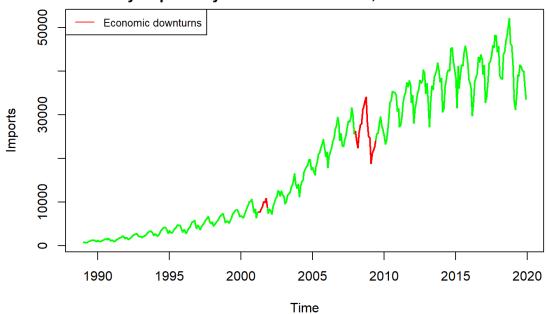
Monthly imports by the U.S. from China

The file ChinaImports2.txt gives monthly imports by the U.S. from China, in millions of dollars, for the period January 1989 through December 2019.

1. Graph both the monthly imports and the logged monthly imports vs. time. Comment in detail on what you can determine from the plots. Mark economic downturns in the plots.

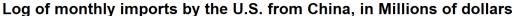
```
> china<-read.csv("ChinaImports2.txt")</pre>
> attach(china)
> head(china)
  Year Month Imports logImports dlogImports
                                                     c348
                                                                s348
           s432
1 1989
           1
               788.9
                        6.670640 -0.03646724 -0.57757270
                                                           0.8163393 -
0.9101060 0.4143756
2 1989
           2
               798.1
                        6.682234 0.01159433 -0.33281954 -0.9429905
0.6565858 -0.7542514
           3
                        6.503240 -0.17899418 0.96202767
3 1989
               667.3
                                                          0.2729519 -
0.2850193 0.9585218
           4
                        6.590988
                                  0.08774791 -0.77846230
                                                           0.6276914 -
4 1989
0.1377903 -0.9904614
           5
                        6.843643
                                  0.25265571 -0.06279052 -0.9980267
5 1989
               937.9
0.5358268 0.8443279
                        6.929223 0.08557985 0.85099448 0.5251746 -
6 1989
           6 1021.7
0.8375280 -0.5463943
> fMonth<-as.factor(Month)</pre>
> china<-data.frame(china,fMonth)</pre>
> downturn<-
c(rep(NA,147),Imports[148:155],rep(NA,73),Imports[229:246],rep(NA,19))
> plot(ts(Imports, start=c(1989, 1), freq=12),
ylab="Imports", main="Monthly imports by the U.S. from China, in
Millions of dollars", col="green", lwd=2)
> lines(ts(downturn,start=c(1989,1),freq=12),col="red",lwd=2)
> legend("topleft",legend="Economic downturns",col="red",lty=1,cex=0.8)
```

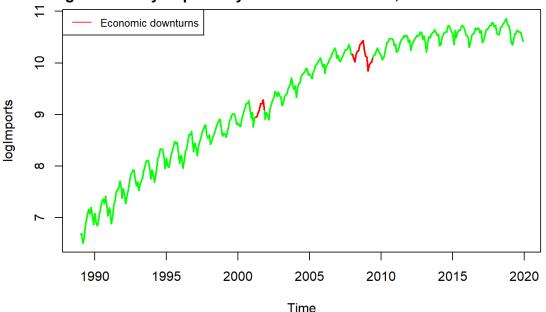
Monthly imports by the U.S. from China, in Millions of dollars



Imports from China have risen substantially since 1989, to the point where they are now ranging from about \$30 billion to \$4 billion per month. According to the Business Cycle Dating Committee of the National Bureau of Economic Research, we mark two periods of economic downturns (2001(4) to 2001(11), 2008(1) to 2009(6)). The second downturn affects imports significantly. In the plot three interruptions in the rise are visible, one during the 2008–2009 recession, a second smaller pause in the climb during 2016, and the third one in late 2019 in response to China–U.S. trade war as well as COVID-19. There is a strong seasonal component, with imports appearing to peak in the fall months, and the seasonal structure is apparently dynamic. With greater volatility evident at higher levels of imports, it is clear that the data should be logged.

```
> logdownturn<-
c(rep(NA,147),logImports[148:155],rep(NA,73),logImports[229:246],rep(NA,19))
> plot(ts(logImports,start=c(1989,1),freq=12),
ylab="logImports",main="Log of monthly imports by the U.S. from China,
in Millions of dollars",col="green",lwd=2)
> lines(ts(logdownturn,start=c(1989,1),freq=12),col="red",lwd=2)
> legend("topleft",legend="Economic downturns",col="red",lty=1,cex=0.8)
```





The log plot reveals that the rate of increase of logImports has been trending upward at a rate which has been slowly declining. The strong seasonality is visible throughout the period from 1989 to the present, and its dynamic structure is visible.

