        -4.043e+03  1.668e+03  -2.424 0.016801 *
c348        -3.306e+03  1.725e+03  -1.917 0.057563 .
```

```
s348         2.160e+03  1.714e+03   1.260 0.209915    
c432         6.634e+02  1.661e+03   0.399 0.690287    
s432         3.585e+02  1.691e+03   0.212 0.832419    
feb44       -4.191e+04  1.594e+04  -2.628 0.009653 ** 
dec44       -5.188e+04  1.515e+04  -3.425 0.000832 ***
jan45        7.176e+04  1.560e+04   4.601 1.02e-05 ***
sep45       -4.263e+04  1.519e+04  -2.807 0.005801 ** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 13790 on 125 degrees of freedom
Multiple R-squared:  0.8822,	Adjusted R-squared:  0.8539 
F-statistic:  31.2 on 30 and 125 DF,  p-value: < 2.2e-16
```

Here we see that the model1a has many insignificant variables. So let us remove the insignificant variables and refit the model1a.

```
> model1a_refit<-lm(msales ~ madv + madvl1 + fmonth + msalesl1 + msalesl2 
+ feb44 + dec44 + jan45 + sep45)

> summary(model1a_refit)

Call:
lm(formula = msales ~ madv + madvl1 + fmonth + msalesl1 + msalesl2 + 
    feb44 + dec44 + jan45 + sep45)

Residuals:
   Min     1Q Median     3Q    Max 
-30671  -9075   -389   6803  41995 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  3.193e+04  7.941e+03   4.021 9.56e-05 ***
madv         2.603e-01  4.370e-02   5.957 2.08e-08 ***
madvl1       8.162e-02  4.850e-02   1.683 0.094698 .  
fmonth2     -2.111e+04  7.666e+03  -2.754 0.006685 ** 
fmonth3     -2.156e+04  6.211e+03  -3.471 0.000696 ***
fmonth4     -3.552e+04  7.130e+03  -4.981 1.88e-06 ***
fmonth5     -3.600e+04  6.774e+03  -5.314 4.27e-07 ***
fmonth6     -3.211e+04  6.354e+03  -5.054 1.37e-06 ***
fmonth7     -1.261e+04  6.052e+03  -2.084 0.039070 *  
fmonth8     -1.757e+04  6.212e+03  -2.828 0.005387 ** 
fmonth9     -1.490e+04  6.537e+03  -2.279 0.024210 *  
fmonth10    -1.199e+04  6.969e+03  -1.720 0.087697 .  
fmonth11    -5.823e+04  7.683e+03  -7.580 4.83e-12 ***
fmonth12    -5.534e+04  7.133e+03  -7.758 1.84e-12 ***
msalesl1     4.755e-01  6.776e-02   7.017 9.72e-11 ***
msalesl2     3.376e-01  6.550e-02   5.154 8.79e-07 ***
feb44       -3.748e+04  1.572e+04  -2.385 0.018470 *  
dec44       -6.185e+04  1.502e+04  -4.117 6.61e-05 ***
jan45        7.597e+04  1.553e+04   4.892 2.78e-06 ***
sep45       -5.334e+04  1.512e+04  -3.527 0.000574 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 14170 on 136 degrees of freedom
```

```
Multiple R-squared:  0.8647,   Adjusted R-squared:  0.8459
F-statistic: 45.76 on 19 and 136 DF,  p-value: < 2.2e-16
```

We have removed the insignificant variable including all the trigonometric variables as well. But to confirm this let us perform the anova test and see if the removed variables are insignificant or not.

```
> anova(model1a_refit, model1a)
Analysis of Variance Table

Model 1: msales ~ madv + madvl1 + fmonth + msalesl1 + msalesl2 + feb44 +
    dec44 + jan45 + sep45
Model 2: msales ~ madv + fmonth + madvl1 + madvl2 + madvl3 + madvl4 +
    msalesl1 + msalesl2 + msalesl3 + msalesl4 + c220 + s220 +
    c348 + s348 + c432 + s432 + feb44 + dec44 + jan45 + sep45
  Res.Df        RSS Df  Sum of Sq      F  Pr(>F)
1    136 2.7298e+10
2    125 2.3776e+10 11 3522206072 1.6835 0.08446 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Here we see that the anova score is 0.08446, which is >0.05. This suggests that we have correctly removed all the insignificant variables and we should continue our analysis with model1a_refit.

(b) Analyze the residuals from your model. (i) Give the residual normal quantile plot and test for normality with the Shapiro–Wilk test. (ii) Plot the residuals vs. time. (iii) Present the residual autocorrelations and partial autocorrelations. (iv) Produce the residual spectral density plot and use it to test for reduction to white noise via two methods: (α) use of the blue line in the upper right part of the plot and (β) Bartlett's version of the Kolmogorov–Smirnov test. To perform Bartlett's test, install the *hwwntest* package and load it from your library:

```
install.packages ("hwwntest")
library("hwwntest")
```

The null hypothesis is white noise structure. Give the command

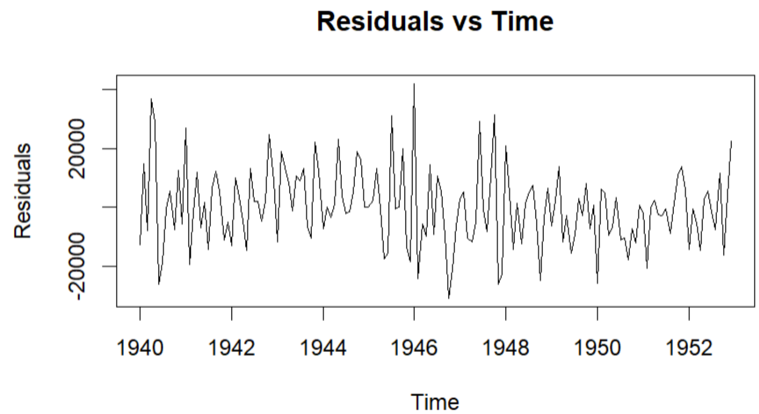
```
bartlettB.test(ts(resid(model)))
```

Reject the null hypothesis if the *p*-value is less than 0.05 for a five percent test. Remember that use of the blue line gives a visual test which provides an approximation. Bartlett's test is more precise.

