Group Assingment

Reece Wooten, Kyle Katzen and Daxi Cheng

12/2/2017

Contents

Question 1(20 pts) Select variables in a parsimonious regression model to estimate the value of lagged

effects of "Consumer Packs" and "Dealer Allowances" on "Case Shipments" for Treat. Limit your analysis to lags of zero, one and two months.

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

```
y<-ts(case_data["Case.Shipments"])
y_tr <- window(y, start=c(13),end=c(48))
y_te <- window(y, start=c(49), end=c(60))
y_reg <-case_data[13:60,"Case.Shipments"]

case_data['lag.cp1']<-slide(case_data, Var = "Consumer.Packs", slideBy = -1,NewVar = 'lag.cp1')['lag.cp
```

```
##
## Remember to put case_data in time order before running.
```

```
##
## Lagging Consumer.Packs by 1 time units.
```

```
case_data['lag.cp2']<-slide(case_data, Var = "Consumer.Packs", slideBy = -2,NewVar = 'lag.cp2')['lag.cp
```

```
##
## Remember to put case_data in time order before running.
```

```
##
## Lagging Consumer.Packs by 2 time units.
```

```
case_data['lag.da1']<-slide(case_data, Var = "Dealer.Allowance", slideBy = -1,NewVar = 'lag.da1')['lag.
```

```
##
## Remember to put case_data in time order before running.
```

```
##
## Lagging Dealer.Allowance by 1 time units.
```

```
case_data['lag.da2']<-slide(case_data, Var = "Dealer.Allowance", slideBy = -2,NewVar = 'lag.da2')['lag.
```

```
##
## Remember to put case_data in time order before running.
```

```
##
## Lagging Dealer.Allowance by 2 time units.
```

```r
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])
```

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

```r
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)       3.001182e+05
## Consumer.Packs    5.670763e-01
## Dealer.Allowance  8.581826e-02
## lag.cp1          -1.857882e-01
## lag.cp2           .
## lag.da1           1.143673e-02
## lag.da2          -9.616145e-03
```

```r
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
```

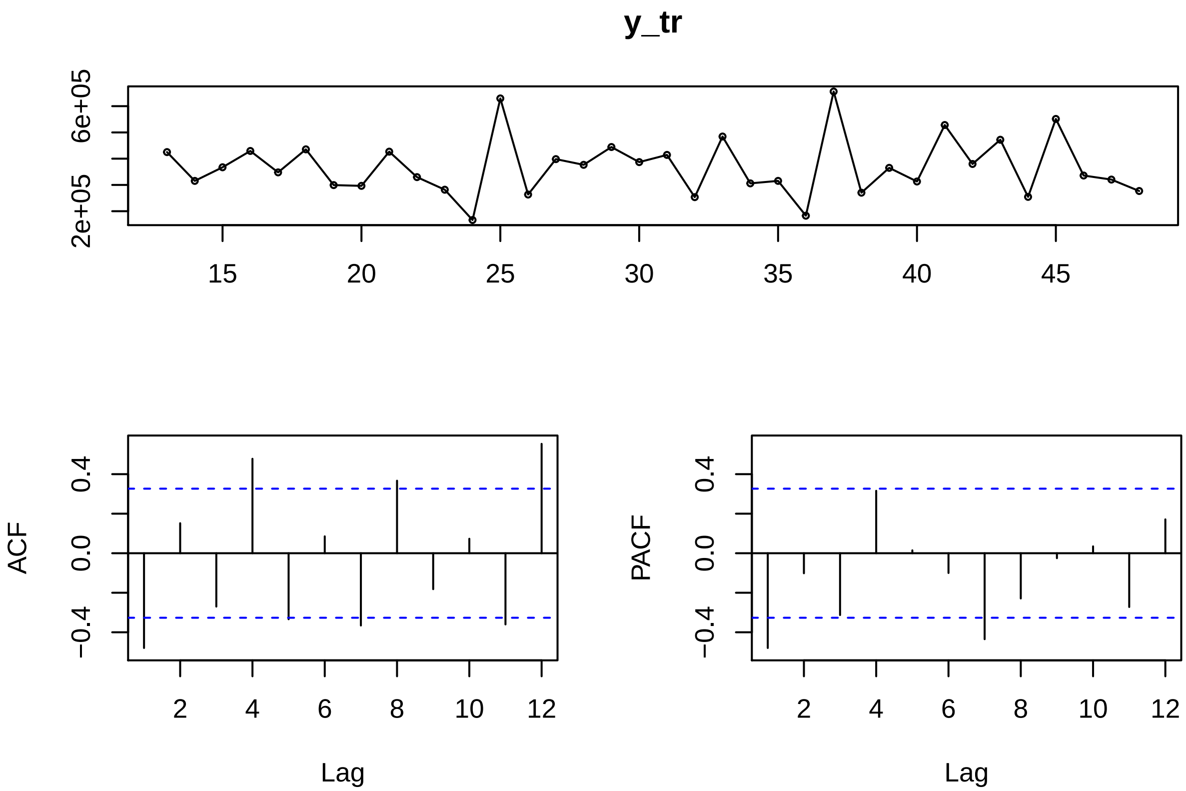
```r
x_red_tr<-X_tr[,c('Consumer.Packs','Dealer.Allowance','lag.cp1')]
x_red_te<-X_te[,c('Consumer.Packs','Dealer.Allowance','lag.cp1')]
```