2. Fit a regression model with seasonal and significant calendar components to the differences of the logged monthly imports.

```
> obs146<-c(rep(0,145),1,rep(0,226))
> obs242<-c(rep(0,241),1,rep(0,130))
> obs315<-c(rep(0,314),1,rep(0,57))</pre>
```

```
> model1<-
lm(dlogImports~fMonth+obs146+obs242+obs315+c348+s348+c432+s432)
> summary(model1)
Call:
lm(formula = dlogImports ~ fMonth + obs146 + obs242 + obs315 +
   c348 + s348 + c432 + s432)
Residuals:
    Min
            10
                 Median
                             3Q
-0.229104 -0.033795 -0.001041 0.033935 0.219363
Coefficients:
          Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.060101 0.011553 5.202 3.34e-07 ***
        fMonth2
         fMonth3
fMonth4
         0.032039 0.016334 1.961 0.05061 .
fMonth5
         0.011557 0.016337 0.707 0.47977
fMonth6
fMonth7
         0.007103 0.016336 0.435 0.66399
fMonth8
         -0.005907 0.016340 -0.361 0.71795
         -0.043952 0.016340 -2.690 0.00749 **
fMonth9
         -0.020609 0.016332 -1.262 0.20783
fMonth10
fMonth11
         -0.160201 0.016344 -9.802 < 2e-16 ***
fMonth12
         obs146
        obs242
obs315
         0.280041 0.065768 4.258 2.65e-05 ***
         -0.008678 0.004732 -1.834 0.06750 .
c348
         -0.022521 0.004755 -4.737 3.15e-06 ***
s348
c432
         -0.011169 0.004764 -2.344 0.01961 *
         0.002499 0.004716 0.530 0.59652
s432
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 0.06429 on 353 degrees of freedom
Multiple R-squared: 0.6494, Adjusted R-squared: 0.6315
F-statistic: 36.32 on 18 and 353 DF, p-value: < 2.2e-16
> model2<-lm(dlogImports~fMonth+obs146+obs242+obs315+c348+s348)
> summary(model2)
Call:
lm(formula = dlogImports \sim fMonth + obs146 + obs242 + obs315 +
   c348 + s348)
Residuals:
            10
                 Median
    Min
                             3Q
-0.217776 -0.035779 -0.000286 0.036271 0.225532
Coefficients:
          Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.060177 0.011612 5.182 3.69e-07 ***
fMonth2
        -0.184954   0.016704   -11.072   < 2e-16 ***
fMonth3
         0.031425 0.016417 1.914 0.05641 .
fMonth4
```

```
        fMonth5
        0.036658
        0.016426
        2.232
        0.02626 *

        fMonth6
        0.011074
        0.016423
        0.674
        0.50057

        fMonth7
        0.007251
        0.016418
        0.442
        0.65902

        fMonth8
        -0.005986
        0.016427
        -0.364
        0.71576

        fMonth9
        -0.044246
        0.016421
        -2.694
        0.00739
        **

fMonth10 -0.020283 0.016419 -1.235 0.21752
fMonth11 -0.160794 0.016427 -9.788 < 2e-16 ***
             fMonth12
             obs146
             obs242
obs315
c348
              -0.008710 0.004757 -1.831 0.06796 .
s348
             Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '' 1
Residual standard error: 0.06463 on 355 degrees of freedom
Multiple R-squared: 0.6436, Adjusted R-squared: 0.6275
F-statistic: 40.07 on 16 and 355 DF, p-value: < 2.2e-16
> anova (model2, model1)
Analysis of Variance Table
Model 1: dlogImports ~ fMonth + obs146 + obs242 + obs315 + c348 + s348
Model 2: dlogImports \sim fMonth + obs146 + obs242 + obs315 + c348 + s348
    c432 + s432
  Res.Df RSS Df Sum of Sq F Pr(>F)
  355 1.4831
      353 1.4591 2 0.023997 2.9029 0.05618 .
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '' 1
```

The partial F test reveals that the calendar pair with frequency 0.432 is significant at 0.1 significance level, and we select the first model.

(a) Describe your fitted model.

The model is fit to the log return data, and no trend is present. The fit includes monthly dummies to estimate static seasonal structure, the calendar trigonometric pair with frequency 0.348 and 0.432, and three dummies to account for an outlier in February 2001, February 2009, and March 2015, an unusually low value. *R* squared is 0.6494, and the residual standard error is 0.06429.

(b) Tabulate and plot the estimated seasonal indices *for the imports series*. Provide careful interpretation of the estimates.

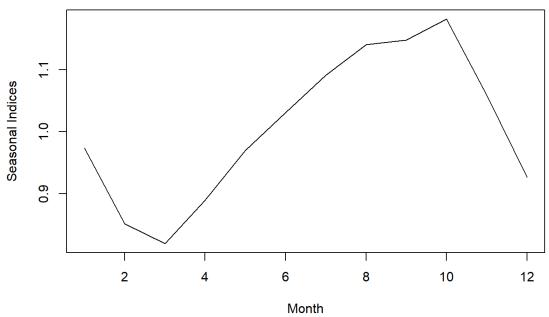
```
> b1<-coef(model1)[1]
> b2<-coef(model1)[2:12]+b1
> b3<-c(b1,b2)
> x<-b3-mean(b3)
> s12<-0
> for(j in 2:12){
```

```
+ xsub<-x[j:12]
+ s12<-s12+sum(xsub)
+ }
> s12<-s12/12
> s<-c(rep(0,times=12))
> s[12]<-s12
> for(j in 1:11) {
+ xsub<-x[1:j]
+ s[j]<-s[12]+sum(xsub)
+ }
> s<-exp(s)
> s

[1] 0.9736367 0.8511472 0.8191299 0.8892265 0.9693605 1.0309782
1.0916394 1.1409302 1.1479318 1.1822537
[11] 1.0589645 0.9260872
```