Discuss each of these residual diagnostic results carefully.

Residual Analysis: for model1a_refit

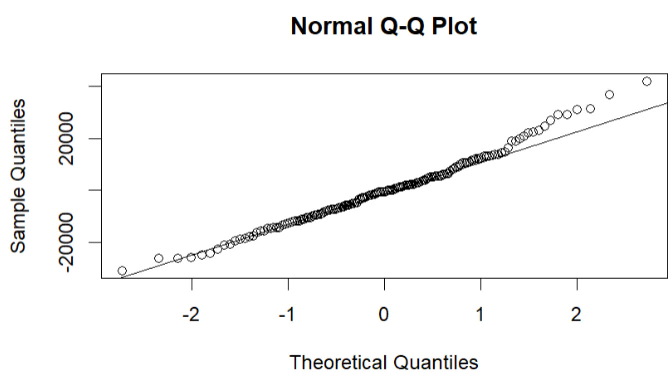
```
plot(ts(resid(model1a_refit), start = c(1940, 1), frequency = 12),
xlab="Time", ylab = "Residuals", main="Residuals vs Time")
```



The model does not fully capture the trend structure. The residuals are very high during 1940-41 and during 1945-48. At least the trend structure is modest.

Normal Quantile Plot:

```
> q<-qqnorm(resid(model1a_refit))
> qqline(resid(model1a_refit))
```



The residuals move away from normality at the upper tail a little. The outliers in the upper tail are 4, 94 and 73 which are 1940-Apr, 1947-Oct and 1946-Jan.

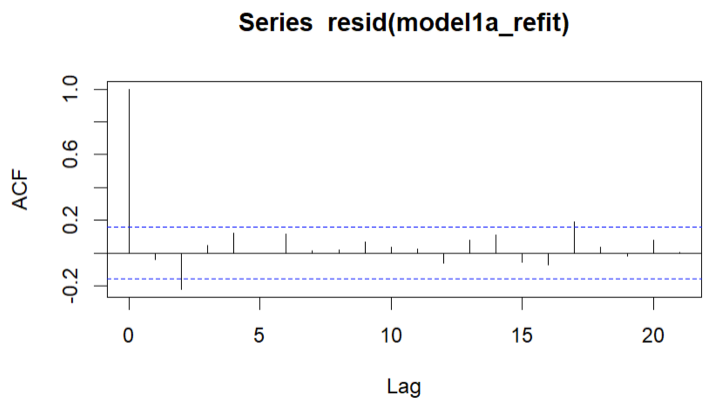
Shapiro Test:

```
> shapiro.test(resid(model1a_refit))

	Shapiro-Wilk normality test

data:  resid(model1a_refit)
W = 0.98617, p-value = 0.1235
```

This shows non significance for rejection of normality.

Autocorrelations (acf):

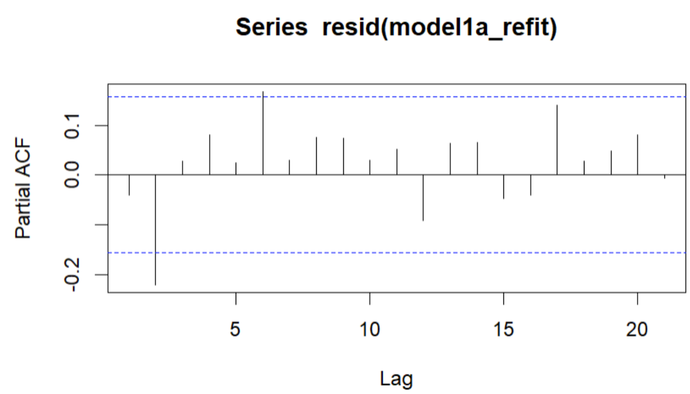
```
> acf(resid(model1a_refit))
```



We see that there is significant residual correlation at lag 2 and lag 17. But these can be neglected. This shows that the model is almost reduced to white noise structure.

Partial Autocorrelations:

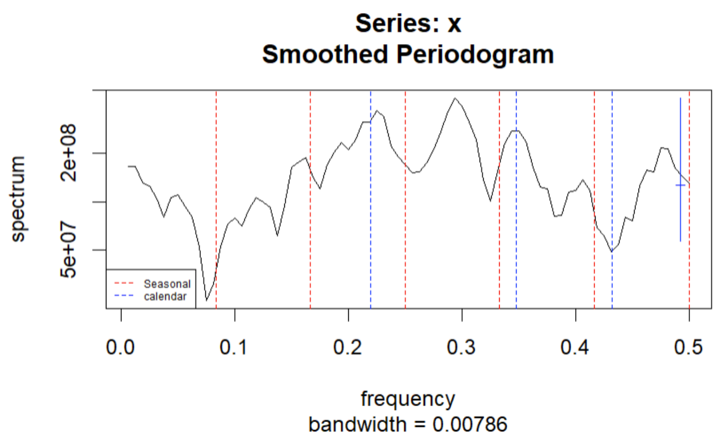
```
> pacf(resid(model1a_refit))
```



We see that there is significant residual partial correlation at lag 2 and lag 6.

