Group Assingment

Reece Wooten, Kyle Katzen and Daxi Cheng

12/2/2017

Contents

Question 1

```r
case_data<-read.csv('Case Shipments.csv')
seas_data<-read.csv('Seasonality Index.csv')
seas_data<- rbind(seas_data,seas_data,seas_data,seas_data,seas_data)

y<-ts(case_data["Case.Shipments"],frequency = 12)
y_tr <- window(y, start=c(2,1),end=c(4,12))
y_te <- window(y, start=c(5,1), end=c(5,12))


y_reg <-case_data[13:60,"Case.Shipments"]

case_data['lag.cp1']<-slide(case_data, Var = "Consumer.Packs", slideBy = -1,NewVar = 'lag.cp1')['lag.cp1
case_data['lag.cp2']<-slide(case_data, Var = "Consumer.Packs", slideBy = -2,NewVar = 'lag.cp2')['lag.cp2
case_data['lag.da1']<-slide(case_data, Var = "Dealer.Allowance", slideBy = -1,NewVar = 'lag.da1')['lag.d
case_data['lag.da2']<-slide(case_data, Var = "Dealer.Allowance", slideBy = -2,NewVar = 'lag.da2')['lag.d

X_tr <- as.matrix(case_data[13:48,3:8])
X_te <- as.matrix(case_data[49:60,3:8])
X<- as.matrix(case_data[13:60,3:8])

mod1_cv<-cv.glmnet(x = X,y = y_reg,nfolds = 10)
best_lam<-mod1_cv$lambda.min

mod1_cv.coef=predict(mod1_cv,type='coefficients',s=best_lam)
predict(mod1_cv,type='coefficients',s=best_lam)

