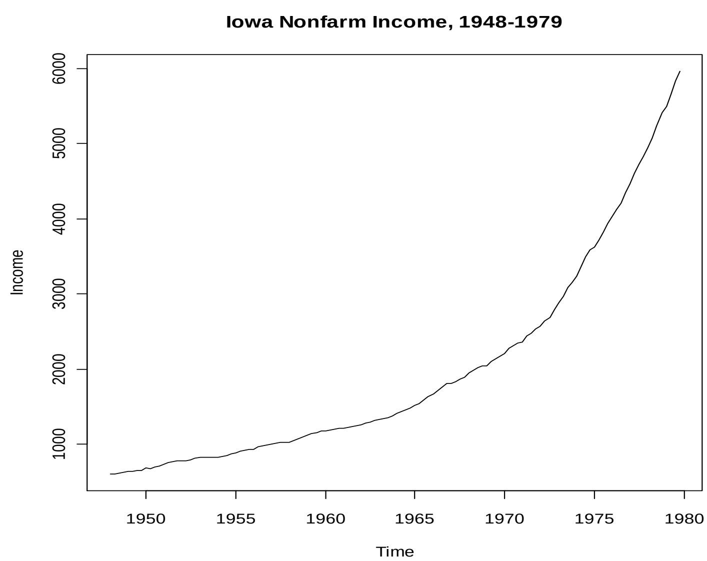
Statistics 435/535/711 13 April 2022
P. Shaman

Seasonal ARIMA Model Fits—More Examples

C. Quarterly Iowa Nonfarm Income, 1948-1979

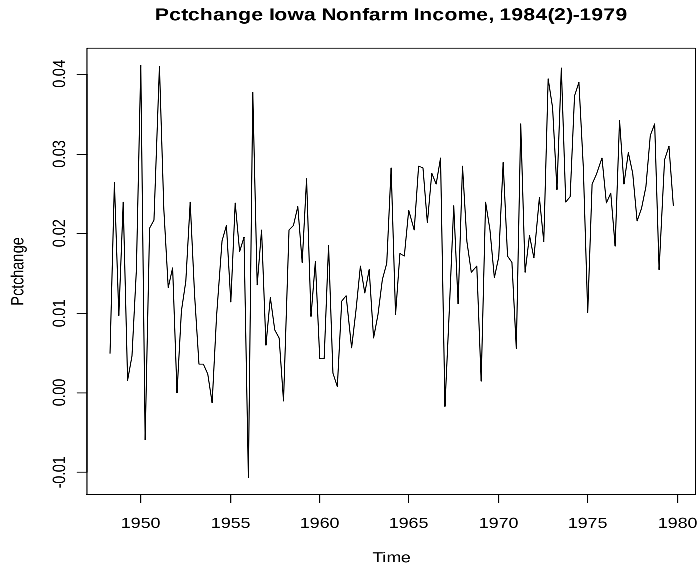
```
> iowa<-read.csv("G:/Stat71122Spring/iowa.txt")
> attach(iowa)
> head(iowa)
  income  pctchange time
1    601         NA    1
2    604 0.00499168    2
3    620 0.02649007    3
4    626 0.00967742    4
5    641 0.02396166    5
6    642 0.00156006    6

>
plot(ts(income,start=c(1948,1),freq=4),xlab="Time",ylab="Income",main="
Iowa Nonfarm Income, 1948-1979")
```



The plot directly below shows percentage change from one quarter to the next. It reveals structure not evident in the graph of income, shown directly above.

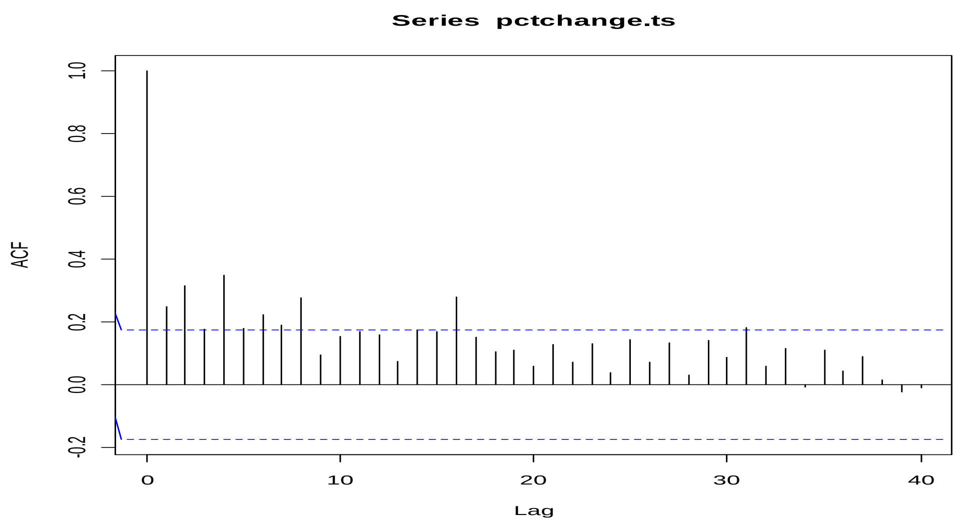
```
> sel<-2:128
> pctchange<-pctchange[sel]
> pctchange.ts<-ts(pctchange,start=c(1948,2),freq=4)
> plot(pctchange.ts,xlab="Time",ylab="Pctchange",main="Pctchange Iowa
```

```
Nonfarm Income, 1984(2)-1979")
```



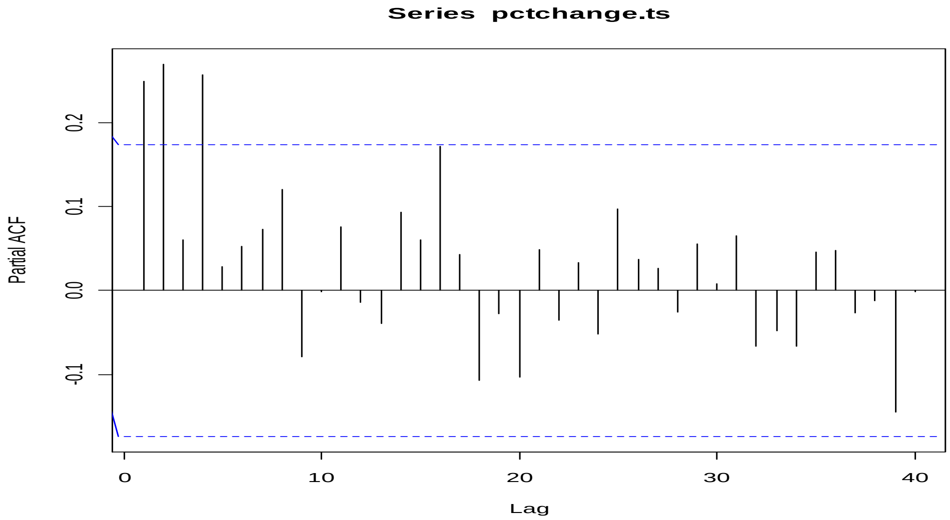
There is a slight upward trend starting about 1962.

```
> pctchange.ts<-ts(pctchange)
> acf(pctchange.ts,40)
```



There are significant autocorrelations at lags 1, 2, 4, 6, 8, and 16. Note that the data are quarterly and thus seasonal activity is expected to manifest itself at multiples of lag 4. However, we will give a different interpretation below to the apparent activity at lag 16. The estimated partial autocorrelations follow.

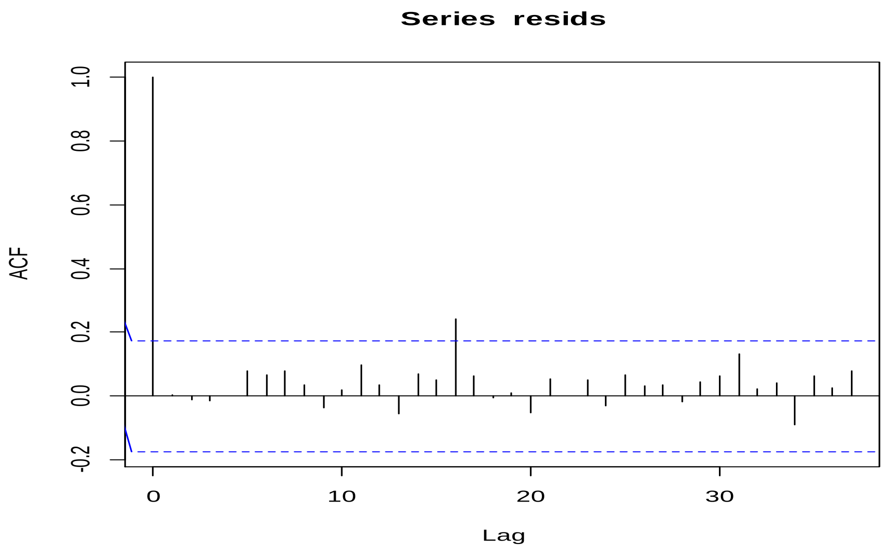
```
> pacf(pctchange.ts,40)
```



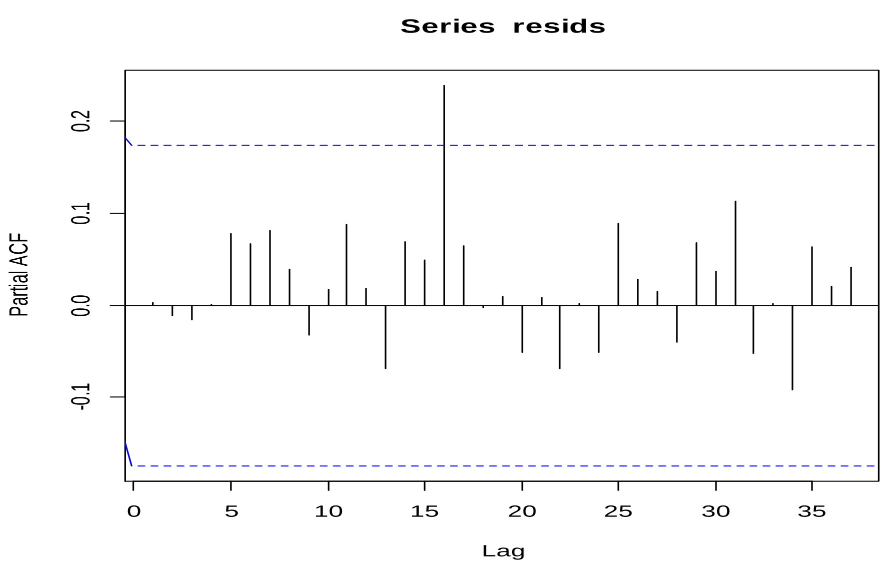
Here the partial correlations are significant at lags 1, 2, and 4, and almost at 16. Let's start with an ARIMA(2,0,0)(0,0,2)$_4$ fit.

```
> model<-
arima(pctchange.ts,order=c(2,0,0),seasonal=list(order=c(0,0,2),period=4
))
> model

Call:
arima(x = pctchange.ts, order = c(2, 0, 0), seasonal = list(order =
c(0, 0,
    2), period = 4))

Coefficients:
         ar1     ar2    sma1    sma2  intercept
      0.1535  0.1982  0.2256  0.1572     0.0184
s.e.  0.0869  0.0903  0.0955  0.0749     0.0018

sigma^2 estimated as 9.085e-05:  log likelihood = 410.48,  aic = -
808.97

> coeftest(model)

z test of coefficients:

           Estimate Std. Error z value Pr(>|z|)
ar1       0.1535437  0.0869430  1.7660  0.07739 .
ar2       0.1982321  0.0902621  2.1962  0.02808 *
sma1      0.2255676  0.0954889  2.3622  0.01816 *
sma2      0.1571697  0.0749369  2.0974  0.03596 *
intercept 0.0183783  0.0017734 10.3633  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

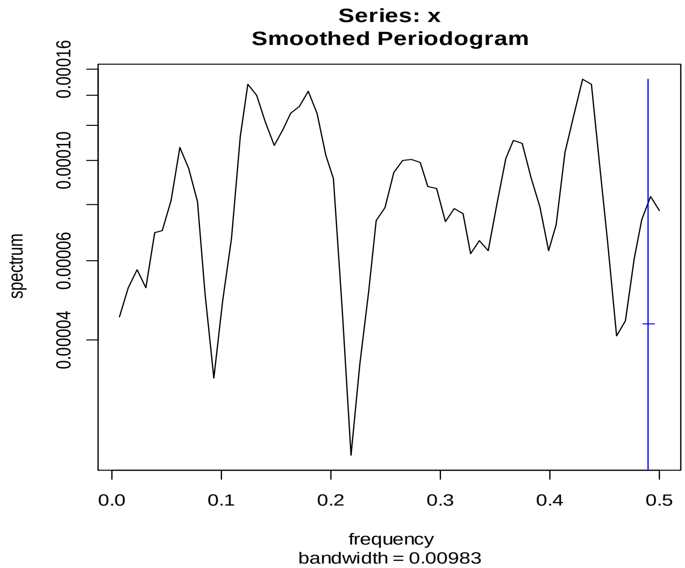
```
> head(resid(model))
[1]           NA -0.012244152  0.010220155 -0.006440052  0.005114639
[6] -0.012752815

> sel<-2:128
> resids<-resid(model)[sel]
> acf(resids,37)
```




```
> pacf(resids,37)
```

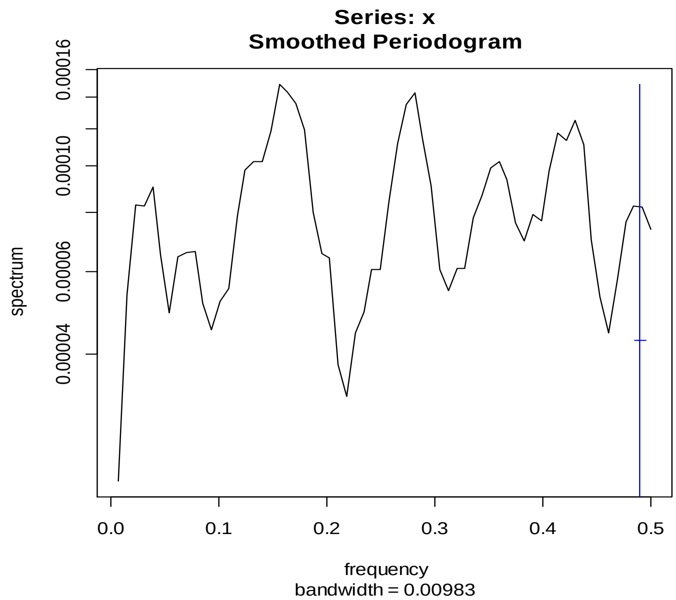


The next page shows the estimate of the spectral density of the residuals from the above ARIMA model fit. The lag 16 correlation and partial correlation for the residuals are both significant. This suggests a component with period 16, which corresponds to frequency 1/16 = 0.0625. The spectral plot shows a local peak at this frequency and at its harmonics. Let's continue by fitting a seasonal AR(1) with period 16 to these residuals.




