HOMEWORK ASSIGNMENT 9

Zachary Lazerick

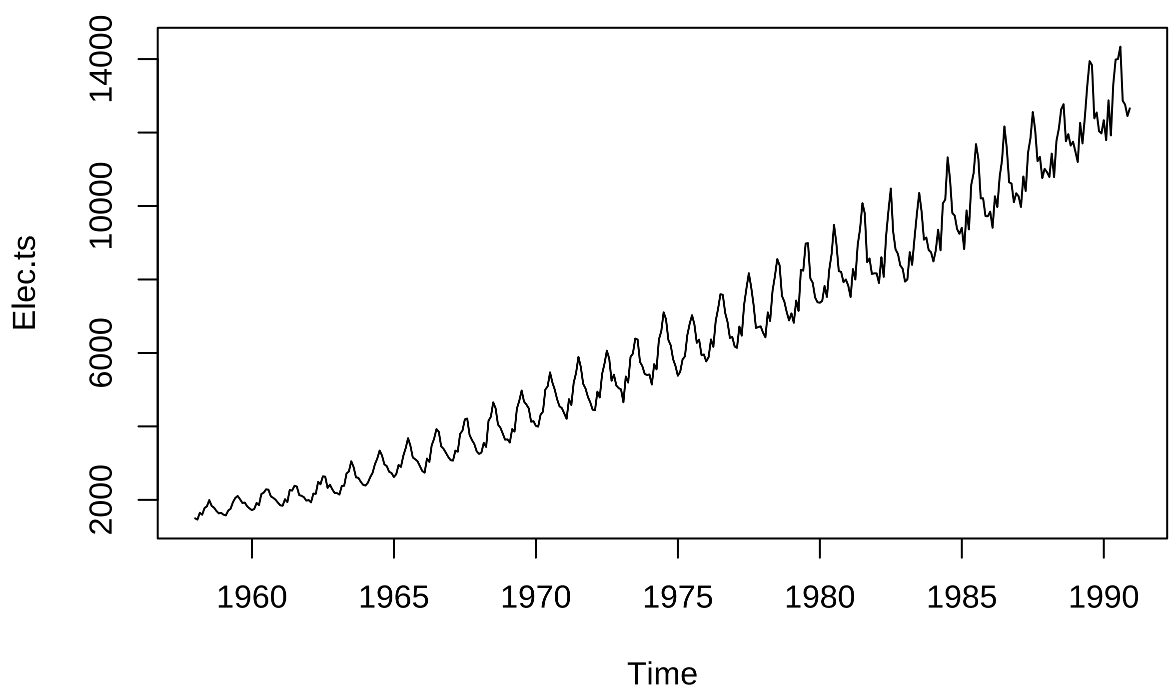
13 April 2023

PROBLEM 1

This question is based on the electricity production series (1958–1990).

a) Read in and plot the electricity time series data. Give one reason why a log-transformation may be appropriate for the electricity series.

```
cbe <- read.table("cbe.dat", header = T)
Elec.ts <- ts(cbe[, 3], start = 1958, freq = 12)
Beer.ts <- ts(cbe[, 2], start = 1958, freq = 12)
Choc.ts <- ts(cbe[, 1], start = 1958, freq = 12)
plot(Elec.ts)
```



One reason why a log-transformation would be appropriate is due to the small degree of curvature (nonlinearity) in the data. Another is due to the growing variation as the series progresses. For the early years, the peaks are about the same size. Later on, the size of the peaks grow with each iteration.

