HW3 Combined

# Problem 1

1. This problem involves examination of monthly Lydia Pinkham data for the period 1938 through 1952. It addresses model construction and estimation of the 90 per cent duration interval. The data are in Lydiamonthly.txt.

m.sales <- read.csv("Lydiamonthly.txt")  
  
Time <- seq(1, nrow(m.sales))  
fMonth<-as.factor(m.sales$month)  
logmsales <- log(m.sales$msales)  
  
m.sales<-data.frame(m.sales, Time, fMonth, logmsales)  
  
attach(m.sales)

## The following objects are masked \_by\_ .GlobalEnv:  
##   
## fMonth, logmsales, Time

1. 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.

## Problem 1.a.1

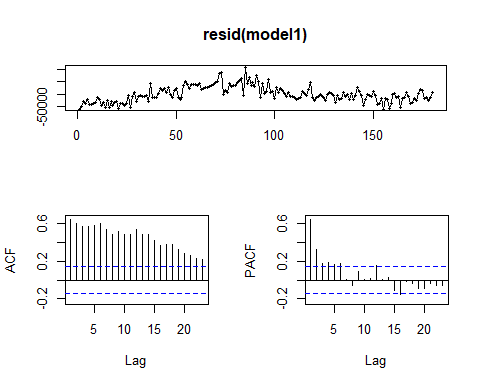
Construct a model relating sales and advertising with sales as the response.

**Discussion:** An initial inspection of model1 suggests that AR3 model is a good place to start.

model1<-lm(msales~madv)  
summary(model1)

##   
## Call:  
## lm(formula = msales ~ madv)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -58909 -20194 -6041 20867 108816   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.212e+05 5.261e+03 23.041 <2e-16 \*\*\*  
## madv 5.901e-01 5.939e-02 9.935 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 31370 on 178 degrees of freedom  
## Multiple R-squared: 0.3567, Adjusted R-squared: 0.3531   
## F-statistic: 98.71 on 1 and 178 DF, p-value: < 2.2e-16

forecast::tsdisplay(resid(model1))



## Problem 1.a.2

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.

**Discussion:** First we fit an AR(3) model with the inclusion of following X variables: advertising, lagged advertising, all dummy fmonth variables, all dummy outlier variables, and all calendar pairs.

The initial residual plot is not flat, and while the autocorrelation plot shows that there remains significant autocorrelation in the model at lag 28, the lags are largely within the blue lines. The spectrum analysis of the residuals show that the model has been reduced to white noise (as indicated by the blue line measurement). The AIC of this model is 3965.163.

We now discuss the significance of the variables.

**Lag Variables:** The initial summary indicates that AR3 (msalesl3) has marginal significance. We will try AR(2) model next.

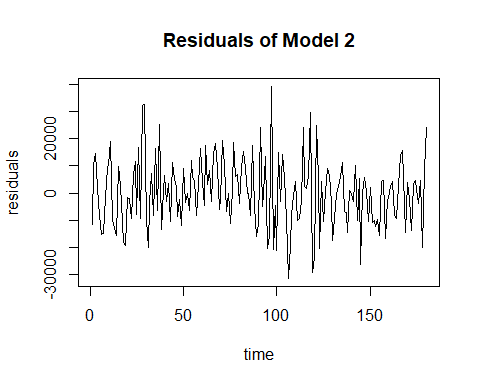
**X Variables:** The preliminary summary below indicates that lagged advertising variables 2, 3, and 4 have marginal significance. We will try reducing the number to 2 lagged advertising variables (including madvl1 and madvl2).

The fMonth dummy variables seem to be significant, will have to conduct a partial f-test to confirm this. Some of the calendar pairs do not seem to be significant and we have to individually inspect their significance through partial f-test. All the outliers seem to be significant, so we will keep them.

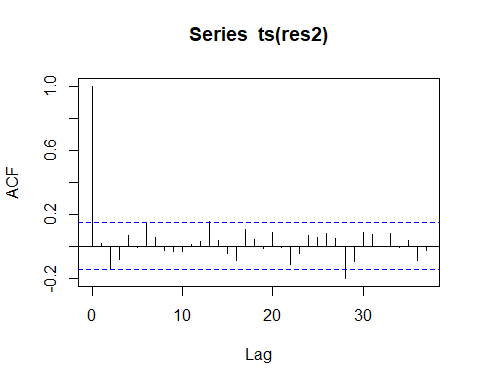
model2 <- lm(msales~madv+madvl1+madvl2+madvl3+madvl4+msalesl1+msalesl2+msalesl3+fMonth+feb44+dec44+jan45+sep45+c220+s220+c348+s348+c432+s432)  
  
summary(model2)

##   
## Call:  
## lm(formula = msales ~ madv + madvl1 + madvl2 + madvl3 + madvl4 +   
## msalesl1 + msalesl2 + msalesl3 + fMonth + feb44 + dec44 +   
## jan45 + sep45 + c220 + s220 + c348 + s348 + c432 + s432)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -31327 -9248 0 6825 39232   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.286e+04 8.019e+03 2.851 0.004975 \*\*   
## madv 2.224e-01 4.394e-02 5.062 1.20e-06 \*\*\*  
## madvl1 9.603e-02 4.878e-02 1.969 0.050831 .   
## madvl2 -1.994e-02 4.888e-02 -0.408 0.683935   
## madvl3 -2.701e-02 4.780e-02 -0.565 0.572910   
## madvl4 1.863e-02 4.540e-02 0.410 0.682084   
## msalesl1 4.580e-01 7.184e-02 6.375 2.13e-09 \*\*\*  
## msalesl2 2.812e-01 7.414e-02 3.792 0.000216 \*\*\*  
## msalesl3 1.140e-01 6.861e-02 1.662 0.098593 .   
## fMonth2 -1.786e+04 8.200e+03 -2.178 0.030972 \*   
## fMonth3 -8.487e+03 8.715e+03 -0.974 0.331688   
## fMonth4 -2.898e+04 7.563e+03 -3.832 0.000186 \*\*\*  
## fMonth5 -2.879e+04 7.258e+03 -3.966 0.000113 \*\*\*  
## fMonth6 -2.891e+04 6.076e+03 -4.757 4.58e-06 \*\*\*  
## fMonth7 -1.100e+04 6.091e+03 -1.806 0.072957 .   
## fMonth8 -1.228e+04 6.735e+03 -1.824 0.070171 .   
## fMonth9 -7.649e+03 7.527e+03 -1.016 0.311183   
## fMonth10 -3.829e+03 7.679e+03 -0.499 0.618732   
## fMonth11 -4.764e+04 8.216e+03 -5.798 3.83e-08 \*\*\*  
## fMonth12 -4.987e+04 6.990e+03 -7.135 3.86e-11 \*\*\*  
## feb44 -4.314e+04 1.524e+04 -2.830 0.005289 \*\*   
## dec44 -5.426e+04 1.464e+04 -3.707 0.000294 \*\*\*  
## jan45 7.683e+04 1.510e+04 5.089 1.06e-06 \*\*\*  
## sep45 -4.385e+04 1.470e+04 -2.983 0.003328 \*\*   
## c220 3.371e+03 1.468e+03 2.296 0.023065 \*   
## s220 9.777e+02 1.509e+03 0.648 0.518115   
## c348 6.544e+02 1.566e+03 0.418 0.676638   
## s348 -3.798e+03 1.549e+03 -2.452 0.015354 \*   
## c432 -9.077e+02 1.519e+03 -0.598 0.551055   
## s432 1.703e+02 1.476e+03 0.115 0.908284   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 13550 on 150 degrees of freedom  
## Multiple R-squared: 0.8988, Adjusted R-squared: 0.8793   
## F-statistic: 45.95 on 29 and 150 DF, p-value: < 2.2e-16

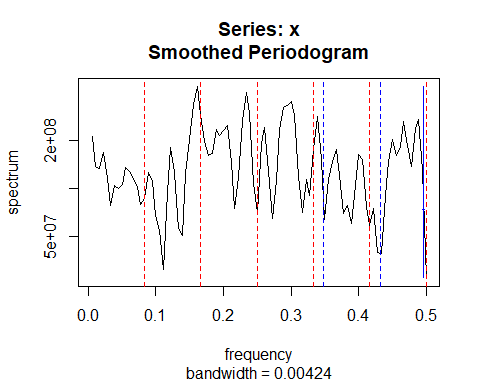
res2 <- resid(model2)  
resid2 <- ts(res2) #,start=c(1992,1),freq=12)  
plot(resid2, xlab="time",ylab="residuals",main="Residuals of Model 2")



acf(ts(res2), 37)



spectrum(res2, span=3)  
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)



AIC(model2)

## [1] 3965.163

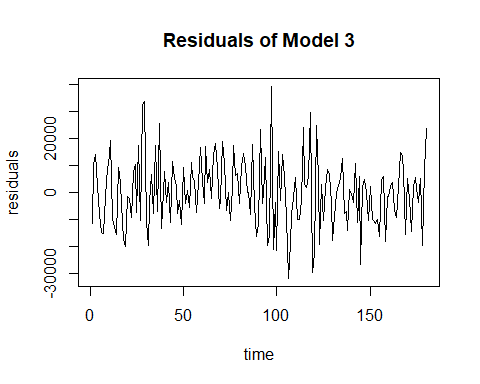
**Discussion:** This model 3 shows a little bit of improvement in terms of AIC: the AIC IS now 3961.624, down from 3965.16 previously. This is in part due to the reduction of lagged advertising variables (we include in model3 only the first two lagged advertising variables compared to all four lagged advertising variables in the previous model).

The summary shows that madvl2 has marginal significance in this model. We conducted anova test to see the significance of madvl1 and madvl2 together, and they show a marginal significance of .1195. This suggests that madvl2 may not be necessary. We will exclude it in our next fit.

