Statistics 5350/7110 Fall 2023
P. Shaman

Assignment 3 solution

1. This problem involves examination of monthly Lydia Pinkham data for the period 1946 through 1960(6). It addresses model construction and estimation of the 90 per cent duration interval. The data are in Lydiamonthly4660.txt.

```
> lydiam<-read.csv("F:/Stat71121Fall/Lydiamonthly.txt")
> attach(lydiam)
> head(lydiam)
  yrmon month jan46 oct47 jan48 dec57 jan58   madv msales madvl1 madvl2 madvl3
1  4601     1     1     0     0     0     0  79300 226900  19800  97300 155800
2  4602     2     0     0     0     0     0  67200 167500  79300  19800  97300
3  4603     3     0     0     0     0     0 107700 194500  67200  79300  19800
4  4604     4     0     0     0     0     0 118700 175500 107700  67200  79300
5  4605     5     0     0     0     0     0  64500 185900 118700 107700  67200
6  4606     6     0     0     0     0     0  46800 155900  64500 118700 107700
  madvl4 msalesl1 msalesl2 msalesl3 msalesl4       c220       s220        c348
1 135600   144600   183500   254900   189300  0.1873813  0.9822873 -0.57757270
2 155800   226900   144600   183500   254900 -0.9297765  0.3681246 -0.33281954
3  97300   167500   226900   144600   183500 -0.5358268 -0.8443279  0.96202767
4  19800   194500   167500   226900   144600  0.7289686 -0.6845471 -0.77846230
5  79300   175500   194500   167500   226900  0.8090170  0.5877853 -0.06279052
6  67200   185900   175500   194500   167500 -0.4257793  0.9048271  0.85099448
        s348
1  0.8163393
2 -0.9429905
3  0.2729519
4  0.6276914
5 -0.9980267
6  0.5251746
```

(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.

```
> options(digits=10)

> model1<-
lm(msales~fmonth+jan46+oct47+jan48+dec57+jan58+madv+madvl1+msalesl1+msalesl2+
c220+s220+c348+s348);summary(model1)

Call:
lm(formula = msales ~ fmonth + jan46 + oct47 + jan48 + dec57 +
    jan58 + madv + madvl1 + msalesl1 + msalesl2 + c220 + s220 +
    c348 + s348)

Residuals:
        Min           1Q      Median           3Q          Max
-27927.4230  -6220.4592   -549.4605   6555.1153  24574.4432
```

```
Coefficients:
                 Estimate    Std. Error  t value   Pr(>|t|)
(Intercept)  3.565673e+04  6.811838e+03  5.23452 5.5301e-07 ***
fmonth2     -1.744719e+04  5.570723e+03 -3.13194 0.00209083 **
fmonth3     -1.943522e+03  5.676609e+03 -0.34237 0.73255207
fmonth4     -2.228692e+04  5.930027e+03 -3.75832 0.00024500 ***
fmonth5     -1.574951e+04  6.090322e+03 -2.58599 0.01066812 *
fmonth6     -9.277141e+03  5.937868e+03 -1.56237 0.12032254
fmonth7      4.392956e+03  5.415842e+03  0.81113 0.41858433
fmonth8      1.394417e+03  4.798395e+03  0.29060 0.77176081
fmonth9     -2.333252e+03  4.877995e+03 -0.47832 0.63312213
fmonth10    -3.370832e+03  5.714488e+03 -0.58987 0.55616817
fmonth11    -3.537813e+04  6.515614e+03 -5.42975 2.2424e-07 ***
fmonth12    -2.245449e+04  6.559967e+03 -3.42296 0.00080000 ***
jan46        7.219502e+04  1.189160e+04  6.07109 1.0077e-08 ***
oct47        4.356592e+04  1.182855e+04  3.68311 0.00032168 ***
jan48        5.345158e+04  1.170575e+04  4.56627 1.0324e-05 ***
dec57        4.650332e+04  1.178132e+04  3.94721 0.00012150 ***
jan58       -6.670709e+04  1.213951e+04 -5.49504 1.6505e-07 ***
madv         2.115962e-01  3.949120e-02  5.35806 3.1311e-07 ***
madvl1       1.389372e-01  4.270569e-02  3.25337 0.00141105 **
msalesl1     5.254221e-01  6.002201e-02  8.75382 4.1504e-15 ***
msalesl2     1.231413e-01  5.585844e-02  2.20452 0.02902087 *
c220         2.922418e+03  1.220180e+03  2.39507 0.01785926 *
s220        -3.211161e+03  1.210079e+03 -2.65368 0.00882598 **
c348        -5.909778e+02  1.249074e+03 -0.47313 0.63681090
s348         3.932936e+03  1.226906e+03  3.20557 0.00164942 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 11065.6 on 149 degrees of freedom
Multiple R-squared:  0.8580157, Adjusted R-squared:  0.8351457
F-statistic: 37.51715 on 24 and 149 DF,  p-value: < 2.2204e-16
```

Monthly sales are modeled with seasonal dummy variables, contemporaneous advertising, advertising lagged one month, sales lagged one and two months, calendar trigonometric pairs for frequencies 0.220 and 0.348, and dummy variables for outliers at January 1946, October 1947, January 1948, December 1957, and January 1958. *R* square is 0.858, and the residual standard error is 11,066.

(b) ***This part is optional. Points will not be deducted if you do not attempt it. And points will not be deducted if you do present it and there are mistakes.*** Find the static seasonal estimates stemming from your model, and tabulate, plot, and interpret them.

Let's calculate estimates of the seasonal indices. This calculation requires care. It is complicated by the presence of lag 1 and lag 2 monthly sales as independent variables in the regression. The fitted model can be written

$$msales_t = 0.5254221 msales_{t-1} + 0.1231413 msales_{t-2} + S_t^* + \text{the rest},$$

or

$$\left(1 - 0.5254221B - 0.1231413B^2\right) msales_t = S_t^* + \text{the rest},$$

or

$$msales_t = \left(1 - 0.5254221B - 0.1231413B^2\right)^{-1} S_t^* + \text{the rest},$$

where the desired seasonal index estimates are given by

(1) $\left(1 - 0.5254221B - 0.1231413B^2\right)^{-1} S_t^* = \left(\delta_0 + \delta_1 B + \delta_2 B^2 + \delta_3 B^3 + \cdots\right) S_t^*$ **.**

First, we calculate estimates of the S_t^* values from the given model output. These values do not make adjustment for inclusion of lag 1 and lag 2 monthly sales in the model.

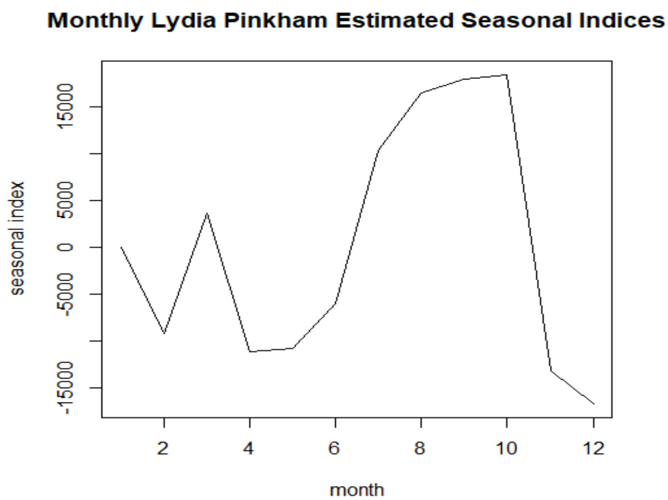
```
> b1<-coef(model1)[1]
> b2<-coef(model1)[2:12]+b1
> b3<-c(b1,b2)
> Sstar<-b3-mean(b3)

> Sstar
  (Intercept)        fmonth2        fmonth3        fmonth4        fmonth5
 10371.134189  -7076.051011   8427.611788 -11915.785678  -5378.373366
      fmonth6        fmonth7        fmonth8        fmonth9       fmonth10
  1093.992998  14764.090655  11765.551226   8037.881786   7000.302551
     fmonth11       fmonth12
-25006.996262 -12083.358877
```

