

Code Appendix

Victor Huang, Parker Johnson, and Elijah Jones

6/6/2022

Code Appendix

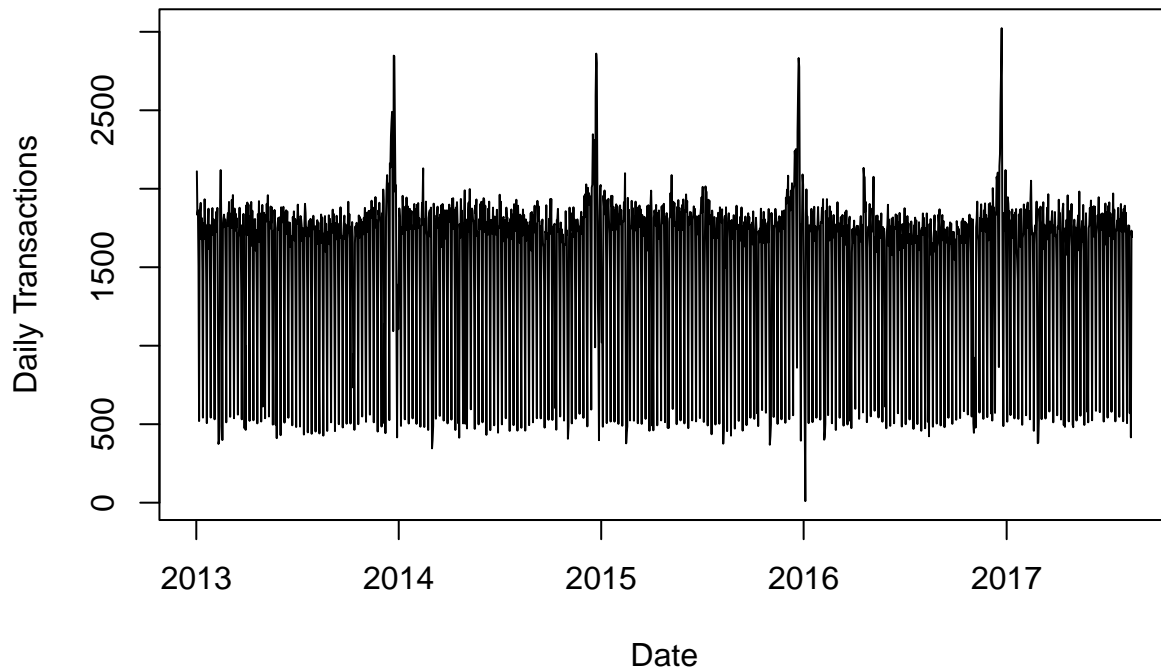
Store 1

```
store_1 <- transactions %>%  
  filter(store_nbr == 1) %>%  
  select(date, transactions) %>%  
  mutate(isSaturday = ifelse(wday(date) %in% 7, "Yes", "No")) %>%  
  mutate(isSunday = ifelse(wday(date) %in% 1, "Yes", "No"))
```

Initial plot of store 1 transactions

```
plot(x = store_1$date, y = store_1$transactions, type = "l", xlab = "Date",  
     ylab = "Daily Transactions", main = "Transactions for Store 1 in Quito vs. Date")
```