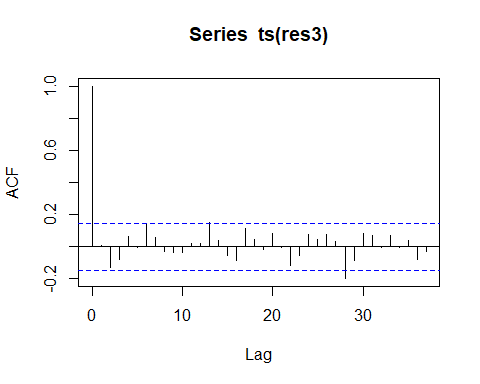
model3 <- lm(msales~madv+madvl1+madvl2+msalesl1+msalesl2+msalesl3+fMonth+feb44+dec44+jan45+sep45+c220+s220+c348+s348+c432+s432)  
  
summary(model3)

##   
## Call:  
## lm(formula = msales ~ madv + madvl1 + madvl2 + msalesl1 + msalesl2 +   
## msalesl3 + fMonth + feb44 + dec44 + jan45 + sep45 + c220 +   
## s220 + c348 + s348 + c432 + s432)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -31794 -9214 0 7656 39252   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.236e+04 7.641e+03 2.927 0.003954 \*\*   
## madv 2.216e-01 4.317e-02 5.133 8.61e-07 \*\*\*  
## madvl1 9.939e-02 4.799e-02 2.071 0.040045 \*   
## madvl2 -3.087e-02 4.529e-02 -0.682 0.496544   
## msalesl1 4.601e-01 7.133e-02 6.450 1.41e-09 \*\*\*  
## msalesl2 2.738e-01 7.253e-02 3.775 0.000229 \*\*\*  
## msalesl3 1.190e-01 6.514e-02 1.827 0.069625 .   
## fMonth2 -1.714e+04 8.027e+03 -2.135 0.034351 \*   
## fMonth3 -6.422e+03 7.918e+03 -0.811 0.418581   
## fMonth4 -2.947e+04 7.069e+03 -4.169 5.12e-05 \*\*\*  
## fMonth5 -2.799e+04 7.039e+03 -3.977 0.000108 \*\*\*  
## fMonth6 -2.893e+04 5.990e+03 -4.830 3.30e-06 \*\*\*  
## fMonth7 -1.055e+04 6.012e+03 -1.755 0.081272 .   
## fMonth8 -1.129e+04 6.499e+03 -1.737 0.084478 .   
## fMonth9 -6.802e+03 7.248e+03 -0.938 0.349534   
## fMonth10 -3.104e+03 7.305e+03 -0.425 0.671451   
## fMonth11 -4.742e+04 7.956e+03 -5.960 1.69e-08 \*\*\*  
## fMonth12 -4.952e+04 6.795e+03 -7.288 1.60e-11 \*\*\*  
## feb44 -4.378e+04 1.512e+04 -2.896 0.004333 \*\*   
## dec44 -5.337e+04 1.447e+04 -3.688 0.000314 \*\*\*  
## jan45 7.568e+04 1.488e+04 5.086 1.06e-06 \*\*\*  
## sep45 -4.380e+04 1.462e+04 -2.996 0.003192 \*\*   
## c220 3.409e+03 1.459e+03 2.337 0.020761 \*   
## s220 9.791e+02 1.501e+03 0.652 0.515198   
## c348 5.389e+02 1.540e+03 0.350 0.726943   
## s348 -3.594e+03 1.501e+03 -2.394 0.017868 \*   
## c432 -1.095e+03 1.479e+03 -0.741 0.460137   
## s432 1.371e+02 1.466e+03 0.094 0.925567   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 13480 on 152 degrees of freedom  
## Multiple R-squared: 0.8986, Adjusted R-squared: 0.8805   
## F-statistic: 49.86 on 27 and 152 DF, p-value: < 2.2e-16

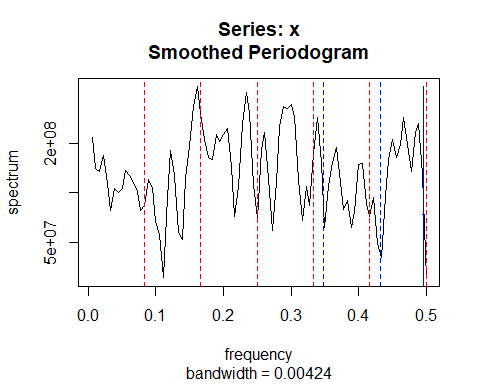
res3 <- resid(model3)  
resid3 <- ts(res3) #,start=c(1992,1),freq=12)  
plot(resid3, xlab="time",ylab="residuals",main="Residuals of Model 3")



acf(ts(res3), 37)



spectrum(res3, span=3)  
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)



AIC(model3)

## [1] 3961.624

model3.1 <- lm(msales~madv+msalesl1+msalesl2+msalesl3+fMonth+feb44+dec44+jan45+sep45+c220+s220+c348+s348+c432+s432)  
  
anova(model3.1, model3)

## Analysis of Variance Table  
##   
## Model 1: msales ~ madv + msalesl1 + msalesl2 + msalesl3 + fMonth + feb44 +   
## dec44 + jan45 + sep45 + c220 + s220 + c348 + s348 + c432 +   
## s432  
## Model 2: msales ~ madv + madvl1 + madvl2 + msalesl1 + msalesl2 + msalesl3 +   
## fMonth + feb44 + dec44 + jan45 + sep45 + c220 + s220 + c348 +   
## s348 + c432 + s432  
## Res.Df RSS Df Sum of Sq F Pr(>F)  
## 1 154 2.8405e+10   
## 2 152 2.7622e+10 2 783058353 2.1545 0.1195

**Discussion:** In this model we exclude the variable madvl2 and noticed an improvement in AIC from 3161.62 to 3960.17. The advertising and lagged advertising variables are now shown to be significant. Next we conduct partial f-test to test the significance of fMonth and calendar pair variables.

model4 <- lm(msales~madv+madvl1+msalesl1+msalesl2+msalesl3+fMonth+feb44+dec44+jan45+sep45+c220+s220+c348+s348+c432+s432)  
  
summary(model4)

##   
## Call:  
## lm(formula = msales ~ madv + madvl1 + msalesl1 + msalesl2 + msalesl3 +   
## fMonth + feb44 + dec44 + jan45 + sep45 + c220 + s220 + c348 +   
## s348 + c432 + s432)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -32067 -8886 0 7254 38490   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.098e+04 7.353e+03 2.853 0.004933 \*\*   
## madv 2.280e-01 4.205e-02 5.424 2.24e-07 \*\*\*  
## madvl1 9.077e-02 4.621e-02 1.964 0.051321 .   
## msalesl1 4.502e-01 6.971e-02 6.458 1.34e-09 \*\*\*  
## msalesl2 2.673e-01 7.176e-02 3.724 0.000275 \*\*\*  
## msalesl3 1.249e-01 6.446e-02 1.937 0.054561 .   
## fMonth2 -1.489e+04 7.307e+03 -2.038 0.043246 \*   
## fMonth3 -5.424e+03 7.768e+03 -0.698 0.486029   
## fMonth4 -2.849e+04 6.908e+03 -4.124 6.09e-05 \*\*\*  
## fMonth5 -2.765e+04 7.009e+03 -3.945 0.000121 \*\*\*  
## fMonth6 -2.877e+04 5.975e+03 -4.815 3.50e-06 \*\*\*  
## fMonth7 -9.567e+03 5.826e+03 -1.642 0.102597   
## fMonth8 -9.900e+03 6.162e+03 -1.607 0.110182   
## fMonth9 -5.163e+03 6.826e+03 -0.756 0.450599   
## fMonth10 -2.057e+03 7.129e+03 -0.289 0.773323   
## fMonth11 -4.656e+04 7.841e+03 -5.938 1.87e-08 \*\*\*  
## fMonth12 -4.951e+04 6.783e+03 -7.298 1.48e-11 \*\*\*  
## feb44 -4.368e+04 1.509e+04 -2.895 0.004348 \*\*   
## dec44 -5.407e+04 1.441e+04 -3.752 0.000248 \*\*\*  
## jan45 7.573e+04 1.485e+04 5.099 9.98e-07 \*\*\*  
## sep45 -4.384e+04 1.459e+04 -3.005 0.003109 \*\*   
## c220 3.431e+03 1.456e+03 2.356 0.019736 \*   
## s220 9.970e+02 1.498e+03 0.665 0.506739   
## c348 7.145e+02 1.516e+03 0.471 0.638107   
## s348 -3.518e+03 1.494e+03 -2.354 0.019838 \*   
## c432 -1.002e+03 1.470e+03 -0.681 0.496758   
## s432 2.409e+02 1.455e+03 0.166 0.868712   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 13460 on 153 degrees of freedom  
## Multiple R-squared: 0.8982, Adjusted R-squared: 0.881   
## F-statistic: 51.95 on 26 and 153 DF, p-value: < 2.2e-16

AIC(model4)

## [1] 3960.173

**Partial F-test testing the significance of fMonth variables**

**Discussion:** Partial f-tests below indicate that fMonth variables are significant.

model5.1 <- lm(msales~madv+madvl1+msalesl1+msalesl2+msalesl3+fMonth+feb44+dec44+jan45+sep45+c220+s220+c348+s348+c432+s432)  
  
model5.2 <- lm(msales~madv+madvl1+msalesl1+msalesl2+msalesl3+feb44+dec44+jan45+sep45+c348+c220+s220+s348+c432+s432)  
  
anova(model5.2, model5.1)

## Analysis of Variance Table  
##   
## Model 1: msales ~ madv + madvl1 + msalesl1 + msalesl2 + msalesl3 + feb44 +   
## dec44 + jan45 + sep45 + c348 + c220 + s220 + s348 + c432 +   
## s432  
## Model 2: msales ~ madv + madvl1 + msalesl1 + msalesl2 + msalesl3 + fMonth +   
## feb44 + dec44 + jan45 + sep45 + c220 + s220 + c348 + s348 +   
## c432 + s432  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 164 5.1831e+10   
## 2 153 2.7706e+10 11 2.4125e+10 12.111 3.567e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

**Partial F-test testing the significance of calendar pair variables**

**Discussion:** Partial f-tests below indicate that calendar pairs 220, 348, and 432 have significances of .05286, 0.05859, and 0.7798 respectively. Based on this we can remove the 432 pair. After removing the pair, the AIC improves to 3956.76 from 3960.17

model5.3 <- lm(msales~madv+madvl1+msalesl1+msalesl2+msalesl3+fMonth+feb44+dec44+jan45+sep45+c348+s348+c432+s432)  
  
anova(model5.3, model5.1)

## Analysis of Variance Table  
##   
## Model 1: msales ~ madv + madvl1 + msalesl1 + msalesl2 + msalesl3 + fMonth +   
## feb44 + dec44 + jan45 + sep45 + c348 + s348 + c432 + s432  
## Model 2: msales ~ madv + madvl1 + msalesl1 + msalesl2 + msalesl3 + fMonth +   
## feb44 + dec44 + jan45 + sep45 + c220 + s220 + c348 + s348 +   
## c432 + s432  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 155 2.8792e+10   
## 2 153 2.7706e+10 2 1085535089 2.9973 0.05286 .  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

model5.4 <- lm(msales~madv+madvl1+msalesl1+msalesl2+msalesl3+fMonth+feb44+dec44+jan45+sep45+c220+s220+c432+s432)  
  
anova(model5.4, model5.1)