Next, we obtain the desired seasonal index estimates by calculating the right-hand side of (1)**. We first calculate the deltas using**

$$1 = \left(1 - 0.5254221B - 0.1231413B^2\right)\left(\delta_0 + \delta_1 B + \delta_2 B^2 + \delta_3 B^3 + \cdots\right).$$

```
> delta<-c(rep(0,times=60))
> delta[1]<-1;delta[2]<-0.5254221
> for(j in 3:60){
+ j1<-j-1;j2<-j-2
+ delta[j]<-0.5254221*delta[j1]+0.1231413*delta[j2]
+ }

> delta[57:60]
[1] 1.843029099e-09 1.292092046e-09 9.058467155e-10 6.350617778e-10

> s6<-c(rep(Sstar,6))
> for(i in 1:12){
+ ir<-73-i
+ for(j in 2:60){
+ ir2<-ir-j+1
+ s6[ir]<-s6[ir]+delta[j]*s6[ir2]
+ }
+ }
> seas<-s6[61:72]
```

```
> seas
    (Intercept)          fmonth2          fmonth3          fmonth4          fmonth5
   -12.08884761  -9137.88349074   3624.87721562 -11136.44592658 -10783.33607513
        fmonth6          fmonth7          fmonth8          fmonth9         fmonth10
 -5943.16651626  10313.54559398  16452.66675274  17952.49990950  18458.94552263
       fmonth11         fmonth12
-13097.56415375 -16692.04998442

> sum(seas)
[1] 3.177547114e-11
```



The seasonal index estimates show that monthly sales are most prominent in the months July through October, and least prominent in the months November, December, February, April and May, this relative to the base level predicted by advertising expenditures, calendar variables, and adjustments for outlier values.

(c) 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:

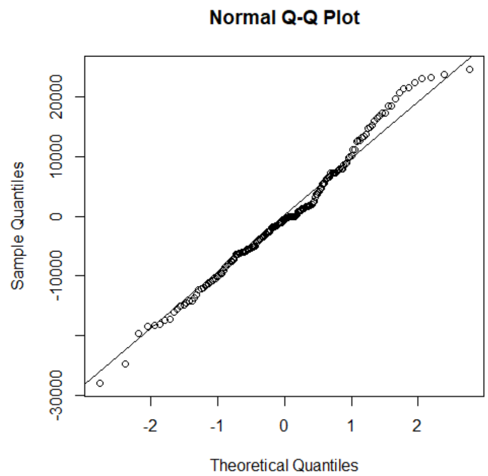
```
install.packages ("hwwntest")
library("hwwntest")
```

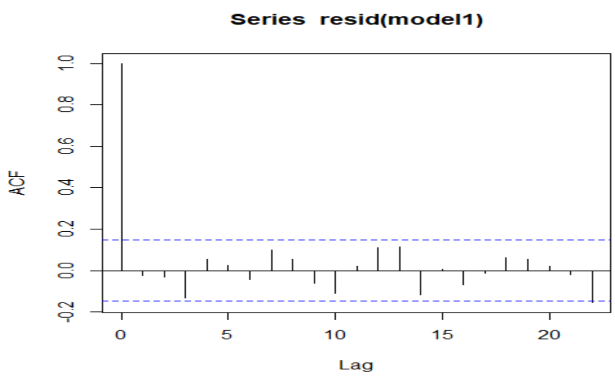
The null hypothesis is white noise structure. Give the command

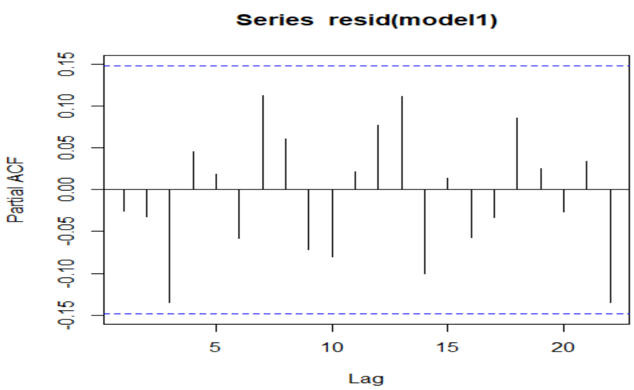
```
bartlettB.test(ts(resid(model)))
```

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.

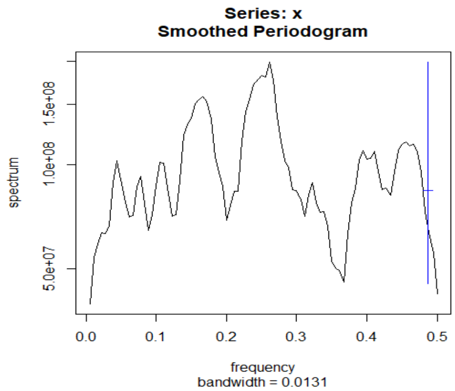






The model residuals deviate slightly from normality—part of the upper tail is too long. The residual acf and pacf plots do show adequate reduction to white noise.

```
> spectrum(resid(model1),span=8)
```




```
> library("hwwntest")
> bartlettB.test(ts(resid(model1)))

        Bartlett B Test for white noise

data:
= 0.64516, p-value = 0.7995
```

Bartlett's test confirms that the model 1 residuals are consistent with reduction to white noise.

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

```
> deltapartial<-delta<-c(rep(0,times=500))
> #deltapartial is the partial sum of the deltas
> delta[1]<-0.211596;delta[2]<-0.250114
> deltapartial[1]<-delta[1]
> deltapartial[2]<-deltapartial[1]+delta[2]
> for(j in 3:500){
+ j1<-j-1;j2<-j-2
+ delta[j]<-0.525422*delta[j1]+0.123141*delta[j2]
+ deltapartial[j]<-deltapartial[j1]+delta[j]
+ }

> deltapartial[500]*0.9
[1] 0.897684438

> deltapartial[1:20]
 [1] 0.2115960000 0.4617100000 0.6191815411 0.7327198413 0.8117665651
 [6] 0.8672806726 0.9061828987 0.9334590468 0.9525809941 0.9659868980
[11] 0.9753853506 0.9819743208 0.9865936455 0.9898321127 0.9921025029
[16] 0.9936942040 0.9948100968 0.9955924152 0.9961408746 0.9965253827
```