## 7 x 1 sparse Matrix of class "dgCMatrix"
##                               1
## (Intercept)      2.993484e+05
## Consumer.Packs   5.749909e-01
## Dealer.Allowance 8.660401e-02
## lag.cp1         -1.891597e-01
## lag.cp2               .
## lag.da1          1.219152e-02
## lag.da2         -1.054635e-02
```

```
x_red<-X[,c('Consumer.Packs','Dealer.Allowance','lag.cp1','lag.da1','lag.da2')]
mod2_lm<-lm(y_reg~x_red)
summary(mod2_lm)
```

```
##
## Call:
## lm(formula = y_reg ~ x_red)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -148759  -28423     906   31731  185129
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)         2.976e+05  1.831e+04  16.256  < 2e-16 ***
## x_redConsumer.Packs 5.926e-01  6.979e-02   8.491 1.17e-10 ***
## x_redDealer.Allowance 8.835e-02 1.029e-02  8.585 8.66e-11 ***
## x_redlag.cp1        -1.966e-01  6.818e-02  -2.884  0.00617 **
## x_redlag.da1         1.387e-02  1.037e-02   1.337  0.18857
## x_redlag.da2        -1.261e-02  1.038e-02  -1.215  0.23126
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 61560 on 42 degrees of freedom
## Multiple R-squared:  0.7693, Adjusted R-squared:  0.7419
## F-statistic: 28.01 on 5 and 42 DF,  p-value: 2.284e-12
```
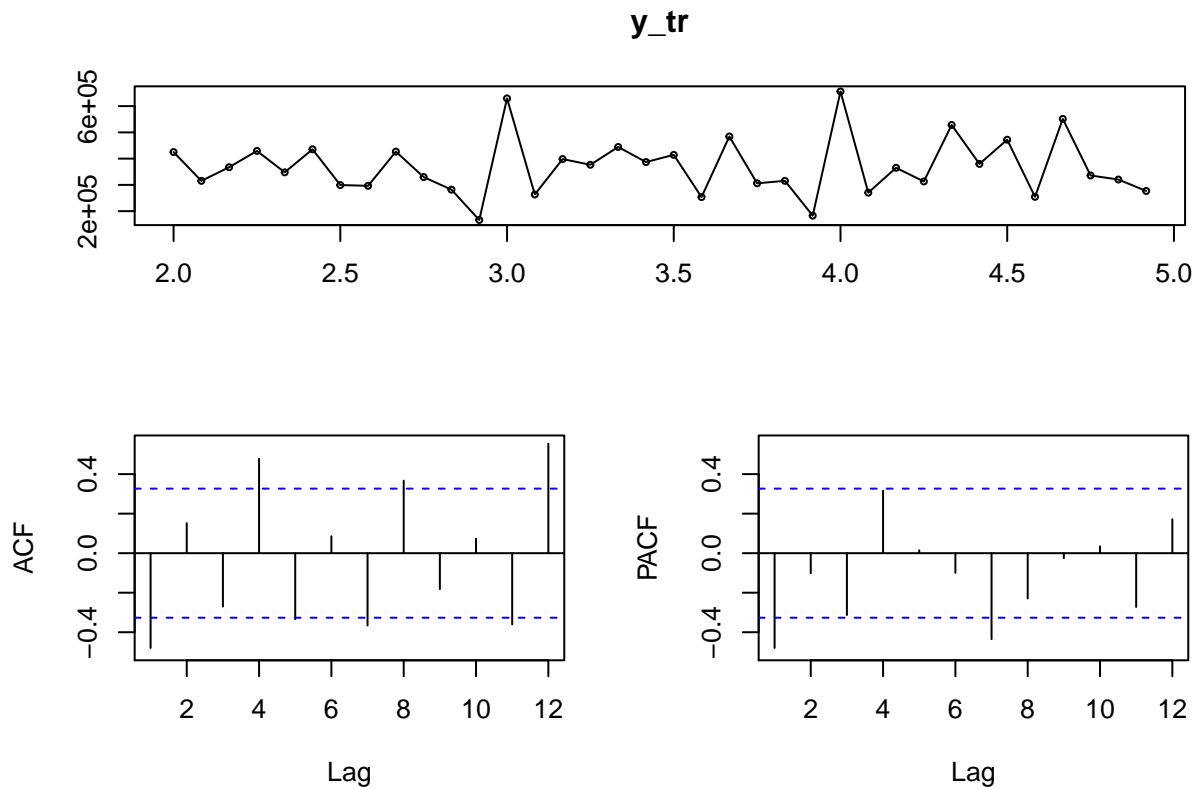
```
x_red_tr<-X_tr[,c('Consumer.Packs','Dealer.Allowance','lag.cp1')]
x_red_te<-X_te[,c('Consumer.Packs','Dealer.Allowance','lag.cp1')]
```