> plot(ts(s),xlab="Month",ylab="Seasonal Indices",main=" Estimated
Seasonal Indices from Model 1")

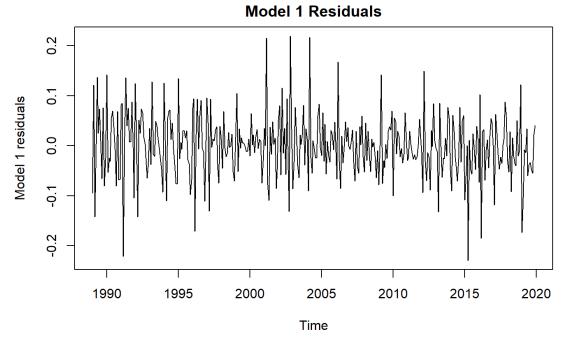
Estimated Seasonal Indices from Model 1



The seasonal estimation shows that imports are annually estimated to be at a low point in March, rise steadily to a peak in October, then fall in November and December, rise slightly in January, and then resume falling in February and March. Imports are estimated to be 18.2 percent higher than trend level in October, and estimates of the levels above trend in July, August, September, and November are 9.2, 14.1, 14.8, and 5.9 percent, respectively. In March the estimate is 18.1 percent below trend, and in February, April, and December the levels are estimated at 14.9, 11.1, and 7.4 percent, respectively, below trend.

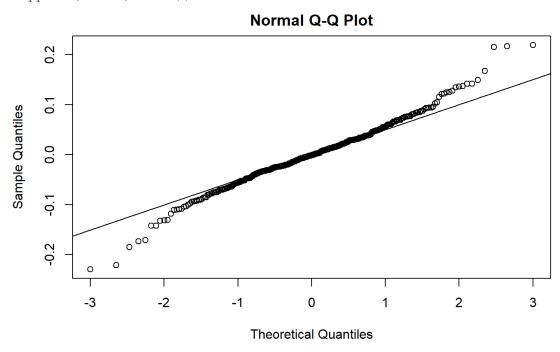
(c) Perform a residual analysis for your model, examining the plot of the residuals vs. time, a residual normal quantile plot, the residual acf and pacf plots, and the residual spectral density (along with Bartlett's test). Discuss all.

> plot(ts(resid(model1), start=c(1989,1), freq=12), ylab="Model 1
residuals", main="Model 1 Residuals")



The residual plot does show some short bursts of elevated volatility. There is no trending—the differencing operation has eliminated trend structure.

- > qqnorm(resid(model1))
- > qqline(resid(model1))



> shapiro.test(resid(model1))

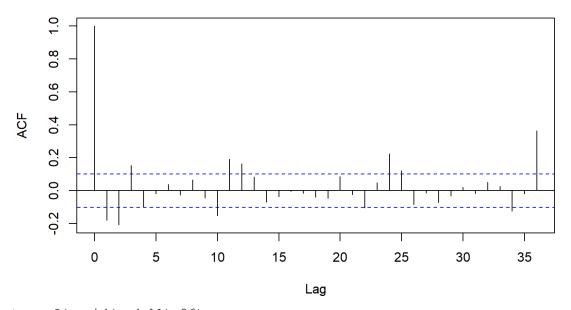
Shapiro-Wilk normality test

```
data: resid(model1)
W = 0.98303, p-value = 0.0002316
```

Residuals from the model are decidedly nonnormal. The model underpredicts many large monthly import values, and it overpredicts some small monthly import amounts. The residual distribution is somewhat positively skewed.

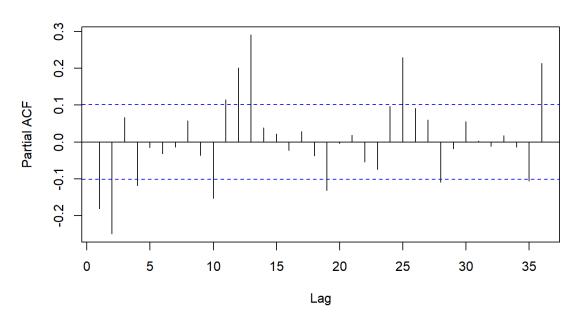
> acf(resid(model1),36)

Series resid(model1)



> pacf(resid(model1),36)

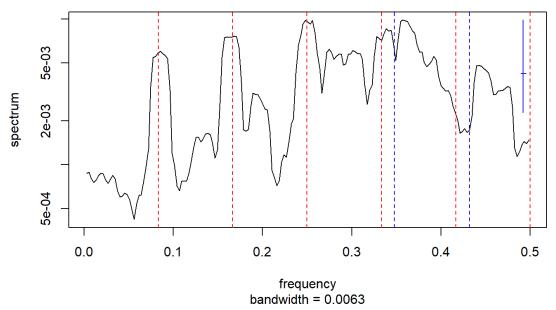
Series resid(model1)