```
> model2<-
arima(ts(resids),order=c(0,0,0),seasonal=list(order=c(1,0,0),period=16)
)
> model2

Call:
arima(x = ts(resids), order = c(0, 0, 0), seasonal = list(order = c(1,
0, 0),
    period = 16))

Coefficients:
        sar1  intercept
      0.2990     0.0001
s.e.  0.0933     0.0011

sigma^2 estimated as 8.328e-05:  log likelihood = 415.52,  aic = -
825.05

> coeftest(model2)

z test of coefficients:

           Estimate Std. Error z value Pr(>|z|)
sar1      0.2989610  0.0933474  3.2027 0.001362 **
intercept 0.0001250  0.0011018  0.1134 0.909679
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

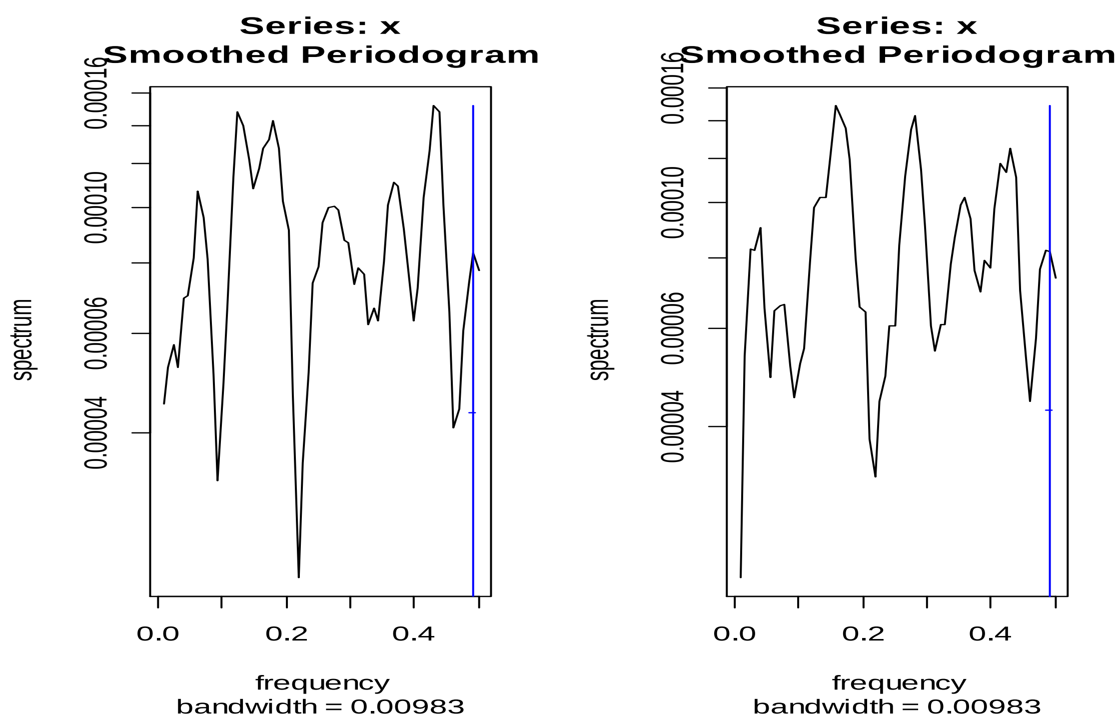
The residual acf and pacf plots from this fit are consistent with white noise. And the spectral plot of these residuals shows attenuation in the vicinity of frequency 0.0625 and (to some extent) its overtones.




```
> bartlettB.test(resid(model2))

        Bartlett B Test for white noise
data:
= 0.44022, p-value = 0.9902
```

The Bartlett test confirms adequate reduction to white noise.

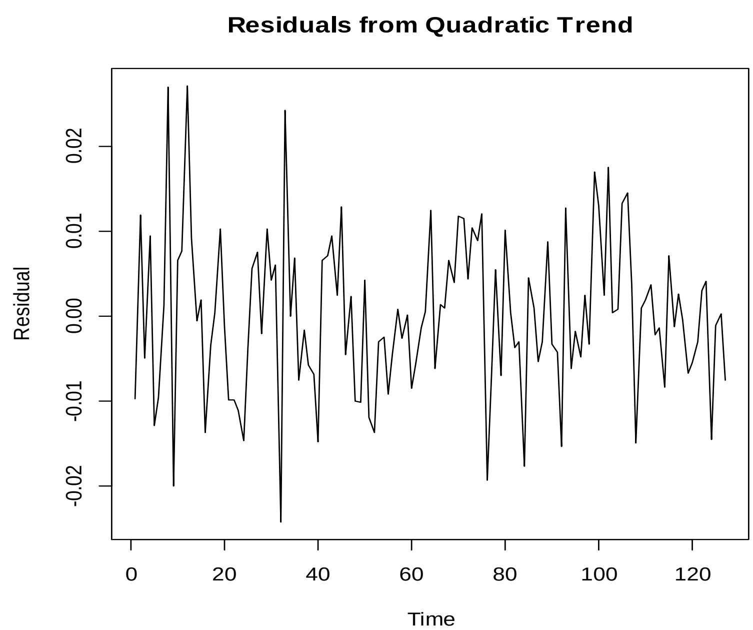
Additionally, here are the two residual spectral plots side-by-side. The plot for the first model is on the left.

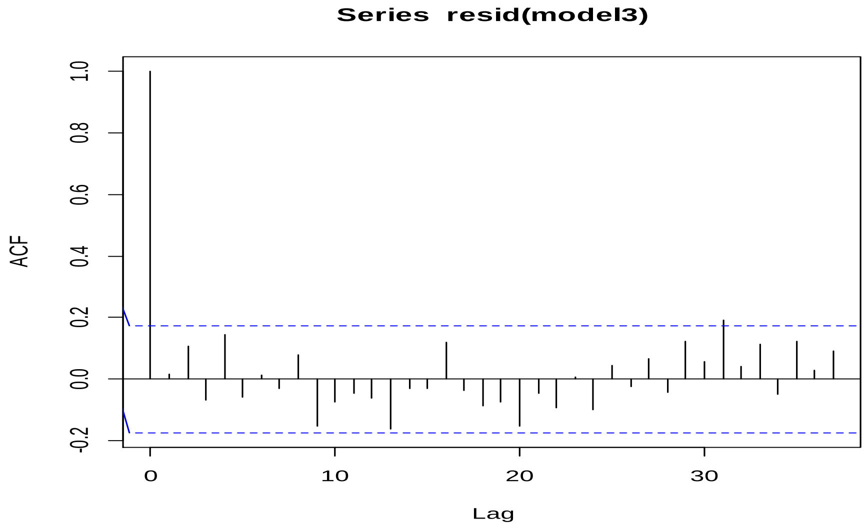


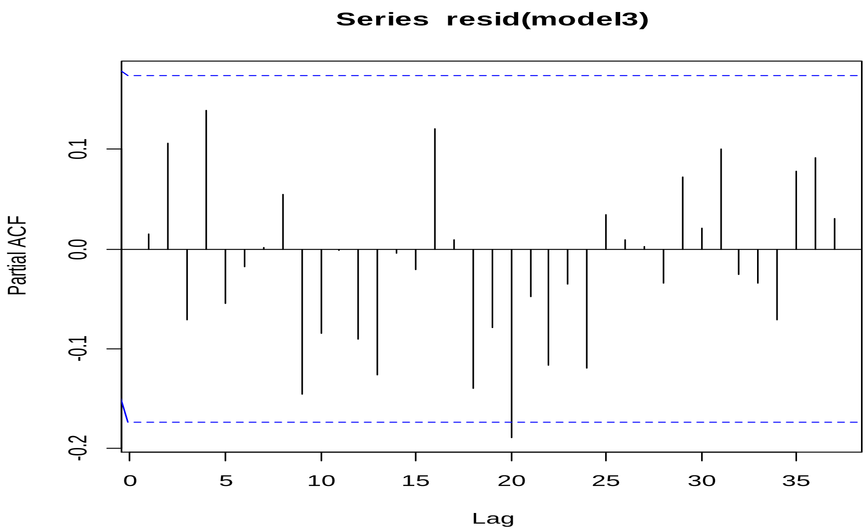
As an alternative to the above analysis, let's first remove a quadratic trend from the percentage change data before fitting an ARIMA model. Recall the plot of the percentage change data given on page 2.

```
> model3<-lm(pctchange~time+I(time^2));summary(model3)

Call:
lm(formula = pctchange ~ time + I(time^2))

Residuals:
       Min         1Q     Median         3Q        Max 
-0.0242534 -0.0057803  0.0001451  0.0062663  0.0270562 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  1.497e-02  2.618e-03   5.717 7.64e-08 ***
time        -1.004e-04  9.287e-05  -1.081   0.2817    
I(time^2)    1.769e-06  6.927e-07   2.553   0.0119 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 

Residual standard error: 0.009383 on 124 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.2394,    Adjusted R-squared:  0.2272 
F-statistic: 19.52 on 2 and 124 DF,  p-value: 4.273e-08
```

Following are the residuals after removal of the quadratic trend, and their acf and pacf plots.







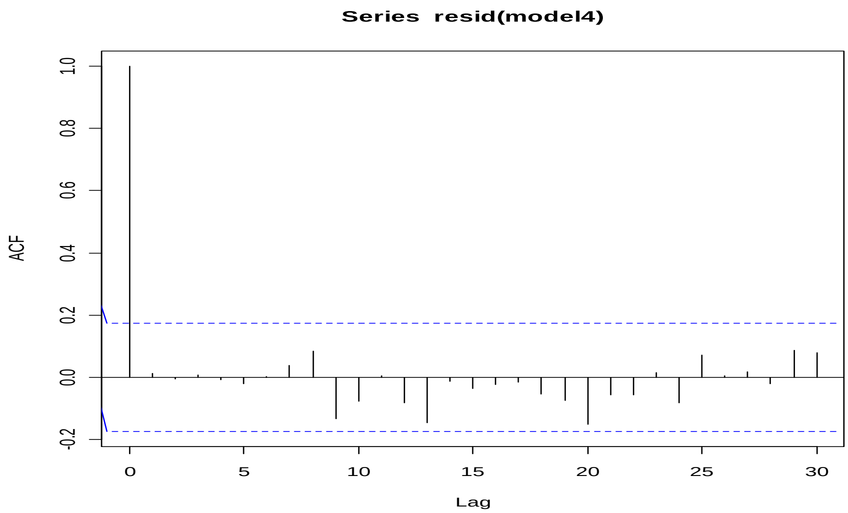
The signal of this residual series is rather weak—the acf and pacf plots show significance only for the correlation at lag 31 and the partial correlation at lag 20, and only barely so.

After some experimentation, I settled on an ARIMA(4,0,0)(1,0,0)$_{16}$ model. This will pick up activity at lags 4 and 16, and also, in fact, at lag 20.