Linear interpolation gives the 90 per cent duration interval,

$6 + (0.897684 - 0.867281)/(0.906183 - 0.867281) = 6.78$ **months.**

An alternative model follows.

```
> model2<-
lm(msales~fmonth+jan46+oct47+jan48+dec57+jan58+madv+madvl1+madvl2+msalesl1+ms
alesl2+c220+s220+c348+s348);summary(model2)

Call:
lm(formula = msales ~ fmonth + jan46 + oct47 + jan48 + dec57 +
    jan58 + madv + madvl1 + madvl2 + msalesl1 + msalesl2 + c220 +
    s220 + c348 + s348)

Residuals:
        Min           1Q      Median          3Q         Max
-24087.6606  -5873.0314   -387.5342   6313.3675  26000.4870

Coefficients:
                 Estimate    Std. Error  t value   Pr(>|t|)
(Intercept)  3.670950e+04  6.755788e+03  5.43379 2.2187e-07 ***
fmonth2     -1.277607e+04  5.948587e+03 -2.14775 0.03336173 *
fmonth3     -3.925549e+03  5.694233e+03 -0.68939 0.49165711
fmonth4     -2.470635e+04  5.978731e+03 -4.13237 5.9859e-05 ***
fmonth5     -1.819447e+04  6.136651e+03 -2.96489 0.00353059 **
fmonth6     -1.111475e+04  5.938434e+03 -1.87166 0.06322685 .
fmonth7      4.960016e+03  5.363134e+03  0.92484 0.35655703
fmonth8      4.232073e+03  4.937363e+03  0.85715 0.39274636
fmonth9      8.562132e+02  5.061616e+03  0.16916 0.86590331
fmonth10    -2.271350e+03  5.676187e+03 -0.40015 0.68961987
fmonth11    -3.672616e+04  6.476317e+03 -5.67084 7.2234e-08 ***
fmonth12    -2.505261e+04  6.606638e+03 -3.79204 0.00021707 ***
jan46        6.996350e+04  1.180940e+04  5.92439 2.1088e-08 ***
oct47        4.261426e+04  1.170725e+04  3.63999 0.00037614 ***
jan48        5.253649e+04  1.158521e+04  4.53479 1.1825e-05 ***
dec57        4.593507e+04  1.165480e+04  3.94130 0.00012458 ***
jan58       -6.735829e+04  1.200993e+04 -5.60855 9.7277e-08 ***
madv         2.162820e-01  3.912119e-02  5.52851 1.4218e-07 ***
madvl1       1.193237e-01  4.327311e-02  2.75746 0.00655952 **
madvl2       8.640550e-02  4.149581e-02  2.08227 0.03903892 *
msalesl1     5.089222e-01  5.988773e-02  8.49794 1.9105e-14 ***
msalesl2     9.815086e-02  5.653208e-02  1.73620 0.08460993 .
c220         2.891021e+03  1.206841e+03  2.39553 0.01784635 *
s220        -3.163579e+03  1.196974e+03 -2.64298 0.00910267 **
c348        -6.479857e+02  1.235626e+03 -0.52442 0.60077166
s348         3.548407e+03  1.227370e+03  2.89107 0.00441805 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10943.77 on 148 degrees of freedom
Multiple R-squared:  0.8620569, Adjusted R-squared:  0.8387557
F-statistic: 36.99624 on 25 and 148 DF,  p-value: < 2.2204e-16 >
```

```
> deltapartial<-delta<-c(rep(0,times=500))
> delta[1]<-0.216282;delta[2]<-0.229395;delta[3]<-0.224378
> deltapartial[1]<-delta[1]
> deltapartial[2]<-deltapartial[1]+delta[2]
> deltapartial[3]<-deltapartial[2]+delta[3]
> for(j in 4:500){
+ j1<-j-1;j2<-j-2
+ delta[j]<-0.508922*delta[j1]+0.098151*delta[j2]
+ deltapartial[j]<-deltapartial[j1]+delta[j]
+ }
> deltapartial[500]*0.9
[1] 0.966618959

> deltapartial[1:20]
 [1] 0.2162820000 0.4456770000 0.6700550000 0.8067612492 0.8983569920
 [6] 0.9583899357 0.9979322352 1.0239484748 1.0410698277 1.0523367808
[11] 1.0597512591 1.0646305129 1.0678414109 1.0699544112 1.0713449164
[16] 1.0722599682 1.0728621376 1.0732584082 1.0735191825 1.0736907906
```

For this model the estimate of the 90 per cent duration interval is

6 + (0.966619 −0.958390)/(0.997932 −0.958390) = 6.21 **months.**

(e) Investigate Granger causality for sales vs. advertising, and for advertising vs. sales. Discuss your results.

Does advertising Granger cause sales?

```
> models1<-
lm(msales~fmonth+jan46+oct47+jan48+dec57+jan58+msalesl1+msalesl2+msalesl3+msa
lesl4+madvl1+madvl2+madvl3+madvl4);summary(models1)

Call:
lm(formula = msales ~ fmonth + jan46 + oct47 + jan48 + dec57 + 
    jan58 + msalesl1 + msalesl2 + msalesl3 + msalesl4 + madvl1 + 
    madvl2 + madvl3 + madvl4)

Residuals:
   Min     1Q Median     3Q    Max 
-31001  -7738   -944   8021  31883 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  4.200e+04  8.444e+03   4.973 1.79e-06 ***
fmonth2     -1.150e+04  7.442e+03  -1.545 0.124384    
fmonth3     -2.853e+03  8.476e+03  -0.337 0.736850    
fmonth4     -2.981e+04  8.121e+03  -3.671 0.000336 ***
fmonth5     -2.600e+04  6.788e+03  -3.830 0.000188 ***
fmonth6     -2.544e+04  5.537e+03  -4.595 9.17e-06 ***
fmonth7     -9.249e+03  5.258e+03  -1.759 0.080643 .  
fmonth8     -8.926e+02  6.072e+03  -0.147 0.883337    
fmonth9      1.507e+03  7.433e+03   0.203 0.839603    
fmonth10    -5.322e+03  8.476e+03  -0.628 0.531029    
fmonth11    -5.202e+04  8.057e+03  -6.456 1.43e-09 ***
fmonth12    -4.627e+04  6.279e+03  -7.368 1.10e-11 ***
```

```
jan46        5.824e+04  1.364e+04   4.269 3.48e-05 ***
oct47        4.270e+04  1.309e+04   3.261 0.001375 **
jan48        3.750e+04  1.365e+04   2.746 0.006769 **
dec57        4.690e+04  1.309e+04   3.584 0.000458 ***
jan58       -6.436e+04  1.350e+04  -4.768 4.38e-06 ***
msalesl1     4.237e-01  7.332e-02   5.779 4.25e-08 ***
msalesl2     8.264e-02  7.158e-02   1.154 0.250167
msalesl3     9.686e-02  7.207e-02   1.344 0.181020
msalesl4     1.406e-01  6.332e-02   2.221 0.027886 *
madvl1       1.566e-01  4.670e-02   3.353 0.001015 **
madvl2       1.182e-01  4.837e-02   2.443 0.015749 *
madvl3      -7.111e-03  4.934e-02  -0.144 0.885588
madvl4      -1.074e-01  4.822e-02  -2.227 0.027468 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 12350 on 149 degrees of freedom
Multiple R-squared: 0.8231,     Adjusted R-squared: 0.7946
F-statistic: 28.88 on 24 and 149 DF,  p-value: < 2.2e-16

> models2<-
lm(msales~fmonth+jan46+oct47+jan48+dec57+jan58+msalesl1+msalesl2+msalesl3+msa
lesl4)

> anova(models2,models1)
Analysis of Variance Table

Model 1: msales ~ fmonth + jan46 + oct47 + jan48 + dec57 + jan58 + msalesl1 +
    msalesl2 + msalesl3 + msalesl4
Model 2: msales ~ fmonth + jan46 + oct47 + jan48 + dec57 + jan58 + msalesl1 +
    msalesl2 + msalesl3 + msalesl4 + madvl1 + madvl2 + madvl3 +
    madvl4
  Res.Df        RSS Df  Sum of Sq      F    Pr(>F)
1    153 2.7354e+10
2    149 2.2734e+10  4 4620350605 7.5706 1.402e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Yes, advertising does Granger cause sales.

Now consider whether sales Granger causes advertising. The November 1950 value for monthly advertising is an outlier, and we define a dummy variable for it.

```
> nov50<-c(rep(0,58),1,rep(0,115))

> modela1<-
lm(madv~fmonth+nov50+msalesl1+msalesl2+msalesl3+msalesl4+madvl1+madvl2+madvl3
+madvl4);summary(modela1)

Call:
lm(formula = madv ~ fmonth + nov50 + msalesl1 + msalesl2 + msalesl3 +
    msalesl4 + madvl1 + madvl2 + madvl3 + madvl4)