The correlations and partial correlations of the residuals reveal failure of the model to reduce the data to white noise. In the acf plot there are significant residual correlations at lags 1, 2, 3, 10, 11, 12, 24, 34, and 36. The results at lags 12, 24 and 36 indicate there is remaining dynamic seasonality in the residuals. Significant partial correlations are seen at lags 1, 2, 12, 13, 25, and 36.

```
> spectrum(resid(model1), span=8)
> abline(v=c(1/12,2/12,3/12,4/12,5/12,6/12),col="red",lty=2)
> abline(v=c(0.348,0.432),col="blue",lty=2)
```





> bartlettB.test(ts(resid(model1)))

Bartlett B Test for white noise

data: = 2.5879, p-value = 3.046e-06

Call:

In the residual spectral density plot, we see that there are peaks at the seasonal frequencies 1/12, 2,12, 3/12, and 4/12, also indicative of remaining seasonal structure. Moreover, the upward slope in the plot shows that higher frequency activity is more prevalent in the residuals than slow movements. Bartlett's test clearly rejects the null hypothesis of reduction to white noise by the model.

3. (a) Fit a regression model with just the calendar variables to the logged monthly imports series. [The calendar pairs will not be significant in this model. That's okay—do so and proceed as below.]

```
> model3<-lm(logImports~c348+s348+c432+s432);summary(model3)</pre>
```

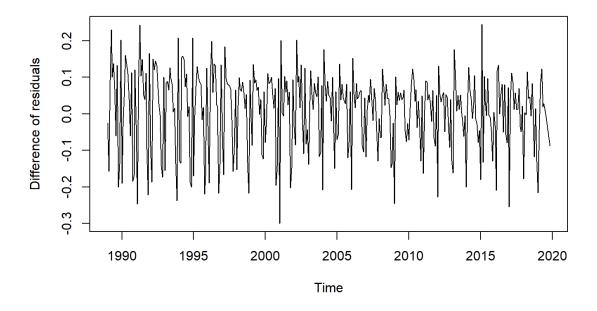
```
lm(formula = logImports \sim c348 + s348 + c432 + s432)
Residuals:
    Min
            10 Median
                             3Q
                                   Max
-2.8392 -0.9070 0.3619 1.0704 1.5057
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
                                          <2e-16 ***
(Intercept) 9.346891
                       0.061860 151.096
c348
             0.004698
                       0.087472
                                  0.054
                                            0.957
                                 -0.192
s348
            -0.016804
                       0.087499
                                            0.848
c432
            0.000234
                       0.087665
                                 0.003
                                            0.998
s432
            -0.004481
                        0.087309 -0.051
                                            0.959
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '' 1
Residual standard error: 1.193 on 367 degrees of freedom
Multiple R-squared: 0.0001158,
                                    Adjusted R-squared: -0.01078
F-statistic: 0.01063 on 4 and 367 DF, p-value: 0.9998
```

(b) Fit a seasonal ARIMA model to the residuals from the model in part (a).

Firstly, difference of ordinary residuals.

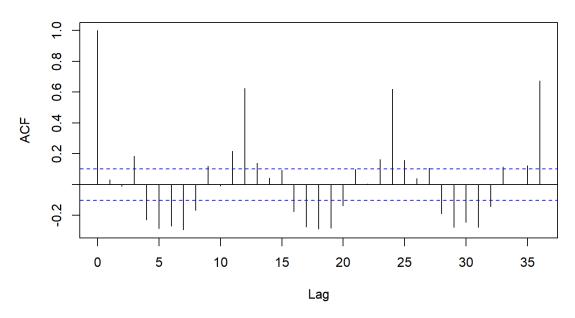
> plot(ts(diff(resid(model3)), start=c(1989,1), freq=12), ylab="Difference
of residuals", main="Difference of Residuals")

Difference of Residuals



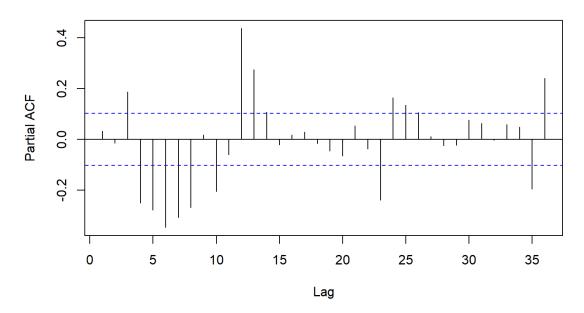
> acf(diff(resid(model3)),36)

Series diff(resid(model3))



> pacf(diff(resid(model3)),36)

Series diff(resid(model3))

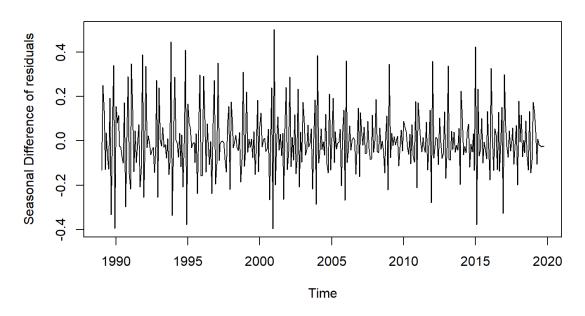


The acf plot shows there is very strong seasonal structure.