Transactions for Store 1 in Quito vs. Date

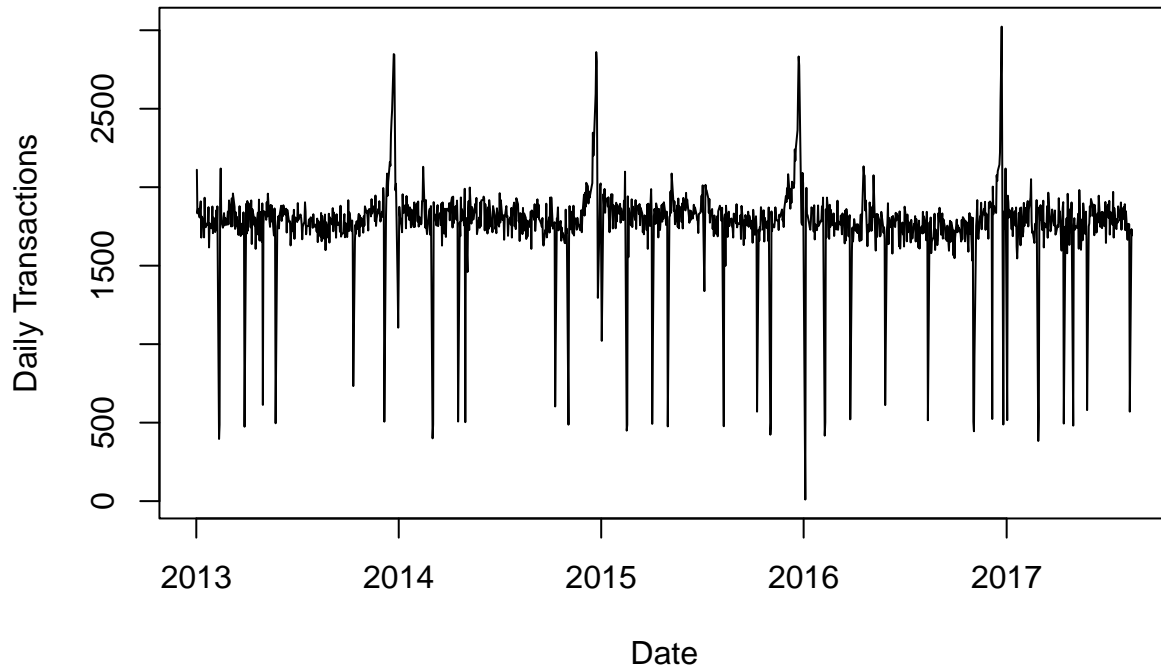


Plot of store 1 transactions with weekends temporarily removed

```
store_1_testing <- store_1 %>%  
  filter(isSaturday == "No" & isSunday == "No")
```

```
plot(x = store_1_testing$date, y = store_1_testing$transactions, type = "l",
     xlab = "Date", ylab = "Daily Transactions",
     main = "Weekday Transactions for Store 1 in Quito vs. Date")
```

Weekday Transactions for Store 1 in Quito vs. Date



Adding factor variables

```
store_1 <- transactions %>%
  filter(store_nbr == 1) %>%
  select(date, transactions) %>%
  mutate(isSaturday = ifelse(wday(date) %in% 7, "Yes", "No")) %>%
  mutate(isSunday = ifelse(wday(date) %in% 1, "Yes", "No"))

store_1_holidays1 <- store_1 %>%
  filter(isSaturday == "No" & isSunday == "No") %>%
  filter(transactions < 1000) %>% #This was done to catch most of the holidays
  filter(date != "2016-01-04" & date != "2016-11-04" & date != "2016-12-06" &
         date != "2017-01-02") #Removing the non-holiday dates

store_1_holidays2 <- store_1 %>%
  filter(date == "2013-08-10" | date == "2013-11-01" | date == "2013-11-02" |
         date == "2013-11-04" | date == "2014-05-24" | date == "2014-08-10" |
         date == "2014-11-02" | date == "2015-05-24" | date == "2016-05-01" |
         date == "2016-10-09") #The rest of the holidays

store_1_holidays <- rbind(store_1_holidays1, store_1_holidays2) %>%
  arrange(date) #Merging the datasets into 1 that has all of the holidays

store_1_december <- store_1 %>%
  filter(date == "2013-12-18" | date == "2013-12-19" | date == "2013-12-20" |
         date == "2013-12-21" | date == "2013-12-22" | date == "2013-12-23" |
```

```

      date == "2013-12-24" | date == "2014-12-18" | date == "2014-12-19" |
      date == "2014-12-20" | date == "2014-12-21" | date == "2014-12-22" |
      date == "2014-12-23" | date == "2014-12-24" | date == "2015-12-18" |
      date == "2015-12-19" | date == "2015-12-20" | date == "2015-12-21" |
      date == "2015-12-22" | date == "2015-12-23" | date == "2015-12-24" |
      date == "2016-12-18" | date == "2016-12-19" | date == "2016-12-20" |
      date == "2016-12-21" | date == "2016-12-22" | date == "2016-12-23" |
      date == "2016-12-24") #December dates

store_1 <- store_1 %>%
  mutate(isHoliday = ifelse(date %in% store_1_holidays$date, "Yes", "No")) %>%
  mutate(isChristmasWeek = ifelse(date %in% store_1_december$date, "Yes", "No"))

store_1 <- store_1 %>% #Turning all variables into factors
  mutate(decimal_date = decimal_date(date)) %>%
  mutate(isSaturday = ifelse(isSaturday == "Yes", 1, 0)) %>%
  mutate(isSunday = ifelse(isSunday == "Yes", 1, 0)) %>%
  mutate(isHoliday = ifelse(isHoliday == "Yes", 1, 0)) %>%
  mutate(isChristmasWeek = ifelse(isChristmasWeek == "Yes", 1, 0))