b) Fit a seasonal indicator model with a quadratic trend to the (natural) logarithm of the series. Use stepwise regression to select the best model based on the AIC.

```
attach(cbe)
length(elec)
```

```
## [1] 396
```

```
length(elec)/12
```

```
## [1] 33
```

```
imth = rep(1:12,33)
T1 = 1:396
T2 = T1^2
elec.lm = lm(log(elec)~T1+T2+factor(imth))
elec.lm.best = step(elec.lm)
```

```
## Start:  AIC=-2717.56
## log(elec) ~ T1 + T2 + factor(imth)
##
##                Df Sum of Sq     RSS     AIC
## <none>                       0.3860 -2717.6
## - factor(imth) 11    2.4917  2.8777 -1944.1
## - T2            1    2.5629  2.9489 -1914.4
## - T1            1   20.3972 20.7833 -1141.1
```

```
elec.lm.best
```

```
##
## Call:
## lm(formula = log(elec) ~ T1 + T2 + factor(imth))
##
## Coefficients:
##    (Intercept)              T1              T2  factor(imth)2  factor(imth)3
##      7.271e+00       7.960e-03      -6.883e-06     -1.991e-02      6.598e-02
##  factor(imth)4   factor(imth)5   factor(imth)6  factor(imth)7  factor(imth)8
##      3.288e-02       1.462e-01       1.777e-01      2.375e-01      1.994e-01
##  factor(imth)9  factor(imth)10  factor(imth)11 factor(imth)12
##      1.074e-01       9.044e-02       4.278e-02      2.350e-02
```

c) Fit a harmonic model with a quadratic trend to the logarithm of the series. Use stepwise regression to select the best model based on the AIC.

```
c1=cos(2*pi*T1/12); s1=sin(2*pi*T1/12)
c2=cos(2*2*pi*T1/12); s2=sin(2*2*pi*T1/12)
c3=cos(3*2*pi*T1/12); s3=sin(3*2*pi*T1/12)
c4=cos(4*2*pi*T1/12); s4=sin(4*2*pi*T1/12)
```

```r
c5=cos(5*2*pi*T1/12); s5=sin(5*2*pi*T1/12)
c6=cos(6*2*pi*T1/12); s6=sin(6*2*pi*T1/12)
elec.harm = lm(log(elec)~T1+T2+c1+s1+c2+s2+c3+s3+c4+s4+c5+s5+c6+s6)
elec.harm.best = step(elec.harm)
```

```
## Start:  AIC=-2715.56
## log(elec) ~ T1 + T2 + c1 + s1 + c2 + s2 + c3 + s3 + c4 + s4 +
##     c5 + s5 + c6 + s6
##
##        Df Sum of Sq     RSS     AIC
## - s6    1    0.0000  0.3860 -2717.6
## - s4    1    0.0002  0.3863 -2717.3
## - c3    1    0.0003  0.3863 -2717.2
## - c4    1    0.0005  0.3866 -2717.0
## <none>               0.3860 -2715.6
## - c6    1    0.0195  0.4056 -2698.0
## - c5    1    0.0198  0.4059 -2697.7
## - c2    1    0.0442  0.4303 -2674.6
## - s2    1    0.0451  0.4312 -2673.8
## - s3    1    0.0469  0.4329 -2672.2
## - s5    1    0.0949  0.4809 -2630.5
## - s1    1    0.6664  1.0525 -2320.4
## - c1    1    1.5469  1.9329 -2079.7
## - T2    1    2.5620  2.9480 -1912.5
## - T1    1   20.3924 20.7784 -1139.2
##
## Step:  AIC=-2717.56