Then we explore the seasonal difference.

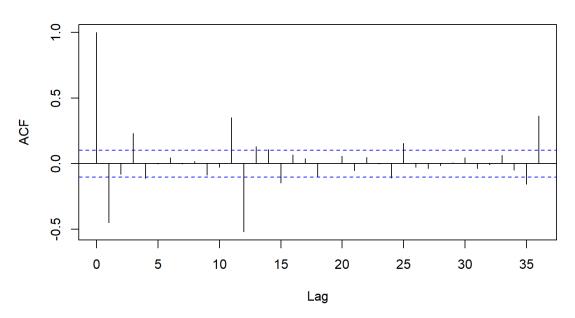
plot(ts(diff(diff(resid(model3))), start=c(1989,1), freq=12), ylab="Season
al Difference of residuals", main="Seasonal Difference of Residuals")

Seasonal Difference of Residuals



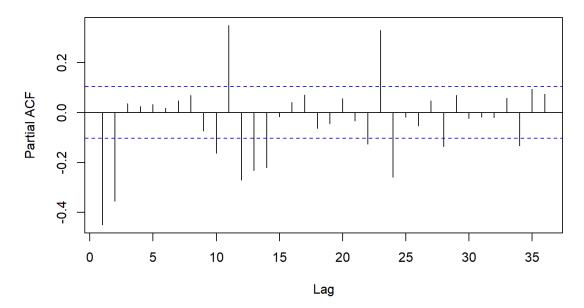
> acf(diff(diff(resid(model3)),12),36)

Series diff(diff(resid(model3)), 12)



> pacf(diff(diff(resid(model3)),12),36)

Series diff(diff(resid(model3)), 12)



We apply both ordinary and seasonal differencing to the log imports data. The resulting acf and pacf plots show significant results at lags 11 and 12, and some activity at lag 36. We use autoregressive structure for both the nonseasonal and seasonal parts of the ARIMA model. After inspection of the residual diagnostics from an initial model, we determine that AR(4) structure is called for in the seasonal component. The model chosen is $ARIMAX(11,1,0)(4,1,0)_{12}$, with the dummy for the outlier in February 2001, February 2009, and March 2015 included.

```
> df<-data.frame(obs146,obs242,obs315)</pre>
> arima1<-
arima(resid(model3),order=c(11,1,0),seasonal=list(order=c(4,1,0),period
=12),xreg=df)
> arima1
Call:
arima(x = resid(model3), order = c(11, 1, 0), seasonal = list(order = c(11, 1, 0))
c(4, 1,
    0), period = 12), xreg = df)
Coefficients:
           ar1
                     ar2
                              ar3
                                       ar4
                                                ar5
                                                          ar6
                                                                   ar7
ar8
      -0.3178
                -0.2080
                          0.1163
                                   -0.0619
                                             0.0429
                                                      -0.0279
                                                                0.0781
0.0062
s.e.
       0.0576
                 0.0562
                          0.0583
                                    0.0582
                                             0.0580
                                                       0.0589
                                                                0.0577
0.0582
          ar9
                  ar10
                           ar11
                                     sar1
                                               sar2
                                                         sar3
                                                                   sar4
      0.0073
               -0.0627
                         0.1135
                                  -0.7066
                                            -0.5415
                                                      -0.2277
                                                                -0.3544
      0.0578
                0.0570
                         0.0557
                                   0.0554
                                             0.0702
                                                       0.0675
                                                                 0.0574
s.e.
       obs146
                 obs242
                          obs315
      -0.1292
               -0.0144
                          0.1789
```

```
s.e. 0.0338 0.0335 0.0360
sigma^2 estimated as 0.002308: log likelihood = 572.52, aic = -
1107.04
> coeftest(arima1)
z test of coefficients:
       Estimate Std. Error z value Pr(>|z|)
ar1
     ar2
     -0.2079700 0.0561708 -3.7025 0.0002135 ***
ar3
     ar4
     -0.0619430 0.0582123 -1.0641 0.2872894
ar5
     0.0429203 0.0580049 0.7399 0.4593350
     -0.0279216 0.0589118 -0.4740 0.6355312
ar6
     0.0781057 0.0576931 1.3538 0.1757956
ar7
ar8
     0.0062083 0.0581867 0.1067 0.9150294
     0.0073284 0.0577602 0.1269 0.8990385
ar9
ar10 -0.0627270 0.0570165 -1.1002 0.2712649
     0.1134894 0.0556539 2.0392 0.0414302 *
sar1
     -0.7065944 0.0554004 -12.7543 < 2.2e-16 ***
     -0.5415300 0.0702423 -7.7095 1.264e-14 ***
sar2
     -0.2276516   0.0675362   -3.3708   0.0007495 ***
sar3
     -0.3544220 0.0573663 -6.1782 6.482e-10 ***
sar4
obs242 -0.0144285 0.0334668 -0.4311 0.6663747
obs315 0.1788911 0.0360381 4.9639 6.908e-07 ***
Signif. codes: 0 \***' 0.001 \**' 0.01 \*' 0.05 \.' 0.1 \' 1
```

(i) Describe your fitted ARIMA model.