Residuals:
   Min     1Q Median     3Q    Max
```

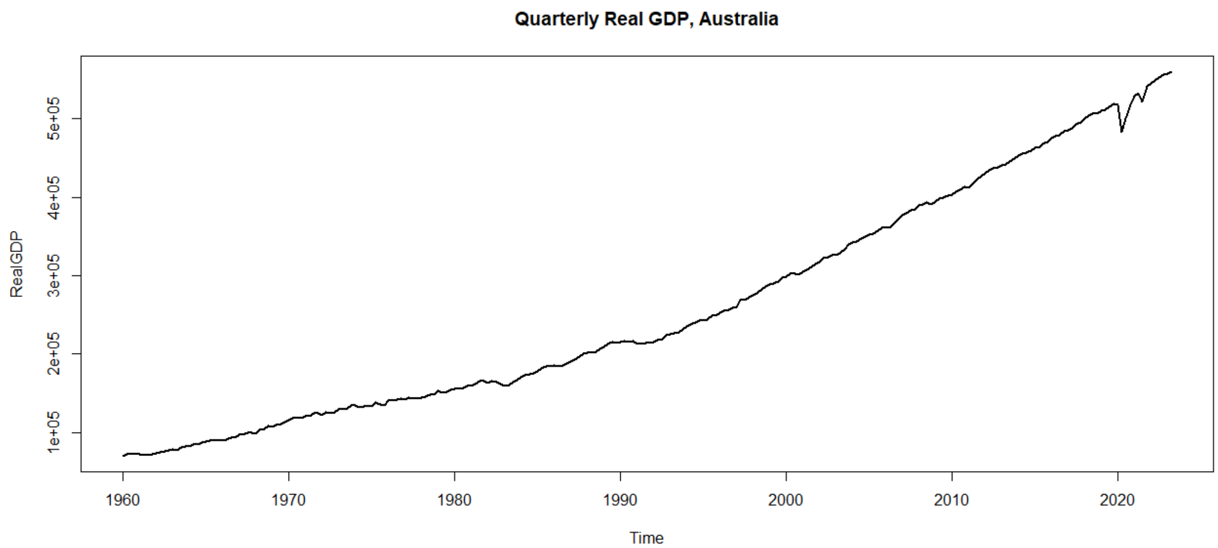
```
-52344 -13260   -836  12758  55265

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  6.023e+04  1.434e+04   4.200 4.51e-05 ***
fmonth2     -1.253e+04  1.291e+04  -0.971 0.333325    
fmonth3     -2.199e+04  1.470e+04  -1.496 0.136651    
fmonth4     -3.636e+04  1.386e+04  -2.624 0.009580 ** 
fmonth5     -6.417e+04  1.175e+04  -5.460 1.89e-07 ***
fmonth6     -8.381e+04  9.388e+03  -8.927 1.28e-15 ***
fmonth7     -7.221e+04  9.037e+03  -7.991 3.02e-13 ***
fmonth8     -4.100e+04  1.044e+04  -3.926 0.000130 ***
fmonth9     -2.105e+04  1.284e+04  -1.639 0.103334    
fmonth10    -3.291e+04  1.437e+04  -2.290 0.023375 *  
fmonth11    -9.260e+04  1.402e+04  -6.603 6.25e-10 ***
fmonth12    -1.041e+05  1.073e+04  -9.697  < 2e-16 ***
nov50        8.368e+04  2.311e+04   3.621 0.000399 ***
msalesl1    -7.636e-02  1.233e-01  -0.619 0.536542    
msalesl2     2.248e-01  1.254e-01   1.793 0.075022 .  
msalesl3     1.753e-01  1.196e-01   1.466 0.144818    
msalesl4     8.757e-02  1.115e-01   0.785 0.433620    
madvl1       2.927e-01  8.174e-02   3.581 0.000459 ***
madvl2      -3.658e-02  8.473e-02  -0.432 0.666586    
madvl3      -1.379e-01  8.849e-02  -1.559 0.121057    
madvl4      -1.485e-01  8.381e-02  -1.772 0.078347 .  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 21940 on 153 degrees of freedom
Multiple R-squared: 0.7689,     Adjusted R-squared: 0.7387 
F-statistic: 25.46 on 20 and 153 DF,  p-value: < 2.2e-16

> modela2<-lm(madv~fmonth+nov50+madvl1+madvl2+madvl3+madvl4)

> anova(modela2,modela1)
Analysis of Variance Table

Model 1: madv ~ fmonth + nov50 + madvl1 + madvl2 + madvl3 + madvl4
Model 2: madv ~ fmonth + nov50 + msalesl1 + msalesl2 + msalesl3 + msalesl4 + 
    madvl1 + madvl2 + madvl3 + madvl4
  Res.Df        RSS Df  Sum of Sq      F   Pr(>F)   
1    157 8.2037e+10                                 
2    153 7.3638e+10  4 8398623262 4.3625 0.002279 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Yes, sales does Granger cause advertising.

2. The file GDPAustralia.txt contains quarterly data measuring Real Gross Domestic Product for Australia from 1960 to the second quarter of 2023. The data variables are the GDP figure each quarter and change in GDP from the previous quarter. The adjective "Real" indicates that the data have been adjusted for inflation. The GDP readings have been seasonally adjusted.

```
> ausgdp<-read.csv("F:/Stat711023Fall/GDPAustralia.txt")
> attach(ausgdp)

> head(ausgdp)
     Quarter Time RealGDP dRealGDP
1 1960-01-01    1   69787      345
2 1960-04-01    2   71718     1931
3 1960-07-01    3   71878      160
4 1960-10-01    4   71737     -141
5 1961-01-01    5   71952      215
6 1961-04-01    6   71160     -792
```

(a) Plot the quarterly GDP data and their differences, each vs. time and discuss features of the two plots. (i) Are there any unusual data points? If you do identify unusual data points, comment and discuss their causes. (ii) What can you conclude about the overall behavior of GDP and the quarterly differences of GDP during the period from 1960 to 2023? Explain.



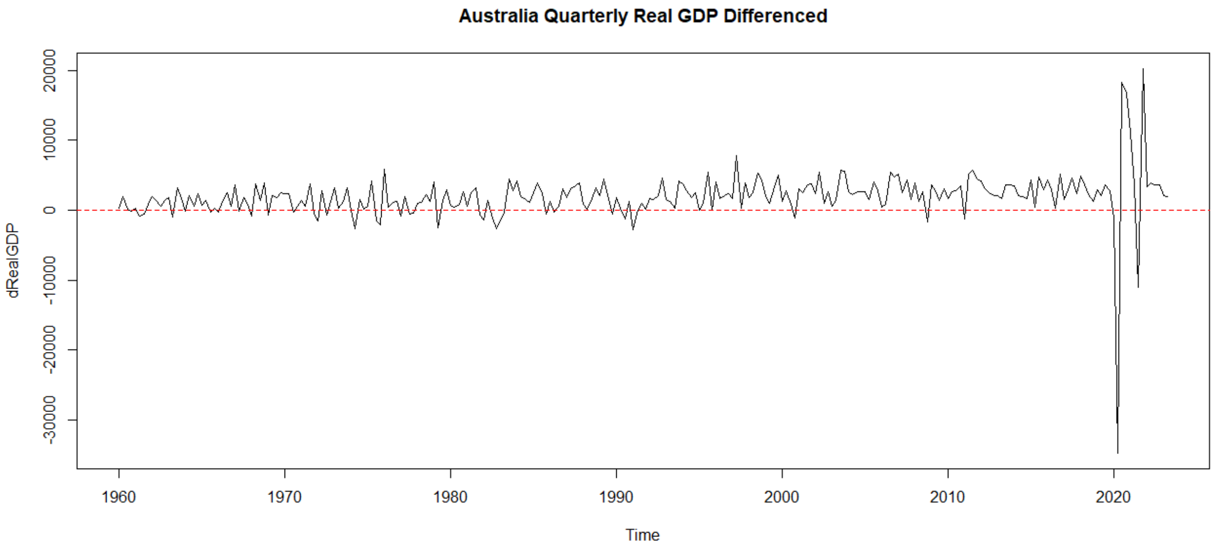
The plot of Australian quarterly Real GDP shows, with the exception of small dips in the early 1980s and early 1990s, steady upward growth until the outbreak of the COVID-19 global pandemic. Much of this was fueled by population growth. Australian government data list the following recessions since 1960:

1961(2) to 1961(3)	**2 quarters**
1965(3) to 1966(1)	**3 quarters**
1971(4) to 1972(1)	**2 quarters**
1975(3) to 1975(4)	**2 quarters**
1977(3) to 1977(4)	**2 quarters**
1981(4) to 1983(2)	**7 quarters**
1990(3) to 1991(2)	**4 quarters**
2020(3) to 2020(4)	**2 quarters**

As the plot shows, the 2020 recession caused a notable decline in Real GDP.