## log(elec) ~ T1 + T2 + c1 + s1 + c2 + s2 + c3 + s3 + c4 + s4 +
##     c5 + s5 + c6
##
##        Df Sum of Sq     RSS     AIC
## - s4    1    0.0002  0.3863 -2719.3
## - c3    1    0.0003  0.3863 -2719.2
## - c4    1    0.0005  0.3866 -2719.0
## <none>               0.3860 -2717.6
## - c5    1    0.0199  0.4059 -2699.7
## - c6    1    0.0252  0.4113 -2694.5
## - c2    1    0.0443  0.4303 -2676.6
## - s2    1    0.0451  0.4312 -2675.8
## - s3    1    0.0472  0.4332 -2673.9
## - s5    1    0.0949  0.4810 -2632.5
## - s1    1    0.6664  1.0525 -2322.4
## - c1    1    1.5472  1.9332 -2081.6
## - T2    1    2.5629  2.9489 -1914.4
```

```
## - T1    1   20.3972 20.7833 -1141.1
##
## Step:  AIC=-2719.31
## log(elec) ~ T1 + T2 + c1 + s1 + c2 + s2 + c3 + s3 + c4 + c5 +
##     s5 + c6
##
##        Df Sum of Sq     RSS     AIC
## - c3    1    0.0003  0.3866 -2721.0
## - c4    1    0.0005  0.3868 -2720.8
## <none>               0.3863 -2719.3
## - c5    1    0.0199  0.4062 -2701.4
## - c6    1    0.0252  0.4115 -2696.2
## - c2    1    0.0443  0.4306 -2678.3
## - s2    1    0.0451  0.4314 -2677.5
## - s3    1    0.0472  0.4335 -2675.7
## - s5    1    0.0949  0.4812 -2634.3
## - s1    1    0.6664  1.0527 -2324.3
## - c1    1    1.5472  1.9335 -2083.6
## - T2    1    2.5629  2.9492 -1916.3
## - T1    1   20.3974 20.7837 -1143.1
##
## Step:  AIC=-2720.99
## log(elec) ~ T1 + T2 + c1 + s1 + c2 + s2 + s3 + c4 + c5 + s5 +
##     c6
##
##        Df Sum of Sq     RSS     AIC
## - c4    1    0.0005  0.3871 -2722.4
## <none>               0.3866 -2721.0
## - c5    1    0.0199  0.4065 -2703.1
## - c6    1    0.0252  0.4118 -2697.9
## - c2    1    0.0443  0.4309 -2680.0
## - s2    1    0.0452  0.4317 -2679.2
## - s3    1    0.0472  0.4338 -2677.4
## - s5    1    0.0949  0.4815 -2636.0
## - s1    1    0.6664  1.0530 -2326.2
## - c1    1    1.5472  1.9338 -2085.5
## - T2    1    2.5629  2.9495 -1918.3
## - T1    1   20.3977 20.7842 -1145.1
##
## Step:  AIC=-2722.43
## log(elec) ~ T1 + T2 + c1 + s1 + c2 + s2 + s3 + c5 + s5 + c6
##
##        Df Sum of Sq     RSS     AIC
## <none>               0.3871 -2722.4
## - c5    1    0.0199  0.4070 -2704.6
```

```
## - c6    1    0.0252  0.4124 -2699.4
## - c2    1    0.0443  0.4314 -2681.5
## - s2    1    0.0452  0.4323 -2680.8
## - s3    1    0.0472  0.4343 -2678.9
## - s5    1    0.0949  0.4821 -2637.6
## - s1    1    0.6664  1.0536 -2328.0
## - c1    1    1.5472  1.9343 -2087.4
## - T2    1    2.5629  2.9500 -1920.2
## - T1    1   20.3980 20.7852 -1147.1
```

```
elec.harm.best
```