## Analysis of Variance Table  
##   
## Model 1: msales ~ madv + madvl1 + msalesl1 + msalesl2 + msalesl3 + fMonth +   
## feb44 + dec44 + jan45 + sep45 + c220 + s220 + c432 + s432  
## Model 2: msales ~ madv + madvl1 + msalesl1 + msalesl2 + msalesl3 + fMonth +   
## feb44 + dec44 + jan45 + sep45 + c220 + s220 + c348 + s348 +   
## c432 + s432  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 155 2.8753e+10   
## 2 153 2.7706e+10 2 1046886000 2.8905 0.05859 .  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

model5.5 <- lm(msales~madv+madvl1+msalesl1+msalesl2+msalesl3+fMonth+feb44+dec44+jan45+sep45+c220+s220+c348+s348)  
  
anova(model5.5, model5.1)

## Analysis of Variance Table  
##   
## Model 1: msales ~ madv + madvl1 + msalesl1 + msalesl2 + msalesl3 + fMonth +   
## feb44 + dec44 + jan45 + sep45 + c220 + s220 + c348 + s348  
## Model 2: msales ~ madv + madvl1 + msalesl1 + msalesl2 + msalesl3 + fMonth +   
## feb44 + dec44 + jan45 + sep45 + c220 + s220 + c348 + s348 +   
## c432 + s432  
## Res.Df RSS Df Sum of Sq F Pr(>F)  
## 1 155 2.7797e+10   
## 2 153 2.7706e+10 2 90227241 0.2491 0.7798

model6<- lm(msales~madv+madvl1+msalesl1+msalesl2+msalesl3+fMonth+feb44+dec44+jan45+sep45+c220+s220+c348+s348)  
  
AIC(model6)

## [1] 3956.758

**final model**

The final model includes as variables: - contemporary advertising - lag 1 advertising variables (madvl1) - all the lag sales variables (msalesl1-msalesl3) - the fmonth dummy variables to account for seasonality - all four outlier variables(feb44, dec44, jan45, sep45) - calendar pairs 220 and 348

We conduct a final partial f-test to determine the significance of lagged advertising variables (madvl2, madvl3, madvl4) and the test concludes that they do not Granger-cause msales with p-value of 0.8088.

A look at the seasonal indices plot indicate that, annually and on average, sales drop in April to June and rises from August to October before dropping in November and December. Thereafter sales would pick up in January to March before dropping in April.

model6<- lm(msales~madv+madvl1+msalesl1+msalesl2+msalesl3+fMonth+feb44+dec44+jan45+sep45+c220+s220+c348+s348)  
  
model6.1<- lm(msales~madv+madvl1+madvl2+madvl3+madvl4+msalesl1+msalesl2+msalesl3+fMonth+feb44+dec44+jan45+sep45+c220+s220+c348+s348)  
  
anova(model6.1, model6)

## Analysis of Variance Table  
##   
## Model 1: msales ~ madv + madvl1 + madvl2 + madvl3 + madvl4 + msalesl1 +   
## msalesl2 + msalesl3 + fMonth + feb44 + dec44 + jan45 + sep45 +   
## c220 + s220 + c348 + s348  
## Model 2: msales ~ madv + madvl1 + msalesl1 + msalesl2 + msalesl3 + fMonth +   
## feb44 + dec44 + jan45 + sep45 + c220 + s220 + c348 + s348  
## Res.Df RSS Df Sum of Sq F Pr(>F)  
## 1 152 2.7621e+10   
## 2 155 2.7797e+10 -3 -176060182 0.323 0.8088

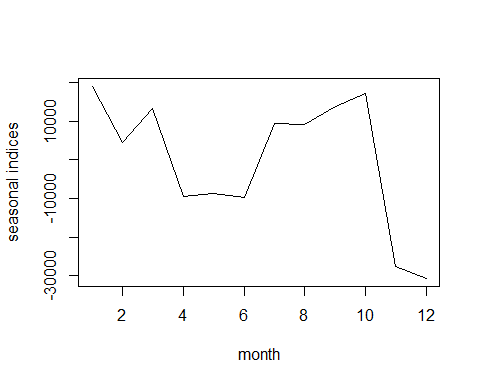
b1<-coef(model6)[1]  
b2<-coef(model6)[7:17]+b1  
b3<-c(b1,b2)  
seas<-b3-mean(b3)  
  
seas.ts<-ts(seas)  
seas

## (Intercept) fMonth2 fMonth3 fMonth4 fMonth5 fMonth6   
## 18954.498 4501.648 13321.164 -9409.932 -8671.434 -9826.754   
## fMonth7 fMonth8 fMonth9 fMonth10 fMonth11 fMonth12   
## 9447.281 9113.859 13779.909 16997.011 -27539.025 -30668.224

seas.ts

## Time Series:  
## Start = 1   
## End = 12   
## Frequency = 1   
## (Intercept) fMonth2 fMonth3 fMonth4 fMonth5 fMonth6   
## 18954.498 4501.648 13321.164 -9409.932 -8671.434 -9826.754   
## fMonth7 fMonth8 fMonth9 fMonth10 fMonth11 fMonth12   
## 9447.281 9113.859 13779.909 16997.011 -27539.025 -30668.224

plot(seas.ts,ylab="seasonal indices",xlab="month")



## Problem 1.b.1

Analyze the residuals from your model. Give a normal quantile plot, plot the residuals vs. time, present the residual autocorrelations and partial autocorrelations, and produce the residual spectral density plot. Discuss the results carefully.

**Discussion:** The qqplot indicates that normality has been roughly achieved, as most observations fall on the qqline. There are a few deviations at the tails, however they are a few compared to the the vast majority of observations which fall on the line.

The residuals plot show that it is largely flat and variance is relatively constant around the mean of 0. We cannot reject that homoscedacticity is not achieved based on this plot.

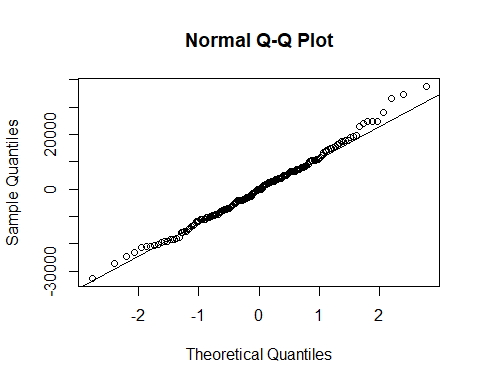
The ACF and PACF plots show no significant autocorrelation except for lag 28 in the ACF plot. There are a few lags in the two plots which seem to reach the blue line but they do not excessively cross the line. I decided to leave them as is as they do not constitute a major issue for the model.

The residual spectral density plot indicates that white noise reduction has been achieved as the distance between highest and lowest peaks of the plot is less than twice the length of the blue measurement line above the notch.

res6 <- resid(model6)

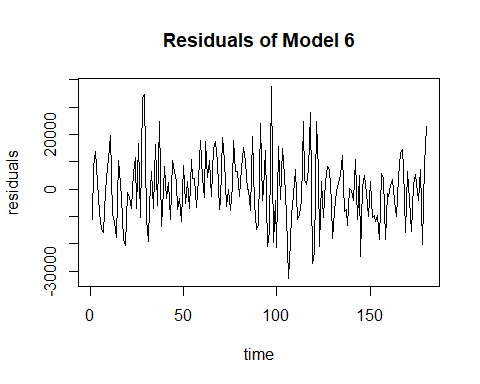
**normal quantile plot**

qqnorm(res6)  
qqline(res6)



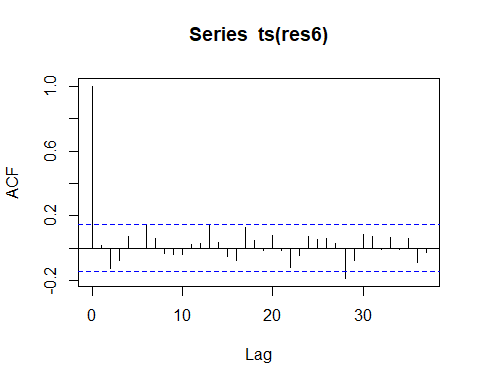
**plot the residuals vs. time**

plot(ts(res6), xlab="time",ylab="residuals",main="Residuals of Model 6")

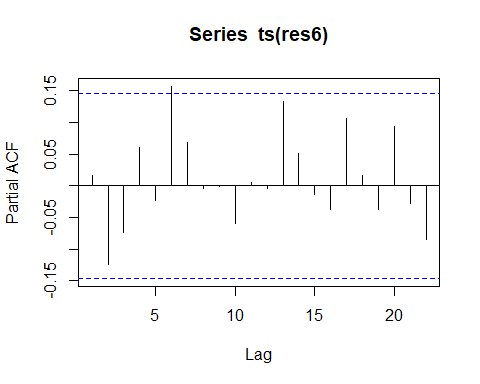


**present the residual autocorrelations and partial autocorrelations**

acf(ts(res6), 37)

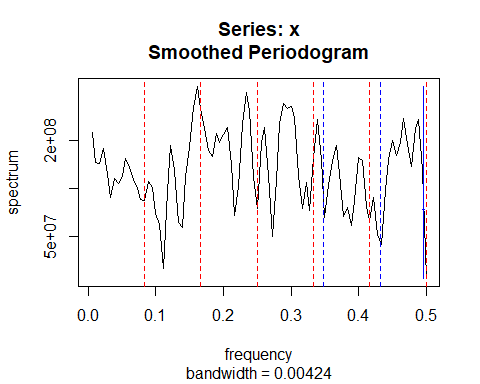


pacf(ts(res6))



**produce the residual spectral density plot**

spectrum(res6, span=3)  
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)



summary(model6)