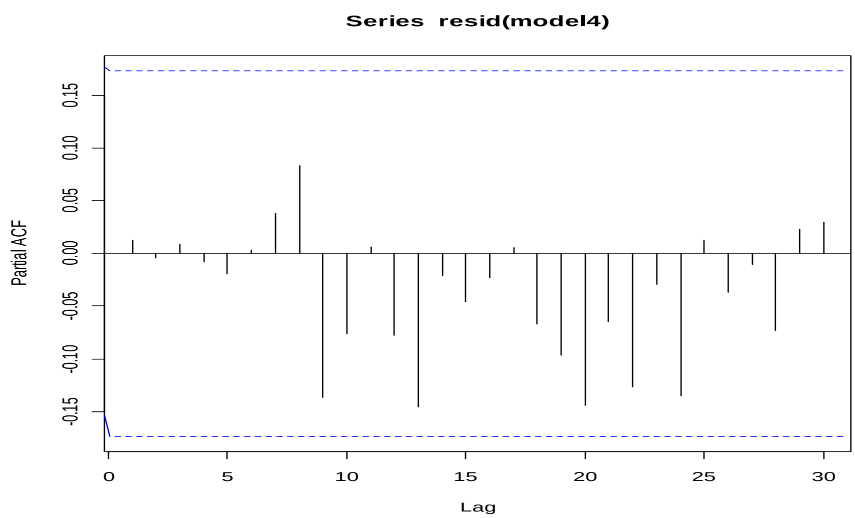
```
> model4<-
arima(resid3.ts,order=c(4,0,0),seasonal=list(order=c(1,0,0),period=16))
> model4

Call:
arima(x = resid3.ts, order = c(4, 0, 0), seasonal = list(order = c(1,
0, 0),
    period = 16))



Coefficients:
         ar1     ar2      ar3     ar4    sar1  intercept
      0.0381  0.1133  -0.0372  0.1862  0.2067    -0.0001
s.e.  0.0873  0.0876   0.0899  0.0896  0.1011     0.0014

sigma^2 estimated as 7.963e-05:  log likelihood = 418.67,  aic = -
823.35

> coeftest(model4)

z test of coefficients:

             Estimate  Std. Error z value Pr(>|z|)
ar1        3.8144e-02  8.7288e-02  0.4370  0.66212
ar2        1.1333e-01  8.7619e-02  1.2934  0.19586
ar3       -3.7213e-02  8.9919e-02 -0.4138  0.67899
ar4        1.8619e-01  8.9611e-02  2.0777  0.03774 *
sar1       2.0674e-01  1.0105e-01  2.0459  0.04077 *
intercept -5.8844e-05  1.3729e-03 -0.0429  0.96581
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Residual analysis follows.

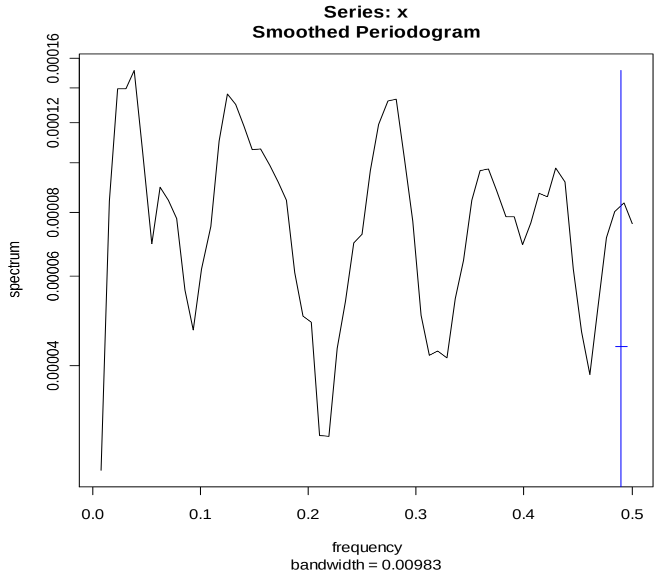
```
> acf(resid(model4),30)
```




```
> pacf(resid(model4),30)
```



Let's also examine the residual spectral density.




```
> bartlettB.test(resid(model4))

        Bartlett B Test for white noise

data:
= 0.35854, p-value = 0.9995
```

The residual diagnostics confirm reduction to white noise.

Overall, this estimated residual spectrum appears to be flatter than the estimated residual spectrum for the combination of the first two models—see the plot at the top of page 6.

Let's finish by examining the two fits. The first is (y_t is the percentage change)

$$(1 - 0.1535B - 0.1982B^2)(y_t - 0.0184) = (1 + 0.2256B^4 + 0.1572B^8)\varepsilon_t,$$

$$(1 - 0.2990B^{16})\varepsilon_t = \eta_t.$$

After some manipulation we can write this as

$$(1 - 0.2990B^{16})(1 - 0.1535B - 0.1982B^2)\, y_t = 0.0084 + (1 + 0.2256B^4 + 0.1572B^8)\eta_t,$$

or

$$y_t = 0.0084 + 0.1535y_{t-1} + 0.1982y_{t-2} + 0.2990y_{t-16} - 0.0459y_{t-17} - 0.0593y_{t-18}$$

$$+ \eta_t + 0.2256\eta_{t-4} + 0.1572\eta_{t-8}.$$

In autoregressive form this is

$$(1 - 0.1535B - 0.1982B^2 - 0.2990B^{16} + 0.0459B^{17} + 0.0593B^{18})$$

$$\cdot(1 + 0.2256B^4 + 0.1572B^8)^{-1} (y_t - 0.0084)$$

$$= \eta_t \, .$$

We can further write this as

$$(1 - 0.1535B - 0.1982B^2 - 0.2256B^4 - 0.1572B^8 - 0.2990B^{16}$$

$$+ \text{ small terms }) (y_t - 0.0084) = \eta_t \, .$$

The second fit is

$$w_t = y_t - 0.01497 + 0.0001004t - 0.00000177t^2,$$

$$(1 - 0.0381B - 0.1133B^2 + 0.0372B^3 - 0.1862B^4)(1 - 0.2067B^{16})w_t = \varepsilon_t,$$

The last equation gives

$$w_t = 0.0381w_{t-1} + 0.1133w_{t-2} - 0.0372w_{t-3} + 0.1862w_{t-4} + 0.2067w_{t-16}$$

$$- 0.0079w_{t-17} - 0.0234w_{t-18} + 0.0077w_{t-19} - 0.03852w_{t-20}$$

$$+ \eta_t \, .$$

Note that use of the quadratic trend renders forecasting with the second model more problematic.

Addendum—Still Another Model

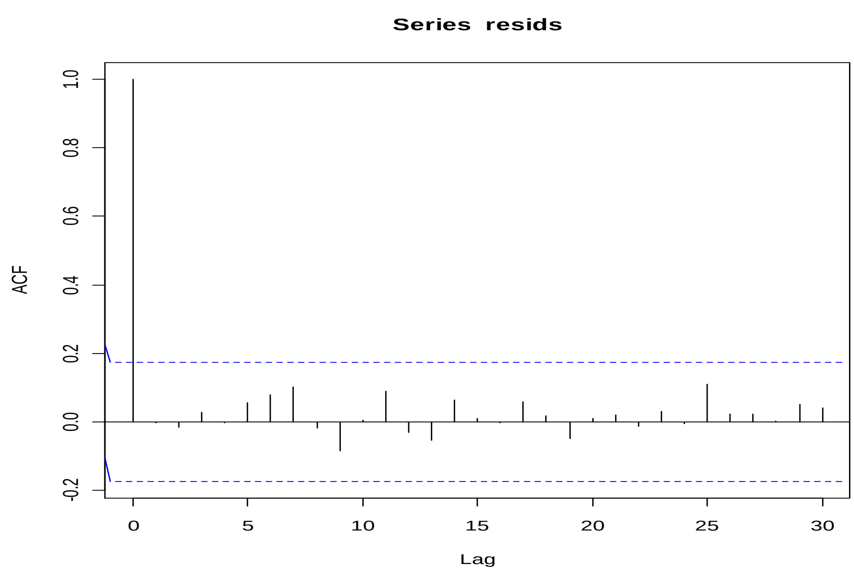
An ARIMA(2,0,0)(0,0,4)$_4$ model fit follows. This does not address the quadratic trend in the data directly, but does pick up the activity at lag 16 in one step.

```
> model5<-
arima(ts(pctchange),order=c(2,0,0),seasonal=list(order=c(0,0,4),period=
4))
> model5

Call:
arima(x = ts(pctchange), order = c(2, 0, 0), seasonal = list(order =
c(0, 0,
    4), period = 4))
```

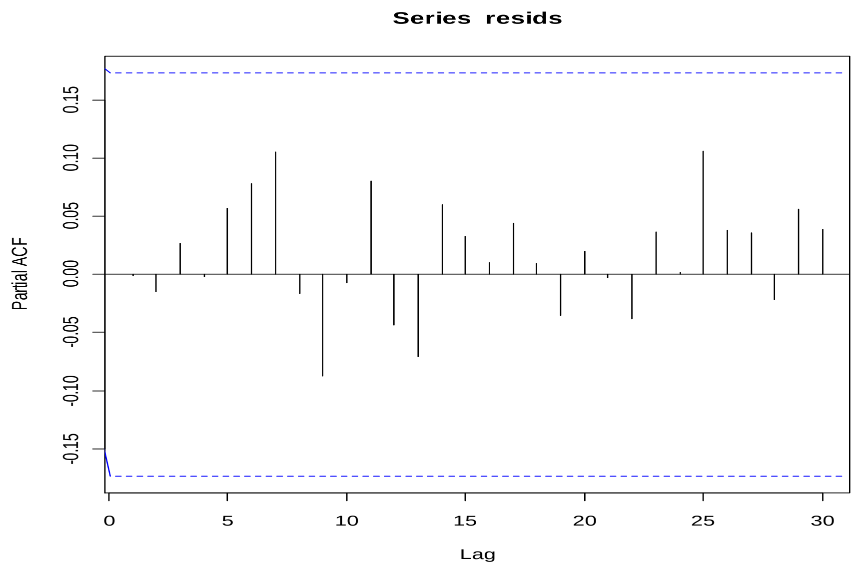
```
Coefficients:
         ar1     ar2    sma1    sma2    sma3    sma4  intercept
      0.1454  0.1737  0.2258  0.2234  0.1023  0.3090     0.0186
s.e.  0.0878  0.0904  0.0888  0.0972  0.1058  0.0878     0.0021

sigma^2 estimated as 8.259e-05:  log likelihood = 415.75,  aic = -815.5

> coeftest(model5)

z test of coefficients:

           Estimate Std. Error z value  Pr(>|z|)    
ar1       0.1453516  0.0878198  1.6551 0.0979018 .  
ar2       0.1737345  0.0904011  1.9218 0.0546285 .  
sma1      0.2257847  0.0888111  2.5423 0.0110125 *  
sma2      0.2233673  0.0972192  2.2976 0.0215867 *  
sma3      0.1023392  0.1058011  0.9673 0.3334042    
sma4      0.3089613  0.0877831  3.5196 0.0004322 ***
intercept 0.0186041  0.0021179  8.7841 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

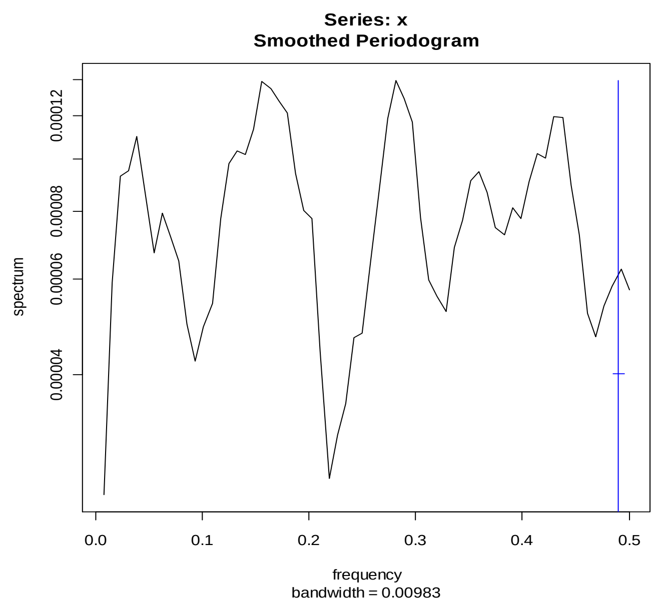
The residual acf and pacf indicate reduction to white noise.

```
> resids<-resid(model5)[2:128]
> acf(resids,30)
```



```
> pacf(resids,30)
```






```
> bartlettB.test(resids)

        Bartlett B Test for white noise

data:
= 0.29932, p-value = 1
```

This spectral plot is similar to the spectral plots for the residuals from the previous models.

The representation of the model fit is

$$(1 - 0.1454B - 0.1737B^2)(y_t - 0.0186)$$
$$= (1 + 0.2258B^4 + 0.2234B^8 + 0.1023B^{12} + 0.3090B^{16})\,\varepsilon_t.$$

In autoregressive form it is

$$(1 - 0.1454B - 0.1737B^2 - 0.2258B^4 - 0.2234B^8 - 0.1023B^{12}$$
$$- 0.3090B^{16} + \text{small terms}) \cdot (y_t - 0.0186) = \varepsilon_t.$$

This is very similar to the representation of the first model, shown on page 11.