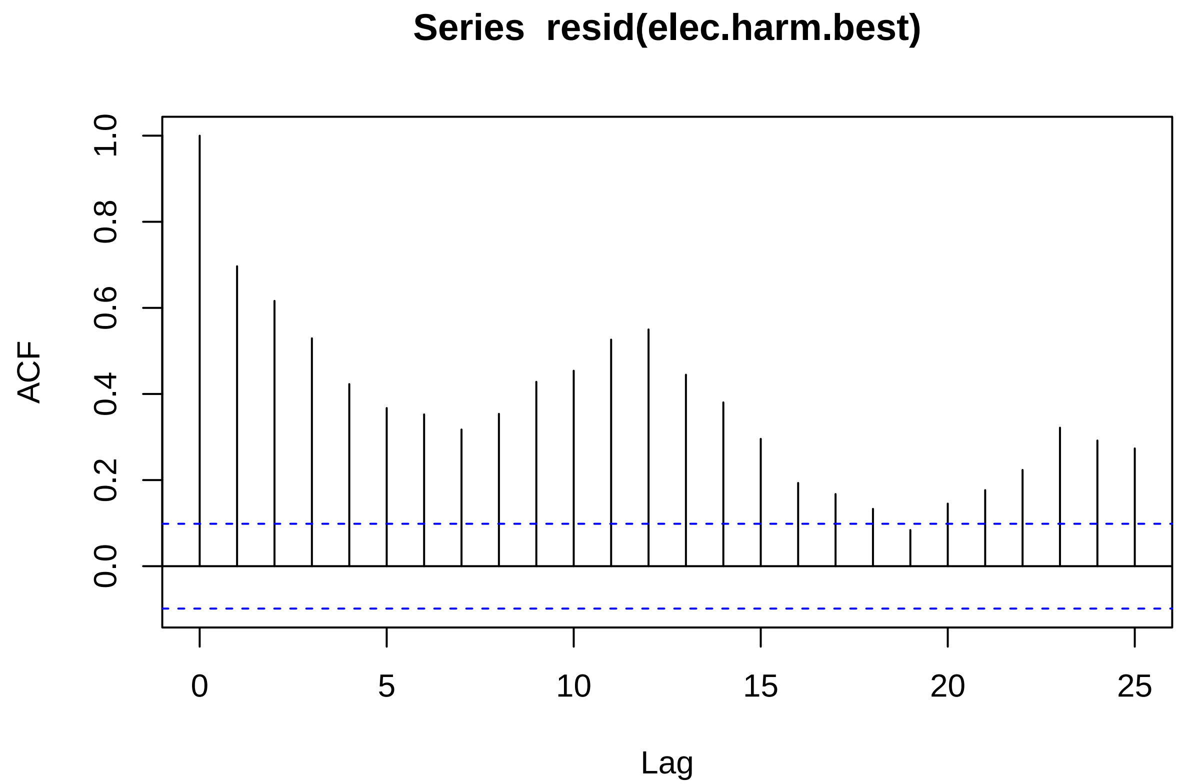
```
##
## Call:
## lm(formula = log(elec) ~ T1 + T2 + c1 + s1 + c2 + s2 + s3 + c5 +
##     s5 + c6)
##
## Coefficients:
## (Intercept)           T1           T2           c1           s1           c2
##   7.363e+00    7.961e-03   -6.883e-06   -8.840e-02   -5.803e-02    1.496e-02
##          s2           s3           c5           s5           c6
##   1.510e-02   -1.544e-02    1.003e-02    2.190e-02   -7.985e-03
```

```
coef(elec.harm.best)/sqrt(diag(vcov(elec.harm.best)))
```