There was no Australian recession at the time of the U.S. 2008–09 Great Financial Crisis, because Australian banks had little exposure to the U.S. housing market and to U.S. banks.

The list of recessions shows continuous growth from 1991 to 2019. However, the Federal Reserve Bank of St. Louis has issued a paper disputing this assertion.[1] The disagreement relates to how one defines a recession.



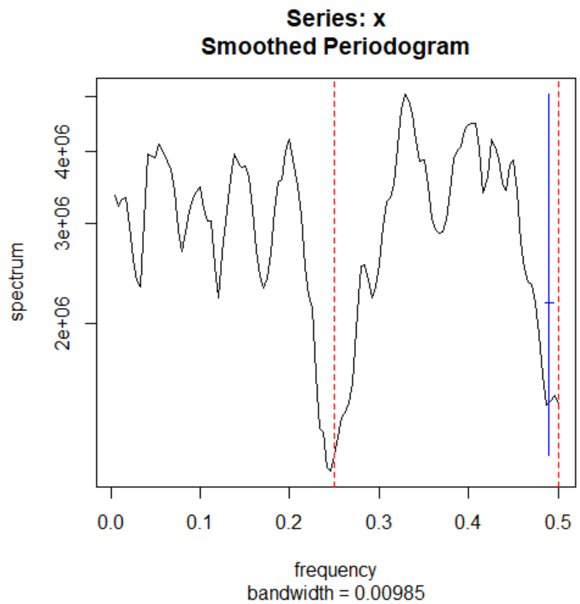
The plot of differenced RealGDP shows the differencing operation has not fully eliminated trend structure. It shows a gentle upward trend, especially after 1991. There is a drop in growth leading up to and during the 1990–91 recession. The effect of the COVID-19 pandemic is quite severe. Although the quarterly data have been deseasonalized, there appears to be some remaining seasonal behavior. Volatility is relatively stable prior to 2020.

[1] https://www.stlouisfed.org/on-the-economy/2019/september/australia-28-year-expansion

For the remainder of part 2, restrict attention to the years 1960 to 2019 for all calculations and discussion.

(b) Form the estimated spectral density for the differenced series. Discuss carefully what the plot reveals.

```
> spectrum(dRealGDP[1:240],8)
> abline(v=c(1/4,1/2),lty=2,col="red")
```




```
> bartlettB.test(dRealGDP)

        Bartlett B Test for white noise

data:
= 1.2484, p-value = 0.08857
```

The spectral plot suggests the differenced Real GDP data have a weak signal, as the estimated spectrum is rather flat. The seasonal frequencies 1/4 and 2/4 are marked, and it is clear that the data have been subjected to deseasonalization. Despite the evidence of some trend in the second plot in part (a), the spectral calculation does not show a peak of note at frequency 0. There is a spectral peak near frequency 0.32 which may indicate some feature in the data. Bartlett's test gives marginal significance for rejection of the hypothesis that the time series has white noise structure. We conclude that the differenced RealGDP data do show some trending and have structure which can be modelled.

(c) An ARX model has the form

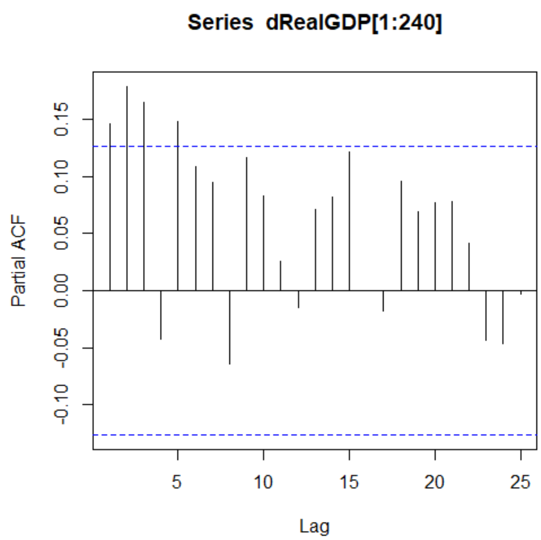
$$y_t = \mu + \phi_1(y_{t-1} - \mu) + \cdots + \phi_p(y_{t-p} - \mu) + \beta_1 x_{1t} + \cdots + \beta_k x_{kt} + \varepsilon_t.$$

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 differences, if the chosen order of the AR structure is *p*, are as follows:

```
df<-data.frame(d1,d2)
arxmodel<-arima(dRealGDP.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 as indicated.

Fit an ARX model to the data. Explain in detail how you arrived at your model fit, and describe your model. You may want to experiment with two different choices of *p*.



The pacf plot suggests an AR(5) fit. However, we also want to include Time as a regression explanatory variable, and the lag 5 estimate in the resulting ARX(5) model fit is not significant. We change to an ARX(4) model, which follows.

```
> arx4<-arima(dRealGDP[1:240],order=c(4,0,0),xreg=Time[1:240])
> arx4

Call:
arima(x = dRealGDP[1:240], order = c(4, 0, 0), xreg = Time[1:240])

Coefficients:
          ar1     ar2     ar3      ar4  intercept  Time[1:240]
      -0.0118  0.0490  0.0535  -0.1452   574.6516      10.7719
s.e.   0.0637  0.0637  0.0635   0.0635   204.9022       1.4751

sigma^2 estimated as 2765526:  log likelihood = -2120.52,  aic = 4255.05
```

```
> coeftest(arx4)

z test of coefficients:

               Estimate Std. Error z value  Pr(>|z|)
ar1           -0.011841   0.063698 -0.1859  0.852531
ar2            0.048998   0.063660  0.7697  0.441488
ar3            0.053478   0.063506  0.8421  0.399736
ar4           -0.145216   0.063506 -2.2867  0.022216 *
intercept    574.651579 204.902177  2.8045  0.005039 **
Time[1:240]   10.771875   1.475129  7.3023 2.828e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Let's also fit an ARX(8) model—perhaps there is some remaining seasonality in the data.

```
> arx8<-arima(dRealGDP[1:240],order=c(8,0,0),xreg=Time[1:240])
> arx8

Call:
arima(x = dRealGDP[1:240], order = c(8, 0, 0), xreg = Time[1:240])

Coefficients:
          ar1     ar2     ar3      ar4     ar5     ar6     ar7      ar8
      -0.0034  0.0516  0.0599  -0.1662  0.0495  0.0285  0.0224  -0.1309
s.e.   0.0638  0.0639  0.0637   0.0638  0.0638  0.0637  0.0636   0.0636
      intercept  Time[1:240]
       579.1386      10.7384
s.e.   197.4041       1.4226

sigma^2 estimated as 2708054:  log likelihood = -2118.08,  aic = 4258.16

> coeftest(arx8)

z test of coefficients:

                Estimate  Std. Error z value  Pr(>|z|)
ar1           -0.0034128   0.0638277 -0.0535  0.957359
ar2            0.0515934   0.0638552  0.8080  0.419105
ar3            0.0599135   0.0637237  0.9402  0.347111
ar4           -0.1662405   0.0638069 -2.6054  0.009178 **
ar5            0.0495102   0.0637651  0.7764  0.437485
ar6            0.0284870   0.0637034  0.4472  0.654744
ar7            0.0223846   0.0635826  0.3521  0.724797
ar8           -0.1309051   0.0636468 -2.0567  0.039711 *
intercept    579.1385852 197.4041390  2.9338  0.003349 **
Time[1:240]   10.7383776   1.4226431  7.5482 4.414e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Indeed, both the lag 4 and lag 8 coefficients are significant. There is thus some remaining seasonality in the quarterly Real GDP growth data.

Let's fit a reduced form of this ARX(8) model, one that includes only lag 4 and lag 8 coefficients and the time variable. We accomplish this with the seasonal ARIMAX framework, introduced in the 14 November notes.

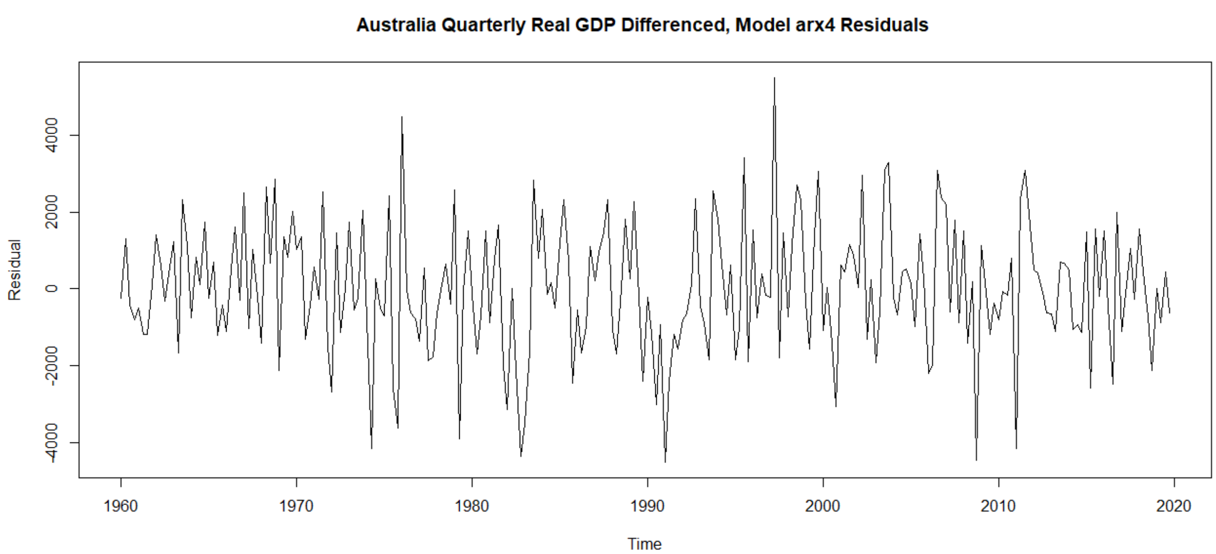
```
> sarx8<-
arima(dRealGDP[1:240],order=c(0,0,0),seasonal=list(order=c(2,0,0),period=4),x
reg=Time[1:240])