D. Monthly U.S. employment for males aged 16-19, in thousands, January 1971—December 1981

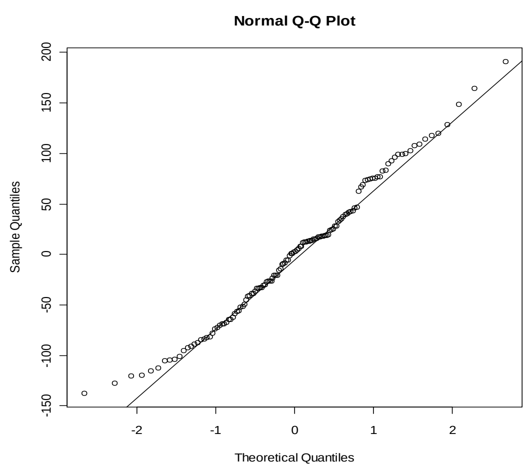
This data set is described briefly in the 12 January class notes.

```
> memp<-read.csv("G:/Stat71122Spring/Memp1619.txt")
> attach(memp)
> head(memp)
  month employment time
1     1        707    1
2     2        655    2
3     3        638    3
4     4        574    4
5     5        552    5
6     6        980    6
```

There are a prominent trend and a strong seasonal pattern. Let's first fit a regression model and estimate seasonal structure. We include a polynomial trend and monthly dummies.

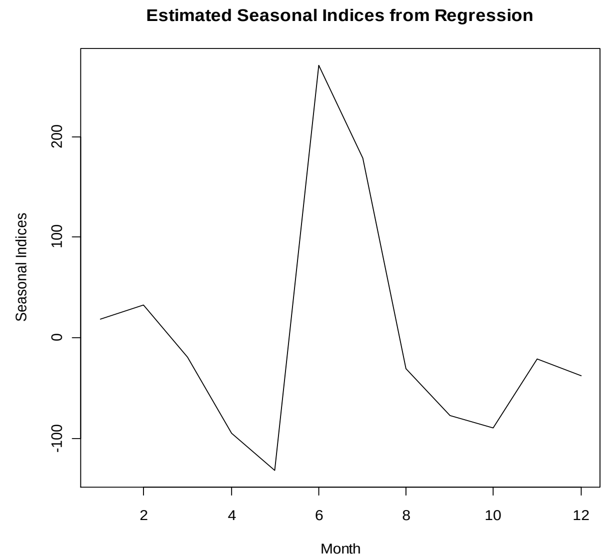
```
> time<-as.numeric(time)
> fmonth<-as.factor(month)
> model<-
lm(employment~time+I(time^2)+I(time^3)+I(time^4)+I(time^5)+fmonth);summ
ary(model)

Call:
lm(formula = employment ~ time + I(time^2) + I(time^3) + I(time^4) +
    I(time^5) + fmonth)

Residuals:
     Min       1Q   Median       3Q      Max
-137.695  -51.645    2.604   40.664  191.284
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  8.275e+02  4.469e+01  18.517  < 2e-16 ***
time        -2.684e+01  6.219e+00  -4.316 3.39e-05 ***
I(time^2)    1.322e+00  2.867e-01   4.611 1.05e-05 ***
I(time^3)   -2.159e-02  5.439e-03  -3.968 0.000126 ***
I(time^4)    1.441e-04  4.501e-05   3.201 0.001771 ** 
I(time^5)   -3.334e-07  1.347e-07  -2.475 0.014790 *  
fmonth2      1.379e+01  3.116e+01   0.443 0.658830    
fmonth3     -3.824e+01  3.117e+01  -1.227 0.222443    
fmonth4     -1.137e+02  3.119e+01  -3.646 0.000402 ***
fmonth5     -1.511e+02  3.122e+01  -4.840 4.08e-06 ***
fmonth6      2.525e+02  3.125e+01   8.081 7.14e-13 ***
fmonth7      1.600e+02  3.128e+01   5.115 1.27e-06 ***
fmonth8     -4.973e+01  3.132e+01  -1.588 0.115044    
fmonth9     -9.607e+01  3.136e+01  -3.064 0.002725 ** 
fmonth10    -1.086e+02  3.141e+01  -3.457 0.000767 ***
fmonth11    -3.978e+01  3.147e+01  -1.264 0.208846    
fmonth12    -5.692e+01  3.155e+01  -1.804 0.073778 .  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 73.07 on 115 degrees of freedom
Multiple R-squared: 0.8224,     Adjusted R-squared: 0.7976 
F-statistic: 33.27 on 16 and 115 DF,  p-value: < 2.2e-16

> qqnorm(resid(model))
> qqline(resid(model))
```



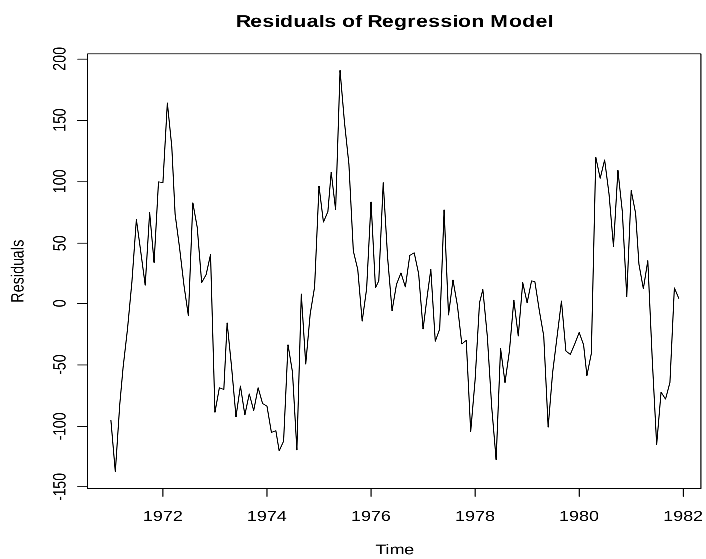
Calculate and plot the estimated seasonal indices.

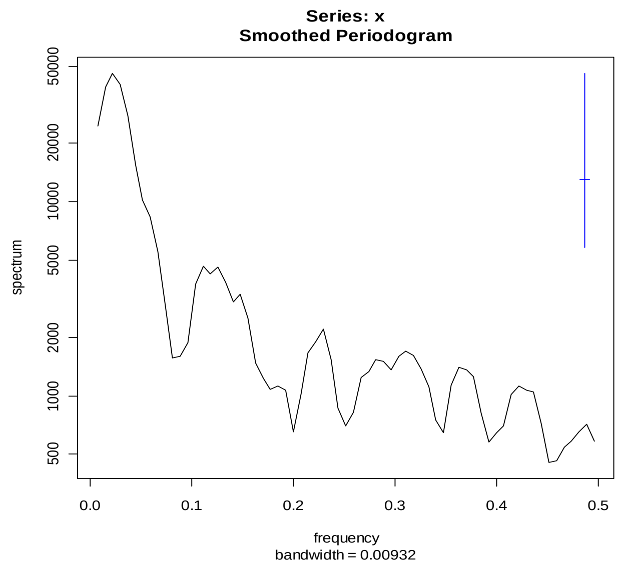
```
> b1<-coef(model)[1]
> b2<-coef(model)[7:17]+b1
> b3<-c(b1,b2)
> seasreg<-b3-mean(b3)
> seasreg
(Intercept)     fmonth2     fmonth3     fmonth4     fmonth5     fmonth6 
   18.98644    32.78083   -19.25450   -94.73557  -132.09754   271.49629 
    fmonth7     fmonth8     fmonth9    fmonth10    fmonth11    fmonth12 
  178.97266   -30.74252   -77.08773   -89.59312   -20.78953   -37.93571 
```



The high level of male youth employment relative to the trend is estimated to be concentrated in June and July, especially in June. The low months are April, May, September, and October, which are all below trend level.

Let's examine the residual time series plot and the residual spectral density.

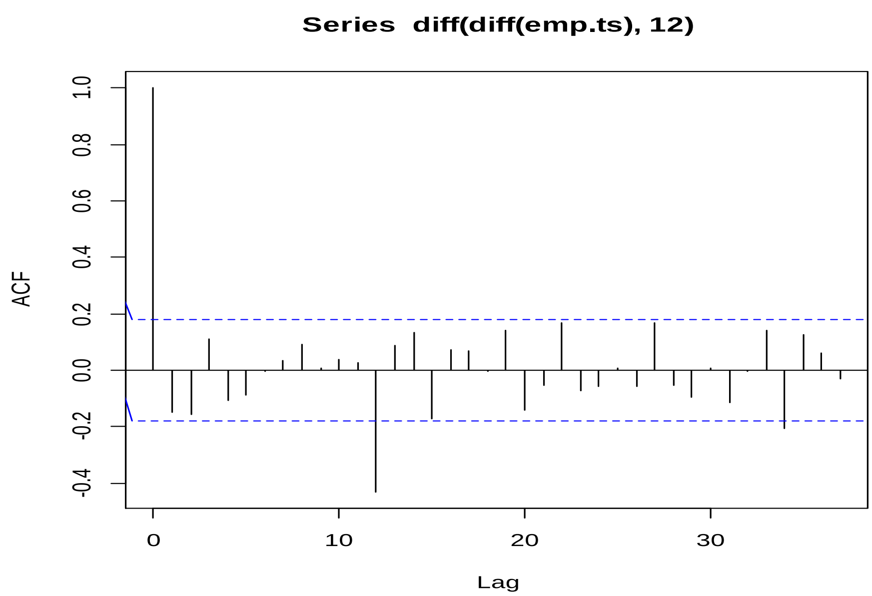




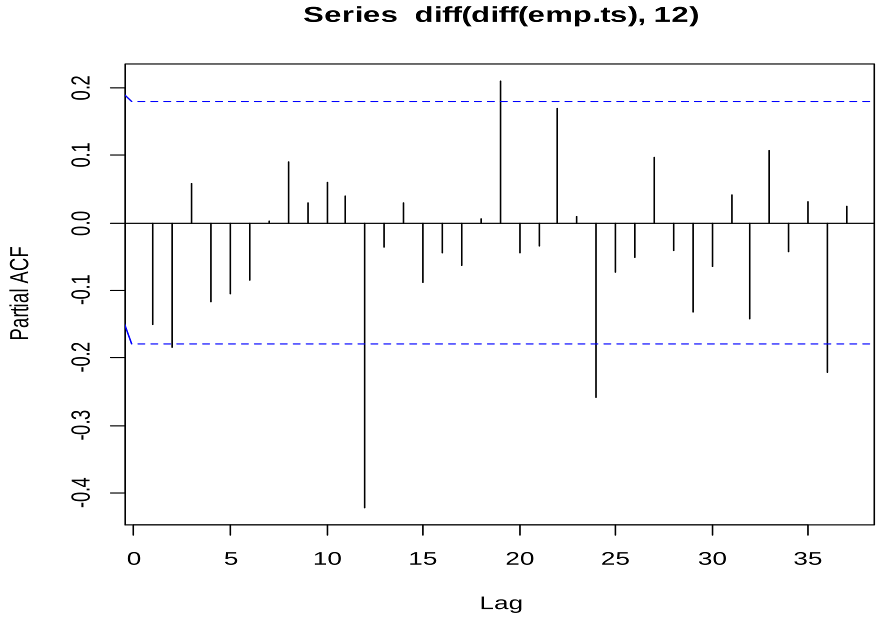
Clearly, the regression model has not provided reduction to white noise. Primarily, the regression has failed to capture trend structure. However, we do believe the model has adequately estimated seasonal structure, because the residual spectrum does not contain peaks at the seasonal frequencies.

Let's turn to ARIMA estimation of the employment series. The plot on page 15 indicates that the series requires both regular and seasonal differencing. The acf and pacf for the series so differenced follow.

```
> emp.ts<-ts(employment)
> acf(diff(diff(emp.ts),12),37)
```



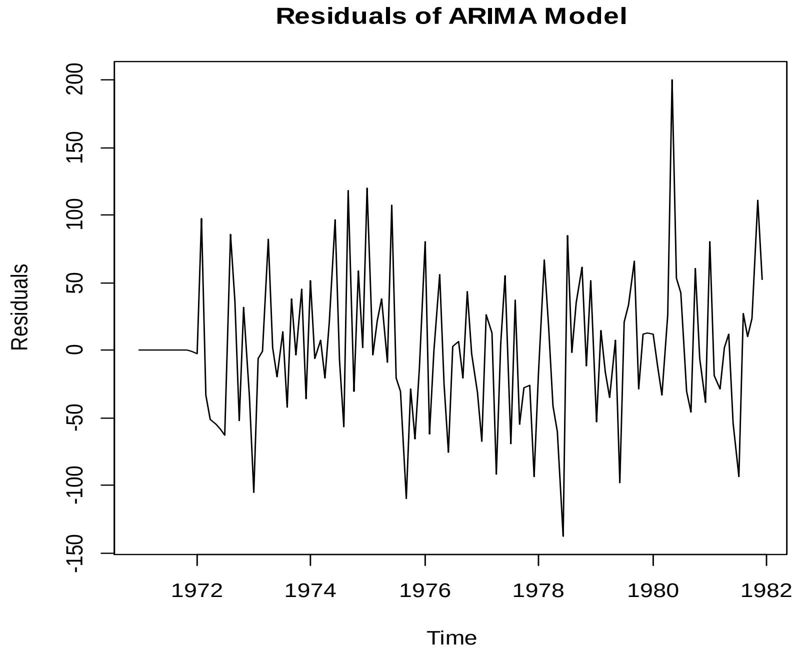
```
> pacf(diff(diff(emp.ts),12),37)
```



The pacf plot suggests trying an ARIMA(2,1,0)(3,1,0)$_{12}$ fit.