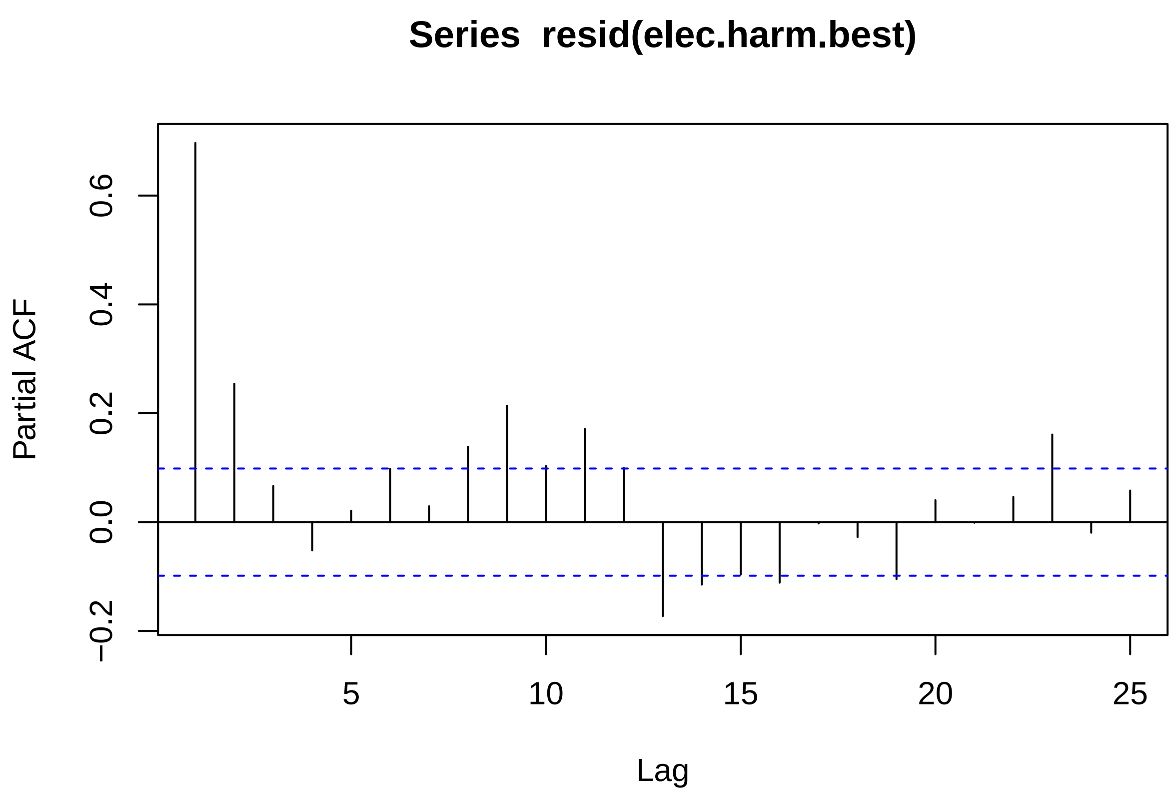
```
## (Intercept)          T1          T2          c1          s1          c2
## 1532.290700  142.426743  -50.485062  -39.225472  -25.743878    6.636567
##          s2          s3          c5          s5          c6
##    6.700863   -6.849451    4.448522    9.715774   -5.011018
```