```

Creating the OLS model for store 1

```

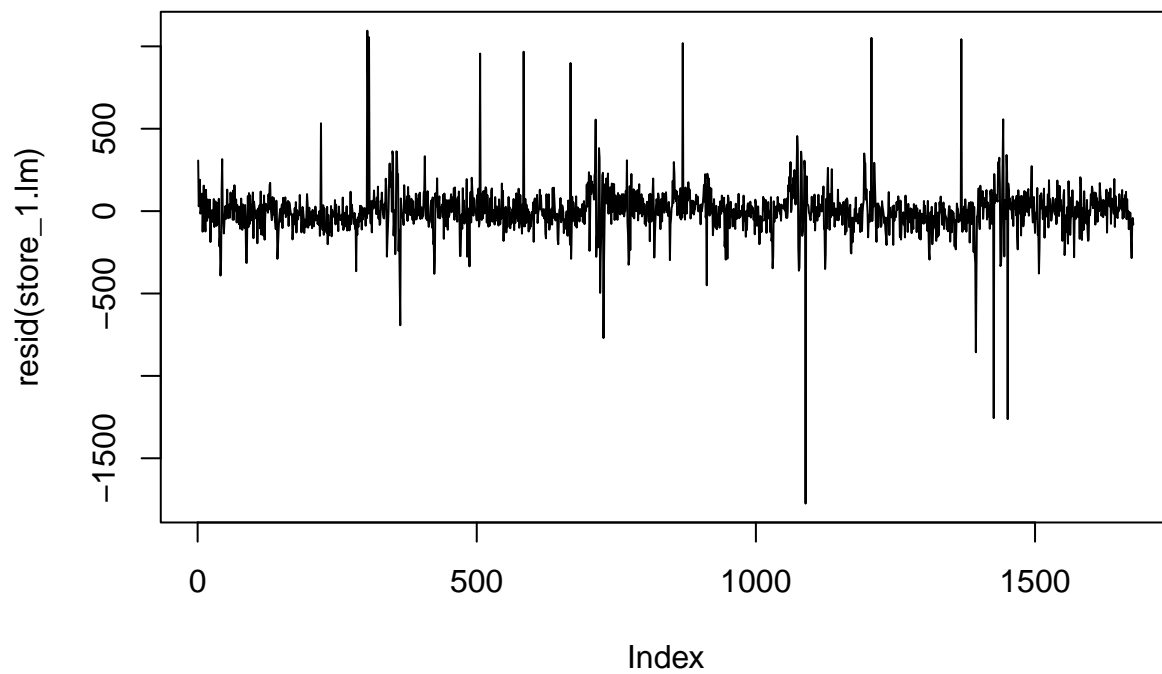
store_1.lm <- lm(transactions ~ decimal_date + isSaturday + isSunday +
                 isHoliday + isChristmasWeek, data = store_1)
summary(store_1.lm)