Question 2 (20 pts) Using only the variables that you identified in the question above, develop and justify a

regression model with ARIMA errors to forecast monthly demand for treat twelve months into the future. To accomplish this make the following assumptions:

Figure 2.0

```r
tsdisplay(y_tr)
```




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

```
## Series: y_tr
## Regression with ARIMA(1,2,2) errors
##
## Coefficients:
##           ar1      ma1      ma2  Consumer.Packs  Dealer.Allowance  lag.cp1
##       -0.9234  -0.4150  -0.5850          0.6577            0.1563  -0.0226
## s.e.   0.0886   0.2258   0.2148          0.0527            0.0334   0.0496
##
## sigma^2 estimated as 3.896e+09:  log likelihood=-422.72
## AIC=859.43   AICc=863.74   BIC=870.12
##
## Training set error measures:
##                    ME    RMSE      MAE        MPE     MAPE      MASE
## Training set 4456.495 55047.3 44286.31 -0.8617088 13.26598 0.3086417
##                      ACF1
```

```
## Training set -0.001466898
```

Figure 2.1

```r
tsdiag(mod3)
```

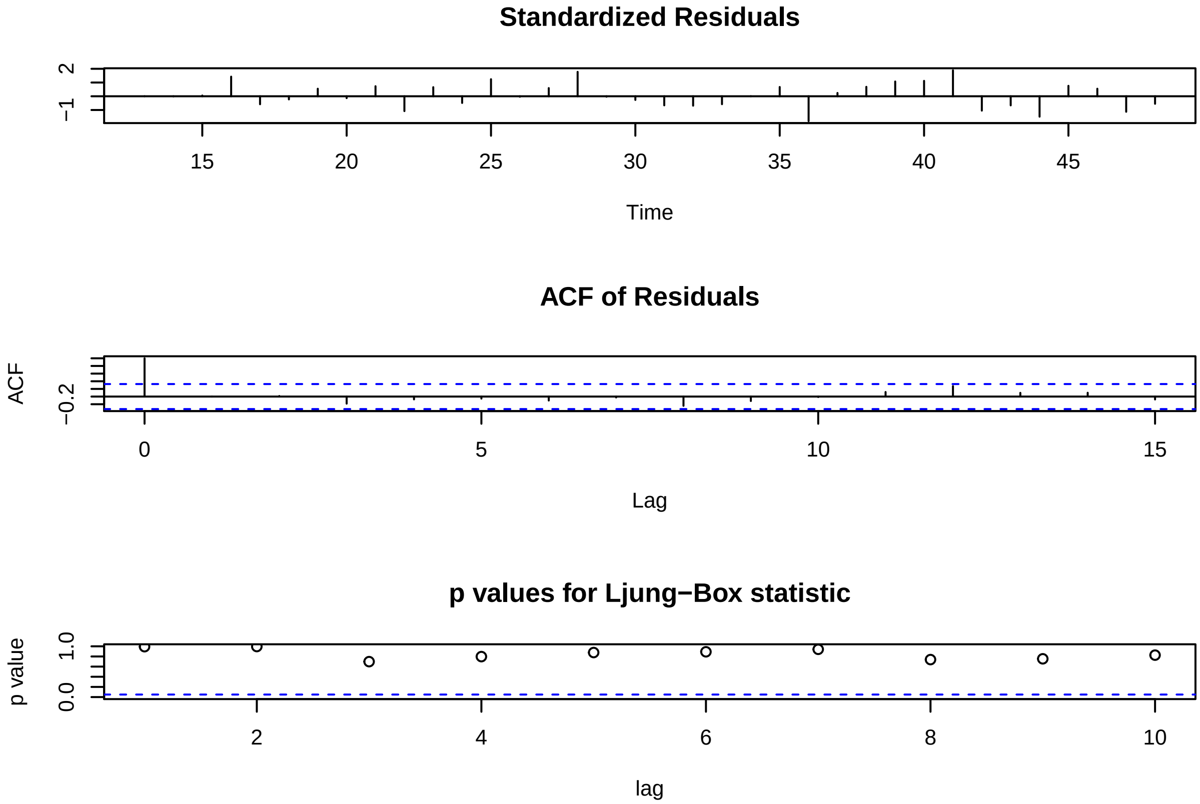


Figure 2.2

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

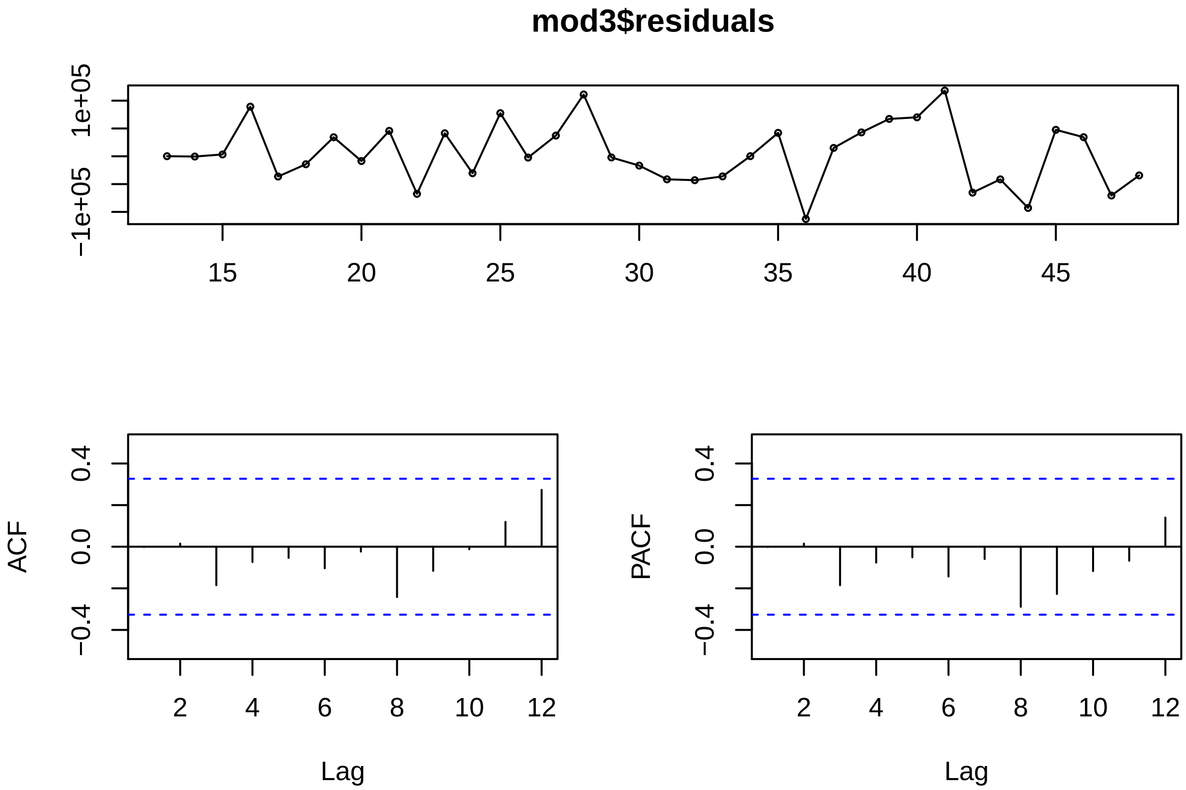
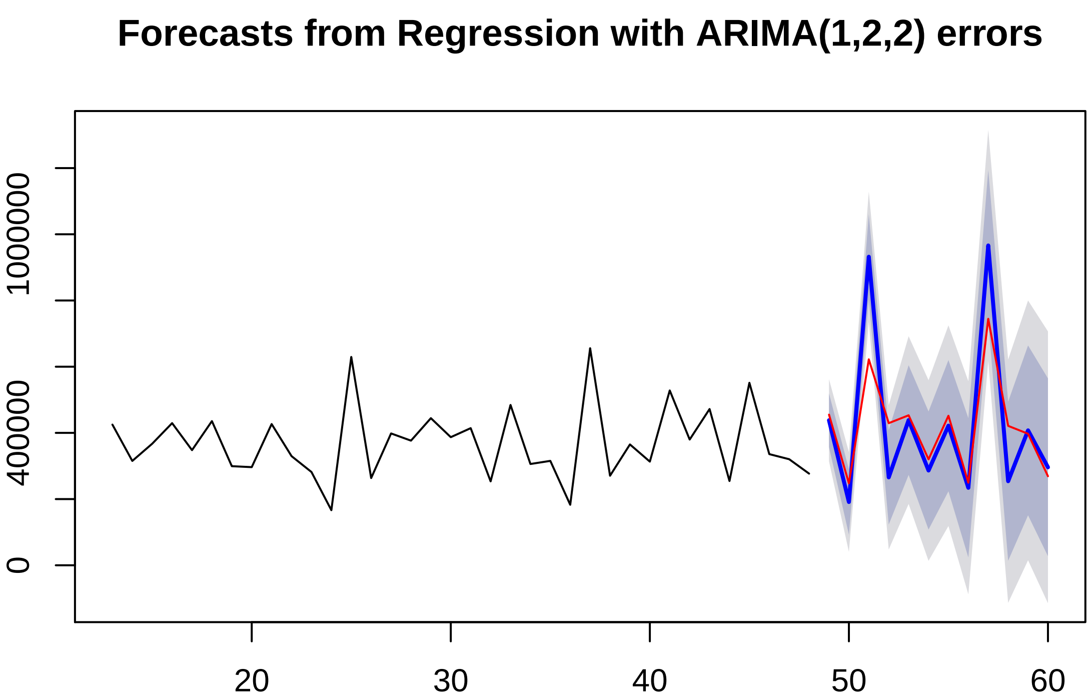