Analysis

Question 2

Figure 2.0

```
tsdisplay(y_tr)
```

y_tr


```r
mod3<-Arima(y_tr,order = c(1,2,2),seasonal = c(0,0,2),xreg = x_red_tr)
summary(mod3)
```

```
## Series: y_tr
## Regression with ARIMA(1,2,2)(0,0,2)[12] errors
##
## Coefficients:
##           ar1      ma1     ma2     sma1    sma2  Consumer.Packs
##       -0.6054  -1.2112  0.2112  1.5964  0.9998          0.4967
## s.e.   0.2182   0.3439  0.3400  1.4182  1.4501          0.0850
##       Dealer.Allowance  lag.cp1
##                 0.1428  -0.0322
## s.e.            0.0297   0.0599
##
## sigma^2 estimated as 1.249e+09:  log likelihood=-418.39
## AIC=854.78   AICc=862.28   BIC=868.52
##
## Training set error measures:
##                     ME      RMSE      MAE        MPE      MAPE       MASE
## Training set 2070.435 30038.34 21582.25 -0.8361139 6.096753 0.4143205
##                    ACF1
## Training set -0.03810906
```
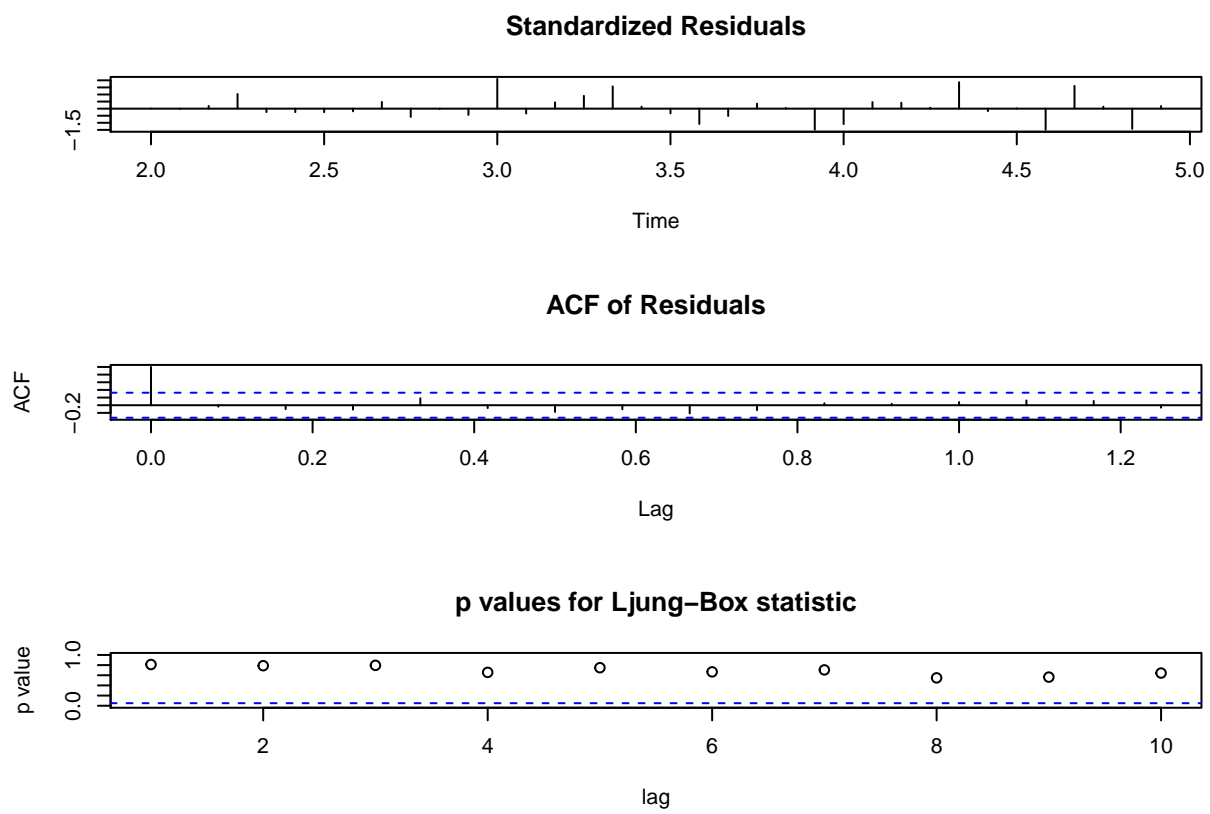
Figure 2.1

```r
tsdiag(mod3)
```