##   
## Call:  
## lm(formula = msales ~ madv + madvl1 + msalesl1 + msalesl2 + msalesl3 +   
## fMonth + feb44 + dec44 + jan45 + sep45 + c220 + s220 + c348 +   
## s348)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -32653 -8676 0 7311 37616   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.092e+04 7.311e+03 2.861 0.004806 \*\*   
## madv 2.296e-01 4.142e-02 5.544 1.24e-07 \*\*\*  
## madvl1 8.894e-02 4.558e-02 1.951 0.052830 .   
## msalesl1 4.458e-01 6.910e-02 6.452 1.34e-09 \*\*\*  
## msalesl2 2.766e-01 7.006e-02 3.948 0.000119 \*\*\*  
## msalesl3 1.202e-01 6.374e-02 1.885 0.061245 .   
## fMonth2 -1.445e+04 7.243e+03 -1.996 0.047740 \*   
## fMonth3 -5.633e+03 7.724e+03 -0.729 0.466896   
## fMonth4 -2.836e+04 6.869e+03 -4.129 5.93e-05 \*\*\*  
## fMonth5 -2.763e+04 6.970e+03 -3.963 0.000113 \*\*\*  
## fMonth6 -2.878e+04 5.938e+03 -4.847 3.02e-06 \*\*\*  
## fMonth7 -9.507e+03 5.796e+03 -1.640 0.102996   
## fMonth8 -9.841e+03 6.129e+03 -1.606 0.110408   
## fMonth9 -5.175e+03 6.792e+03 -0.762 0.447329   
## fMonth10 -1.957e+03 7.091e+03 -0.276 0.782882   
## fMonth11 -4.649e+04 7.796e+03 -5.964 1.61e-08 \*\*\*  
## fMonth12 -4.962e+04 6.736e+03 -7.367 9.73e-12 \*\*\*  
## feb44 -4.508e+04 1.488e+04 -3.030 0.002869 \*\*   
## dec44 -5.374e+04 1.425e+04 -3.771 0.000230 \*\*\*  
## jan45 7.543e+04 1.471e+04 5.129 8.59e-07 \*\*\*  
## sep45 -4.439e+04 1.439e+04 -3.086 0.002405 \*\*   
## c220 3.460e+03 1.448e+03 2.389 0.018105 \*   
## s220 1.014e+03 1.490e+03 0.680 0.497375   
## c348 7.749e+02 1.505e+03 0.515 0.607476   
## s348 -3.511e+03 1.486e+03 -2.363 0.019381 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 13390 on 155 degrees of freedom  
## Multiple R-squared: 0.8979, Adjusted R-squared: 0.8821   
## F-statistic: 56.8 on 24 and 155 DF, p-value: < 2.2e-16

## Problem 1.c

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

**Discussion:** The calculation below shows a 90% duration interval of 6.10 years. This indicates that current advertising continues to have an impact upon sales for approximately 6.10 years.

summary(model6)

##   
## Call:  
## lm(formula = msales ~ madv + madvl1 + msalesl1 + msalesl2 + msalesl3 +   
## fMonth + feb44 + dec44 + jan45 + sep45 + c220 + s220 + c348 +   
## s348)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -32653 -8676 0 7311 37616   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.092e+04 7.311e+03 2.861 0.004806 \*\*   
## madv 2.296e-01 4.142e-02 5.544 1.24e-07 \*\*\*  
## madvl1 8.894e-02 4.558e-02 1.951 0.052830 .   
## msalesl1 4.458e-01 6.910e-02 6.452 1.34e-09 \*\*\*  
## msalesl2 2.766e-01 7.006e-02 3.948 0.000119 \*\*\*  
## msalesl3 1.202e-01 6.374e-02 1.885 0.061245 .   
## fMonth2 -1.445e+04 7.243e+03 -1.996 0.047740 \*   
## fMonth3 -5.633e+03 7.724e+03 -0.729 0.466896   
## fMonth4 -2.836e+04 6.869e+03 -4.129 5.93e-05 \*\*\*  
## fMonth5 -2.763e+04 6.970e+03 -3.963 0.000113 \*\*\*  
## fMonth6 -2.878e+04 5.938e+03 -4.847 3.02e-06 \*\*\*  
## fMonth7 -9.507e+03 5.796e+03 -1.640 0.102996   
## fMonth8 -9.841e+03 6.129e+03 -1.606 0.110408   
## fMonth9 -5.175e+03 6.792e+03 -0.762 0.447329   
## fMonth10 -1.957e+03 7.091e+03 -0.276 0.782882   
## fMonth11 -4.649e+04 7.796e+03 -5.964 1.61e-08 \*\*\*  
## fMonth12 -4.962e+04 6.736e+03 -7.367 9.73e-12 \*\*\*  
## feb44 -4.508e+04 1.488e+04 -3.030 0.002869 \*\*   
## dec44 -5.374e+04 1.425e+04 -3.771 0.000230 \*\*\*  
## jan45 7.543e+04 1.471e+04 5.129 8.59e-07 \*\*\*  
## sep45 -4.439e+04 1.439e+04 -3.086 0.002405 \*\*   
## c220 3.460e+03 1.448e+03 2.389 0.018105 \*   
## s220 1.014e+03 1.490e+03 0.680 0.497375   
## c348 7.749e+02 1.505e+03 0.515 0.607476   
## s348 -3.511e+03 1.486e+03 -2.363 0.019381 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 13390 on 155 degrees of freedom  
## Multiple R-squared: 0.8979, Adjusted R-squared: 0.8821   
## F-statistic: 56.8 on 24 and 155 DF, p-value: < 2.2e-16

St = 20920 + 0.4458St-1 + 0.2766St-2 + 0.1202St-3 + 0.2296At + .08894At-1 + …

(1 – 0.4458BSt-1 - 0.2766B^2St-2 - 0.1202B^3St-3) St = 20920 + 0.2296At + .08894At-1 + …

St = (1 – 0.4458BSt-1 - 0.2766B^2St-2 - 0.1202B3St-3)-1 \* 21030 + (1 – 0.4458BSt-1 - 0.2766B^2St-2 - 0.1202B3St-3)-1 \* (0.2296 + .08894B)At + …

We have a case where p = 3, r = 1, so p >= r,

Now we calculate deltas

delta0 = .2296 delta1 = .08894 + (0.4458 \* .2296) = .1913

deltapartial<-delta<-c(rep(0,times=500))  
#deltapartial is the partial sum of the deltas  
delta[1]<-0.2296  
delta[2]<-0.1913  
  
deltapartial[1]<-delta[1]  
deltapartial[2]<-deltapartial[1]+delta[2]  
for(j in 3:500){  
j1<-j-1;j2<-j-2  
delta[j]<-0.4458\*delta[j1]+0.2766\*delta[j2]  
deltapartial[j]<-deltapartial[j1]+delta[j]  
}  
  
deltapartial[500]\*0.9

## [1] 1.032745

deltapartial[1:20]

## [1] 0.2296000 0.4209000 0.5696889 0.6889326 0.7832464 0.8582743 0.9178090  
## [8] 0.9651022 1.0026529 1.0324742 1.0561551 1.0749606 1.0898943 1.1017533  
## [15] 1.1111707 1.1186492 1.1245879 1.1293040 1.1330491 1.1360231

90 percent duration interval = 2 + (1.033 - .4209) / (.570 - .4209) = 6.10

# Problem 2

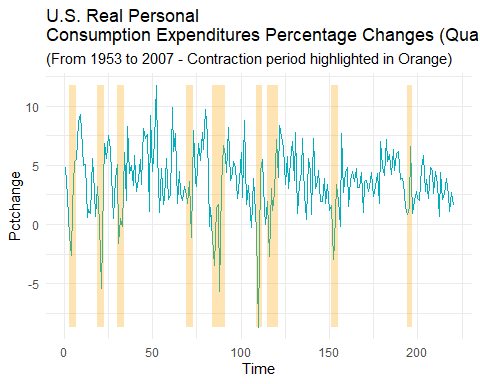
The file qconsumption2.txt contains quarterly data measuring U.S. Real Personal Consumption Expenditures percentage changes for the period 1953.1 to 2019.4. The data are seasonally adjusted.

consumption <- read.csv("qconsumption2.txt")  
attach(consumption)

1. Work with the data spanning the period 1953.1 to 2007.4.
2. Plot the data vs. time and discuss features of the plot. Give a list of the economic downturns as determined by the Business Cycle Dating Committee of NBER. Can you relate these recessions to movements in the plot?

**Discussion:** Personal consumption expenditures percentage changes drop during recessions. Recession periods are highlighted in orange below and as we can see, percent changes consistently drops during recession periods.

Personal consumption expenditures percentage changes are mostly positive from 1953 to 2007, the only exception being negative percentage changes during most of the highlighted recession periods. This indicates that recession largely drives consumers to reduce their personal consumption expenditures.



1. Begin by identifying outliers and form dummies for them. [Form the dummies to have length 220, so they cover the time span from 1953.1 to 2007.4.]

jan58<-rep(0, nrow(consumption.a))  
oct65<-rep(0, nrow(consumption.a))  
oct74<-rep(0, nrow(consumption.a))  
apr80<-rep(0, nrow(consumption.a))  
  
jan58[21] <- 1  
oct65[52] <- 1  
oct74[88] <- 1  
apr80[110] <- 1  
  
consumption.a$jan58 <- jan58  
consumption.a$oct65 <- oct65  
consumption.a$oct74 <- oct74  
consumption.a$apr80 <- apr80

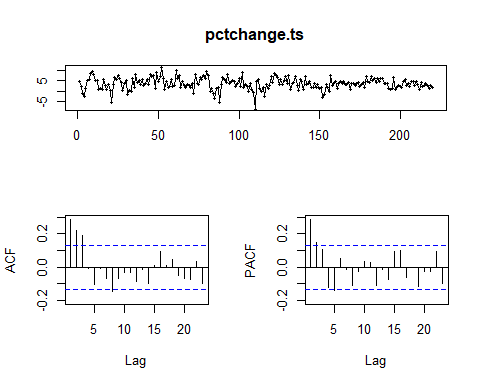
1. An ARX model has the form. Then R commands to fit such a model to the percentage changes, if the chosen order of the AR structure is p, are as follows:

df<-data.frame(d1,d2) arxmodel<-arima(Pctchange.ts[1:220],order=c(p,0,0),xreg=df)

Fit an ARX model to the data covering the period 1953.1 to 2007.4. Explain how you arrived at your model fit.

**Discussion:** The initial ACF suggests MA(3) and PACF suggests AR(2) are good places to start.

pctchange.ts<-ts(consumption.a$Pctchange)  
forecast::tsdisplay(pctchange.ts)

 **Discussion:** We fit MA(3) and AR(2) models below, including the outliers. Both models show that white noise reduction was successfully achieved, however MA(3) has better/smaller AIC than AR(2) (AIC is 1009.97 for MA(3) and 1019.01 for AR(2)). Both models show that all four of the outlier variables are statistically significant.

We next try an ARMA(2,3), and MA(4) and AR(3) for comparison.