Figure 2.3

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



Question 3

```
X0=seas_data['Seasonality.Index']
X1 <- rbind(X0,X0)
X1 <- rbind(X1,X0)
X1 <- rbind(X1,X0)
X1 <- rbind(X1,X0)
y_adj1<-case_data['Case.Shipments']/X1
y_adj<-ts(case_data['Case.Shipments']/X1)
y_tr_adj <- window(y_adj, start=c(13),end=c(48))
y_te_adj <- window(y_adj, start=c(49), end=c(60))
y_reg_adj <-y_adj1[13:60,"Case.Shipments"]
```

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

```
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)       3.152382e+03
## Consumer.Packs    4.146839e-03
## Dealer.Allowance  6.983026e-04
## lag.cp1          -1.575064e-03
## lag.cp2           1.423401e-04
## lag.da1           1.095228e-04
## lag.da2          -1.034004e-04
```

```
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)
```

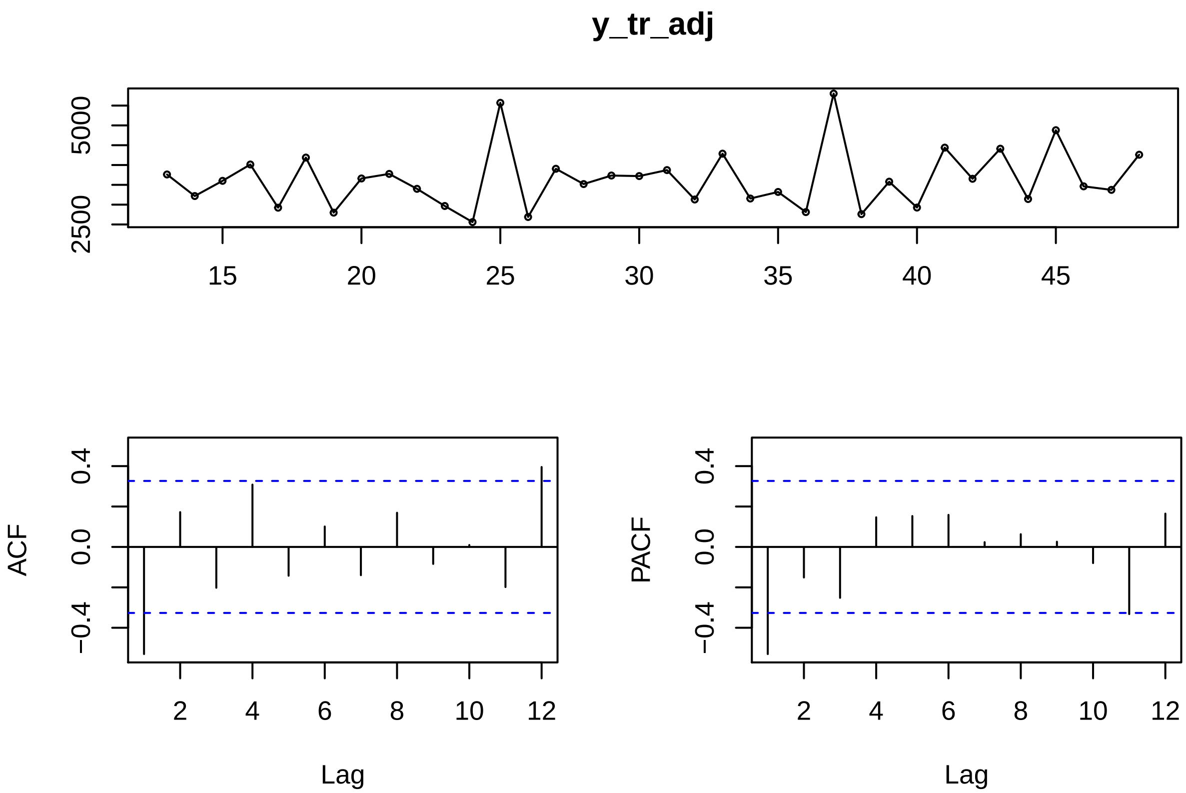
```
##
## Call:
## lm(formula = y_reg_adj ~ x_red)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -736.60 -234.47   -1.83  207.83  940.90
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)             3.141e+03  1.200e+02  26.178  < 2e-16 ***