3

Standardized Residuals

ACF of Residuals

p values for Ljung–Box statistic

Figure 2.2

```r
tsdisplay(mod3$residuals)
```
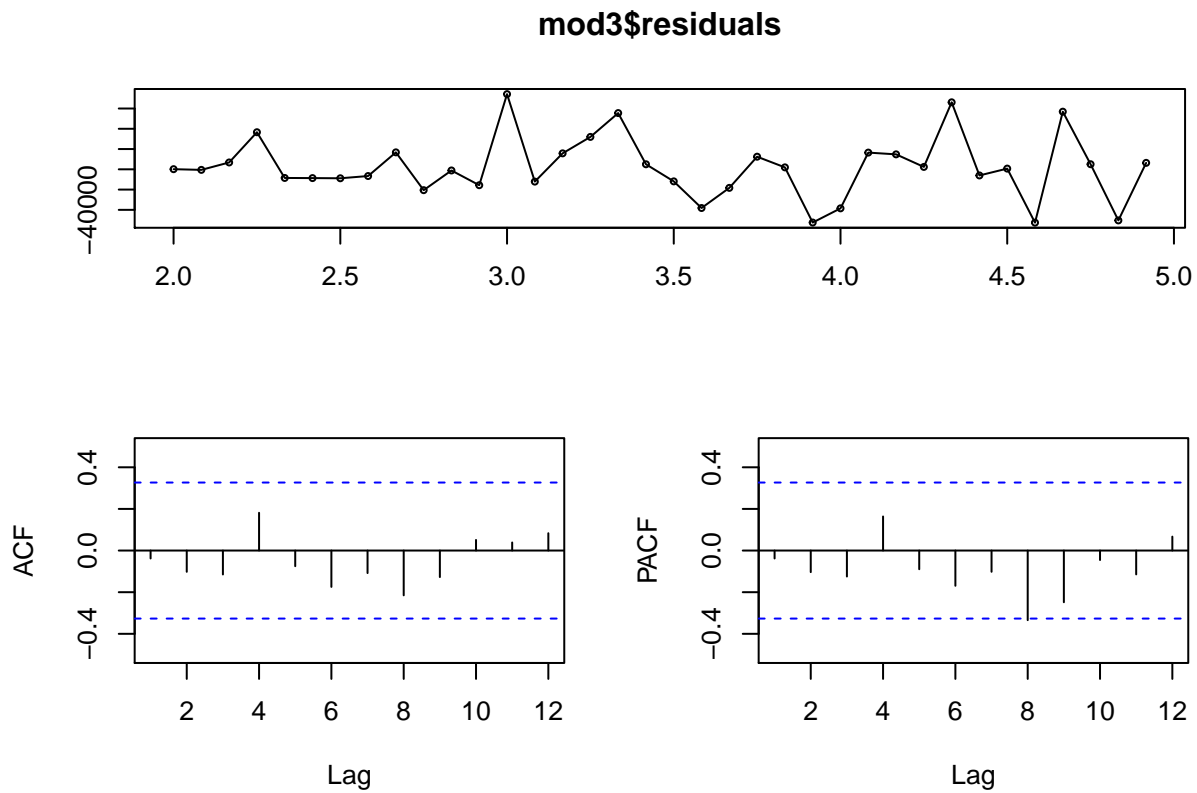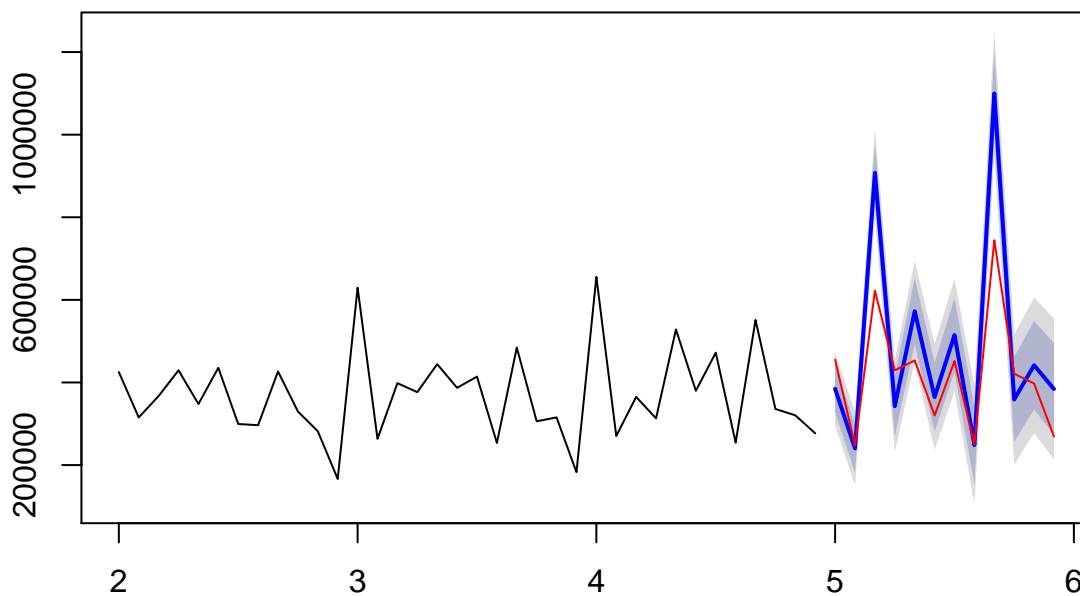
4

mod3$residuals

ACF

PACF

Lag

Lag

Figure 2.3

```
for_non_seas<-forecast(mod3,xreg =x_red_te,h=12 )
plot(for_non_seas)
lines(y_te,col='red')
```