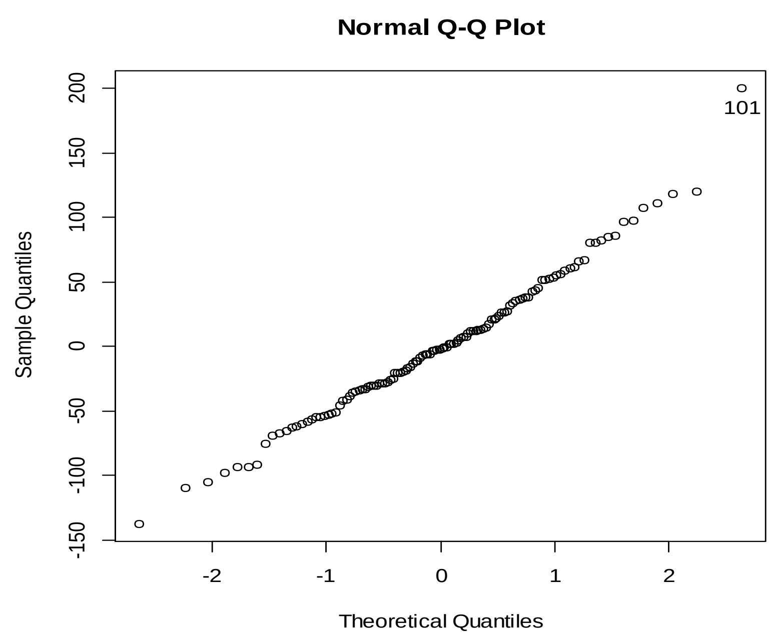
```
> modelarima<-
arima(emp.ts,order=c(2,1,0),seasonal=list(order=c(3,1,0),period=12))
> modelarima

Call:
arima(x = emp.ts, order = c(2, 1, 0), seasonal = list(order = c(3, 1,
0), period = 12))

Coefficients:
          ar1      ar2     sar1     sar2     sar3
      -0.2591  -0.1526  -0.7498  -0.5442  -0.2363
s.e.   0.0927   0.0927   0.0965   0.1168   0.1051

sigma^2 estimated as 3012:  log likelihood = -650.05,  aic = 1312.1

> coeftest(modelarima)

z test of coefficients:

      Estimate Std. Error z value  Pr(>|z|)    
ar1  -0.259145   0.092719 -2.7950  0.005191 ** 
ar2  -0.152616   0.092721 -1.6460  0.099770 .  
sar1 -0.749797   0.096477 -7.7718 7.737e-15 ***
sar2 -0.544200   0.116847 -4.6574 3.203e-06 ***
sar3 -0.236315   0.105143 -2.2476  0.024604 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
>
plot(ts(resid(modelarima),start=c(1971,1),freq=12),xlab="Time",ylab="Re
siduals",main="Residuals of ARIMA Model")
```



This residual plot shows that we need to exclude the first twelve residuals, and also that there is an outlier which needs to be addressed.

```
> sel<-1:12
> qq<-qqnorm(resid(modelarima)[-sel])
> identify(qq)
```

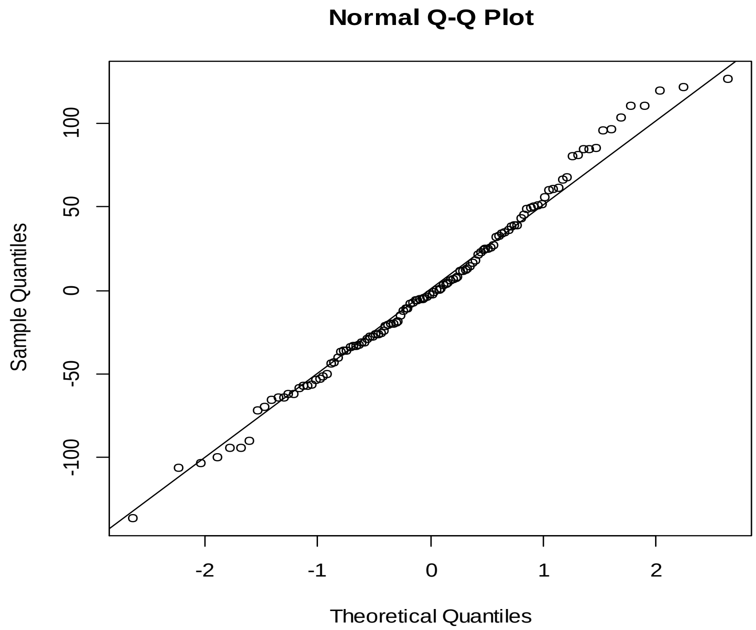


The outlier is at location 113 before removal of the first twelve residuals. Thus, it is the May 1980 data point, and the employment value for this month is 848 (in thousands). Let's smooth it to 775 and refit the same ARIMA model.

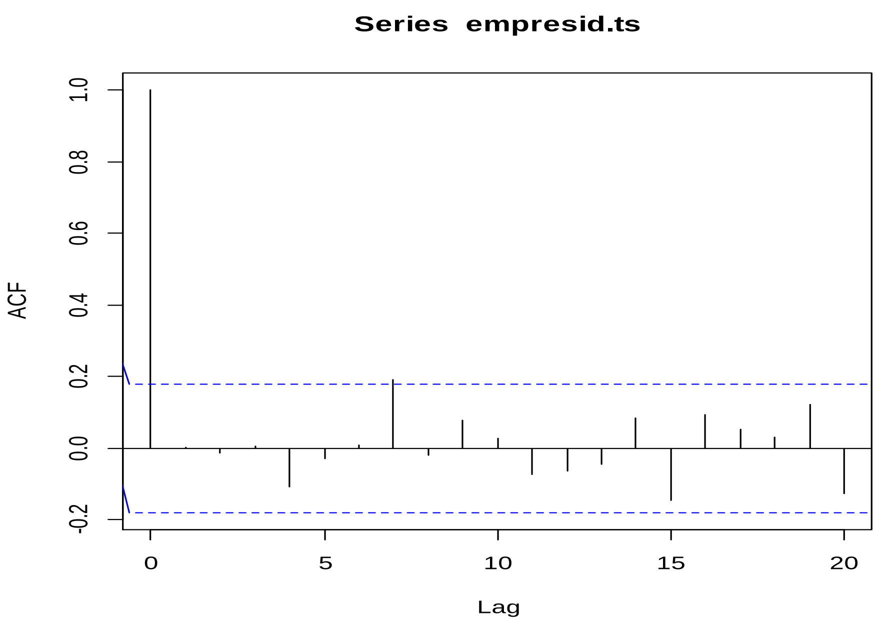
```
> emp.ts[113]<-775
> modelarima2<-
arima(emp.ts,order=c(2,1,0),seasonal=list(order=c(3,1,0),period=12))
> modelarima2

Call:
arima(x = emp.ts, order = c(2, 1, 0), seasonal = list(order = c(3, 1,
0), period = 12))

Coefficients:
          ar1      ar2     sar1     sar2     sar3
      -0.2499  -0.1533  -0.7754  -0.5624  -0.2373
s.e.   0.0921   0.0930   0.0974   0.1166   0.1047

sigma^2 estimated as 2894:  log likelihood = -647.91,  aic = 1307.82

> coeftest(modelarima2)

z test of coefficients:

      Estimate Std. Error z value  Pr(>|z|)
ar1  -0.249948   0.092137 -2.7128  0.006672 **
ar2  -0.153283   0.092994 -1.6483  0.099290 .
sar1 -0.775380   0.097361 -7.9640 1.666e-15 ***
sar2 -0.562382   0.116567 -4.8245 1.403e-06 ***
sar3 -0.237287   0.104746 -2.2654  0.023490 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

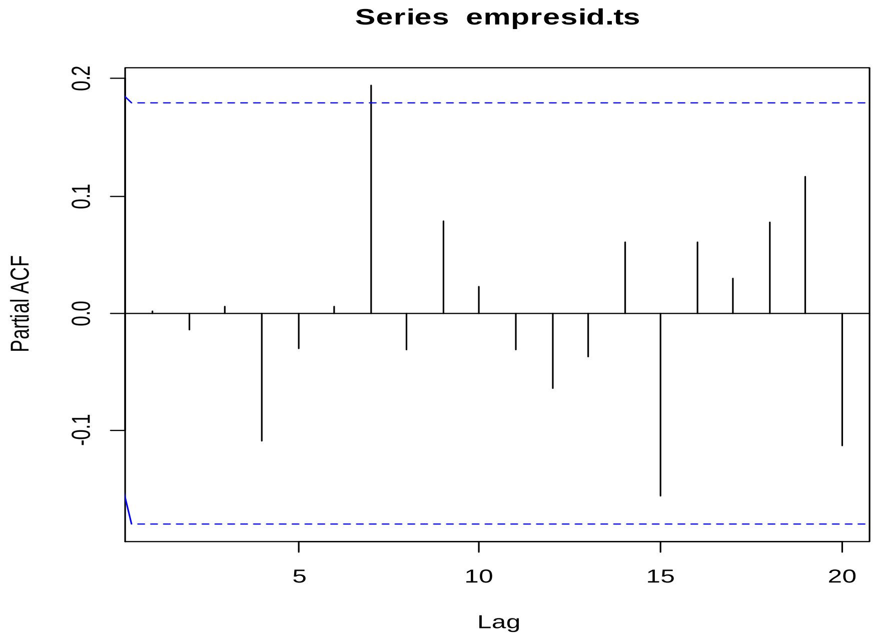
This model fit is very close to the fit given by the first ARIMA model without smoothing the outlier. Residual analysis follows.

```
> qqnorm(resid(modelarima2)[-sel])
> qqline(resid(modelarima2)[-sel])
```

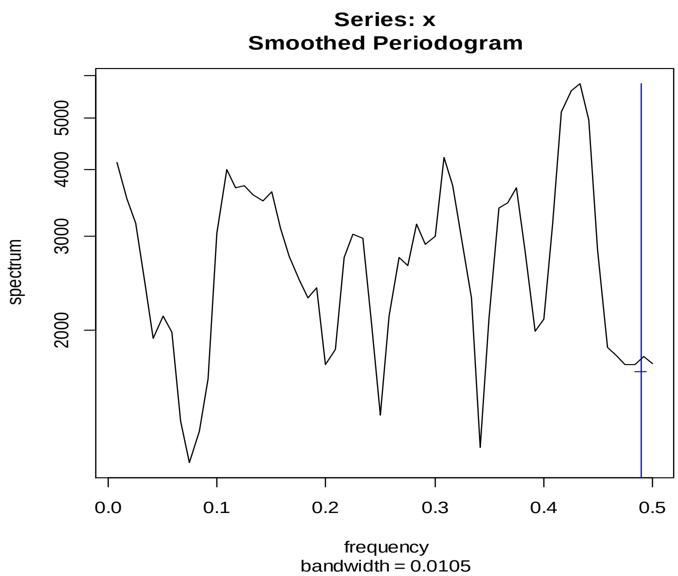


The residual acf and pacf both have a mildly significant lag 8 result.






```
> spectrum(empresid.ts,span=4)
```



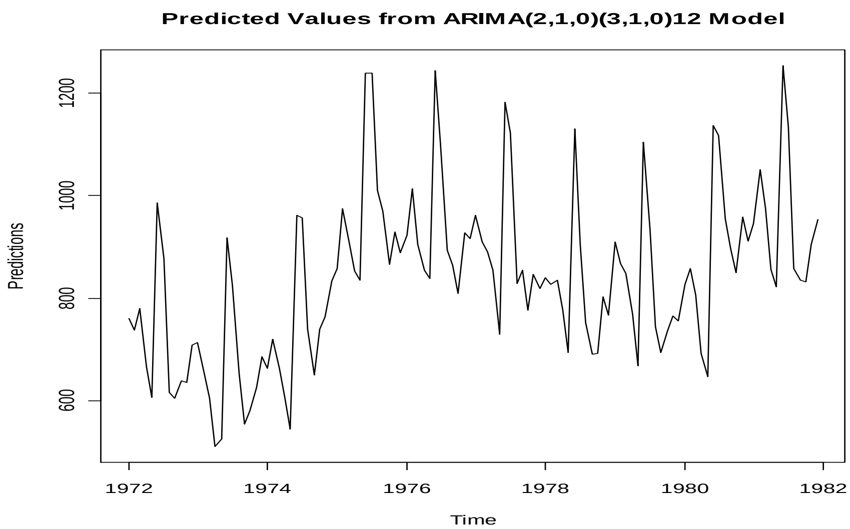

```
> bartlettB.test(empresid.ts)

        Bartlett B Test for white noise

data:
= 0.39704, p-value = 0.9975
```

Bartlett's Kolmogorov–Smirnov test does not reject the white noise hypothesis. The spectrum is reasonably flat except for a modest peak near frequency 0.44. I've tried adjusting for calendar frequency 0.432, but there is no significance.

To continue study of the employment data, let’s obtain seasonal index estimates from the ARIMA model fit. The first step is to save and plot the predicted values from the ARIMA model fit. In doing so, we eliminate the first year of data, because the ARIMA residuals for 1971 are not proper.