```

```

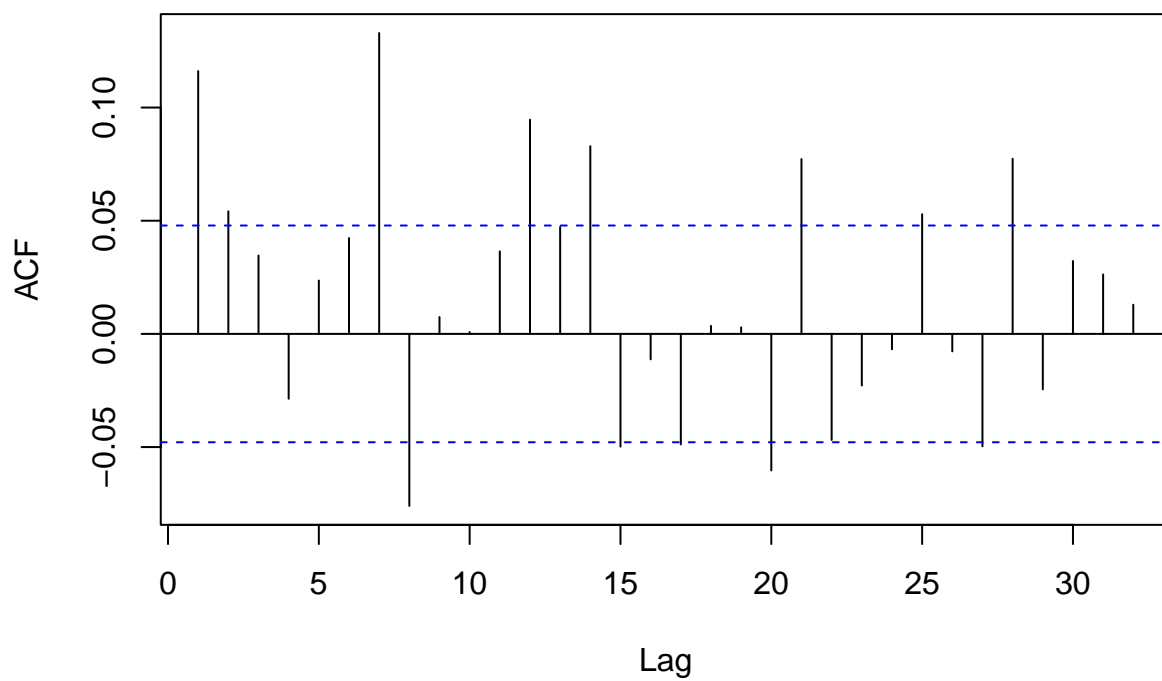
##
## Call:
## lm(formula = transactions ~ decimal_date + isSaturday + isSunday +
##     isHoliday + isChristmasWeek, data = store_1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1774.87   -55.86    -2.37    59.22   1095.06
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   14613.582   5520.868     2.647  0.0082 **
## decimal_date     -6.363     2.739    -2.323  0.0203 *
## isSaturday    -485.620    10.597   -45.826 <2e-16 ***
## isSunday     -1268.512    10.628  -119.353 <2e-16 ***
## isHoliday    -1015.773    22.651   -44.844 <2e-16 ***
## isChristmasWeek  687.081    28.545    24.070 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 149.7 on 1670 degrees of freedom
## Multiple R-squared:  0.9119, Adjusted R-squared:  0.9116
## F-statistic: 3455 on 5 and 1670 DF, p-value: < 2.2e-16
plot(resid(store_1.lm), type = 'l')