Forecasts from Regression with ARIMA(1,2,2)(0,0,2)[12] errors

Analysis

Question 3

```r
seas_index<-seas_data['Seasonality.Index']/100
y_adj<-ts(case_data['Case.Shipments']/seas_index,frequency = 12)
y_adj1<-case_data['Case.Shipments']/seas_index

y_tr_adj <- window(y_adj, start=c(2,1),end=c(4,12))
y_te_adj <- window(y_adj, start=c(5,1), end=c(5,12))

y_reg_adj <-y_adj1[13:60,"Case.Shipments"]
```

```r
mod2_cv<-cv.glmnet(x = X,y = y_reg_adj,nfolds = 10)
best_lam2<-mod2_cv$lambda.min
```

```r
mod2_cv.coef=predict(mod2_cv,type='coefficients',s=best_lam2)
predict(mod2_cv,type='coefficients',s=best_lam2)
```

```
## 7 x 1 sparse Matrix of class "dgCMatrix"
##                             1
## (Intercept)       3.161237e+05
## Consumer.Packs    4.086228e-01
## Dealer.Allowance  6.937565e-02
## lag.cp1          -1.552403e-01
## lag.cp2           1.142473e-02
## lag.da1           1.034309e-02
## lag.da2          -9.743211e-03
```

```r
x_red<-X[,c('Consumer.Packs','Dealer.Allowance','lag.cp1','lag.cp2','lag.da1','lag.da2')]
mod3_lm<-lm(y_reg_adj~x_red)
summary(mod3_lm)
```

```
##
## Call:
## lm(formula = y_reg_adj ~ x_red)
##
## Residuals:
##     Min     1Q Median     3Q     Max
## -73660 -23447    -183  20783  94090
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)          3.141e+05  1.200e+04  26.178  < 2e-16 ***
## x_redConsumer.Packs  4.228e-01  4.199e-02  10.068 1.20e-12 ***
## x_redDealer.Allowance 7.044e-02  6.385e-03  11.032 7.61e-14 ***
## x_redlag.cp1        -1.605e-01  4.083e-02  -3.932 0.000317 ***
## x_redlag.cp2         1.799e-02  4.219e-02   0.426 0.671976
## x_redlag.da1         1.177e-02  6.270e-03   1.877 0.067696 .
## x_redlag.da2        -1.114e-02  6.308e-03  -1.766 0.084903 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 36790 on 41 degrees of freedom
## Multiple R-squared:  0.8496, Adjusted R-squared:  0.8276
```

```
## F-statistic: 38.61 on 6 and 41 DF,  p-value: 2.401e-15
x_red_tr<-X_tr[,c('Consumer.Packs','Dealer.Allowance','lag.cp1')]
x_red_te<-X_te[,c('Consumer.Packs','Dealer.Allowance','lag.cp1')]
```