```
> empresid.ts<-resid(modelarima2)
> arimapred<-emp.ts[-sel]-empresid.ts[-sel]
```

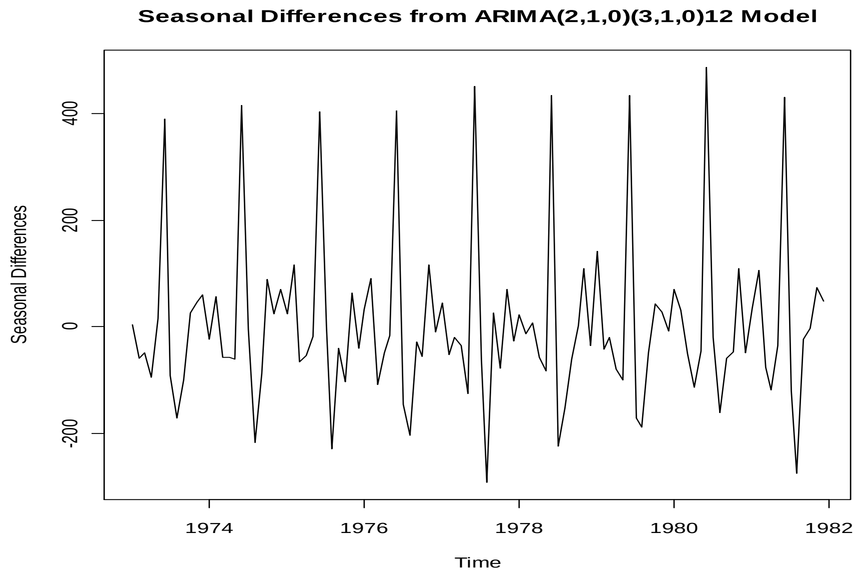
The plot of ARIMA predicted values follows. They run from January 1972 through December 1981.



The plot suggests that the ARIMA predicted values are given by a trend plus a seasonal pattern. Based on the visual evidence and the context of the time series, it is unlikely that any additional structural components are included in these ARIMA predicted values. Thus, to isolate seasonal structure, we need to eliminate the trend structure. Rather than attempt to remove the trend by regression, we’ll difference the predicted values. This will result in loss of the value for January 1972, and thus we’ll consider only the full years 1973 through 1981 subsequently.

```
> darimapred<-diff(arimapred)
> length(darimapred)
[1] 119
> sel2<-1:11
> darimapred<-darimapred[-sel2]
> length(darimapred)
[1] 108
```

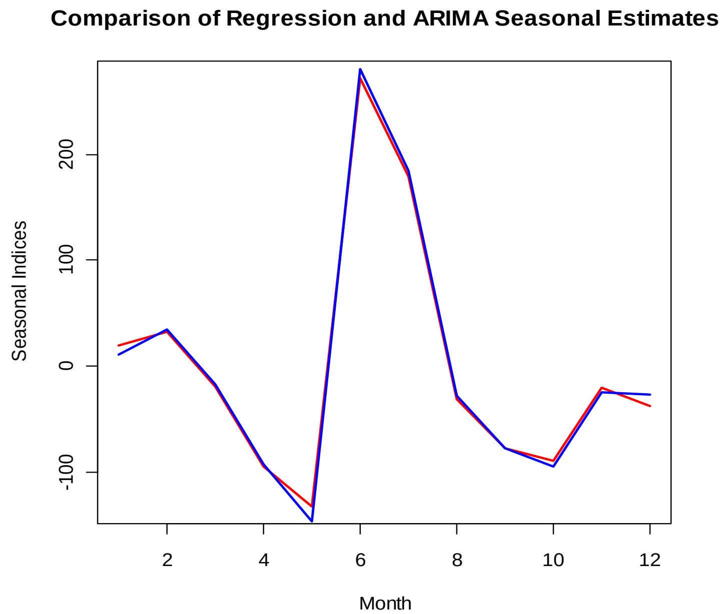
The plot after differencing and removing the year 1972, as in the code directly above, follows.



This result is reasonable—there is no remaining trend.

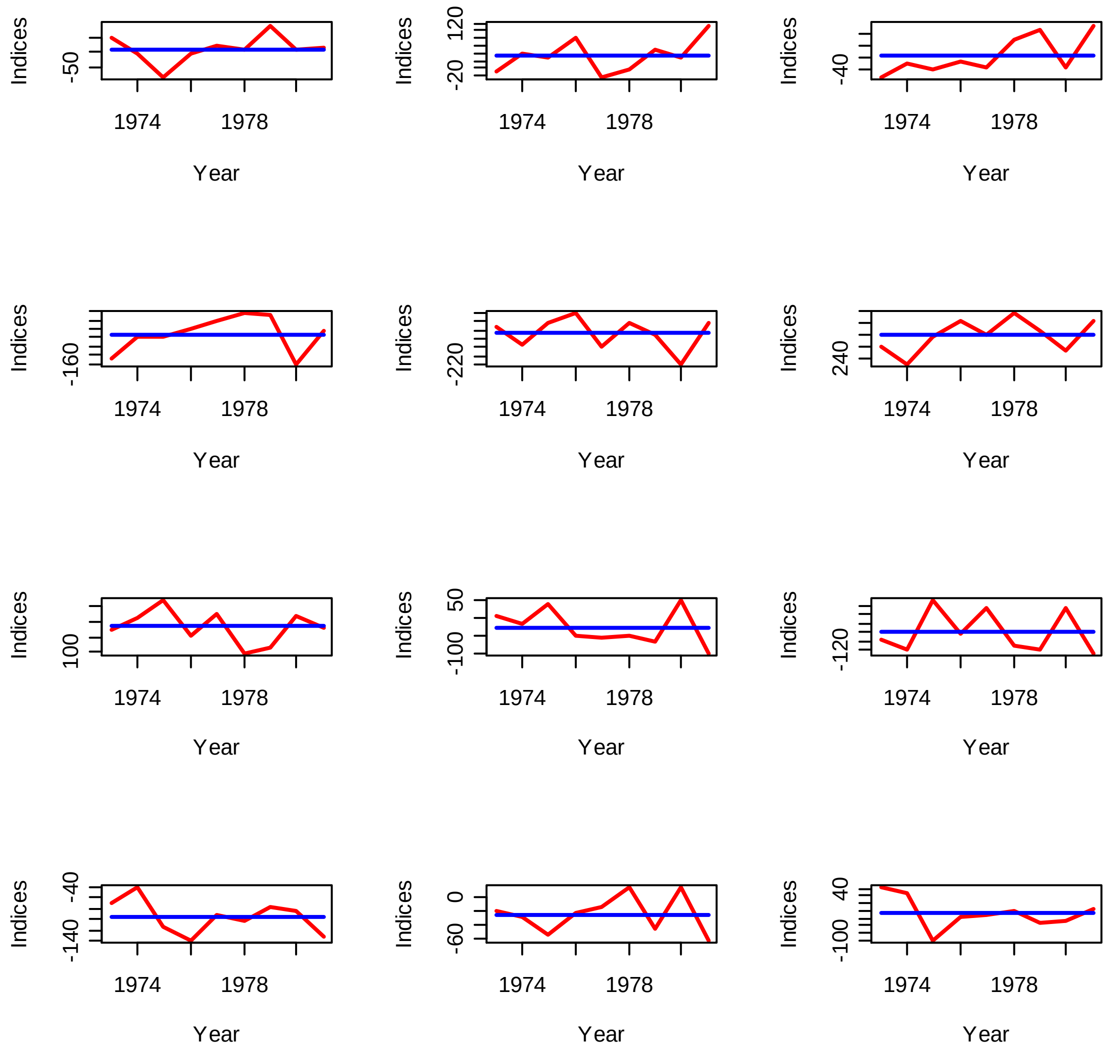
Next, we need to construct the static and dynamic seasonal estimates from the fitted ARIMA model. We need to convert the above estimated seasonal index differences into estimated seasonal indices. The methodology to do this when we are working with the log of the response is on page 35 of the 24 January notes. Here, however, the response, employment, is not logged. The required methodology is on pages 37–38 of the 24 January notes. We begin with R code to undo the differencing. This has to be done for each of the nine years from 1973 through 1981. After the differencing is undone, the static seasonal estimates are constructed.

```
> y<-darimapred
> #length 108
> #to start, for each year adjust the differences to add to 0
> #use the adjusted differences to construct seasonal estimates
> seasm<-matrix(rep(0,108),ncol=9)
> j<--11
> for(ii in 1:9){
+ j<-j+12;j2<-j+11
+ y[j:j2]<-y[j:j2]-mean(y[j:j2])
+ #construct S12
+ j1<-j+1
+ seasm[12,ii]<-0
+ for(i in j1:j2){
+ sub<-y[i:j2]
+ seasm[12,ii]<-seasm[12,ii]+sum(sub)
+ }
+ seasm[12,ii]<-seasm[12,ii]/12
+ #find other S values
+ j3<-j+10
+ ir<-0
+ for(i in j:j3){
+ ir<-ir+1
+ sub<-y[j:i]
+ seasm[ir,ii]<-seasm[12,ii]+sum(sub)
+ }
+ }
> #static seasonal
> seasstatic<-rowMeans(seasm)-mean(rowMeans(seasm))
> seasstatic
 [1]   10.43599   34.23354  -16.88956  -92.52506 -146.82391  279.81821
 [7]  183.99315  -28.44697  -77.76469  -94.36584  -25.18238  -26.48247

 > cbind(1:12,seasreg,seasstatic)

           month seasreg   seasstatic
             1   18.98644   10.43599
             2   32.78083   34.23354
             3  -19.25450  -16.88956
             4  -94.73557  -92.52506
             5 -132.09754 -146.82391
             6  271.49629  279.81821
             7  178.97266  183.99315
             8  -30.74252  -28.44697
             9  -77.08773  -77.76469
            10  -89.59312  -94.36584
            11  -20.78953  -25.18238
            12  -37.93571  -26.48247
```



The red curve is for the regression estimation and the blue curve for the ARIMA estimation. Next, let's look at dynamic estimation of the seasonal structure from the ARIMA estimation.

```
> year<-seq(1973,1981)
> seasstaticm<-matrix(rep(seasstatic,9),ncol=9)
> par(mfrow=c(4,3))
> for(i in 1:12){
+
plot(year,seasm[i,],xlab="Year",ylab="Indices",type="l",lwd=2,col="red"
)
+ lines(year,seasstaticm[i,],lty=1,lwd=2,col="blue")
+ }
```



The estimated seasonal indices do change somewhat from year to year, but no prominent changes during the period 1973 to 1981 are evident.

E. Monthly Australian beer production, January 1956—August 1995.

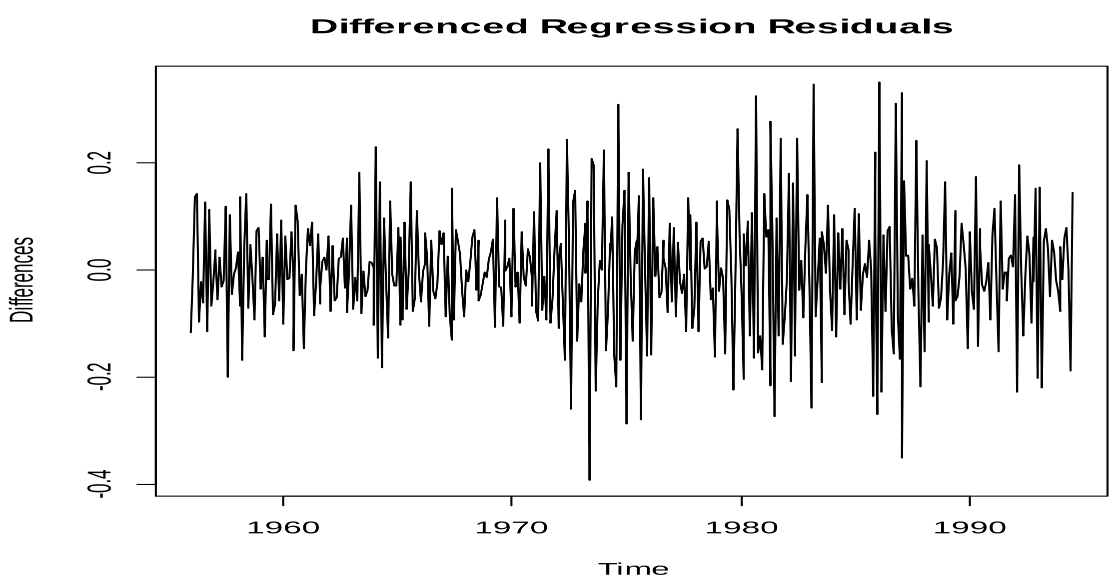
A regression model was fit to this data set in the 24 January class notes. Let's fit an ARIMA model to the logged data.

The May 1982 production figure is unusually low. In addition, the time series contains significant calendar effects at frequencies 0.348 and 0.432. We begin with a regression analysis fit to the log production data to adjust the outlier and remove the calendar effects. This will leave trend and seasonal structure, and an ARIMA model will be fit to the regression residuals.