```



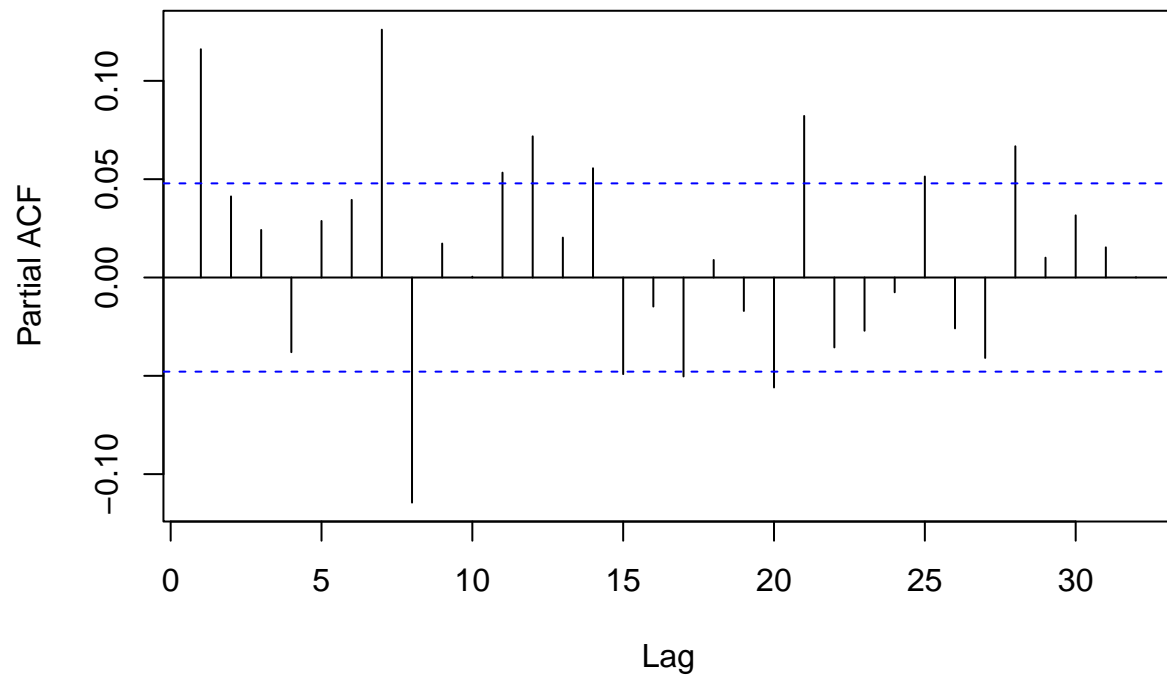
```
acf(resid(store_1.lm))
```

Series resid(store_1.lm)



```
pacf(resid(store_1.lm))
```

Series resid(store_1.lm)



Model Testing for store 1

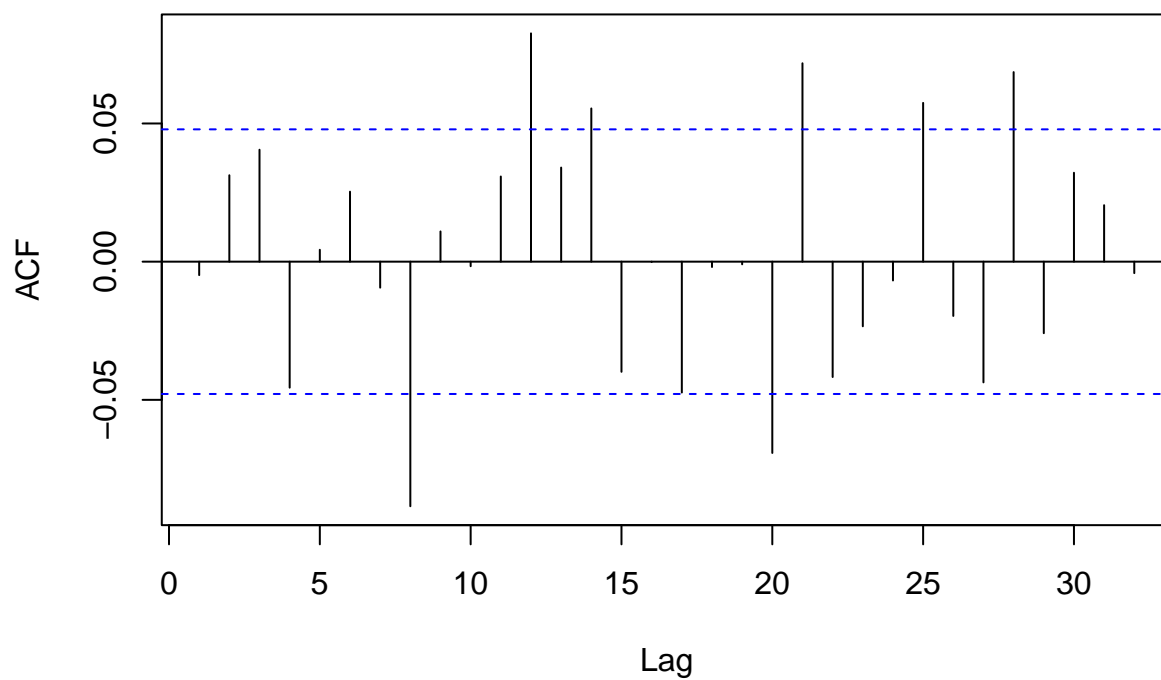
```
store_1.arima <- arima(store_1$transactions, order = c(1, 0, 0),  
  seasonal = list(order = c(1, 0, 0), period = 7),  
  xreg = cbind(store_1$isSaturday,  
    store_1$isSunday,  
    store_1$isHoliday,  
    store_1$isChristmasWeek,  
    store_1$decimal_date))
```

```
AIC(store_1.arima)
```

```
## [1] 21500.53
```