The model has significant nonseasonal autoregressive coefficients at lags 1, 2, 3, and 11, and significant seasonal autoregressive coefficients at lags 12, 24, 36, and 48.

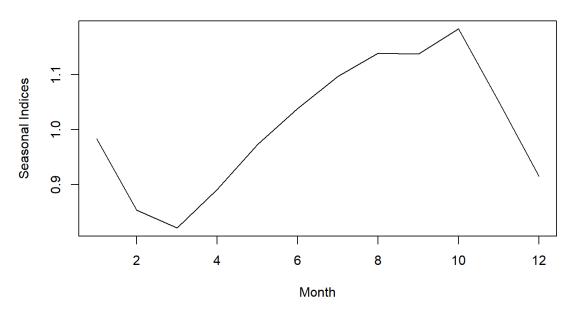
(ii) Tabulate and plot estimated static seasonal indices. Compare the static estimates to those obtained via regression in part 2(b), with a table and a plot. Present the dynamic seasonal index estimates graphically as with the examples in the notes. Discuss the dynamic estimates. What information do they provide about changes over time of seasonal structure?

```
> sel<-c(1:12)
> arimaresid<-resid(arima1)
> arimapred<-resid(model3)[-sel]-arimaresid[-sel]
> darimapred<-diff(arimapred)
> sel2<-1:11
> darimapred<-darimapred[-sel2]
> y<-darimapred
> seasm<-matrix(rep(0,348),ncol=29)
> j<--11
> for(ii in 1:29){
    j<-j+12;j2<-j+11
    y[j:j2]<-exp(y[j:j2]-mean(y[j:j2]))</pre>
```

```
j1<-j+1
    seasm[12,ii]<-1
    for(i in j1:j2) {
      sub<-y[i:j2]</pre>
+
      seasm[12,ii]<-seasm[12,ii]*prod(sub)</pre>
+
    seasm[12,ii] < -(seasm[12,ii])^(1/12)
+
    j3<-j+10
+
    ir<-0
    for(i in j:j3){
      ir < -ir + 1
+
      sub<-y[j:i]</pre>
      seasm[ir,ii]<-seasm[12,ii]*prod(sub)</pre>
+ }
> seasstatic<-apply(seasm,1,prod)^(1/29)</pre>
> seasstatic
 [1] 0.9831267 0.8534084 0.8206552 0.8905525 0.9729753 1.0382493
 [7] 1.0964423 1.1391943 1.1373760 1.1830761 1.0498881 0.9148851
> plot(ts(seasstatic),xlab="Month",ylab="Seasonal Indices",main="
```

Estimated Seasonal Indices from ARIMA Model 1")

Estimated Seasonal Indices from ARIMA Model 1

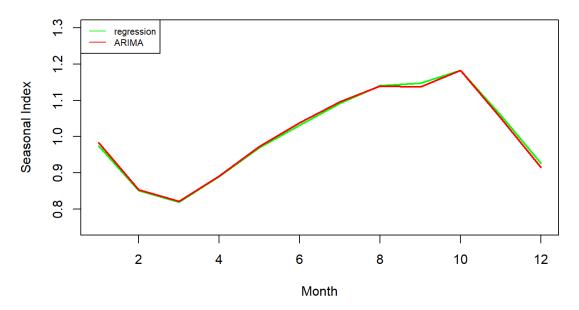


```
> cbind(s,seasstatic)
              s seasstatic
 [1,] 0.9736367
                 0.9831267
 [2,] 0.8511472
                 0.8534084
 [3,] 0.8191299
                 0.8206552
 [4,] 0.8892265
                 0.8905525
 [5,] 0.9693605
                 0.9729753
 [6,] 1.0309782
                1.0382493
 [7,] 1.0916394
                1.0964423
 [8,] 1.1409302 1.1391943
```

```
[9,] 1.1479318  1.1373760
[10,] 1.1822537  1.1830761
[11,] 1.0589645  1.0498881
[12,] 0.9260872  0.9148851

> plot(ts(s),xlab="Month",ylab="Seasonal Index",main="Comparison of Seasonal Index Estimates",ylim=c(0.75,1.3),col="green",lwd=2)
> lines(ts(seasstatic),col="red",lwd=2)
> legend("topleft",legend=c("regression","ARIMA"),col=c("green","red"),lt y=1,cex=0.7)
```