```
> time<-1:length(beer)
> c348<-cos(0.696*pi*time);s348<-sin(0.696*pi*time)
> c432<-cos(0.864*pi*time);s432<-sin(0.864*pi*time)
> model<-lm(log(beer)~obs317+c348+s348+c432+s432);summary(model)

Call:
lm(formula = log(beer) ~ obs317 + c348 + s348 + c432 + s432)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.68216 -0.18118  0.06064  0.17982  0.52662 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  4.882513   0.012126 402.657   <2e-16 ***
obs317      -0.104520   0.265669  -0.393   0.6942    
c348        -0.021569   0.017154  -1.257   0.2092    
s348        -0.034353   0.017143  -2.004   0.0457 *  
c432        -0.012245   0.017192  -0.712   0.4767    
s432         0.006673   0.017105   0.390   0.6966    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 

Residual standard error: 0.2643 on 470 degrees of freedom
Multiple R-squared:  0.01371,   Adjusted R-squared:  0.003214 
F-statistic: 1.306 on 5 and 470 DF,  p-value: 0.26
```

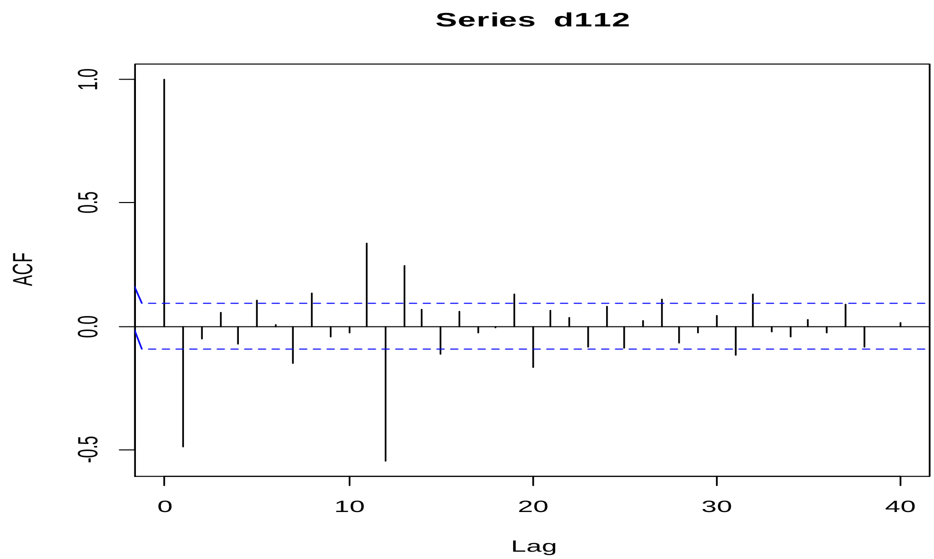
Next, we plot and examine the regression residuals with both ordinary and seasonal differencing.

```
> d112<-diff(diff(resid(model)),12)
> d112.ts<-ts(d112,start=c(1956,1),freq=12)
> plot(d112.ts,xlab="Time",ylab="Differences",main="Differenced
Regression Residuals")
```

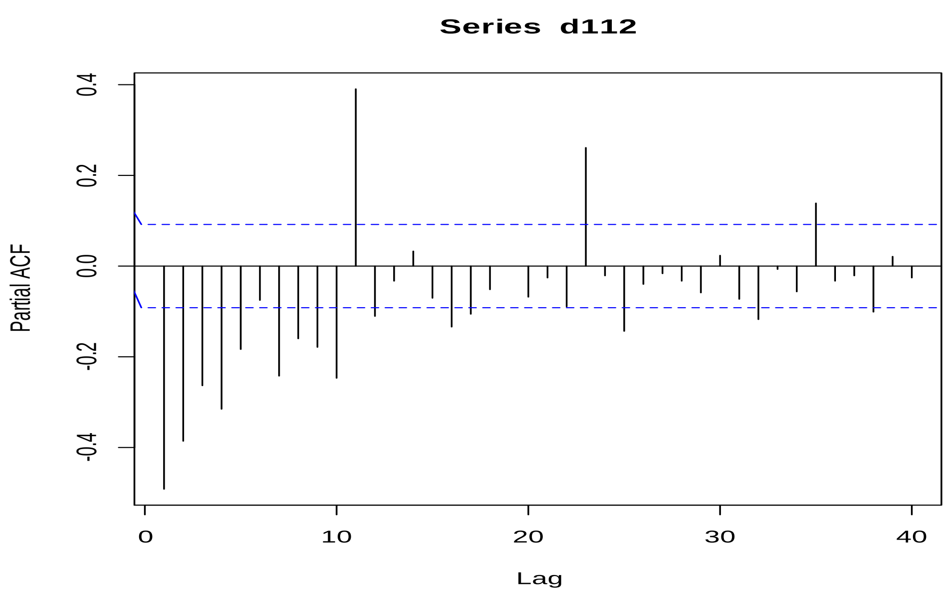


The residuals exhibit some changing volatility, but it is not severe.

```
> acf(d112,40)
```




```
> pacf(d112,40)
```

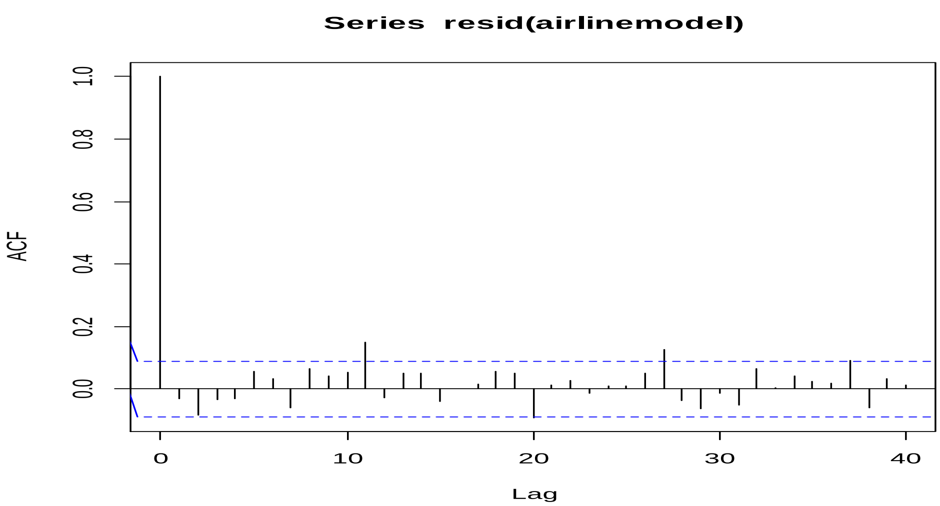


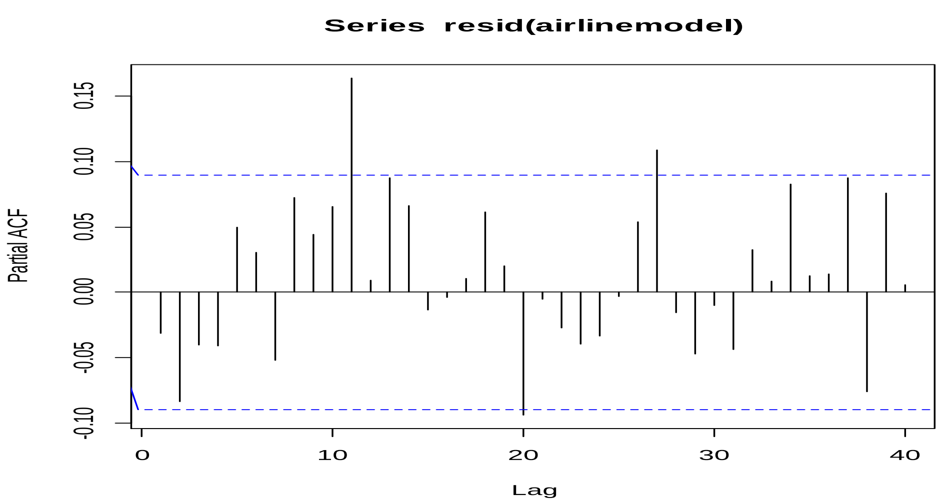
The following ARIMA fit is the airline model applied to the regression residuals.

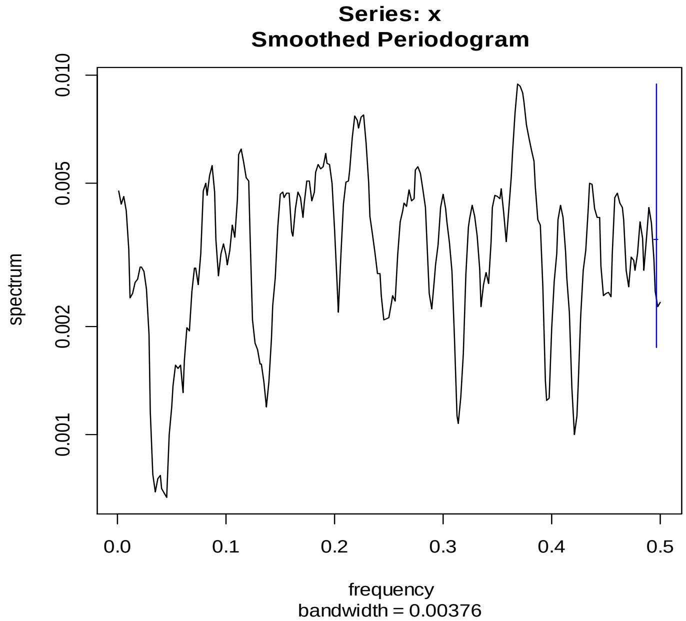
```
> airlinemodel

Call:
arima(x = ts(resid(model)), order = c(0, 1, 1), seasonal = list(order =
c(0, 1, 1), period = 12))
```

```
Coefficients:
          ma1     sma1
      -0.8814  -0.8270
s.e.   0.0182   0.0305

sigma^2 estimated as 0.0036:  log likelihood = 637.78,  aic = -1269.55

> coeftest(airlinemodel)

z test of coefficients:

      Estimate Std. Error z value  Pr(>|z|)
ma1  -0.881405   0.018211 -48.400 < 2.2e-16 ***
sma1 -0.827008   0.030515 -27.102 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Some residual diagnostics follow.







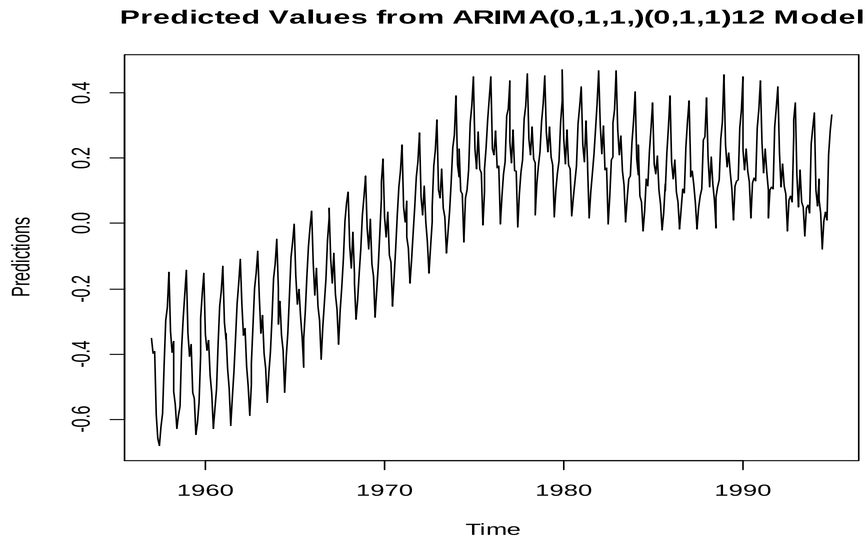
The significant residual correlation and partial correlation at lag 11 is a little worrisome. However, the spectral plot is less troubling, and Bartlett's test does not reject the hypothesis of reduction to white noise.

```
> bartlettB.test(resid(airlinemodel))

        Bartlett B Test for white noise

data:
= 0.93961, p-value = 0.3404
```

We proceed to use the predicted values from this ARIMA model to construct seasonal index estimates. We delete the years 1956 and 1995, the latter with data only through August.