> coeftest(sarx8)

z test of coefficients:

              Estimate Std. Error z value  Pr(>|z|)
sar1         -0.162666   0.063808 -2.5493 0.0107932 *
sar2         -0.123101   0.063687 -1.9329 0.0532468 .
intercept   576.125467 168.874304  3.4116 0.0006459 ***
Time[1:240]  10.773063   1.218113  8.8441 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The coefficient estimates are seen to be close in value to those given by the ARX(8) model.

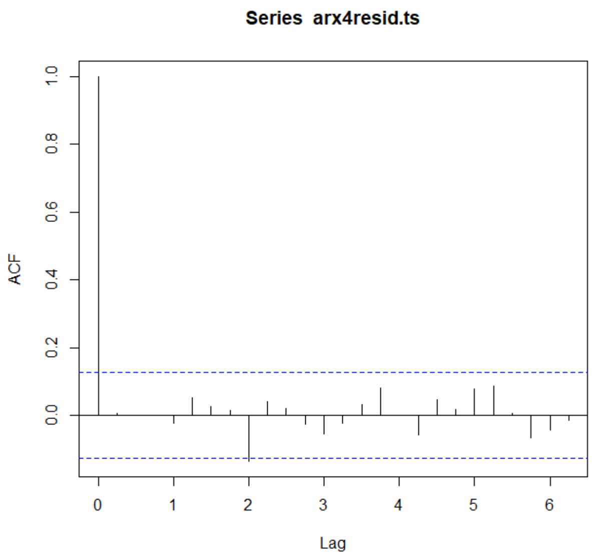
(d) Examine the residuals to investigate whether your selected model has achieved reduction to white noise. For this purpose present the plot of the residuals vs. time, the residual autocorrelations and partial autocorrelations, and the residual spectral density. Perform Bartlett's test to determine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

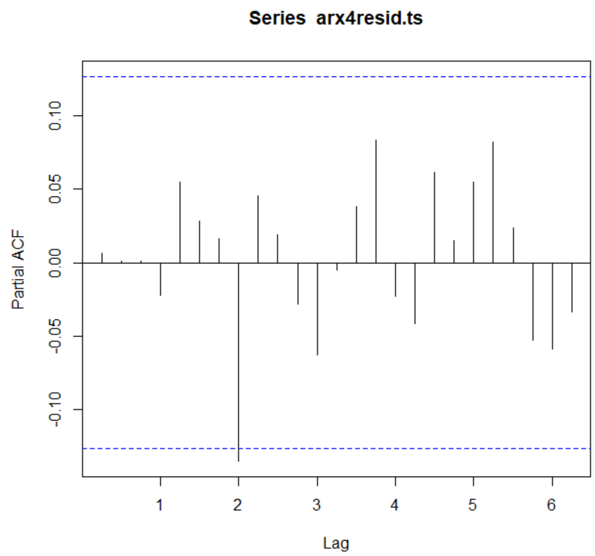
Residual diagnostics for the ARX(4) model follow.

```
> arx4resid.ts<-ts(resid(arx4)[1:240],start=c(1960,1),freq=4)
> plot(arx4resid.ts,ylab="Residual",main="Australia Quarterly Real GDP
Differenced, Model arx4 Residuals ")
```

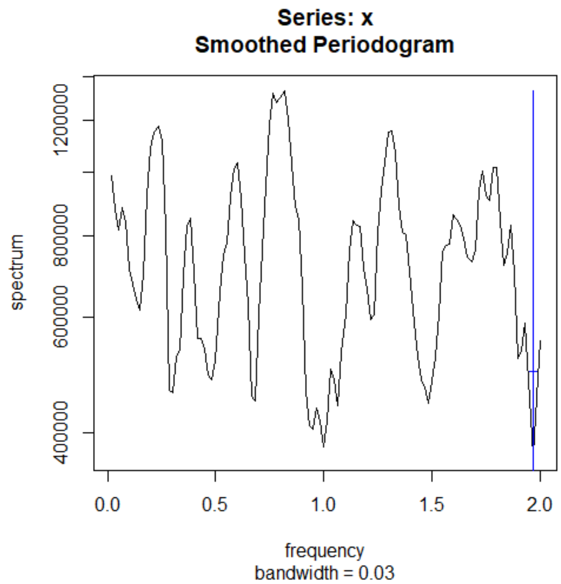


The residual plot has a dip leading up to and during the 1990–91 recession—the model has overestimated quarterly growth during that period of time. Otherwise there is no remaining trend, and the volatility is relatively constant.





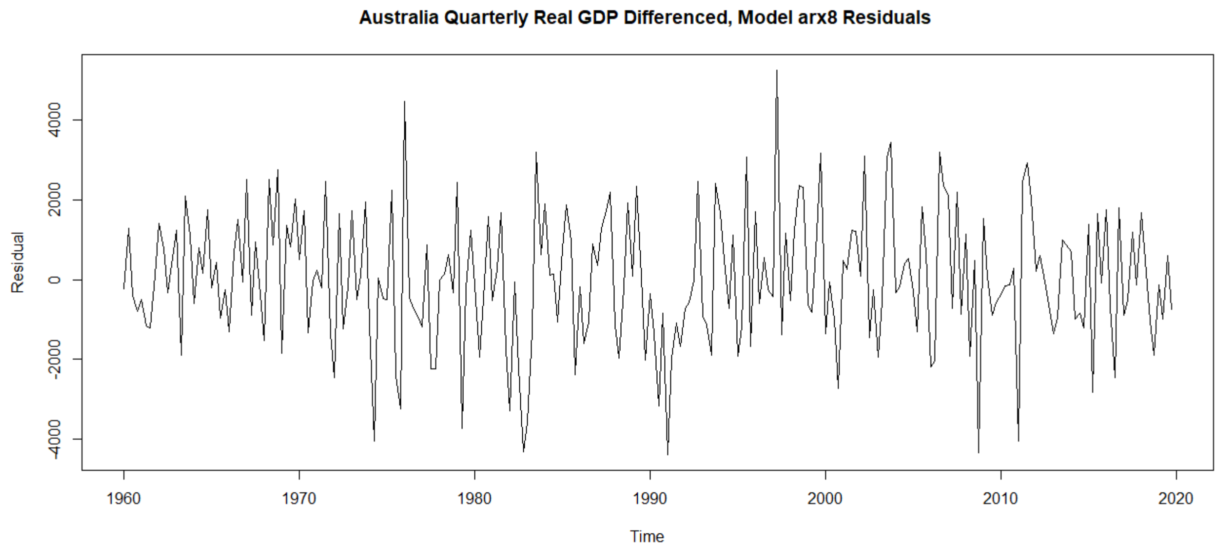
The residual acf and pacf values are very mildly significant at lag 8