Spectral Density Plot:

```
> spectrum(resid(model1a_refit), span=5)
> abline(v=c(1/12,2/12,3/12,4/12,5/12,6/12),col="red",lty=2)
> abline(v=c(0.220,0.348,0.432),col="blue",lty=2)
> legend('bottomleft', legend = c('Seasonal', 'calendar'),col = c('red',
'blue'), lty = 2, cex=0.5)
```



This spectral plot is not flat, indicating that the residuals do not conform to white noise structure. The spectral peak at low frequency is for slow movement of the time series, essentially remaining trend structure. The spectral plot red lines clearly show prominent spectral activity at frequencies 2/12 and perhaps at frequencies 4/12 and 5/12, precisely indicating the presence of little unmodeled seasonal structure in the model residuals to be captured still.
There is also some spectral activity around calendar frequencies 0.220 and 0.348, unclear at frequency 0.432, perhaps signaling modest calendar effects.

Using the blue line in the upper part of the plot: if we double the part of the blue line above the notch, it is not greater than the full vertical range. So here the residuals are not reduced to white noise.

Using Bartlett's test:

```
> bartlettB.test(ts(resid(model1a_refit)))

      Bartlett B Test for white noise

data:
= 1.0405, p-value = 0.2291
```

We need to reject the null hypothesis if the p value < 0.05. But here the p value is 0.2291, so this indicates that this is white noise structure according to Bartlett's Test.

(c) Calculate the estimate of the 90 per cent duration interval and discuss the result.

From our model above, we can write the model equation as,

$S_t = 3.193x10^4 + (4.755x10^{-1}S_{t-1}) + (3.376x10^{-1}S_{t-2}) + 2.603x10^{-1}A_t + (8.162x10^{-2}A_{t-1})$
$S_t - 4.755x10^{-1}BS_t - 3.376x10^{-1}B^2S_t = 3.193x10^4 + 2.603x10^{-1}A_t + 8.162x10^{-2}BA_t$
$S_t = (1-4.755x10^{-1}B - 3.376x10^{-1}B^2)^{-1}.3.193x10^4 + (1-4.755x10^{-1}B -$
$3.376x10^{-1}B^2)^{-1}.A_t.(2.603x10^{-1} + 8.162x10^{-2}B)$

As r<=p,

$2.603x10^{-1} + 8.162x10^{-2}B = (1-4.755x10^{-1}B - 3.376x10^{-1}B^2)(\boldsymbol{\delta} + \boldsymbol{\delta}_1 B + \boldsymbol{\delta}_2 B^2 +...)$

$\boldsymbol{\delta} = 2.603x10^{-1}$
$\boldsymbol{\delta}1 = 8.162x10^{-2} + 4.755x10^{-1}(2.603x10^{-1}) = 0.2053$

```
> deltapartial<-delta<-c(rep(0, times=500))
> delta[1]<-0.2603
> delta[2]<--0.2053
> deltapartial[1]<-delta[1]
> deltapartial[2]<-deltapartial[1] + delta[2]
> for (j in 3:500){
+ k1<-j-1
+ k2<-j-2
+ delta[j]<-0.4755*delta[k1]+0.3376*delta[k2]
+ deltapartial[j]<-deltapartial[k1]+delta[j]
+ }
> deltapartial[500]*0.9
[1] -0.3311685
> deltapartial[1:20]
 [1]  0.26030000  0.05500000  0.04525713 -0.02868488 -0.06713351
-0.11037865
 [7] -0.14392197 -0.17447138 -0.20032185 -0.22292723 -0.24240320
-0.25929560
[13] -0.27390303 -0.28655174 -0.29749766 -0.30697266 -0.31517336
-0.32227155
[19] -0.32841530 -0.33373300
```

So the 90% interval can be calculated as,

19 + (-0.33116 + 0.32841)/(-0.33373 + 0.32841) ~ 19.5 months

2. The file RPCEGoods5219.txt contains quarterly data measuring U.S. Real Personal Consumption Expenditures for goods. The data are percentage changes for the years 1952 to 2019. The adjective “Real” indicates that the data are adjusted for inflation. The data are percentage changes from the previous quarter, and they represent seasonally adjusted annual rates.

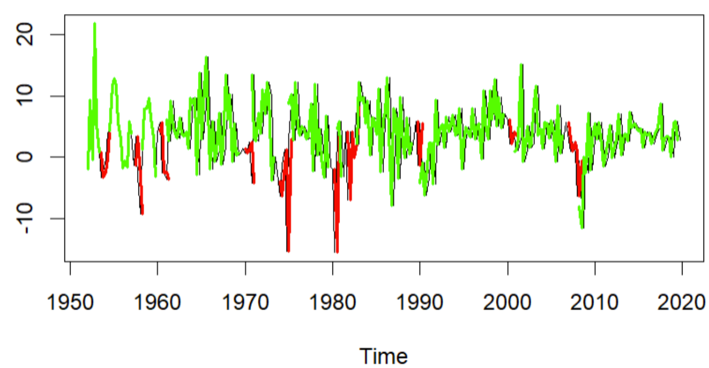
(a) For the years 1952 to 2019 make a list of the economic downturns as defined by the Business Cycle Dating Committee of the National Bureau of Economic Research.