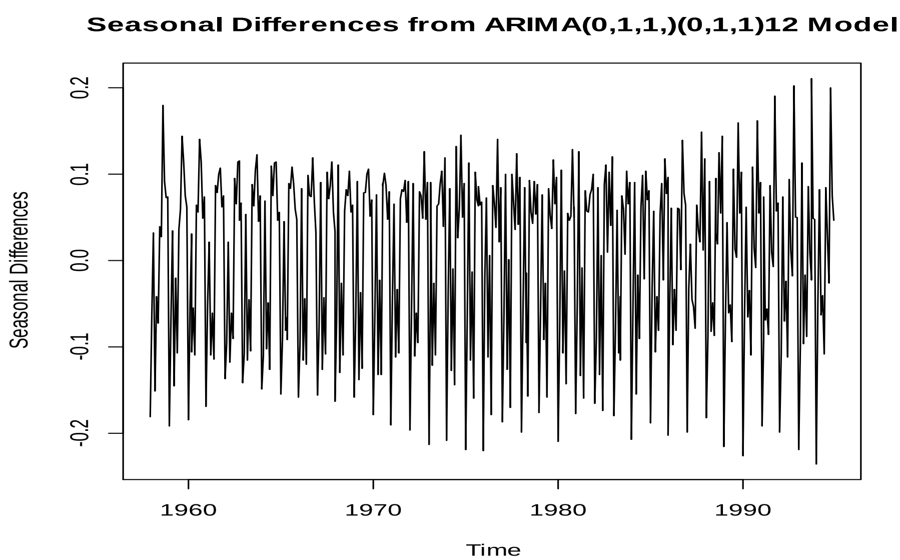
```
> sel<-c(1:12,469:476)
> ausresid.ts<-resid(airlinemodel)
> arimapred<-ts(resid(model)[-sel])-ausresid.ts[-sel]
```

The plot of predicted values for the years 1957 to 1994 follows.



Next, construct the plot of differences of the predicted values, to remove the trend. This series is constructed to span the years 1958 to 1994.

```
> darimapred<-diff(arimapred)
> length(darimapred)
[1] 455
> sel2<-1:11
> darimapred<-darimapred[-sel2]
> length(darimapred)
[1] 444
```



This analysis for the Australian beer production data has employed the log of the response. Thus, we apply the methodology and code on pages 35–36 of the 24 January notes. A few modifications are required, though.

```
> y<-darimapred
> #length 444
> #for each year adjust the differences to add to 0
> #use the adjusted differences to construct seasonal estimates
> seasm<-matrix(rep(0,444),ncol=37)
> j<--11
> for(ii in 1:37){
+ j<-j+12;j2<-j+11
+ y[j:j2]<-exp(y[j:j2]-mean(y[j:j2]))
+ #construct S12
+ j1<-j+1
+ seasm[12,ii]<-1
+ for(i in j1:j2){
+ sub<-y[i:j2]
+ seasm[12,ii]<-seasm[12,ii]*prod(sub)
+ }
+ seasm[12,ii]<-(seasm[12,ii])^(1/12)
+ #find the other S values
+ j3<-j+10
+ ir<-0
+ for(i in j:j3){
+ ir<-ir+1
+ sub<-y[j:i]
+ seasm[ir,ii]<-seasm[12,ii]*prod(sub)
+ }
+ }
> #static seasonal, constructed as a geometric mean across the 37 years
> seasstatic<-apply(seasm,1,prod)^(1/37)

> seasstatic
 [1] 1.0482775 0.9799439 1.0524820 0.9440552 0.9123043 0.8111210 0.8799175
 [8] 0.9313196 0.9812142 1.1100266 1.1714895 1.2662188
```

Regression estimation of the seasonal indices is shown in the 24 January notes.

```
 [1] 1.0387 0.9723 1.0540 0.9488 0.9194 0.8198 0.8871 0.9380 0.9660 1.1076
[11] 1.1715 1.2596
```

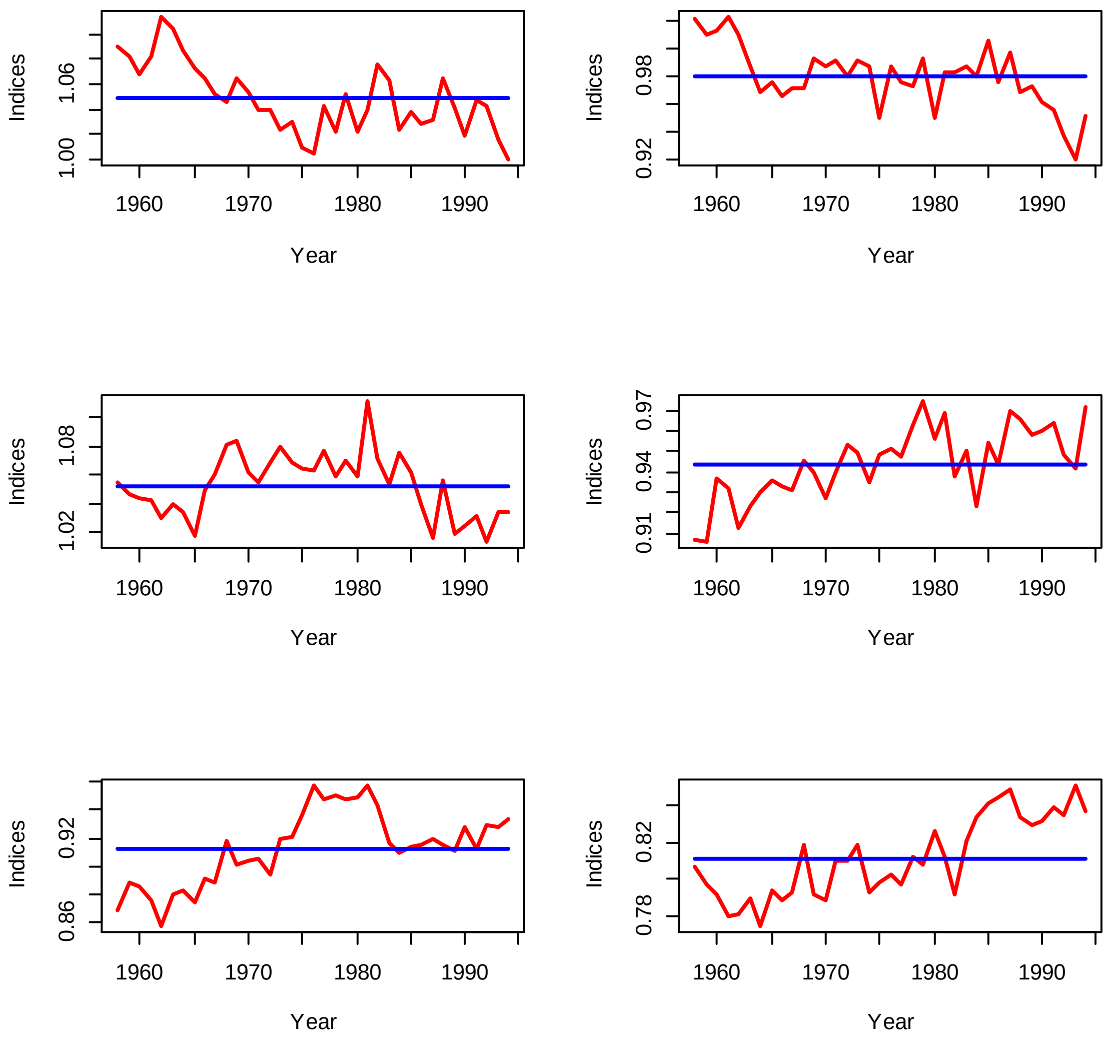
The two sets of estimates are very close to each other.

In the above code, a geometric mean is used to calculate the estimated static seasonal estimates. In fact, an arithmetic mean can also be used, and it gives essentially the same estimates. Here is the calculation:

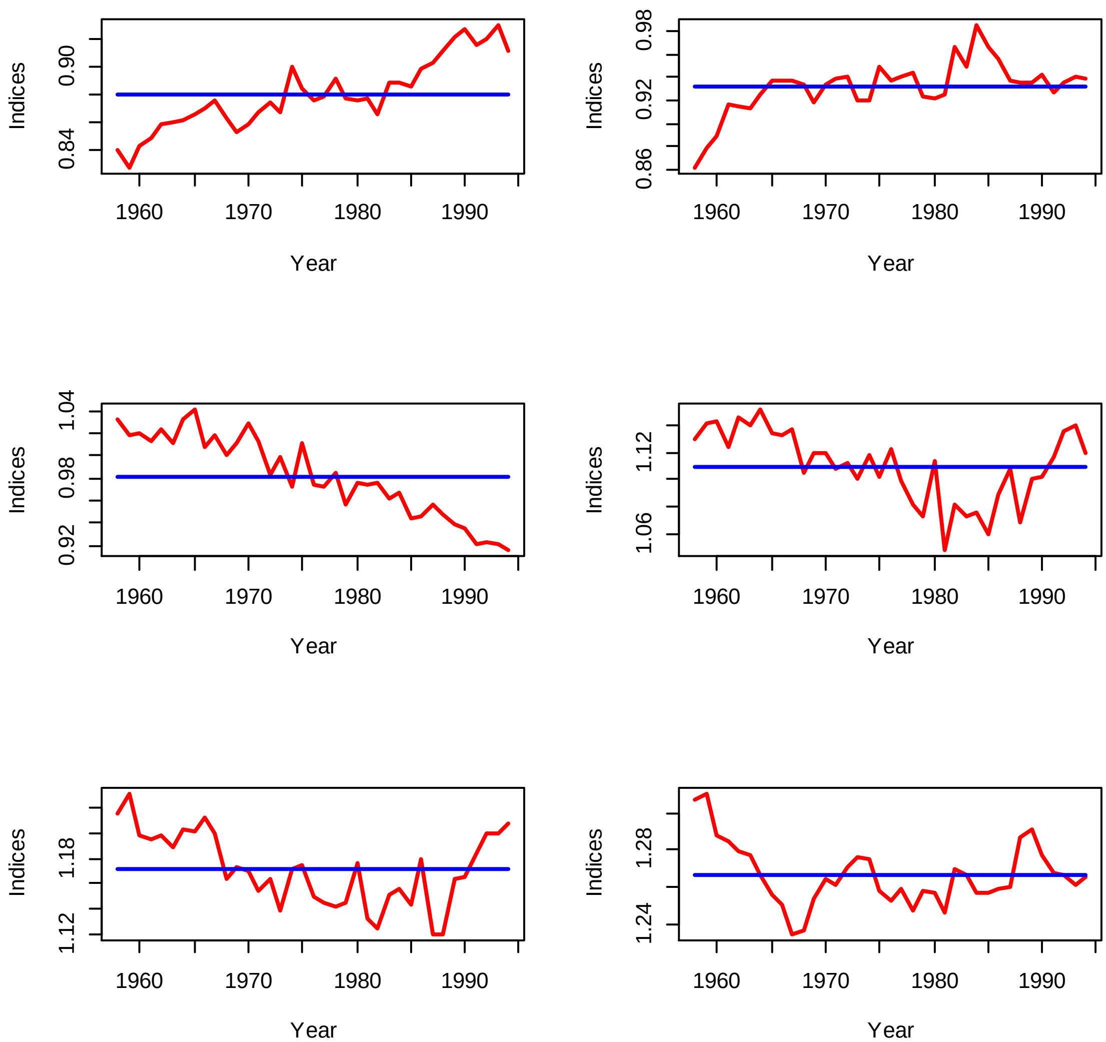
```
> seasstatic2<-apply(seasm,1,mean)
> seasstatic2
 [1] 1.0486267 0.9801908 1.0527069 0.9442158 0.9126766 0.8113932
0.8802791
 [8] 0.9315860 0.9818778 1.1103427 1.1718335 1.2663264
```

Dynamic estimation of the seasonal structure is portrayed on the next two pages. The first matrix plot is for January through June, and the second is for July through December.

```
> year<-seq(1958,1994)
> seasstaticm<-matrix(rep(seasstatic,37),ncol=37)
> par(mfrow=c(3,2))
> for(i in 1:6){
+
plot(year,seasm[i,],xlab="Year",ylab="Indices",type="l",lwd=2,col="red"
)
+ lines(year,seasstaticm[i,],lty=1,lwd=2,col="blue")
+ }
```



```
> par(mfrow=c(3,2))
> for(i in 7:12){
+
plot(year,seasm[i,],xlab="Year",ylab="Indices",type="l",lwd=2,col="red"
)
+ lines(year,seasstaticm[i,],lty=1,lwd=2,col="blue")
+ }
```



The plots show that the seasonal has dynamic structure. Notably, the seasonal index is estimated to have increased over the years during April, May, June, and July, fall and winter months. And the seasonal index is estimated to have decreased over time during January, February, September, October, and November, summer and spring months.

Summary and additional remarks

These notes continue the exploration of seasonal ARIMA model fitting.

1. The quarterly Iowa nonfarm income time series is fit with several models. One feature of the series is a significant four-year cycle to account for state-wide elections.

2. The monthly employment series for males aged 16 to 19 is studied. Static seasonal index estimates for regression and ARIMA modelling are compared and are found to be very similar, and an ARIMA model fit is used to produce dynamic estimation of seasonal structure.

3. For the Australian beer series, a regression analysis is employed initially to remove calendar structure, and then the airline ARIMA model applied to the regression residuals provides a good fit. This fit is used to obtain both static and dynamic seasonal index estimates, and the static estimates are seen to be very similar to the estimates provided by a regression approach.

4. We have seen that the regression and ARIMA approaches lead to similar estimations of static seasonal structure for the employment and Australian beer series. We expect the degree of agreement between the two estimations to vary to some extent among different data sets, and, of course, the amount of agreement will also depend on how well the model estimations match the data.