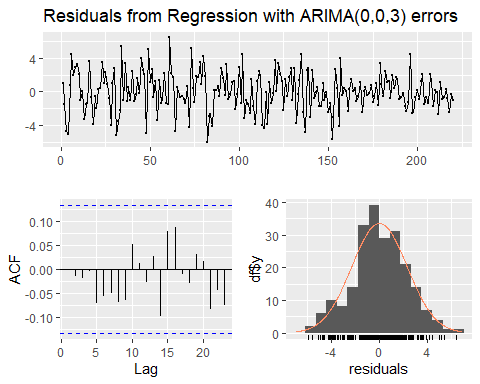
outliers<- consumption.a %>% select(jan58, oct65,oct74,apr80)  
outliers<- as.matrix(outliers)  
MA3<-Arima(pctchange.ts,order=c(0,0,3), xreg=outliers)  
AR2<-Arima(pctchange.ts,order=c(2,0,0), xreg=outliers)

MA(3) analysis

summary(MA3)

## Series: pctchange.ts   
## Regression with ARIMA(0,0,3) errors   
##   
## Coefficients:  
## ma1 ma2 ma3 intercept jan58 oct65 oct74 apr80  
## 0.1876 0.3569 0.2632 3.7032 -7.8517 6.4354 -9.1051 -12.7053  
## s.e. 0.0665 0.0652 0.0667 0.2802 2.1540 2.1435 2.0978 2.1089  
##   
## sigma^2 = 5.507: log likelihood = -495.98  
## AIC=1009.97 AICc=1010.83 BIC=1040.51  
##   
## Training set error measures:  
## ME RMSE MAE MPE MAPE MASE ACF1  
## Training set -0.007702607 2.303681 1.820543 -Inf Inf 0.6780595 0.0006599653

checkresiduals(MA3)



##   
## Ljung-Box test  
##   
## data: Residuals from Regression with ARIMA(0,0,3) errors  
## Q\* = 5.3177, df = 3, p-value = 0.15  
##   
## Model df: 8. Total lags used: 11

coeftest(MA3)

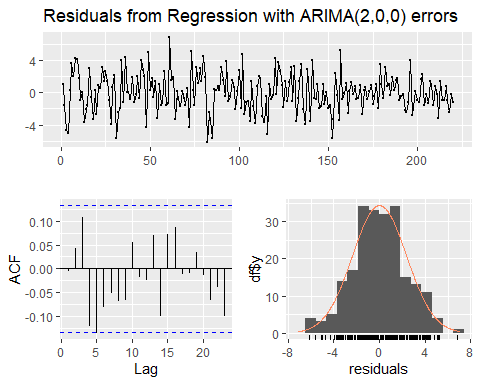
##   
## z test of coefficients:  
##   
## Estimate Std. Error z value Pr(>|z|)   
## ma1 0.187575 0.066508 2.8204 0.0047971 \*\*   
## ma2 0.356894 0.065172 5.4762 4.345e-08 \*\*\*  
## ma3 0.263220 0.066707 3.9459 7.949e-05 \*\*\*  
## intercept 3.703234 0.280235 13.2147 < 2.2e-16 \*\*\*  
## jan58 -7.851720 2.154031 -3.6451 0.0002673 \*\*\*  
## oct65 6.435412 2.143473 3.0023 0.0026792 \*\*   
## oct74 -9.105138 2.097773 -4.3404 1.422e-05 \*\*\*  
## apr80 -12.705262 2.108946 -6.0245 1.697e-09 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

AR2 analysis

summary(AR2)

## Series: pctchange.ts   
## Regression with ARIMA(2,0,0) errors   
##   
## Coefficients:  
## ar1 ar2 intercept jan58 oct65 oct74 apr80  
## 0.2007 0.2473 3.6854 -9.1316 7.6634 -9.2379 -11.7142  
## s.e. 0.0663 0.0661 0.2878 2.2747 2.2810 2.2580 2.2582  
##   
## sigma^2 = 5.77: log likelihood = -501.51  
## AIC=1019.01 AICc=1019.69 BIC=1046.16  
##   
## Training set error measures:  
## ME RMSE MAE MPE MAPE MASE ACF1  
## Training set -0.001426658 2.36358 1.878447 -Inf Inf 0.6996258 -0.004950909

checkresiduals(AR2)



##   
## Ljung-Box test  
##   
## data: Residuals from Regression with ARIMA(2,0,0) errors  
## Q\* = 15.174, df = 3, p-value = 0.001674  
##   
## Model df: 7. Total lags used: 10

coeftest(AR2)

##   
## z test of coefficients:  
##   
## Estimate Std. Error z value Pr(>|z|)   
## ar1 0.200700 0.066329 3.0258 0.0024797 \*\*   
## ar2 0.247291 0.066076 3.7426 0.0001822 \*\*\*  
## intercept 3.685352 0.287767 12.8067 < 2.2e-16 \*\*\*  
## jan58 -9.131588 2.274745 -4.0143 5.961e-05 \*\*\*  
## oct65 7.663437 2.280988 3.3597 0.0007803 \*\*\*  
## oct74 -9.237895 2.257953 -4.0913 4.290e-05 \*\*\*  
## apr80 -11.714199 2.258159 -5.1875 2.131e-07 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

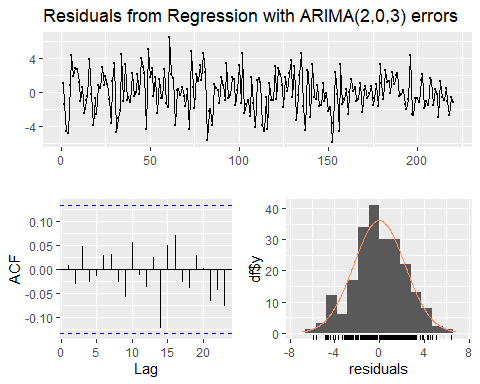
**Discussion:** MA(3) still is the best fit with the smallest AIC of 1009.97 after fitting ARMA(2,3), MA(4), and AR(3), with AICs of 1011.2, 1011.96, 1020.87 respectively. We see also in MA(4) and AR(3) that the added lags in each model are insignificant and thus overfit. We can now confidently use MA(3).

ARMA23<-Arima(pctchange.ts,order=c(2,0,3), xreg=outliers)  
MA4<-Arima(pctchange.ts,order=c(0,0,4), xreg=outliers)  
AR3<-Arima(pctchange.ts,order=c(3,0,0), xreg=outliers)

summary(ARMA23)

## Series: pctchange.ts   
## Regression with ARIMA(2,0,3) errors   
##   
## Coefficients:  
## ar1 ar2 ma1 ma2 ma3 intercept jan58 oct65  
## 1.4260 -0.6255 -1.2778 0.7666 -0.2481 3.7251 -8.9727 7.8307  
## s.e. 0.1435 0.1158 0.1539 0.1178 0.0968 0.1884 2.1414 2.1446  
## oct74 apr80  
## -9.2791 -13.0882  
## s.e. 2.1468 2.1599  
##   
## sigma^2 = 5.489: log likelihood = -494.6  
## AIC=1011.2 AICc=1012.47 BIC=1048.53  
##   
## Training set error measures:  
## ME RMSE MAE MPE MAPE MASE ACF1  
## Training set -0.01314106 2.288967 1.840667 -Inf Inf 0.6855546 0.00813564

checkresiduals(ARMA23)



##   
## Ljung-Box test  
##   
## data: Residuals from Regression with ARIMA(2,0,3) errors  
## Q\* = 3.5635, df = 3, p-value = 0.3126  
##   
## Model df: 10. Total lags used: 13

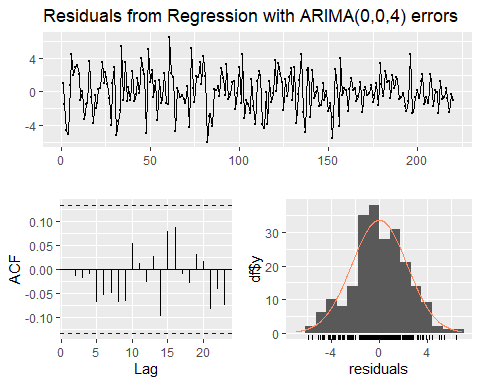
coeftest(ARMA23)

##   
## z test of coefficients:  
##   
## Estimate Std. Error z value Pr(>|z|)   
## ar1 1.425975 0.143539 9.9344 < 2.2e-16 \*\*\*  
## ar2 -0.625527 0.115842 -5.3998 6.670e-08 \*\*\*  
## ma1 -1.277779 0.153906 -8.3023 < 2.2e-16 \*\*\*  
## ma2 0.766639 0.117809 6.5075 7.641e-11 \*\*\*  
## ma3 -0.248118 0.096786 -2.5636 0.0103601 \*   
## intercept 3.725072 0.188367 19.7756 < 2.2e-16 \*\*\*  
## jan58 -8.972699 2.141407 -4.1901 2.788e-05 \*\*\*  
## oct65 7.830662 2.144578 3.6514 0.0002608 \*\*\*  
## oct74 -9.279090 2.146781 -4.3223 1.544e-05 \*\*\*  
## apr80 -13.088195 2.159946 -6.0595 1.365e-09 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

summary(MA4)

## Series: pctchange.ts   
## Regression with ARIMA(0,0,4) errors   
##   
## Coefficients:  
## ma1 ma2 ma3 ma4 intercept jan58 oct65 oct74  
## 0.1892 0.3585 0.2638 0.0066 3.7030 -7.8770 6.4889 -9.1071  
## s.e. 0.0695 0.0682 0.0671 0.0778 0.2818 2.1753 2.2353 2.0991  
## apr80  
## -12.6942  
## s.e. 2.1144  
##   
## sigma^2 = 5.533: log likelihood = -495.98  
## AIC=1011.96 AICc=1013.01 BIC=1045.9  
##   
## Training set error measures:  
## ME RMSE MAE MPE MAPE MASE ACF1  
## Training set -0.007714926 2.303654 1.821641 -Inf Inf 0.6784683 -0.000930194

checkresiduals(MA4)



##   
## Ljung-Box test  
##   
## data: Residuals from Regression with ARIMA(0,0,4) errors  
## Q\* = 5.4602, df = 3, p-value = 0.141  
##   
## Model df: 9. Total lags used: 12

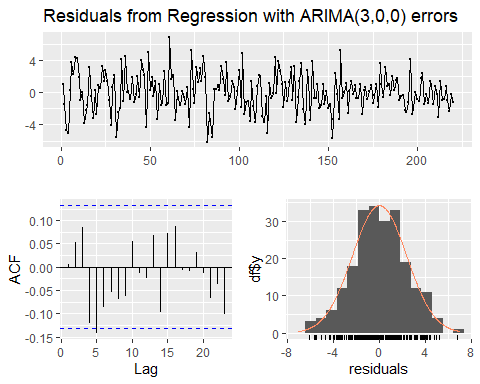
coeftest(MA4)