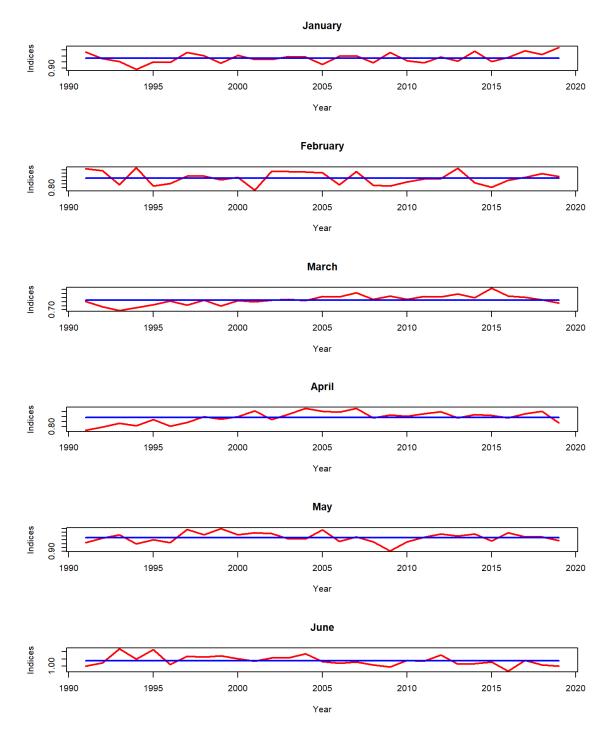
Comparison of Seasonal Index Estimates

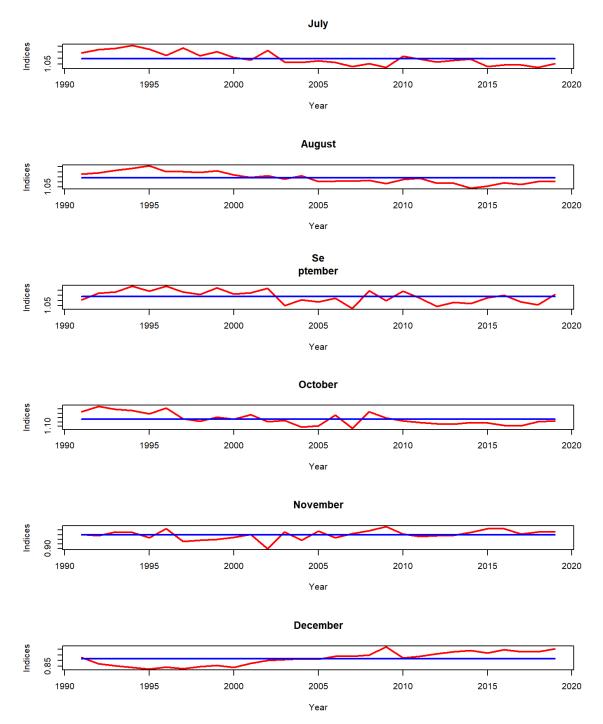


The static seasonal estimates obtained from the regression and ARIMAX models are very similar, as the table and the plot above show. There are small discrepancies between the two in January and September, 1 percent respectively. Otherwise, there is closer agreement.

Monthly plots for dynamic seasonal estimation follow.

```
> year<-seq(1991,2019)
> seasstatic<-matrix(rep(seasstatic,29),ncol=29)
> name<-
c("January", "February", "March", "April", "May", "June", "July", "August", "Se
+ ptember", "October", "November", "December")
> par(mfrow=c(3,1))
> for(i in 1:3){
+
plot(year, seasm[i,], xlab="Year", ylab="Indices", main=name[i], type="l", lw d=2, col="red")
+ lines(year, seasstatic[i,], lty=1, lwd=2, col="blue")
+ }
```





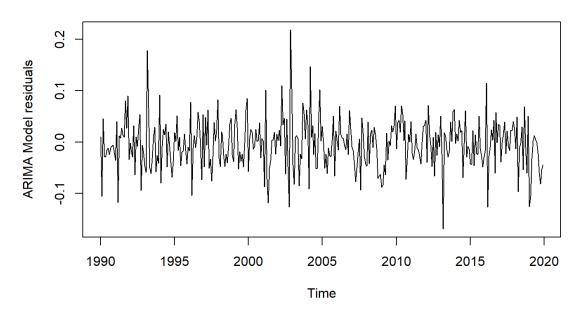
The dynamic estimation covers the years from 1989 through 2019. May shows the most prominent estimated change in the seasonal index, from about 15 percent above trend level in the early years to about 20 percent below trend level in the late years. December has an estimated increase over time, from about 10 percent below trend level trend in 1995 to about 5 percent above trend level in 2019. The index estimates for June, July, August, and September decline over time. For August, the movement is from about 15 percent above trend level to about 10 percent below trend level. For

February, the estimated indices fluctuated around the trend level. Estimated indices for January, March, April, October, and November do not change much over time.