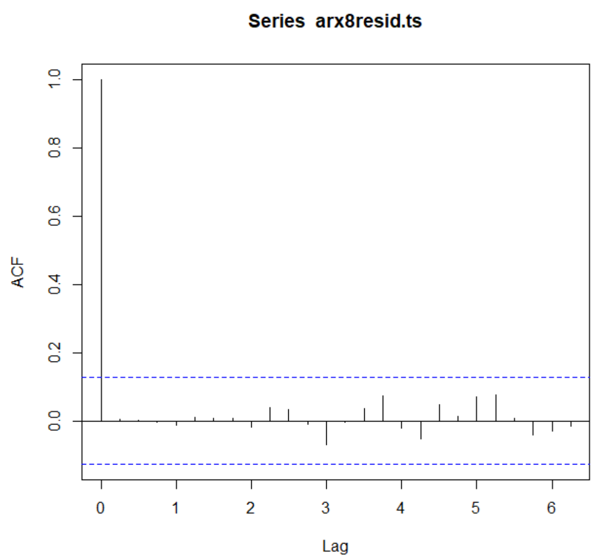

```
> bartlettB.test(arx4resid.ts)

        Bartlett B Test for white noise

data:
= 0.35994, p-value = 0.9995
```

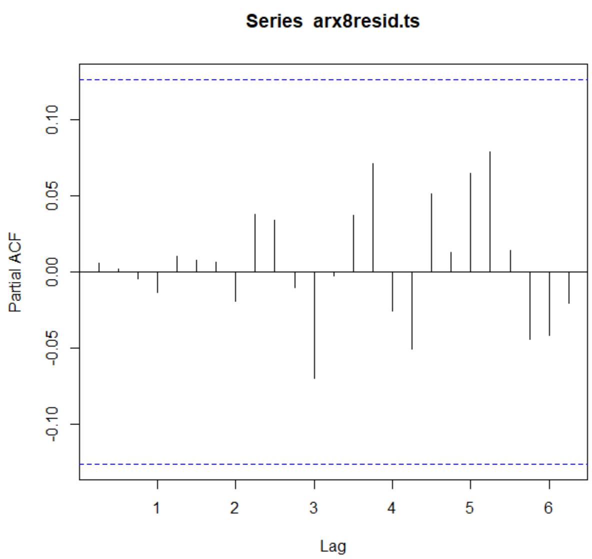
The residual spectrum and Bartlett's test give clear support for reduction to white noise by the ARX(4) model.

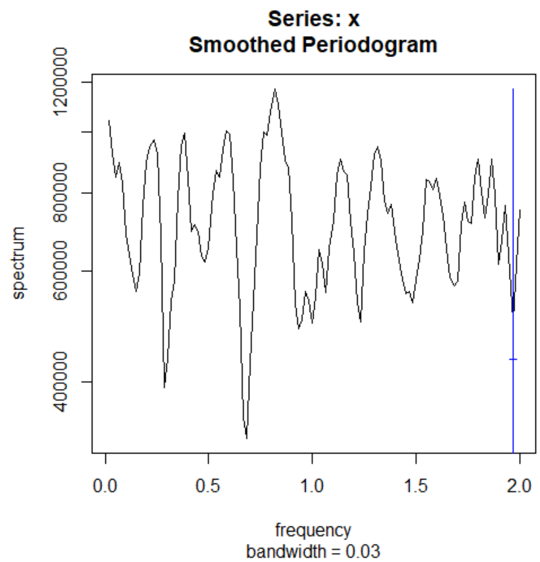
Next residual analysis for the ARX(8) model is given.



The plot of residuals vs. time closely resembles the corresponding plot for the ARX(4) model. And the residual acf and pacf calculations clearly indicate reduction to white noise by the ARX(8) model, as do the residual spectral plot and Bartlett's test.

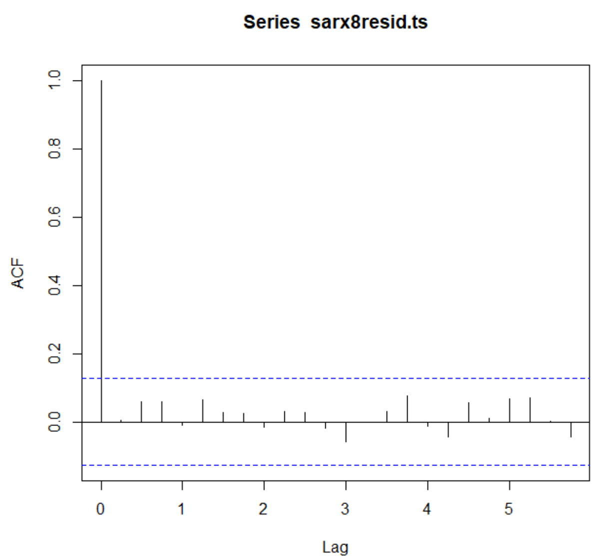


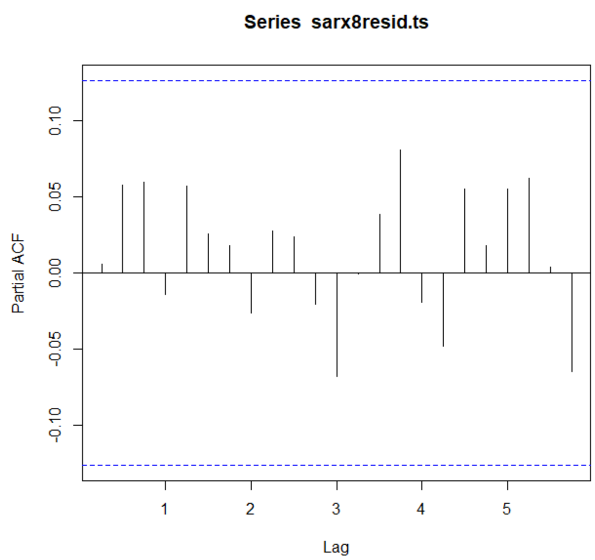


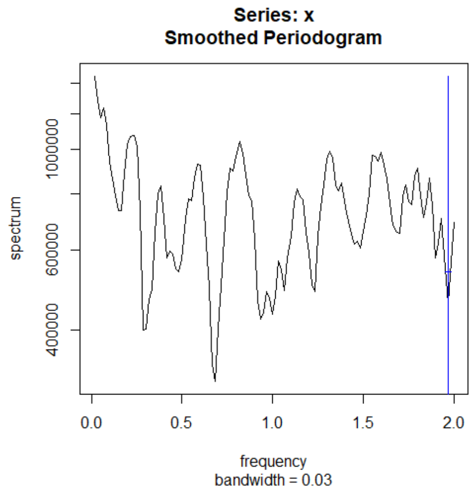



```
> bartlettB.test(arx8resid.ts)

        Bartlett B Test for white noise

data:
= 0.23701, p-value = 1
```

There is one more model for which to display residual diagnostics, the reduced form of the ARX(8) model. We give the acf and pacf plots, the spectrum, and Bartlett's test.








```
> bartlettB.test(sarx8resid.ts)

        Bartlett B Test for white noise

data:
= 0.49228, p-value = 0.9687
```

As with the ARX(8) model, there are no significant residual correlations or partial correlations. And the residual spectral plot and Bartlett's test show very clearly reduction to white noise by the reduced ARX(8) model. However, the residual spectral plot is interesting. It shows a small peak at frequency 0, despite inclusion of the time variable in construction of the model. In other words, the additional (insignificant) lag estimates in the ARX(8) model are probably somewhat beneficial. This suggests the ARX(8) model is preferable. However, let's find and analyze the autoregressive polynomial zeros for all three of the fitted models before making a final choice.