##   
## z test of coefficients:  
##   
## Estimate Std. Error z value Pr(>|z|)   
## ma1 0.1892463 0.0695096 2.7226 0.0064772 \*\*   
## ma2 0.3585037 0.0681844 5.2579 1.457e-07 \*\*\*  
## ma3 0.2637990 0.0671061 3.9311 8.457e-05 \*\*\*  
## ma4 0.0065876 0.0777649 0.0847 0.9324909   
## intercept 3.7030197 0.2818404 13.1387 < 2.2e-16 \*\*\*  
## jan58 -7.8769700 2.1752851 -3.6211 0.0002933 \*\*\*  
## oct65 6.4889154 2.2353366 2.9029 0.0036975 \*\*   
## oct74 -9.1070637 2.0990626 -4.3386 1.434e-05 \*\*\*  
## apr80 -12.6942495 2.1143766 -6.0038 1.928e-09 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

summary(AR3)

## Series: pctchange.ts   
## Regression with ARIMA(3,0,0) errors   
##   
## Coefficients:  
## ar1 ar2 ar3 intercept jan58 oct65 oct74 apr80  
## 0.1948 0.2415 0.0264 3.6827 -9.1138 7.5358 -9.1271 -11.6976  
## s.e. 0.0682 0.0679 0.0691 0.2954 2.2804 2.3108 2.2806 2.2625  
##   
## sigma^2 = 5.793: log likelihood = -501.43  
## AIC=1020.87 AICc=1021.72 BIC=1051.41  
##   
## Training set error measures:  
## ME RMSE MAE MPE MAPE MASE ACF1  
## Training set -0.0006923789 2.362787 1.872081 -Inf Inf 0.6972548 0.006459887

checkresiduals(AR3)



##   
## Ljung-Box test  
##   
## data: Residuals from Regression with ARIMA(3,0,0) errors  
## Q\* = 15.238, df = 3, p-value = 0.001624  
##   
## Model df: 8. Total lags used: 11

coeftest(AR3)

##   
## z test of coefficients:  
##   
## Estimate Std. Error z value Pr(>|z|)   
## ar1 0.194779 0.068158 2.8578 0.0042664 \*\*   
## ar2 0.241465 0.067895 3.5565 0.0003759 \*\*\*  
## ar3 0.026392 0.069138 0.3817 0.7026593   
## intercept 3.682668 0.295351 12.4688 < 2.2e-16 \*\*\*  
## jan58 -9.113808 2.280396 -3.9966 6.426e-05 \*\*\*  
## oct65 7.535807 2.310842 3.2611 0.0011099 \*\*   
## oct74 -9.127150 2.280609 -4.0021 6.279e-05 \*\*\*  
## apr80 -11.697557 2.262474 -5.1703 2.338e-07 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

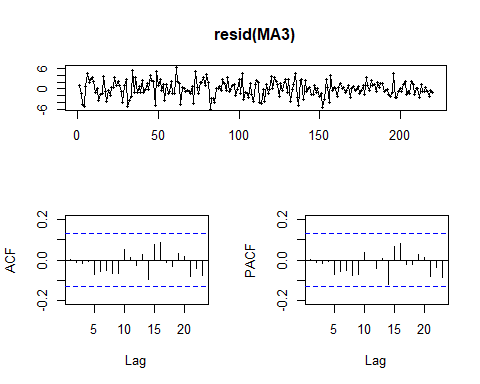
1. Examine the residuals to investigate whether your selected model has achieved reduction to white noise. For this purpose include in your discussion consideration of the residual autocorrelations and partial autocorrelations and the residual spectral density. Also examine and discuss the plot of the residuals vs. time. Perform Bartlett’s test to determine if the fit has produced reduction to white noise. [Bartlett’s test is described and its use is illustrated in the 28 March notes.]

**Discussion:** The ACF, PACF, and spectrum plots all indicate that the MA(3) model has been reduced to white noise. The ACF and PACF plots show no significant lag (all lags are within the blue lines) and the spectrum analysis shows that the highest and lowest peaks are well within the blue line above the notch.

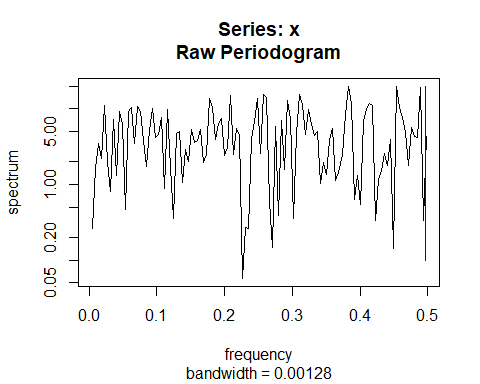
The residuals show constant variance over time, and there is no significant autocorrelation structure that can be seen in the residuals.

Bartlett’s test shows that the model has been reduced to white noise. With a p-value of 0.9949, we cannot reject the null hypothesis that MA(3) fit has been reduced to white noise.

forecast::tsdisplay(resid(MA3))



spectrum(resid(MA3))



bartlettB.test(ts(resid(MA3)))

##   
## Bartlett B Test for white noise  
##   
## data:   
## = 0.41767, p-value = 0.9949

1. Find the zeros of the autoregressive polynomial for your model fit and interpret the results.

**Discussion:** Since we chose an MA(3) model we find first the autoregressive form of the model by calculating deltas. The amplitude of complex zeros is 0.4477, which suggests that the presence of cycle is weak as it is closer to 0 than 1. The zeros of the autoregressive polynomial suggests the presence of a weak cycle of length about 15.45 quarters. This corresponds to peaks in the estimated spectral density at approximately 1/15.45 = 0.0647 cycles per quarter.

coef(MA3)

## ma1 ma2 ma3 intercept jan58 oct65   
## 0.1875751 0.3568942 0.2632198 3.7032337 -7.8517204 6.4354120   
## oct74 apr80   
## -9.1051376 -12.7052624

delta<-c(rep(0,times=10))  
delta[1]<-1  
delta[2]<--0.1875751  
  
for(j in 3:10){  
j1<-j-1  
j2<-j-2  
delta[j]<-- 0.1875751\*delta[j1]-0.3568942\*delta[j2]  
}  
  
delta

## [1] 1.000000000 -0.187575100 -0.321709782 0.127289210 0.090940069  
## [6] -0.062486873 -0.020735002 0.026190573 0.002487503 -0.009813857

# choosing delta[1:4] because MA(3) model has 3 coefficients  
zeros<-1/polyroot(delta[1:4])  
zeros

## [1] 0.4112396-0.1771092i -0.6349042+0.0000000i 0.4112396+0.1771092i

#amplitude. Selecting zero with positive imaginary part.  
Mod(zeros[3])

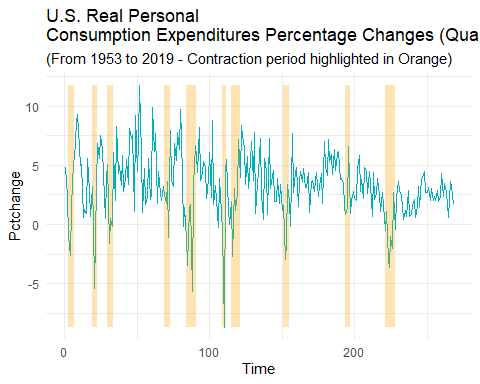
## [1] 0.4477563

#period  
2\*pi/Arg(zeros)[3]

## [1] 15.45053

1. Repeat the analysis in part (a) for the full time period in the data set, 1953.1 to 2019.4. Note that now the time series spans 268 quarters.
2. Plot the data vs. time and discuss features of the plot. Give a list of the economic downturns as determined by the Business Cycle Dating Committee of NBER. Can you relate these recessions to movements in the plot?

**Discussion:** The extended plot now includes the 2008-2009 recession period and in it tthe percentage change of personal consumption expenditures are negative, a continuation of the pattern previously seen in the 1953-2007 analysis. As previously stated, this means consumers spend less on personal consumption during recession periods.

 (ii) Begin by identifying outliers and form dummies for them. [Form the dummies to have length 220, so they cover the time span from 1953.1 to 2007.4.]

**Discussion:** There are no new outliers that can be immediately seen from the plot. We do see a new low point during the 2009 recession but the drop is not so dramatic, as indicated by the percentage change points which is still over -5 (the previous outliers have percentage change points of either less than -5 or greater than 10). If our residuals later indicate that it is indeed an outlier we can refit.

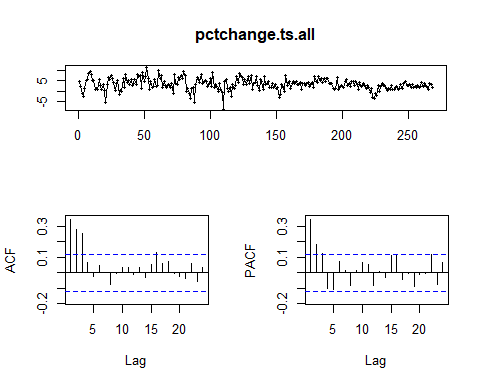
jan58<-rep(0, nrow(consumption))  
oct65<-rep(0, nrow(consumption))  
oct74<-rep(0, nrow(consumption))  
apr80<-rep(0, nrow(consumption))  
  
jan58[21] <- 1  
oct65[52] <- 1  
oct74[88] <- 1  
apr80[110] <- 1  
  
consumption$jan58 <- jan58  
consumption$oct65 <- oct65  
consumption$oct74 <- oct74  
consumption$apr80 <- apr80

1. An ARX model has the form. Then R commands to fit such a model to the percentage changes, if the chosen order of the AR structure is p, are as follows:

Fit an ARX model to the data covering the period 1953.1 to 2007.4. Explain how you arrived at your model fit.

**Discussion:** The initial ACF still suggests MA(3) fit to start, however now the PACF suggests AR(3) could be a better place to start instead of AR(2). We

pctchange.ts.all<-ts(consumption$Pctchange)  
forecast::tsdisplay(pctchange.ts.all)

 **Discussion:** We fit MA(3) and AR(3) models below, including the outliers AND using the entire consumption dataset. Bartlett’s test for both models show that white noise reduction was successfully achieved, however MA(3) has better/smaller AIC than AR(3) (AIC is 1202.46 for MA(3) and 1213.58 for AR(3)). Both models show that all four of the outlier variables are statistically significant.