Peak	Trough	Contraction Peak to Trough	Expansion Trough to Peak
Jul 53	May 54	10	45
Aug 57	Apr 58	8	39
Apr 60	Feb 61	10	24
Dec 69	Nov 70	11	106
Nov 73	Mar 75	16	36
Jan 80	Jul 80	6	58
Jul 81	Nov 82	16	12
Jul 90	Mar 91	8	92

Mar 01	Nov 01	8	120
Dec 07	Jun 09	18	73

(b) Plot the data vs. time and mark the periods of economic downturn from (a). Be careful in doing so, as the data are quarterly, rather than monthly. If a period of downturn spans at least one month of a quarter, include that quarter in the span of the downturn. Discuss features of the plot. How have the periods of downturn affected real personal consumption expenditures for goods?

```
> economic_downturn<-c(rep(NA, 6),PctChange[6:10], rep(NA, 11),
PctChange[22:25], rep(NA, 7), PctChange[33:37], rep(NA, 34),
PctChange[72:76], rep(NA, 11), PctChange[88:93], rep(NA, 19),
PctChange[113:115], rep(NA, 3), PctChange[119:124], rep(NA, 26),
PctChange[151:153], rep(NA, 39), PctChange[193:196], rep(NA, 23),
PctChange[220:226], rep(NA, 46))
> economic_upturn<-c(PctChange[1:6], rep(NA, 4), PctChange[11:21],
rep(NA, 4), PctChange[26:32], rep(NA, 4), PctChange[38:71], rep(NA, 5),
PctChange[77:87], rep(NA, 6), PctChange[94:112], rep(NA, 3),
PctChange[116:118], rep(NA, 6), PctChange[125:150], rep(NA, 3),
PctChange[154:192], rep(NA, 4), PctChange[197:219], rep(NA, 7),
PctChange[227:272])
> plot(ts(PctChange, start = c(1952, 1), frequency = 4))
> lines(ts(economic_downturn, start = c(1952,1), frequency = 4),
col='red', lwd=2)
> lines(ts(economic_upturn, start = c(1952, 1), frequency = 4),
col='green', lwd=2)
```



The percentage changes during the Nov 73 - Mar 75 economic downturn and the Jan 80 - Jul 80 economic downturn have gone down drastically. These both recessions have affected the real personal consumption expenditures for goods significantly. Also the Aug 57 - Apr 58 downturn also got the consumption expenditures down significantly.
The overall trend structure of the time series seems to be modest and constant throughout, not increasing or decreasing. The volatility in the expenditure of goods was high during most of the economic downturns but seems to be decreasing post 2000.
Moreover there is no visible seasonality in the time series data.

(c) An ARX model has the form

$$y_t = \mu + \phi_1(y_{t-1} - \mu) + \cdots + \phi_p(y_{t-p} - \mu) + \beta_1 x_{1t} + \cdots + \beta_k x_{kt} + \varepsilon_t.$$

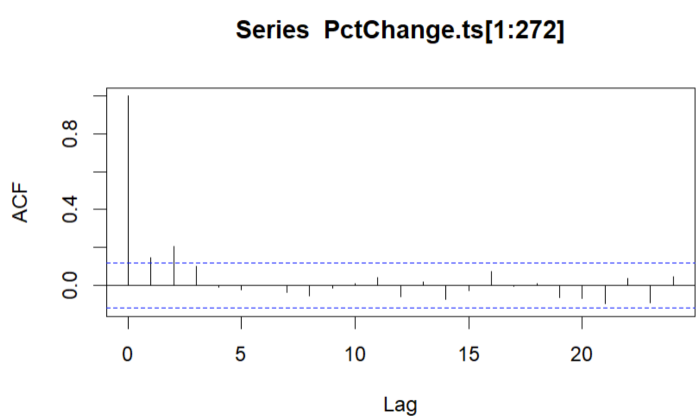
That is, it combines AR and regression structures. As an example, suppose we have identified two outliers and formed dummies for them, *d1* and *d2*. Then R commands to fit such a model to the percentage changes, if the chosen order of the AR structure is *p*, are as follows:

```
df<-data.frame(d1,d2)
arxmodel<-arima(Pctchange.ts,order=c(p,0,0),xreg=df)
```

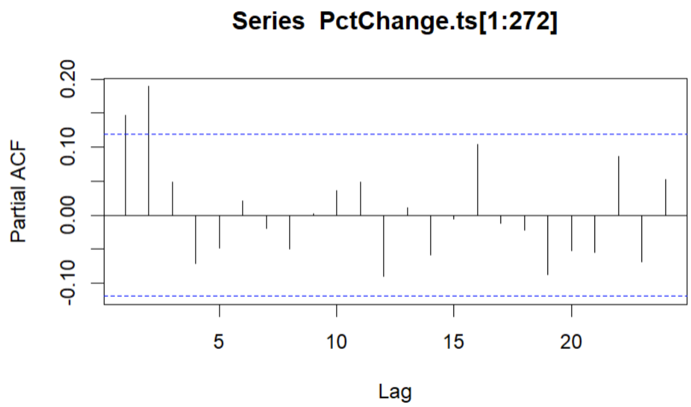
That is, set up a separate data frame to accommodate the outlier dummies, and designate this data frame in the `arima` command.

Fit an ARX model to the data. Explain in detail how you arrived at your model fit. [Hint: You will need to introduce several dummies for outliers. If you don't do so, the fit will be quite different. To verify this, refit the same AR(p) structure—without the use of the outlier dummies.]

```
> PctChange.ts<-ts(PctChange, start = c(1952, 1), frequency = 4)
> acf(PctChange.ts)
```




```
> PctChange.ts<-ts(PctChange, start = c(1952, 1), frequency = 4)
> pacf(PctChange.ts)
```



Looking at the pacf plot we can see that the most apt value for 'p' is 2. But we will use 4 as the p value anyways as we will be considering 2 pseudo cycles each time.