d) Plot the correlogram and partial correlogram of the residuals from the overall best-fitting model and comment on the plots.

```
acf(resid(elec.harm.best))
```




```
pacf(resid(elec.harm.best))
```



The ach shows that the residuals from the best-fitting model have a very high degree of serial correlation. This serial correlation ebbs and flows like a cosine function, whilst remaining mostly throughout the process. The pacf shows that the residuals may be modeled by an AR(2) process, however, there are multiple peaks for several other later lags, not emblematic of an AR(p) process where we would expect that for lags greater than p, the autocorrelation not be significant.

e) Fit an AR model to the residuals of the best-fitting model. Give the order of the best-fitting AR model and the estimated model parameters.

```
fit.ar <- ar(resid(elec.harm.best))
fit.ar$order
```

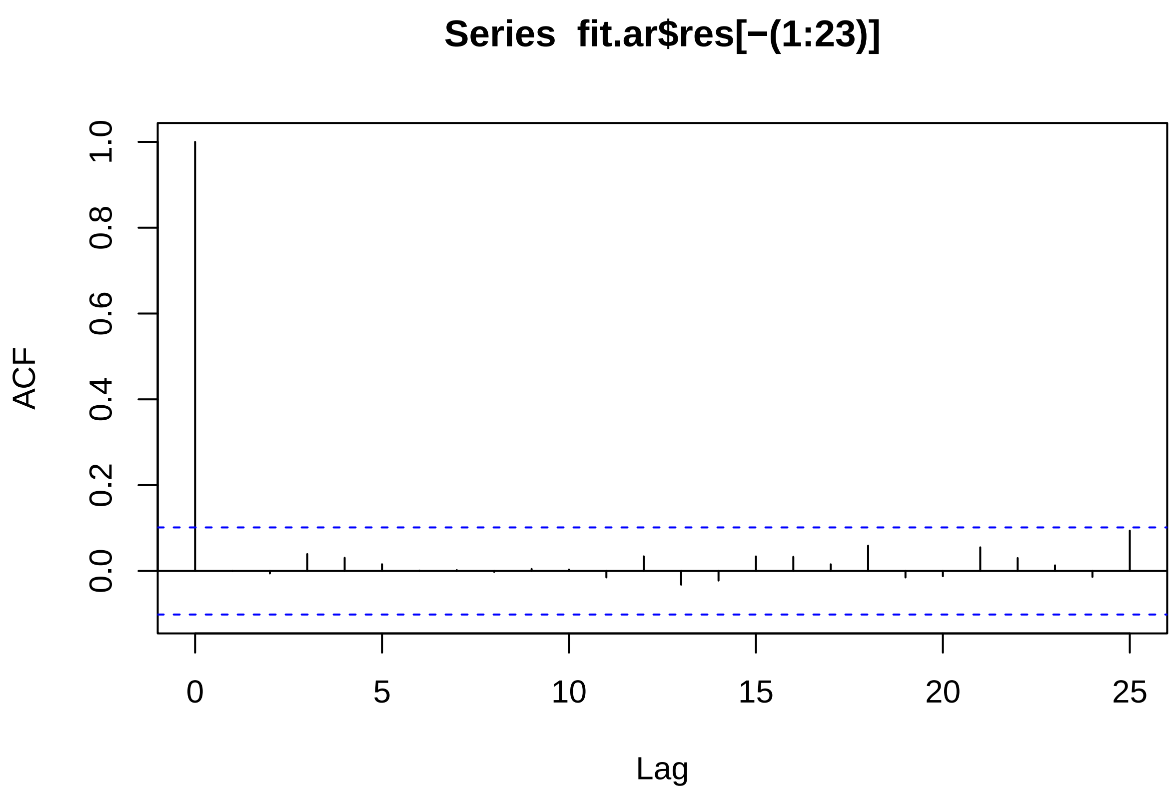
```
## [1] 23
```

```
fit.ar$ar
```

```
##  [1]  0.3811774774  0.2186873536  0.0852889313  0.0274090685 -0.0004494992
##  [6]  0.0349626750 -0.0240764975  0.0071383342  0.1094138047  0.0317226003
## [11]  0.1207815671  0.1714572988 -0.0889351218 -0.0664259881 -0.0458384820
## [16] -0.0895708321  0.0210995246  0.0038074746 -0.1250819716  0.0162358036
## [21] -0.0539162296 -0.0161421778  0.1609273693
```

f) Plot the correlogram of the residuals of the AR model, and comment.

```
acf(fit.ar$res[-(1:23)])
```



The acf is not significant for nonzero lags, therefore, the residuals are well modeled by this AR process.

h) Use the best fitting model to forecast electricity production for the years 1991–2000, making sure you have corrected for any bias due to taking logs.

```
new.t = 397:(397+119)
new.c1 = cos(2*pi*new.t/12); new.s1 = sin(2*pi*new.t/12)
new.c2 = cos(2*2*pi*new.t/12); new.s2 = sin(2*2*pi*new.t/12)
new.c3 = cos(3*2*pi*new.t/12); new.s3 = sin(3*2*pi*new.t/12)
new.c5 = cos(5*2*pi*new.t/12); new.s5 = sin(5*2*pi*new.t/12)
new.c6 = cos(6*2*pi*new.t/12); new.s6 = sin(6*2*pi*new.t/12)
new.t2 = new.t^2
new.dat = data.frame(T1=new.t, T2=new.t2, c1=new.c1, s1=new.s1,
                     c2=new.c2, s2=new.s2, s3=new.s3, c5=new.c5,
                     s5=new.s5, c6=new.c6, s6 = new.s6)
ar.pred = predict(ar(resid(elec.harm.best)), n.ahead=120)
log.pred = predict(elec.harm.best, new.dat)
elec.pred = exp(log.pred + ar.pred$pred + 0.5*0.0003933)
elec.ts = ts(elec, st=1958, fr=12)
elec.pred.ts = ts(elec.pred, st=1991, fr=12)
ts.plot(elec.ts, elec.pred.ts, lty=1:2, col=1:2)
```