Figure 3.0

```
tsdisplay(y_tr_adj)
```
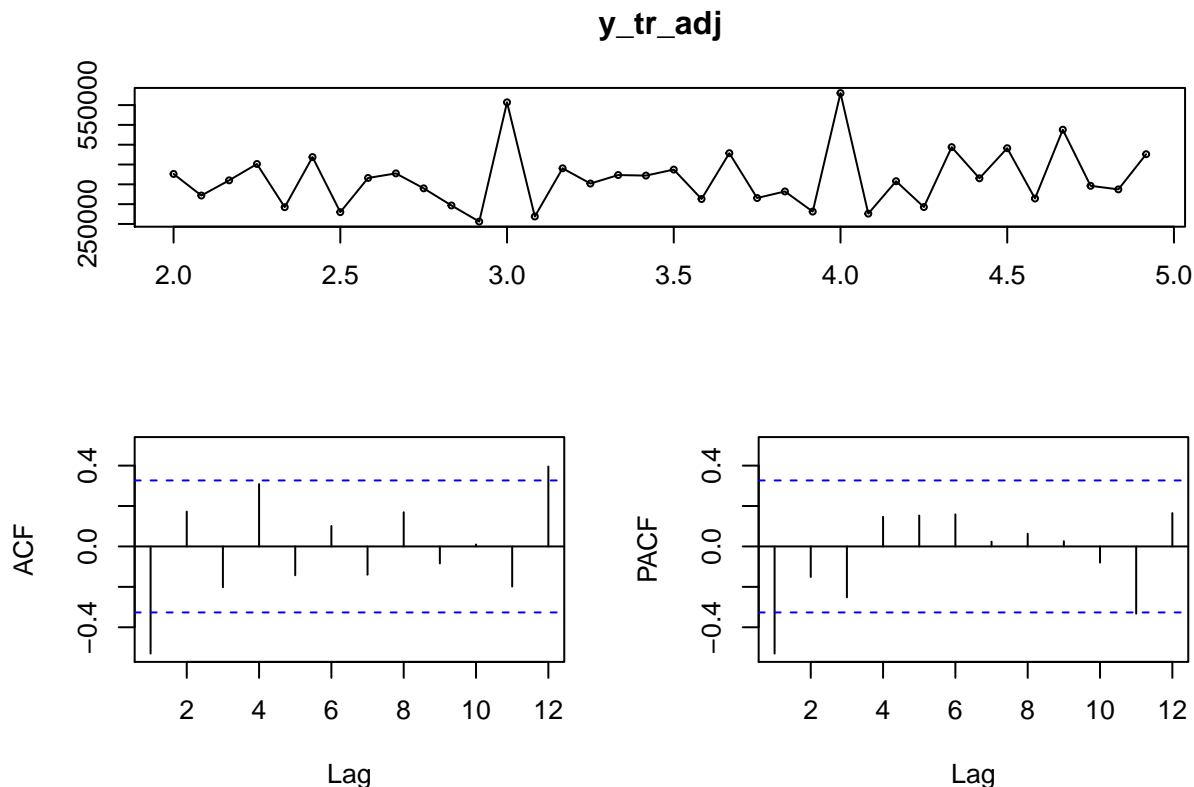


```
mod3<-Arima(y_tr_adj,order = c(2,2,2),seasonal = c(2,0,1),xreg = x_red_tr)
summary(mod3)
```

```
## Series: y_tr_adj
## Regression with ARIMA(2,2,2)(2,0,1)[12] errors
##
## Coefficients:
##           ar1     ar2      ma1     ma2     sar1     sar2    sma1
##       -0.3108  0.2118  -1.0829  0.0831  -0.4325  -0.9214  0.9552
## s.e.   0.7458  0.4013   0.7304  0.7204   0.1771   0.0620  4.0600
##       Consumer.Packs  Dealer.Allowance  lag.cp1
##               0.4391            