## x_redConsumer.Packs     4.228e-03  4.199e-04  10.068 1.20e-12 ***
## x_redDealer.Allowance   7.044e-04  6.385e-05  11.032 7.61e-14 ***
## x_redlag.cp1           -1.605e-03  4.083e-04  -3.932 0.000317 ***
## x_redlag.cp2            1.799e-04  4.219e-04   0.426 0.671976
## x_redlag.da1            1.177e-04  6.270e-05   1.877 0.067696 .
## x_redlag.da2           -1.114e-04  6.308e-05  -1.766 0.084903 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 367.9 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
```

```r
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

```r
tsdisplay(y_tr_adj)
```




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

```
## Series: y_tr_adj
## Regression with ARIMA(2,2,1) errors
##
## Coefficients:
##           ar1      ar2     ma1  Consumer.Packs  Dealer.Allowance  lag.cp1
##       -0.8313  -0.1126  -1.000          0.0041             1e-03  -0.0012
## s.e.   0.1857   0.1812   0.099          0.0004             3e-04   0.0004
##
## sigma^2 estimated as 137170:  log likelihood=-248.88
## AIC=511.77   AICc=516.07   BIC=522.45
##
## Training set error measures:
##                    ME     RMSE      MAE       MPE     MAPE      MASE
## Training set 33.12051 326.6314 230.7431 0.3435914 6.359915 0.2319289
##                     ACF1
## Training set -0.04948407
```

Figure 3.1