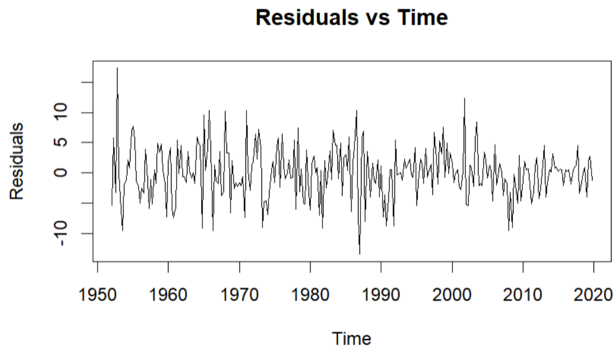
Now let us build the dataframe having the outlier values. From the data vs time plot above, we can see that outlier points are 25, 92 (during Nov 73 - Mar 75 economic downturn), 114 (during Jan 80 - Jul 80 economic downturn) and 228.

```
> out_25<-c(rep(0, 24), 1, rep(0, 247))
> out_92<-c(rep(0, 91), 1, rep(0, 180))
> out_114<-c(rep(0, 113), 1, rep(0, 158))
> out_228<-c(rep(0, 227), 1, rep(0, 44))
> df<-cbind(out_25, out_92, out_114, out_228)
> df<-data.frame(df)


> arxmodel<-arima(PctChange.ts,order=c(4,0,0),xreg=df)
> arxmodel

Call:
arima(x = PctChange.ts, order = c(4, 0, 0), xreg = df)

Coefficients:
         ar1     ar2     ar3      ar4  intercept    out_25    out_92
      0.0707  0.2393  0.0915  -0.1783     3.7086  -12.4014  -19.9490
s.e.  0.0607  0.0602  0.0602   0.0630     0.3385    4.1666    4.1456
       out_114   out_228
      -20.6027  -12.7922
s.e.    4.1985    4.1884

sigma^2 estimated as 18.66:  log likelihood = -784.07,  aic = 1588.14
```

(d) Examine the residuals to investigate whether your selected model has achieved reduction to white noise. For this purpose include in your discussion consideration of the plot of the residuals vs. time, the residual autocorrelations and partial autocorrelations, and the residual spectral density. Perform Bartlett's test to determine if the fit has produced reduction to white noise.

Residuals vs Time:

```
> plot(ts(resid(arxmodel), start = c(1952, 1), frequency = 4),
xlab='Time', ylab='Residuals', main='Residuals vs Time')
```



According to the plot, the model does not fully capture the trend structure. But the residuals are somewhat modest here.

ACF:

```
> acf(resid(arxmodel)[1:272])
```

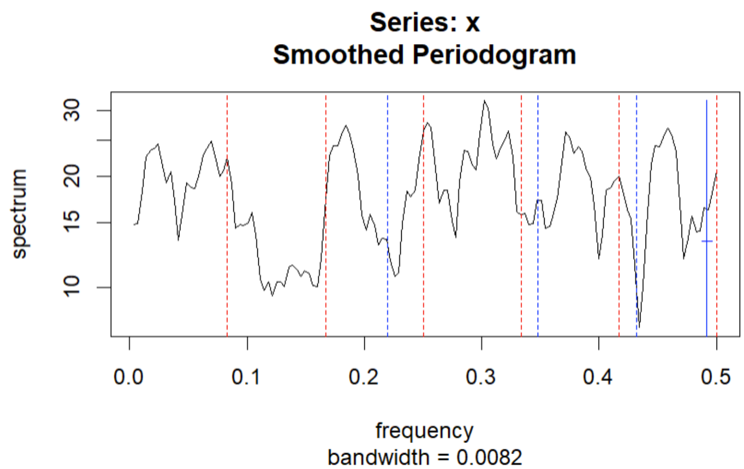
There are no significant autocorrelations at any lags. This is showing that the model has reduced the residuals to white noise.

PACF:

```
> pacf(resid(arxmodel)[1:272)
```

Spectral Density:

```
> spectrum(resid(arxmodel)[1:272], span=8)
> abline(v=c(1/12,2/12,3/12,4/12,5/12,6/12),col="red",lty=2)
> abline(v=c(0.220,0.348,0.432),col="blue",lty=2)
```



This spectral plot is not flat, indicating that the residuals do not conform to white noise structure. The spectral peak at low frequency is for slow movement of the time series, essentially remaining trend structure. The spectral plot red lines clearly show prominent spectral activity at frequencies 3/12, 5/12, 6/12 and perhaps at frequencies 1/12 and 2/12, precisely indicating the presence of unmodeled seasonal structure in the model residuals.
There is also some spectral activity around frequencies 0.348, perhaps signaling modest calendar effects.
On the other hand, if we double the length of the blue line above the notch, it is somewhat longer than the full vertical range of the spectral plot. This shows that the model has reduced the residuals to white noise structure.

BartlettB Test:

```
> bartlettB.test(ts(resid(arxmodel)))

      Bartlett B Test for white noise

data:
= 0.34284, p-value = 0.9998
```

We need to reject the null hypothesis if p-value<0.05. Here the p-value is 0.9998, so we need to accept the null hypothesis. This is white noise structure.