The Coefficient test for AR(3) shows that it is an overfit model (the ar3 variable has a p-value of 0.35).

The ACF and PACF plot for the residuals of MA(3) model now shows that at lag 24 the line exceeds the blue threshold, however it is only slighlty exceeds the blue threshold so it is not necessarily critical that we add a variable to account for it.

We next try an ARMA(3,3), and MA(4) and AR(2) for comparison and see if they may provide a better fit.

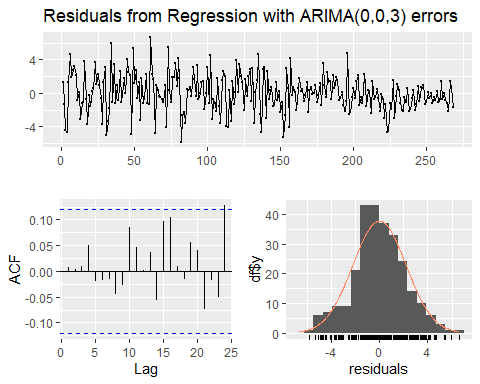
outliers<- consumption %>% select(jan58, oct65,oct74,apr80)  
outliers<- as.matrix(outliers)  
MA3.all<-Arima(pctchange.ts.all,order=c(0,0,3), xreg=outliers)  
AR3.all<-Arima(pctchange.ts.all,order=c(3,0,0), xreg=outliers)

MA(3).all analysis

summary(MA3.all)

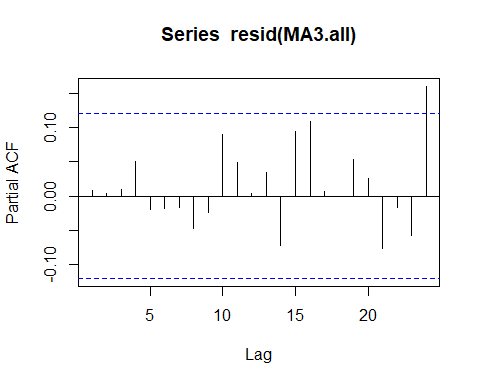
## Series: pctchange.ts.all   
## Regression with ARIMA(0,0,3) errors   
##   
## Coefficients:  
## ma1 ma2 ma3 intercept jan58 oct65 oct74 apr80  
## 0.2257 0.3851 0.2979 3.3677 -7.4515 6.2100 -8.9976 -12.6660  
## s.e. 0.0596 0.0583 0.0580 0.2563 2.0185 2.0141 1.9671 1.9774  
##   
## sigma^2 = 5.003: log likelihood = -592.23  
## AIC=1202.46 AICc=1203.16 BIC=1234.78  
##   
## Training set error measures:  
## ME RMSE MAE MPE MAPE MASE ACF1  
## Training set -0.007967398 2.203086 1.724559 -Inf Inf 0.7061146 0.008520978

checkresiduals(MA3.all)



##   
## Ljung-Box test  
##   
## data: Residuals from Regression with ARIMA(0,0,3) errors  
## Q\* = 4.498, df = 3, p-value = 0.2125  
##   
## Model df: 8. Total lags used: 11

pacf(resid(MA3.all))



coeftest(MA3.all)

##   
## z test of coefficients:  
##   
## Estimate Std. Error z value Pr(>|z|)   
## ma1 0.225721 0.059625 3.7857 0.0001533 \*\*\*  
## ma2 0.385061 0.058312 6.6034 4.018e-11 \*\*\*  
## ma3 0.297876 0.057969 5.1386 2.769e-07 \*\*\*  
## intercept 3.367744 0.256343 13.1376 < 2.2e-16 \*\*\*  
## jan58 -7.451517 2.018525 -3.6916 0.0002229 \*\*\*  
## oct65 6.210025 2.014069 3.0833 0.0020470 \*\*   
## oct74 -8.997600 1.967075 -4.5741 4.783e-06 \*\*\*  
## apr80 -12.666013 1.977356 -6.4055 1.498e-10 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

bartlettB.test(ts(resid(MA3.all)))

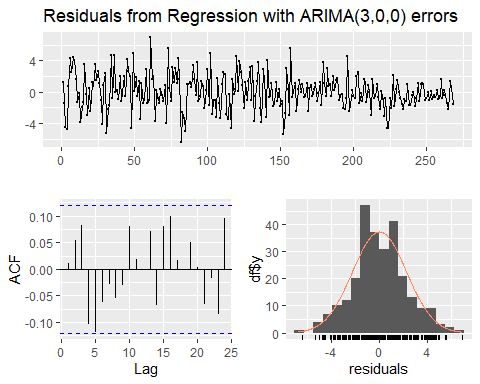
##   
## Bartlett B Test for white noise  
##   
## data:   
## = 0.33371, p-value = 0.9999

AR3.all analysis

summary(AR3.all)

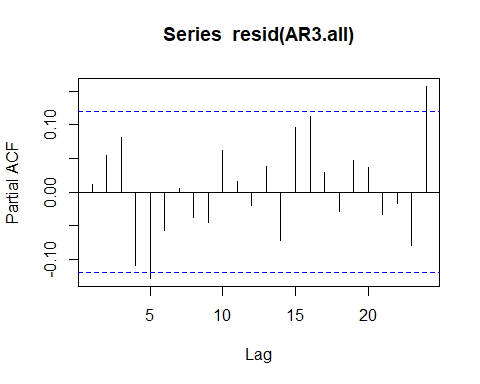
## Series: pctchange.ts.all   
## Regression with ARIMA(3,0,0) errors   
##   
## Coefficients:  
## ar1 ar2 ar3 intercept jan58 oct65 oct74 apr80  
## 0.2335 0.2711 0.0582 3.3466 -9.0298 7.3935 -8.9232 -11.5447  
## s.e. 0.0616 0.0618 0.0623 0.3123 2.1378 2.1644 2.1351 2.1197  
##   
## sigma^2 = 5.219: log likelihood = -597.79  
## AIC=1213.58 AICc=1214.28 BIC=1245.9  
##   
## Training set error measures:  
## ME RMSE MAE MPE MAPE MASE ACF1  
## Training set -0.001896552 2.250262 1.762955 -Inf Inf 0.7218358 0.01093216

checkresiduals(AR3.all)



##   
## Ljung-Box test  
##   
## data: Residuals from Regression with ARIMA(3,0,0) errors  
## Q\* = 13.841, df = 3, p-value = 0.00313  
##   
## Model df: 8. Total lags used: 11

pacf(resid(AR3.all))



coeftest(AR3.all)

##   
## z test of coefficients:  
##   
## Estimate Std. Error z value Pr(>|z|)   
## ar1 0.233460 0.061644 3.7873 0.0001523 \*\*\*  
## ar2 0.271122 0.061756 4.3902 1.132e-05 \*\*\*  
## ar3 0.058220 0.062264 0.9350 0.3497650   
## intercept 3.346593 0.312288 10.7164 < 2.2e-16 \*\*\*  
## jan58 -9.029795 2.137824 -4.2238 2.402e-05 \*\*\*  
## oct65 7.393459 2.164399 3.4159 0.0006356 \*\*\*  
## oct74 -8.923239 2.135095 -4.1793 2.924e-05 \*\*\*  
## apr80 -11.544735 2.119726 -5.4463 5.142e-08 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

bartlettB.test(ts(resid(AR3.all)))

##   
## Bartlett B Test for white noise  
##   
## data:   
## = 0.70538, p-value = 0.7023

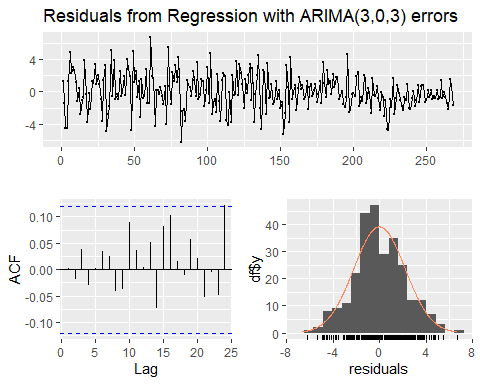
**Discussion:** MA(3) still is the best fit with the smallest AIC of 1202.46 after fitting ARMA(3,3), MA(4), and AR(2), with AICs of 1209.14, 1204.1, 1212.46 respectively. We see also in MA(4) that the added lags in the model are insignificant and thus overfit. MA(3) still has the smallest AIC so we will go forward in our analysis using this model.

ARMA33.all<-Arima(pctchange.ts.all,order=c(3,0,3), xreg=outliers)  
MA4.all<-Arima(pctchange.ts.all,order=c(0,0,4), xreg=outliers)  
AR2.all<-Arima(pctchange.ts.all,order=c(2,0,0), xreg=outliers)

summary(ARMA33.all)

## Series: pctchange.ts.all   
## Regression with ARIMA(3,0,3) errors   
##   
## Coefficients:  
## ar1 ar2 ar3 ma1 ma2 ma3 intercept jan58  
## 1.3199 -0.6453 0.0843 -1.0987 0.7571 -0.2088 3.3716 -8.6502  
## s.e. 0.6266 0.6072 0.2700 0.6228 0.4597 0.3675 0.2513 2.0705  
## oct65 oct74 apr80  
## 7.4678 -9.1561 -12.5575  
## s.e. 2.0198 2.0337 2.0324  
##   
## sigma^2 = 5.075: log likelihood = -592.57  
## AIC=1209.14 AICc=1210.37 BIC=1252.24  
##   
## Training set error measures:  
## ME RMSE MAE MPE MAPE MASE ACF1  
## Training set -0.009005074 2.206117 1.745451 -Inf Inf 0.7146685 0.001522748

checkresiduals(ARMA33.all)



##   
## Ljung-Box test  
##   
## data: Residuals from Regression with ARIMA(3,0,3) errors  
## Q\* = 7.0444, df = 3, p-value = 0.0705  
##   
## Model df: 11. Total lags used: 14

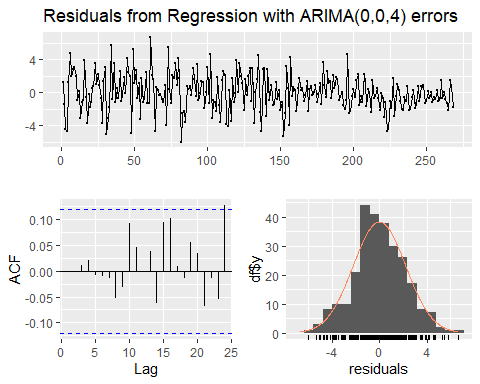
coeftest(ARMA33.all)