0.1008  -0.1561
## s.e.          0.0231            0.0127   0.0200
##
## sigma^2 estimated as 188202179:  log likelihood=-399.44
## AIC=820.87    AICc=832.87    BIC=837.66
##
## Training set error measures:
##                        ME      RMSE       MAE        MPE       MAPE       MASE
```

```
## Training set -924.6516 11201.26 8970.035 -0.3582875 2.560787 0.1746097
##                                ACF1
## Training set -0.06095108
```
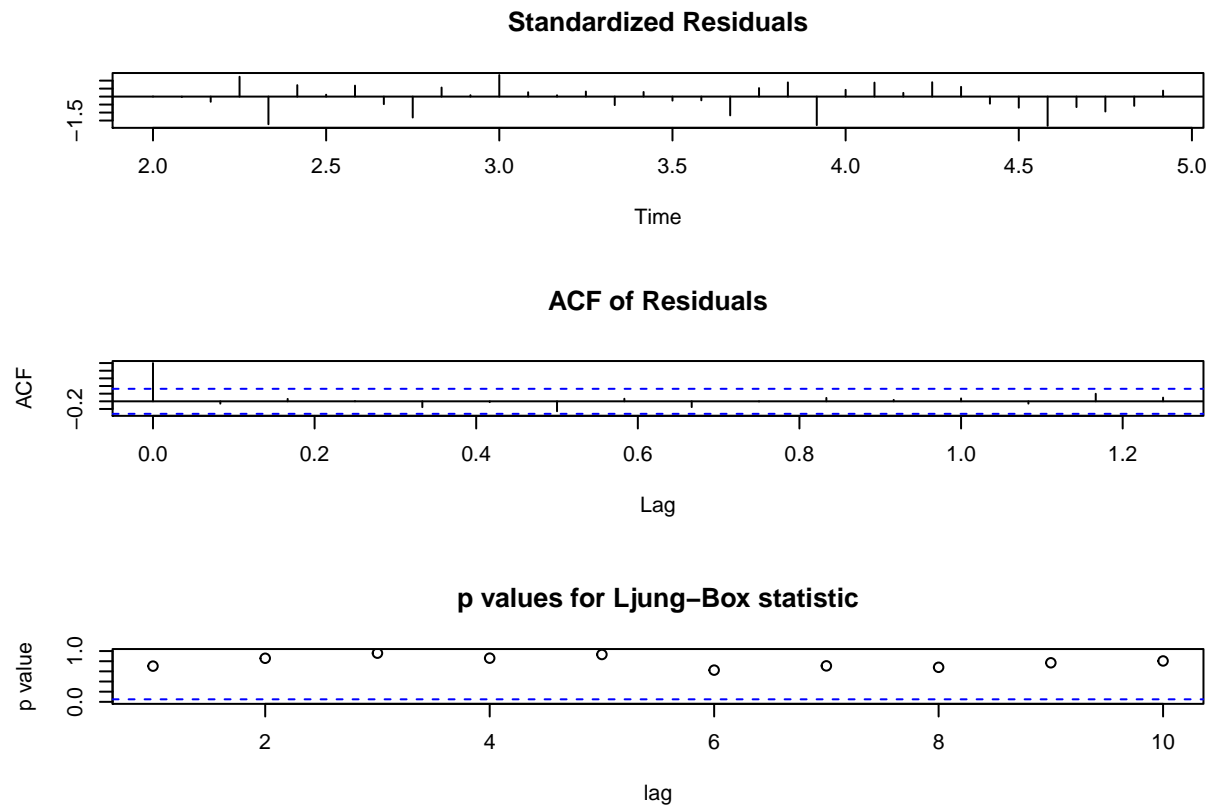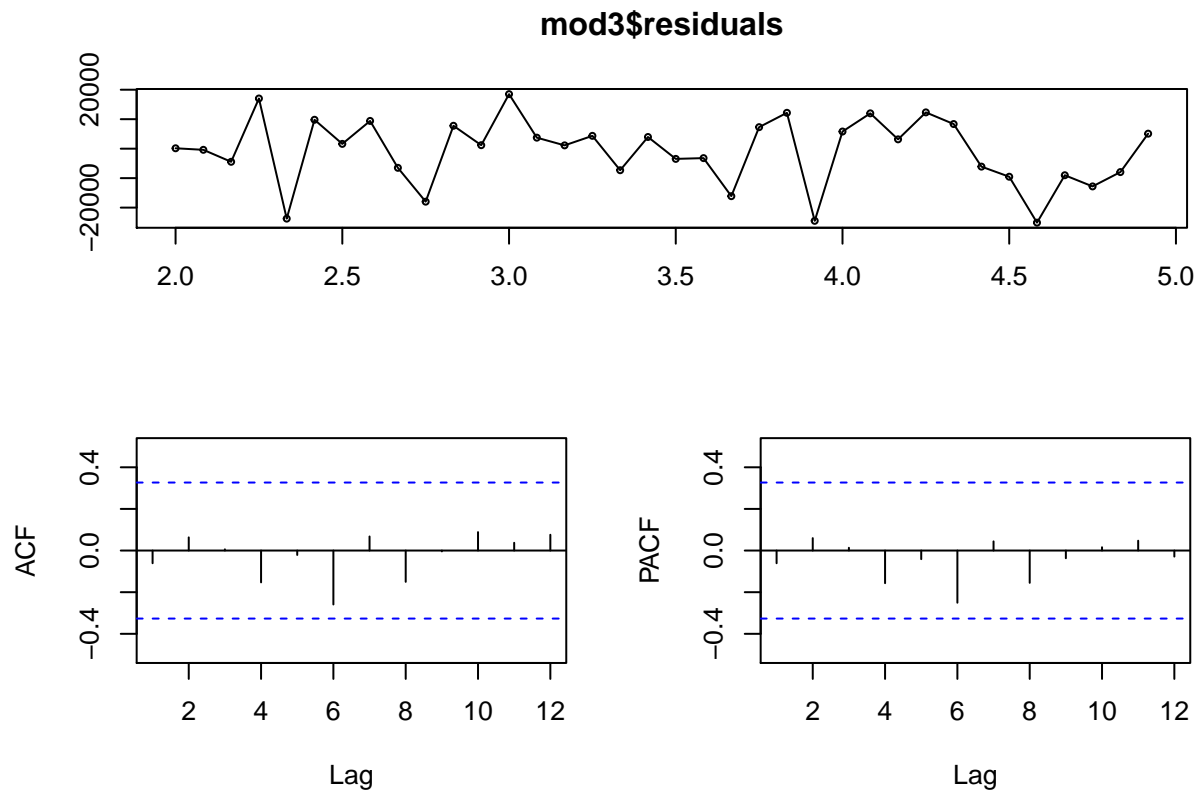
Figure 3.1

```r
tsdiag(mod3)
```