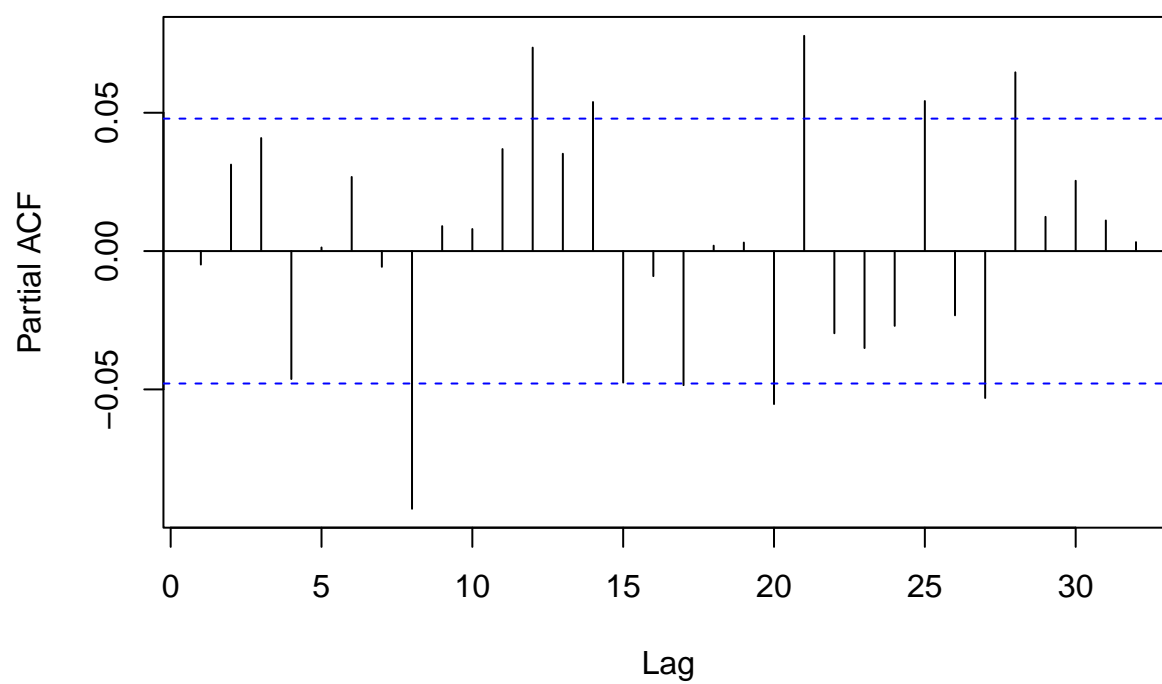
```
acf(resid(store_1.arima))
```

Series resid(store_1.arima)



```
pacf(resid(store_1.arima))
```

Series resid(store_1.arima)



```
#LjungBoxPlot(store_1.arima)
```

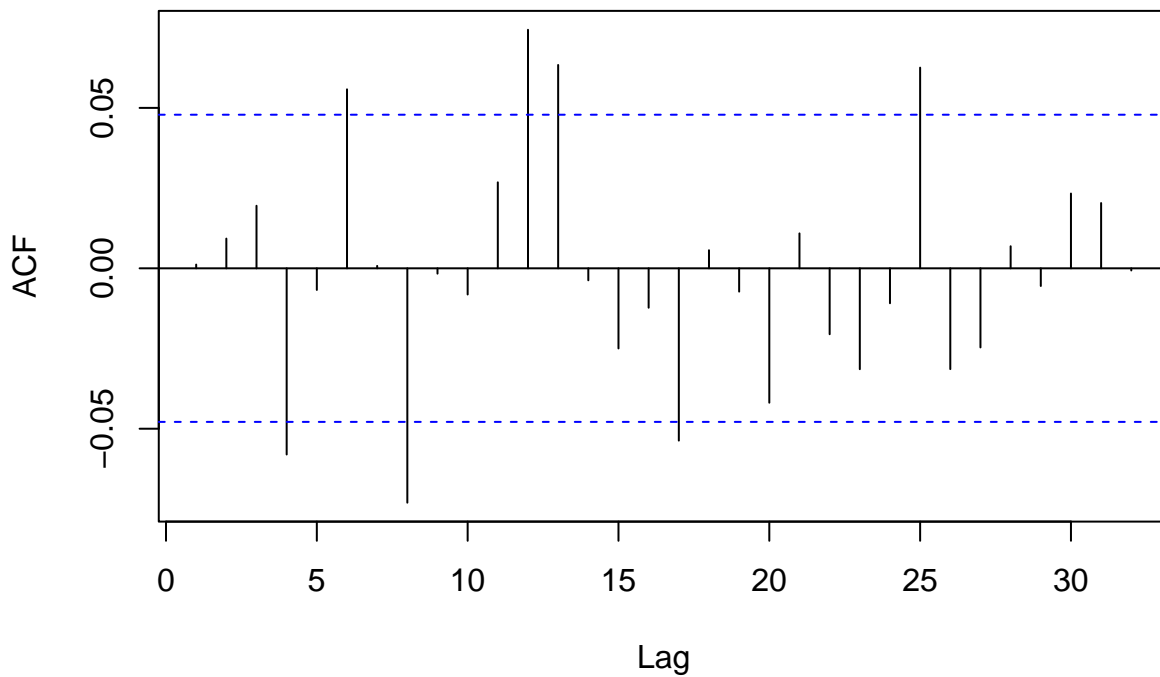
```
store_1.arma <- arima(store_1$transactions, order = c(1, 0, 2),  
  seasonal = list(order = c(1, 0, 2), period = 7),  
  xreg = cbind(store_1$isSaturday,  
    store_1$isSunday,  
    store_1$isHoliday,  
    store_1$isChristmasWeek,  
    store_1$decimal_date))
```

```
AIC(store_1.arma)
```

```
## [1] 21458.1
```