(e) Find the zeros of the autoregressive polynomial for your model fit and interpret the results.

```
> coef(arxmodel)
         ar1          ar2          ar3          ar4    intercept
out_25
  0.07073884   0.23930301   0.09151129  -0.17833248   3.70860915
-12.40141180
```

```
      out_92      out_114      out_228
-19.94899990 -20.60273872 -12.79218964
> zeros<-1/polyroot(c(1,-coef(arxmodel)[1:4]))
> zeros
[1]  0.5390301-0.3213239i -0.5036607-0.4462886i -0.5036607+0.4462886i
[4]  0.5390301+0.3213239i
```

Thus, the zeros are complex conjugate pairs 0.5390±0.321i and -0.5036±0.4462i. Let's find the amplitude and period for the complex conjugate pairs. The calculation uses only the complex zero with positive imaginary part. The phase is calculated, followed by the period.

Amplitude:

```
Mod(zeros[3])
```
For -0.5036607+0.4462886i : 0.6729

```
Mod(zeros[4])
```
For 0.5390301+0.3213239i : 0.6275

Period:

```
2*pi/Arg(zeros)[3]
```
For -0.5036607+0.4462886i : 2.6001

```
2*pi/Arg(zeros)[4]
```
For 0.5390301+0.3213239i : 11.688

(f) How do the results in part (e) compare to those described in the 15 March notes for U.S. real quarterly GNP data for the period 1947(2) to 1991(1)? Discuss in detail.

One pair suggests a stochastic cyclical component with period about 11.68 quarters, and the second pair points to a weaker stochastic cyclical component with period about 2.6 quarters.
As we know this is an AR(4) model fit, and here we are able to also see the shorter (and weaker) cyclical component. We conclude that the AR(4) model reveals features of the time series.
The results here are comparable to GNP data that we say in the March 15th notes. Even there the results were similar.

3. The file RPCEServices5219.txt contains quarterly data measuring U.S. Real Personal Consumption Expenditures for services. The data are percentage changes for the years 1952 to 2019. The adjective "Real" indicates that the data are adjusted for inflation. The data are percentage changes from the previous quarter, and they represent seasonally adjusted annual rates.

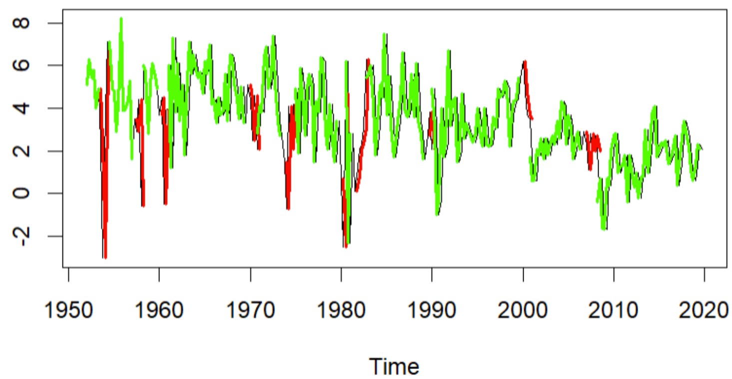
Answer questions (a)–(f) as in part 2. In addition, discuss how the results for services differ from those obtained for goods.

3a)

Peak	Trough	Contraction Peak to Trough	Expansion Trough to Peak
Jul 53	May 54	10	45
Aug 57	Apr 58	8	39
Apr 60	Feb 61	10	24
Dec 69	Nov 70	11	106
Nov 73	Mar 75	16	36
Jan 80	Jul 80	6	58
Jul 81	Nov 82	16	12
Jul 90	Mar 91	8	92
Mar 01	Nov 01	8	120
Dec 07	Jun 09	18	73

3b)

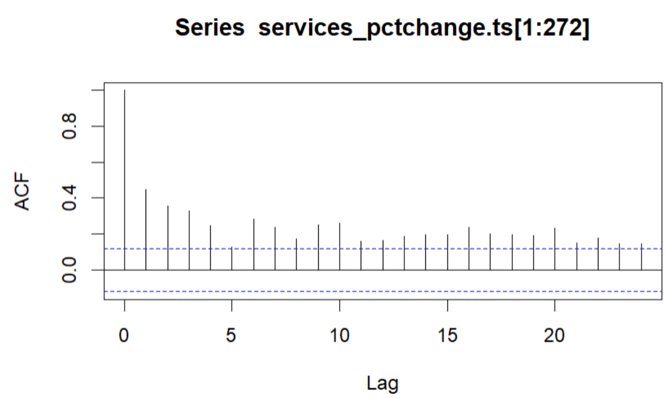
```
> economic_downturn<-c(rep(NA, 6),rpce_services$PctChange[6:10], rep(NA,
11), rpce_services$PctChange[22:25], rep(NA, 7),
rpce_services$PctChange[33:37], rep(NA, 34),
rpce_services$PctChange[72:76], rep(NA, 11),
rpce_services$PctChange[88:93], rep(NA, 19),
rpce_services$PctChange[113:115], rep(NA, 3),
rpce_services$PctChange[119:124], rep(NA, 26),
rpce_services$PctChange[151:153], rep(NA, 39),
rpce_services$PctChange[193:196], rep(NA, 23),
rpce_services$PctChange[220:226], rep(NA, 46))
>
> economic_upturn<-c(rpce_services$PctChange[1:6], rep(NA, 4),
rpce_services$PctChange[11:21], rep(NA, 4),
rpce_services$PctChange[26:32], rep(NA, 4),
rpce_services$PctChange[38:71], rep(NA, 5),
rpce_services$PctChange[77:87], rep(NA, 6),
rpce_services$PctChange[94:112], rep(NA, 3),
rpce_services$PctChange[116:118], rep(NA, 6),
rpce_services$PctChange[125:150], rep(NA, 3),
rpce_services$PctChange[154:192], rep(NA, 4),
rpce_services$PctChange[197:219], rep(NA, 7),
rpce_services$PctChange[227:272])
> plot(ts(rpce_services$PctChange, start = c(1952, 1), frequency = 4))
> lines(ts(economic_downturn, start = c(1952, 1), frequency = 4),
col='red', lwd=2)
> lines(ts(economic_upturn, start = c(1952, 1), frequency = 4),
col='green', lwd=2)
```