##   
## z test of coefficients:  
##   
## Estimate Std. Error z value Pr(>|z|)   
## ar1 1.319897 0.626561 2.1066 0.0351544 \*   
## ar2 -0.645267 0.607183 -1.0627 0.2879073   
## ar3 0.084346 0.269970 0.3124 0.7547143   
## ma1 -1.098707 0.622789 -1.7642 0.0777030 .   
## ma2 0.757114 0.459657 1.6471 0.0995313 .   
## ma3 -0.208778 0.367490 -0.5681 0.5699535   
## intercept 3.371604 0.251322 13.4155 < 2.2e-16 \*\*\*  
## jan58 -8.650167 2.070468 -4.1779 2.942e-05 \*\*\*  
## oct65 7.467840 2.019766 3.6974 0.0002178 \*\*\*  
## oct74 -9.156082 2.033726 -4.5021 6.728e-06 \*\*\*  
## apr80 -12.557453 2.032395 -6.1786 6.465e-10 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

summary(MA4.all)

## Series: pctchange.ts.all   
## Regression with ARIMA(0,0,4) errors   
##   
## Coefficients:  
## ma1 ma2 ma3 ma4 intercept jan58 oct65 oct74  
## 0.2341 0.3927 0.2990 0.0408 3.3673 -7.6296 6.5676 -9.0241  
## s.e. 0.0623 0.0616 0.0586 0.0679 0.2639 2.0467 2.1052 1.9769  
## apr80  
## -12.6048  
## s.e. 1.9906  
##   
## sigma^2 = 5.016: log likelihood = -592.05  
## AIC=1204.1 AICc=1204.96 BIC=1240.01  
##   
## Training set error measures:  
## ME RMSE MAE MPE MAPE MASE ACF1  
## Training set -0.008083565 2.201688 1.726284 -Inf Inf 0.7068209 -0.001065762

checkresiduals(MA4.all)



##   
## Ljung-Box test  
##   
## data: Residuals from Regression with ARIMA(0,0,4) errors  
## Q\* = 4.2435, df = 3, p-value = 0.2363  
##   
## Model df: 9. Total lags used: 12

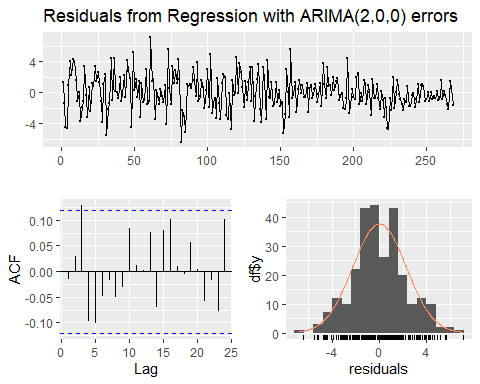
coeftest(MA4.all)

##   
## z test of coefficients:  
##   
## Estimate Std. Error z value Pr(>|z|)   
## ma1 0.234121 0.062279 3.7592 0.0001704 \*\*\*  
## ma2 0.392695 0.061594 6.3755 1.823e-10 \*\*\*  
## ma3 0.298964 0.058631 5.0991 3.413e-07 \*\*\*  
## ma4 0.040805 0.067923 0.6007 0.5480080   
## intercept 3.367307 0.263874 12.7610 < 2.2e-16 \*\*\*  
## jan58 -7.629620 2.046700 -3.7278 0.0001932 \*\*\*  
## oct65 6.567581 2.105241 3.1196 0.0018108 \*\*   
## oct74 -9.024107 1.976893 -4.5648 5.000e-06 \*\*\*  
## apr80 -12.604787 1.990642 -6.3320 2.420e-10 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

summary(AR2.all)

## Series: pctchange.ts.all   
## Regression with ARIMA(2,0,0) errors   
##   
## Coefficients:  
## ar1 ar2 intercept jan58 oct65 oct74 apr80  
## 0.2492 0.2870 3.3493 -9.0651 7.6729 -9.1525 -11.5650  
## s.e. 0.0593 0.0593 0.2953 2.1262 2.1311 2.1115 2.1106  
##   
## sigma^2 = 5.217: log likelihood = -598.23  
## AIC=1212.46 AICc=1213.01 BIC=1241.18  
##   
## Training set error measures:  
## ME RMSE MAE MPE MAPE MASE ACF1  
## Training set -0.002765766 2.253968 1.770341 -Inf Inf 0.7248599 -0.014931

checkresiduals(AR2.all)



##   
## Ljung-Box test  
##   
## data: Residuals from Regression with ARIMA(2,0,0) errors  
## Q\* = 13.956, df = 3, p-value = 0.002965  
##   
## Model df: 7. Total lags used: 10

coeftest(AR2.all)

##   
## z test of coefficients:  
##   
## Estimate Std. Error z value Pr(>|z|)   
## ar1 0.249242 0.059275 4.2049 2.612e-05 \*\*\*  
## ar2 0.286971 0.059257 4.8428 1.280e-06 \*\*\*  
## intercept 3.349277 0.295330 11.3408 < 2.2e-16 \*\*\*  
## jan58 -9.065080 2.126169 -4.2636 2.012e-05 \*\*\*  
## oct65 7.672937 2.131084 3.6005 0.0003176 \*\*\*  
## oct74 -9.152533 2.111454 -4.3347 1.460e-05 \*\*\*  
## apr80 -11.564962 2.110630 -5.4794 4.268e-08 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

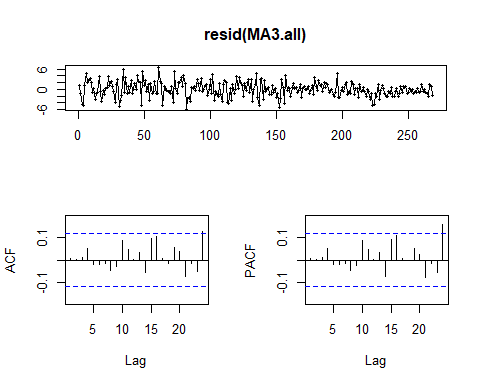
1. Examine the residuals to investigate whether your selected model has achieved reduction to white noise. For this purpose include in your discussion consideration of the residual autocorrelations and partial autocorrelations and the residual spectral density. Also examine and discuss the plot of the residuals vs. time. Perform Bartlett’s test to determine if the fit has produced reduction to white noise. [Bartlett’s test is described and its use is illustrated in the 28 March notes.]

**Discussion:** The ACF, PACF, and spectrum plots all indicate that the MA(3) model has been reduced to white noise. The ACF and PACF plots show some lag at lag 24 but it is not too significant. The spectrum analysis shows that the highest and lowest peaks are well within the blue line above the notch.

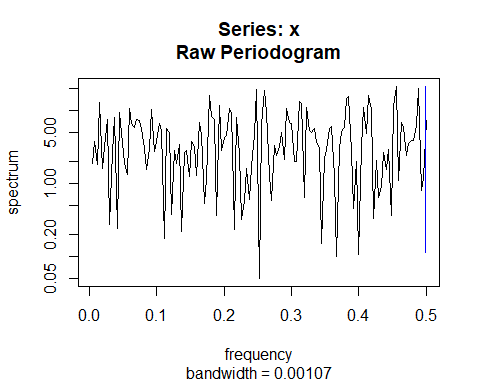
The residuals show constant variance over time, and there is no significant autocorrelation structure that can be seen in the residuals. There may be is a bit of funneling of residuals at later periods but not to such an extent that we can confidently conclude presence of heteroscedacticity in the residuals.

Bartlett’s test overwhelmingly shows that the model has been reduced to white noise with a p-value of 0.9999.

forecast::tsdisplay(resid(MA3.all))



spectrum(resid(MA3.all))



bartlettB.test(ts(resid(MA3.all)))

##   
## Bartlett B Test for white noise  
##   
## data:   
## = 0.33371, p-value = 0.9999

1. Find the zeros of the autoregressive polynomial for your model fit and interpret the results.

**Discussion:** Since we chose an MA(3) model we find first the autoregressive form of the model by calculating deltas. The amplitude of complex zeros of the deltas is now 0.4968268, which suggests that the presence of cycle is weak as it is closer to 0 than 1. The zeros of the autoregressive polynomial suggests the presence of a weak cycle of length about 13.21 quarters. This corresponds to peaks in the estimated spectral density at approximately 1/15.45 = 0.0757 cycles per quarter.

coef(MA3)

## ma1 ma2 ma3 intercept jan58 oct65   
## 0.1875751 0.3568942 0.2632198 3.7032337 -7.8517204 6.4354120   
## oct74 apr80   
## -9.1051376 -12.7052624

coef(MA3.all)

## ma1 ma2 ma3 intercept jan58 oct65   
## 0.2257209 0.3850609 0.2978763 3.3677438 -7.4515169 6.2100247   
## oct74 apr80   
## -8.9976004 -12.6660132

delta<-c(rep(0,times=10))  
delta[1]<-1  
delta[2]<--0.2257209  
  
for(j in 3:10){  
j1<-j-1  
j2<-j-2  
delta[j]<-- 0.2257209\*delta[j1]-0.3850609\*delta[j2]  
}  
  
delta

## [1] 1.000000000 -0.225720900 -0.334110975 0.162332123 0.092011320  
## [6] -0.083276631 -0.016632686 0.035820919 -0.001680933 -0.013413814

# choosing delta[1:4] because MA(3) model has 3 coefficients  
zeros<-1/polyroot(delta[1:4])  
zeros

## [1] 0.4416851-0.2274888i -0.6576493-0.0000000i 0.4416851+0.2274888i

#amplitude. Selecting zero with positive imaginary part.  
Mod(zeros[3])

## [1] 0.4968268

#period  
2\*pi/Arg(zeros)[3]

## [1] 13.21071

1. The quarterly percentage changes for 2020.1 to 2021.4 were −6.9, −33.4, 41.4, 3.4, 11.4, 12.0, 2.0, and 3.1. Comment briefly on the implications of these numbers.

**Discussion:** The introduction of these numbers to fit a model may lead to sub-optimal models. The four previously identified outliers are percentage changes of values of -5.4, 11.7, -5.7, and -8.7. If we use the same standard in selecting these outliers (less than -5 and greater than 10), then five percentage change values from 2020.1 to 2021.4 would classify as outliers (−6.9, −33.4, 41.4, 11.4, 12.0). Two values in particular, -33.4 and 41.4, are vastly different from anything we have seen from 1953.1 to 2019.4, where most percentage changes fall between -5.0 and 10.0. These numbers has the potential to distort the model fitted and should be excluded.