```r
tsdiag(mod3)
```

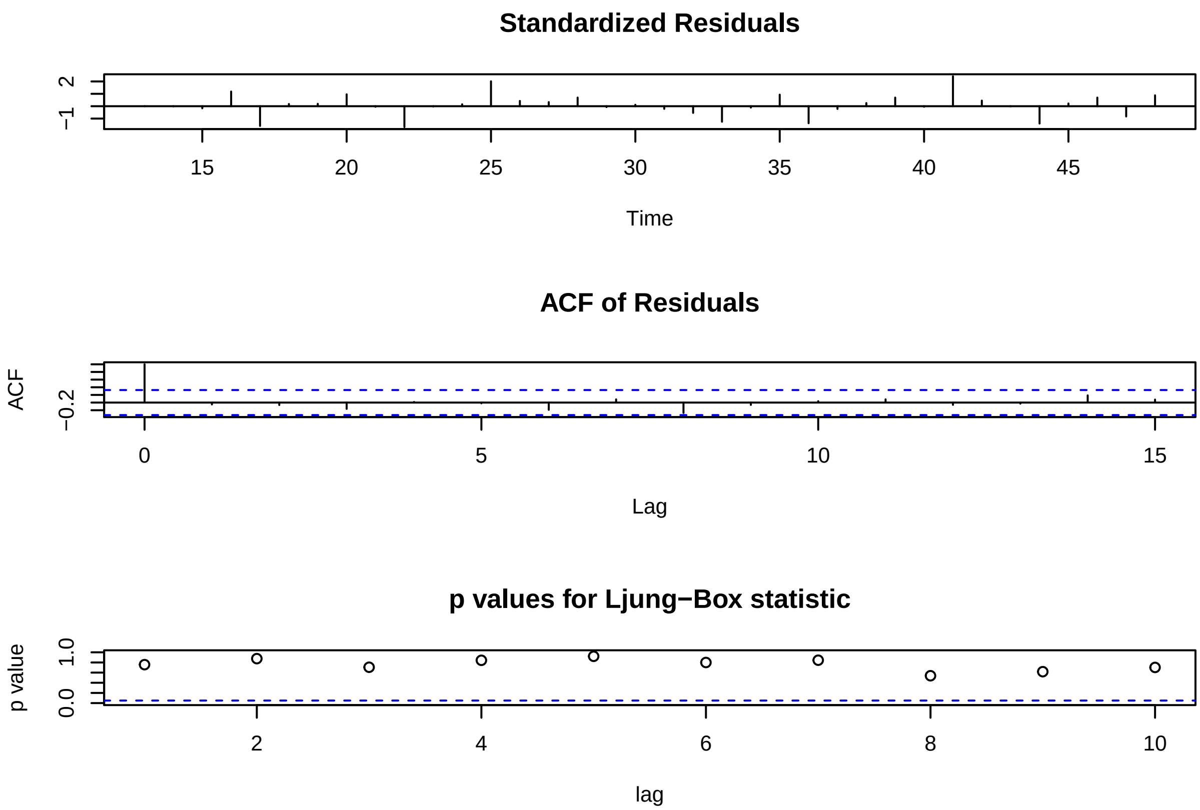
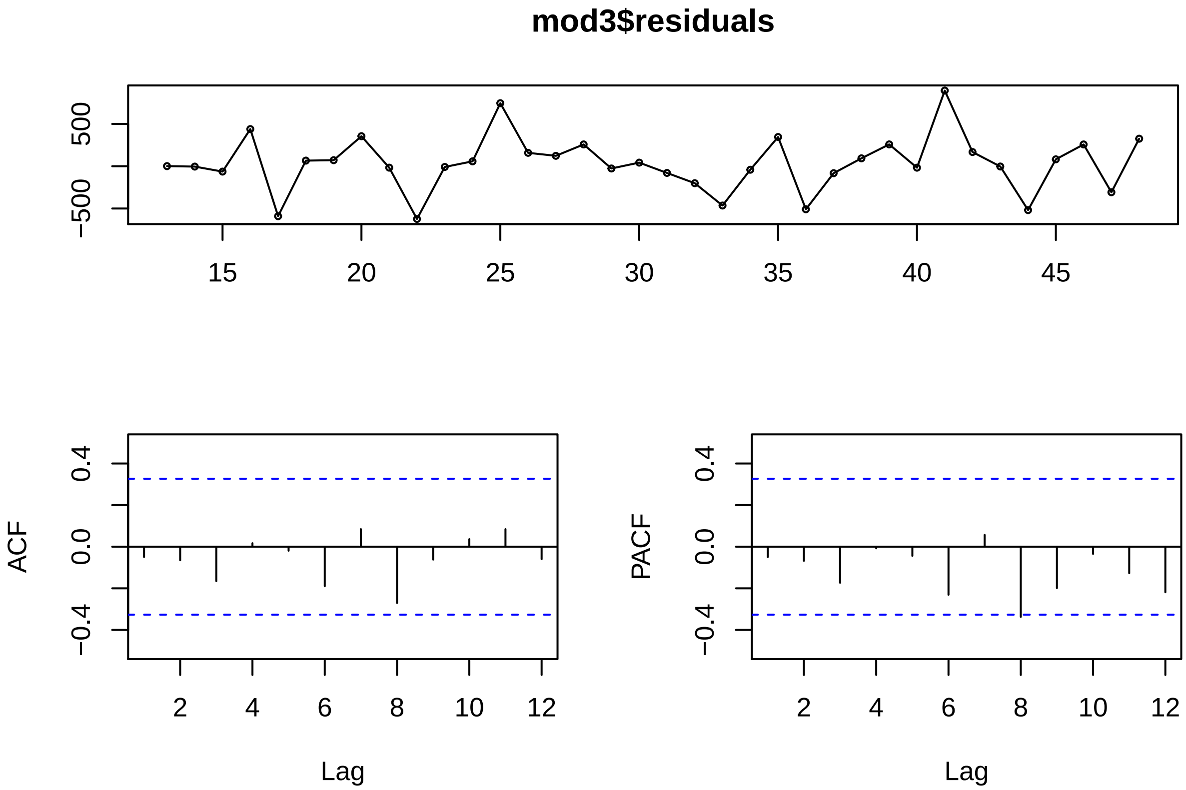


Figure 3.2

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

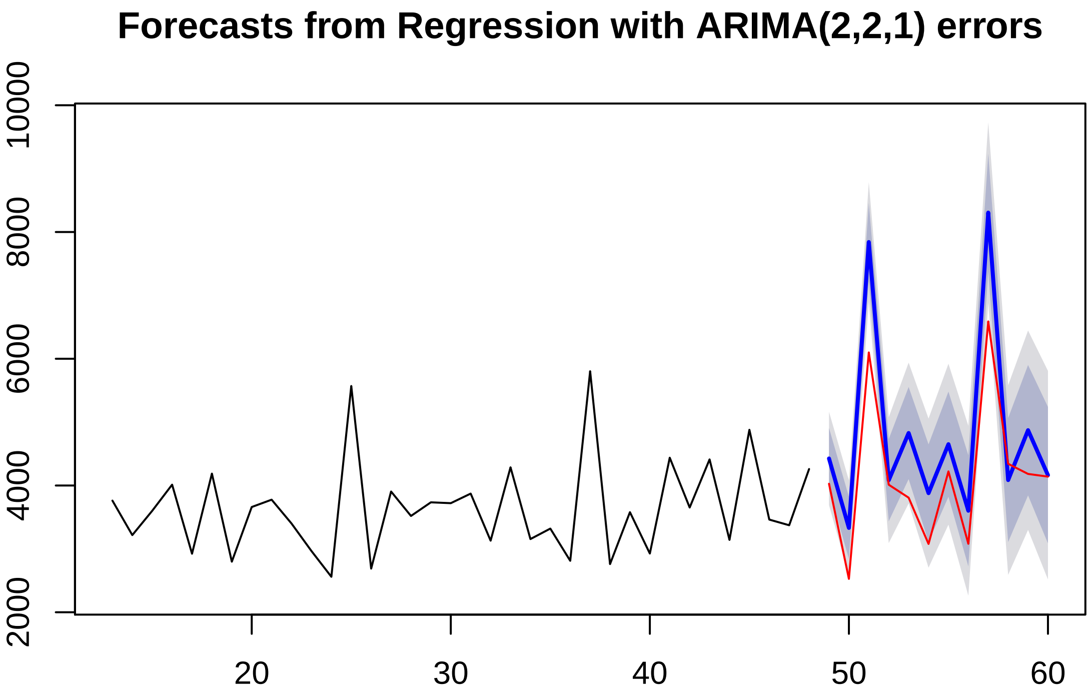



```
seas_data['Seasonality.Index'][49:60,]
```

```
## [1] NA NA NA NA NA NA NA NA NA NA NA NA
```

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(2,2,1) errors
2000
4000
6000
8000
10000
20
30
40
50
60