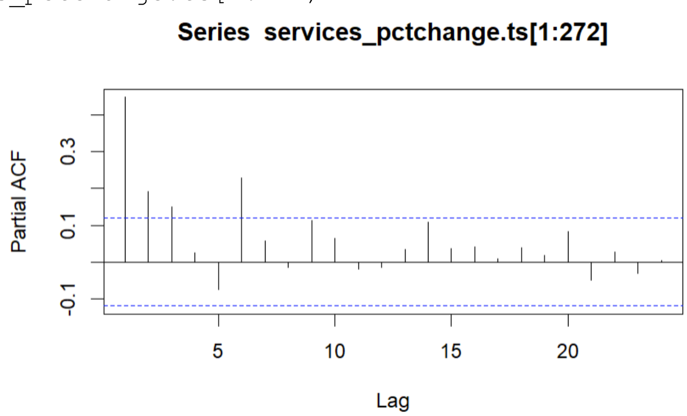
The overall trend of the Real Personal Consumption Expenditures for services is going somewhat down over the years. The volatility is also very high during the early years, during which there were also a couple of economic downturns. but the volatility seems to reduce post 2000 similar to the goods data.

3c) Let us plot the acf and pacf plots to help us determine the value of p for our AR model.

```
> services_pctchange.ts<-ts(rpce_services$PctChange, start = c(1952, 1),
frequency = 4)
> acf(services_pctchange.ts[1:272)
```




```
> pacf(services_pctchange.ts[1:272)
```



Looking at the pacf plot we see that the value of p could be 6, as after lag 6 the other lags seem to be insignificant.

First we need to find out the outliers and create a dataframe using those outliers. The outliers here are 8, 25, 35, 89, 114, 117.

Let us form a dataframe with these 6 outliers.

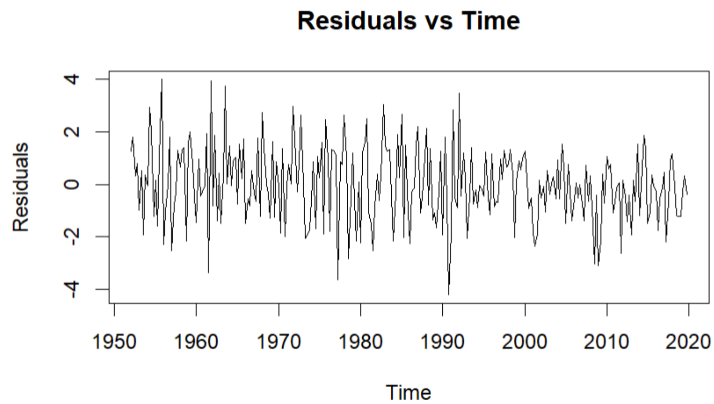
```
> out_8<-c(rep(0, 7), 1, rep(0, 264))
> out_25<-c(rep(0, 24), 1, rep(0, 247))
> out_35<-c(rep(0, 34), 1, rep(0, 237))
> out_89<-c(rep(0, 88), 1, rep(0, 183))
> out_114<-c(rep(0, 113), 1, rep(0, 158))
> out_117<-c(rep(0, 116), 1, rep(0, 155))
> df_services<-cbind(out_8, out_25, out_35, out_89, out_114, out_117)
> df_services<-data.frame(df_services)
```

Lets fit the model now,

```
> arxmodel_services<-arima(services_pctchange.ts, order = c(6,0,0), xreg
= df_services)
> arxmodel_services

Call:
arima(x = services_pctchange.ts, order = c(6, 0, 0), xreg = df_services)

Coefficients:
         ar1     ar2     ar3      ar4      ar5     ar6  intercept
out_8
      0.3957  0.2133  0.0510  -0.0442  -0.0931  0.2480     3.4700
-7.1685
s.e.  0.0597  0.0661  0.0649   0.0657   0.0669  0.0597     0.3639
1.2855
       out_25   out_35  out_89  out_114  out_117
      -5.0861  -4.1268  -2.695  -5.6022  -5.6695
s.e.   1.2759   1.3032   1.289   1.2862   1.2706

sigma^2 estimated as 2.003:  log likelihood = -480.89,  aic = 989.78
```

Residuals vs Time:

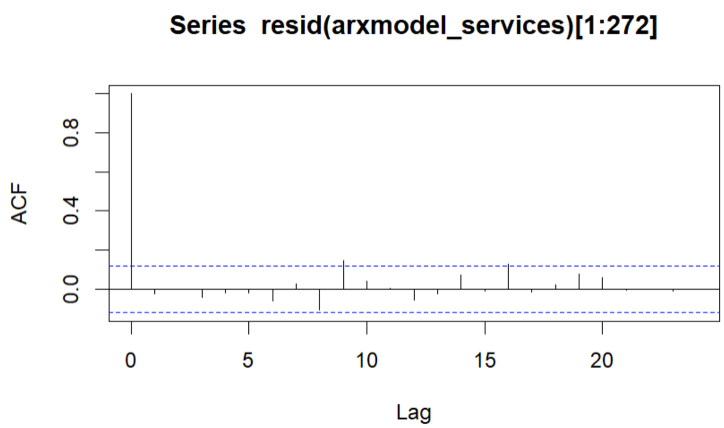
```
> plot(ts(resid(arxmodel_services), start = c(1952, 1), frequency = 4),
xlab='Time', ylab='Residuals', main='Residuals vs Time')
```



According to the plot, the model does not fully capture the trend structure. But the trend structure is modest here.