Standardized Residuals

ACF of Residuals

p values for Ljung–Box statistic

Figure 3.2

```r
tsdisplay(mod3$residuals)
```
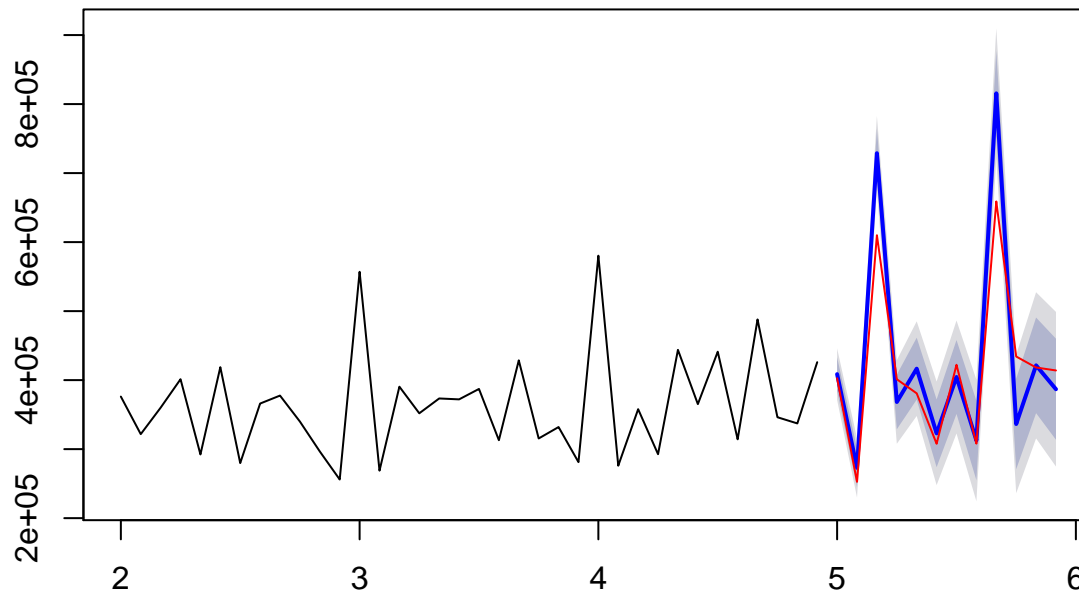
mod3$residuals


```
seas_index_12<-seas_data['Seasonality.Index'][49:60,]
```

Figure 3.2

```
for2<-forecast(mod3,xreg =x_red_te,h=12)
plot(for2)
lines(y_te_adj,col='red')
```

Forecasts from Regression with ARIMA(2,2,2)(2,0,1)[12] errors

Analysis

The third question model has slighlty better p values for the Ljung-Box statistic than the second question model, implying that the third model has a bit mor statistical validity. We see that the bias for the third model is more than half that of the second model in terms of MPE, and the MASE was .17 to the .41 MASE of the second model. Just looking at the RMSE difference where the third model has an RMSE about one third of the RMSE of the second model should indicate the third model is doing much better. All of the Information Criterion are also lower on the third model than the second model. The second model appears worse in every way.

Question 4

Analysis

Question 5

Analysis

- Starting with a lasso regression, none of the variables coefficients converged to zero. So to address their statistical significance they were then put into a linear regression. This showed that only the consumer packs, dealer allowances, and lagged consumer packs were statistically significant. These variables were then put into an arima model which tried to account for its auto correlated errors. The model produced very good forecasts compared to the previous model which wasnt scaled by the seasonal index. The coefficients on consumer packs was .439 and the lagged consumer packs was -.156, indicating a stock up in the subsequent month. The dealer allowance had a smaller yet positive coef. .1008, which didnt have a lagged effect.

- While consumer packs have a larger effect than dealer allowances, consumer packs also have significant stock ups. Given this information, dealer allowances would allow Harmon foods to more stably increase sales, without the harm of stock ups. If Harmon foods wanted to use consumer packs as an advertising

strategy they would need to be conscious of the stock ups and prepare their supply chain for lower demand in the subsequent month. I would recommend Dealer Allowances as a main strategy for Harmon Foods because of its predictable sales increase without the stock up risk.