(iii) Perform a residual analysis for your ARIMA model, examining the plot of the residuals vs. time, the residual normal quantile plot, the residual acf and pacf, and the residual spectral density (along with Bartlett's test). Has the ARIMA model you've fit produced reduction to white noise? Discuss carefully.

```
> plot(ts(resid(arima1)[-sel], start=c(1990,1), freq=12), ylab="ARIMA
Model residuals", main="ARIMA Model Residuals")
```

ARIMA Model Residuals

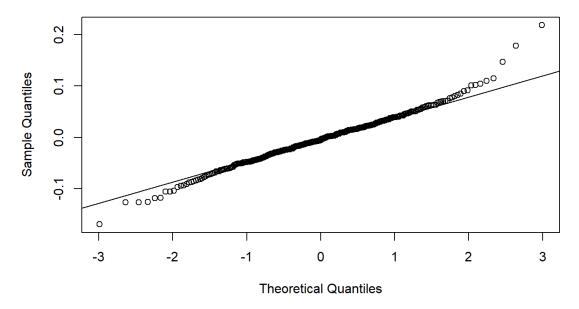


The plot shows the presence of outlier behavior and some decline in volatility over time. There is no noticeable trend.

```
> qqnorm(resid(arima1)[-sel])
```

> qqline(resid(arimal)[-sel])

Normal Q-Q Plot



> shapiro.test(resid(arima1)[-sel])

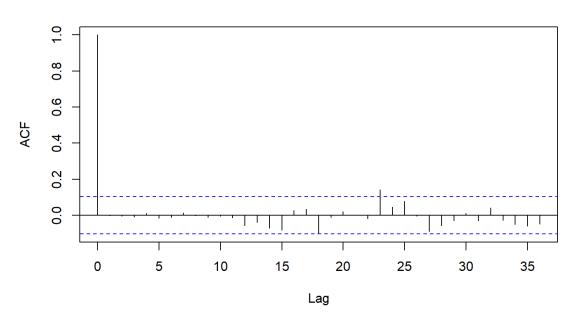
Shapiro-Wilk normality test

data: resid(arima1)[-sel]
W = 0.98328, p-value = 0.0003476

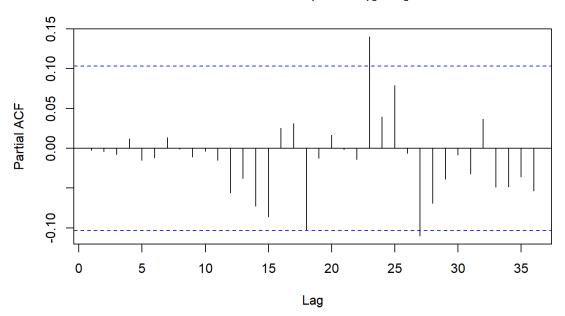
The residuals are nonnormal, with many observations underestimated by the ARIMAX model, and also many overestimated.

> acf(resid(arima1)[-sel],36)

Series resid(arima1)[-sel]

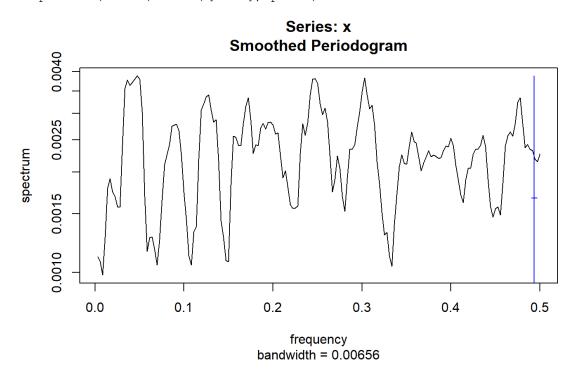


Series resid(arima1)[-sel]



The lag 23 result is significant in both the acf and pacf plots.

> spectrum(resid(arima1)[-sel],span=8)



> bartlettB.test(ts(resid(arimal))[-sel])

Bartlett B Test for white noise

```
data:
= 0.28112, p-value = 1
```

The residual spectral plot and the result of Bartlett's test indicate reduction to white noise by the ARIMAX model.

4. Give some brief concluding remarks. What has the analysis revealed about imports from China to the U.S. during the years 1989 to 2019?

Firstly, we found that the data should be logged to fit the regression model and found several economic downturns during the period. Next, a model was fit to the differences of the logged monthly imports, including monthly dummies, two calendar trigonometric pairs, and three dummies of outliers. According to residual analysis, this model failed in the reduction to white noise.

Then we began to fit an ARIMA model. A regression with calendar trigonometric pairs was initially fit to the time series. Next, a seasonal ARIMA model was fit to the residuals from this regression. Then trend structure was removed via regression from the predicted values from this seasonal ARIMA fit. Centered monthly averages of the modified predicted values were calculated, giving the seasonal index estimates. Finally, the residual analysis indicated successful reduction to white noise by the model.

During 1989 to 2019, imports from China to the U.S. have risen substantially. There is a strong dynamic seasonal component, with imports appearing to peak in the fall months. Three interruptions in the rise are visible. The first one is during the 2008-2009 recession, and it may because consumers will buy less items during hard economic times. The second one is during 2016, the adjustment of industrial structure and tensions in international relations may be the causes. The third one is in late 2019 in response to China–U.S. trade war as well as COVID-19.