ACF:

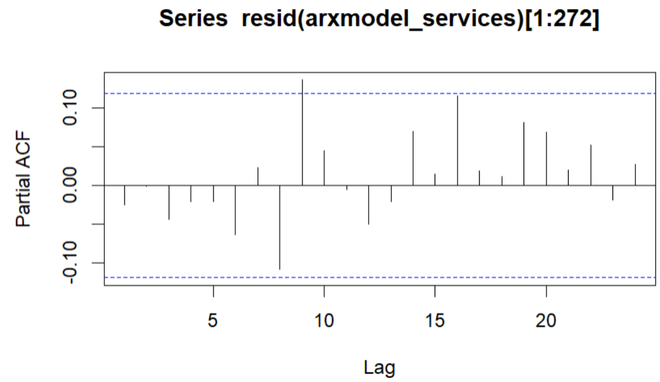
```
> acf(resid(arxmodel_services)[1:272])
```



There are no significantly visible significant lags, meaning that the model has reduced the residuals to white noise.

PACF:

```
> pacf(resid(arxmodel_services)[1:272])
```



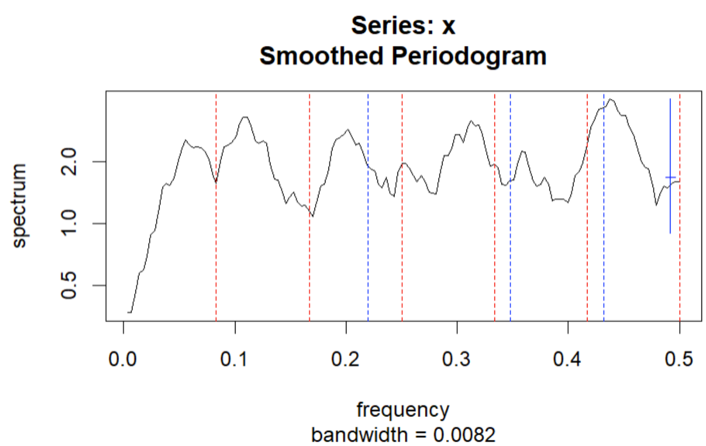
BartlettB Test:

```
> bartlettB.test(ts(resid(arxmodel_services)))

      Bartlett B Test for white noise

data:
= 0.53488, p-value = 0.9372
```

The null hypothesis is white noise structure. We need to reject the null hypothesis if the p-value<0.05. But here the p-value is 0.9372. We need to accept the null hypothesis here. This means this is white noise structure

Spectral Density:

```
> spectrum(resid(arxmodel_services)[1:272], span=8)
> abline(v=c(1/12,2/12,3/12,4/12,5/12,6/12),col="red",lty=2)
> abline(v=c(0.220,0.348,0.432),col="blue",lty=2)
```



The spectral peak at low frequency is for slow movement of the time series, essentially remaining trend structure. The spectral plot red lines clearly show prominent spectral activity at frequencies 3/12 and perhaps at frequencies 5/12. There is not much seasonal structure to be captured.
If we double the length of the blue line above the notch, it is not more than the full vertical range of the spectral plot. This shows that it is not white noise structure.
But this blue line is just a guide.

Looking at all the other analyses above, we can say that this model might be reduced to white noise structure.

Zeros:

```
> coef(arxmodel_services)
        ar1         ar2         ar3         ar4         ar5         ar6 
 0.39567482  0.21330255  0.05104185 -0.04418748 -0.09314537  0.24797432 
  intercept       out_8      out_25      out_35      out_89     out_114 
 3.46997558 -7.16850755 -5.08606061 -4.12684207 -2.69502273 -5.60215013 
    out_117 
-5.66950030 
> zeros<-1/polyroot(c(1,-coef(arxmodel_services)[1:6]))
> zeros
[1]  0.4986627-0.5715416i -0.3610006-0.6908813i -0.3610006+0.6908813i
[4]  0.9045399+0.0000000i -0.7841894-0.0000000i  0.4986627+0.5715416i
```

Amplitudes:

`Mod(zeros[3])`
For -0.3610006+0.6908813i : 0.7795

`Mod(zeros[4])`
For 0.9045399+0.0000000i : 0.9045

`Mod(zeros[6])`
For 0.4986627+0.5715416i : 0.7585

Periods:

`2*pi/Arg(zeros)[3]`
For -0.3610006+0.6908813i : 3.061

`2*pi/Arg(zeros)[4]`
For 0.9045399+0.0000000i : 3.85e+14

`2*pi/Arg(zeros)[6]`
For 0.4986627+0.5715416i : 7.362

3f)

One pair suggests a stochastic cyclical component with period about 7.362 quarters, and the second pair points to a weaker stochastic cyclical component with period about 3.061 quarters.
As we know this is an AR(6) model fit, and here we are able to also see the shorter (and weaker) cyclical component. We conclude that the AR(6) model reveals features of the time series.