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

We start with the autoregressive polynomial zeros for the ARX(4) model.

```
> zeros<-1/polyroot(c(1,-coef(arx4)[1:4]))
> zeros
[1]  0.4481177-0.3866104i -0.4540382-0.4565349i -0.4540382+0.4565349i
[4]  0.4481177+0.3866104i
```

There are two pairs of complex-valued zeros.

```
> #amplitudes
> Mod(zeros[3:4])
[1] 0.6438748 0.5918421

> #periods
> 2*pi/Arg(zeros)[3:4]
[1] 2.669774 8.826612
```

The ARX(4) model estimates two pseudocycles, with period lengths 2.67 and 8.83 quarters.

Next, the ARX(8) model.

```
> zeros<-1/polyroot(c(1,-coef(arx8)[1:8]))
> zeros
[1]  0.3505622-0.6976927i -0.7016804-0.3579496i -0.3511471+0.6933084i
[4]  0.7005588+0.2866309i  0.7005588-0.2866309i -0.3511471-0.6933084i
[7]  0.3505622+0.6976927i -0.7016804+0.3579496i

> Mod(zeros[c(3,4,7,8)])
[1] 0.7771620 0.7569279 0.7808131 0.7877076

> 2*pi/Arg(zeros[c(3,4,7,8)])
[1]  3.080575 16.178518  5.685198  2.353365
```

The ARX(8) model estimates four pseudocycles, with period lengths 2.35, 3.08, 5.69, and 16.18 quarters.

Finally, here are the zero calculations for the reduced ARX(8) model.

```
> zeros<-1/polyroot(c(1,-c(0,0,0,-0.162666,0,0,0,-0.123101)))
> zeros
[1]  0.3355834-0.6926154i -0.6926154-0.3355834i -0.3355834+0.6926154i
[4]  0.6926154+0.3355834i  0.6926154-0.3355834i -0.3355834-0.6926154i
[7]  0.3355834+0.6926154i -0.6926154+0.3355834i
```

There are four pairs of complex zeros.

```
> Mod(zeros)[c(3,4,7,8)]
[1] 0.7696313 0.7696313 0.7696313 0.7696313

> 2*pi/Arg(zeros)[c(3,4,7,8)]
[1]  3.107441 13.925990  5.611930  2.335402
```

Four pseudocycles, with period lengths 2.34, 3.11, 5.61, and 13.93 quarters, are estimated.

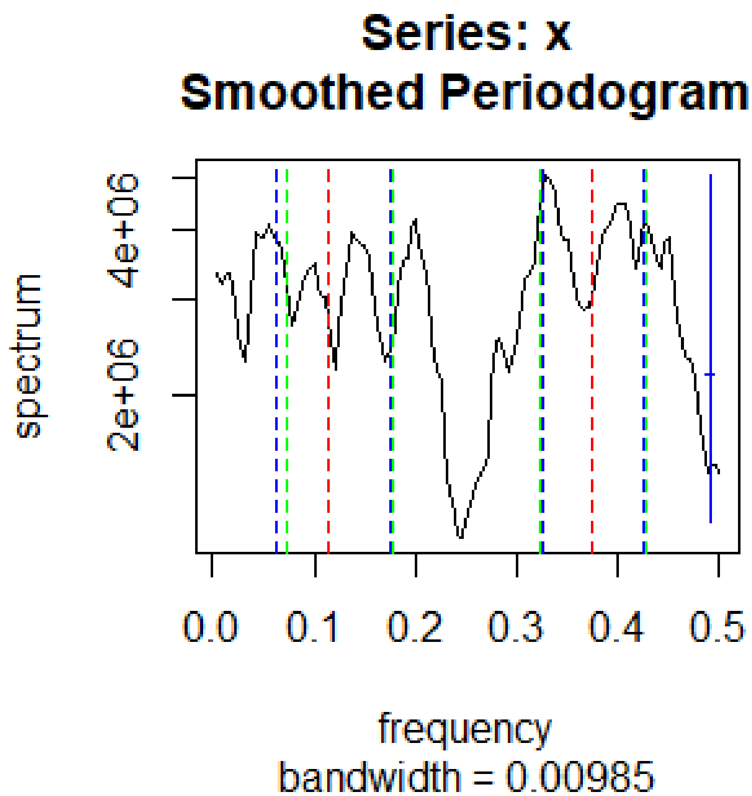
Let's summarize and compare the pseudoperiod estimates from the three models. The frequencies in the table are the reciprocals of the pseudoperiods.

Model	Pseudoperiods	Frequencies
ARX(4)	2.67, 8.83	0.375, 0.113
ARX(8)	2.35, 3.08, 5.69, 16.18	0.426, 0.325, 0.176, 0.062
Reduced ARX(8)	2.34, 3.11, 5.61, 13.93	0.427, 0.322, 0.178, 0.072

As the table shows, the ARX(8) and reduced ARX(8) models give three similar estimated pseudoperiods. The models differ somewhat according to the longest estimated pseudoperiod.

Let's investigate whether the spectrum of the differenced Real GDP data has peaks at any of the frequencies corresponding to the estimated pseudoperiods.

```
> spectrum(dRealGDP[1:240],8)
> abline(v=c(0.113,0.375),lty=2,col="red")
> abline(v=c(0.062,0.176,0.325,0.426),lty=2,col="blue")
> abline(v=c(0.072,0.178,0.322,0.427),lty=2,col="green")
```



The two frequencies from the ARX(4) model fit are marked in red. Neither of these coincides with a peak of the spectrum. The four blue lines are for the ARX(8) model fit. Three of the four frequencies (corresponding to pseudocycle lengths 16.18, 3.08, and 2.35 quarters) are at peaks (for frequency 0.426 the peak is very small). And the four green lines are for the reduced ARX(8) model fit. Two of its estimated frequencies are at spectral peaks, and the estimated frequency 0.072 is slightly off the center of a peak.

What can we conclude? The spectral plot suggests the estimated pseudoperiod lengths 3.08 and 16.18 are most credible. These are determined by the ARX(8) model fit. The reduced ARX(8) fit gives somewhat similar estimates (3.11 and 13.93 quarters). Perhaps these estimates suggest stretches of economic expansion and of economic stagnation.

(f) How do the results in part (e) compare to those described in the 24 October notes for U.S. real quarterly GNP data for the period 1947(2) to 1991(1)? Discuss in detail.

In the 24 October notes an AR(4) model was fit to deseasonalized U.S. quarterly growth of real GNP data (differenced GNP data). Two pseudocycles were estimated, with lengths 2.46 and 10.56 quarters. The data were taken from Ruey Tsay's book. Tsay also used a Markov switching model to obtain estimates of 3.7 and 11.3 quarters. The U.S. data span the years from 1947 to 1991.

There are two notable differences between the Australian data and the U.S. data. The Australian time series is longer, covering the years 1960 to 2022. And, as noted in part (a), Australia had seven recessions between 1960 and 1991, all of short duration, and was then recession-free for 28 years, from 1992 to 2019. For the U.S. there were nine recessions, also of short duration, but there was no long stretch of time which was recession-free.

Certainly, the U.S. estimate of 2.46 quarters is similar to the Australian estimate of 3.08 quarters. The Australian estimate for the average duration of expansion, 16.18 quarters (greater than the U.S. estimate of 10.56 quarters) is a figure inflated by the 28-year recession-free period from 1992 to 2019.

Perhaps we should estimate models for two different time spans for the Australian data, in view of the long recession-free period after 1991.