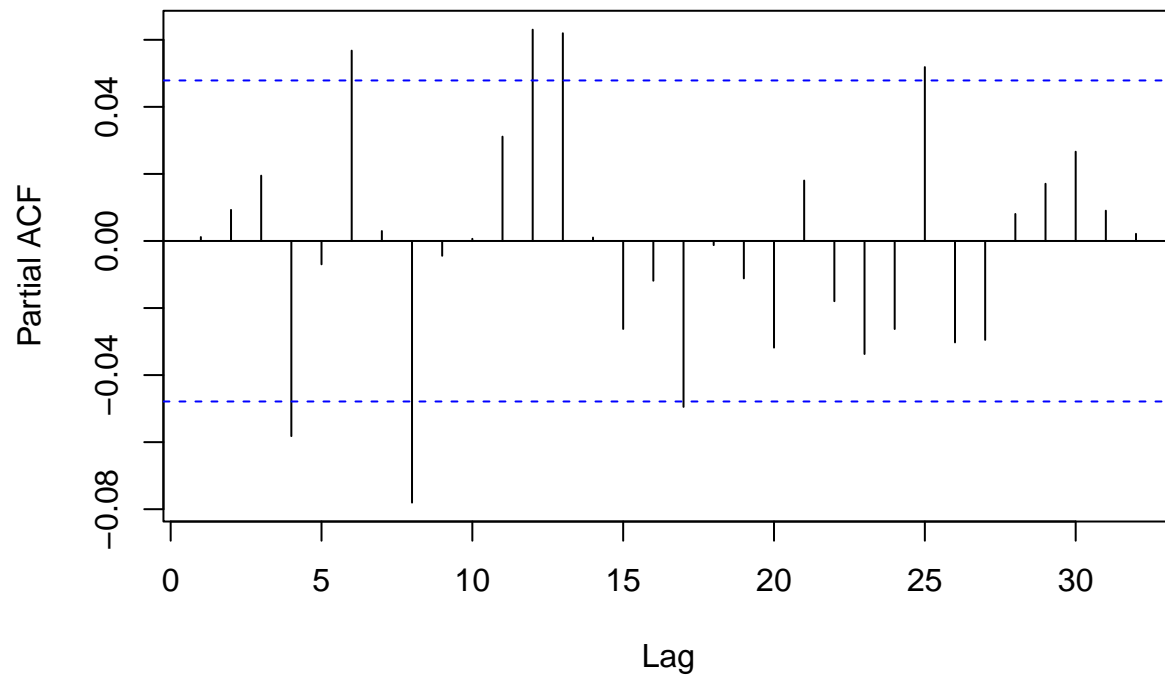
```
acf(resid(store_1.arma))
```

Series resid(store_1.arma)



```
pacf(resid(store_1.arma))
```

Series resid(store_1.arima)



```
#LjungBoxPlot(store_1.arima)
```

Final ARMA Modeling for store 1, and the ACF and PACF of the model

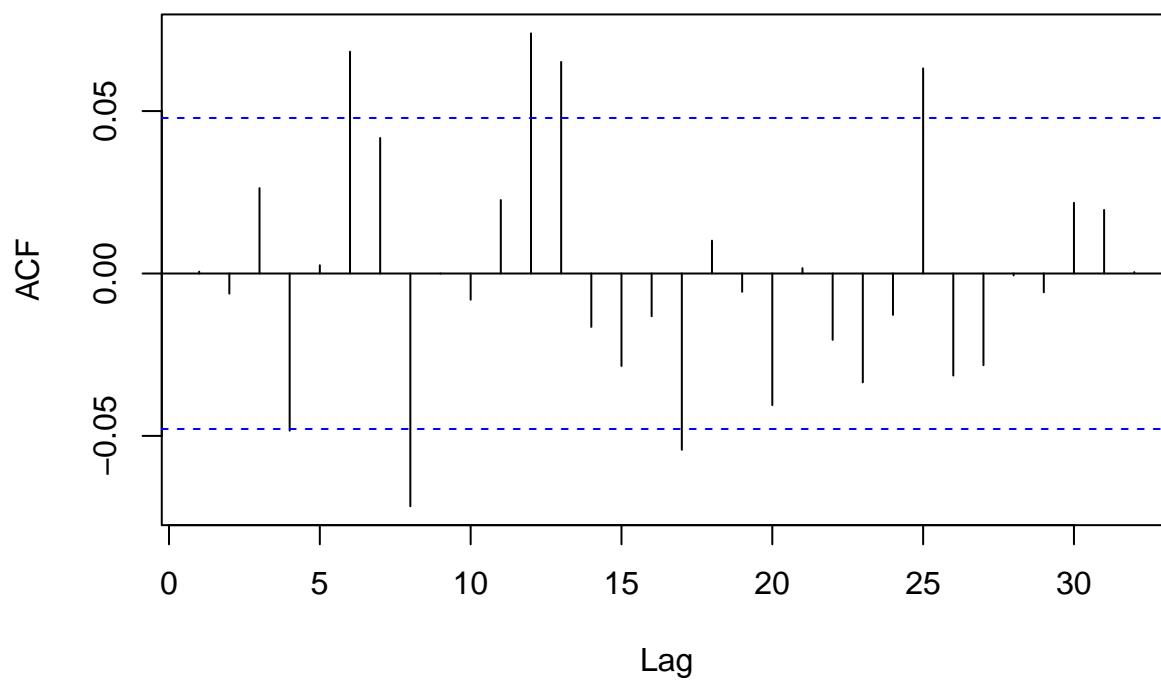
```
store_1.arima <- arima(store_1$transactions, order = c(1, 0, 1),  
  seasonal = list(order = c(1, 0, 1), period = 7),  
  xreg = cbind(store_1$isSaturday,  
    store_1$isSunday,  
    store_1$isHoliday,  
    store_1$isChristmasWeek,  
    store_1$decimal_date))
```

```
AIC(store_1.arima)
```

```
## [1] 21457.53
```

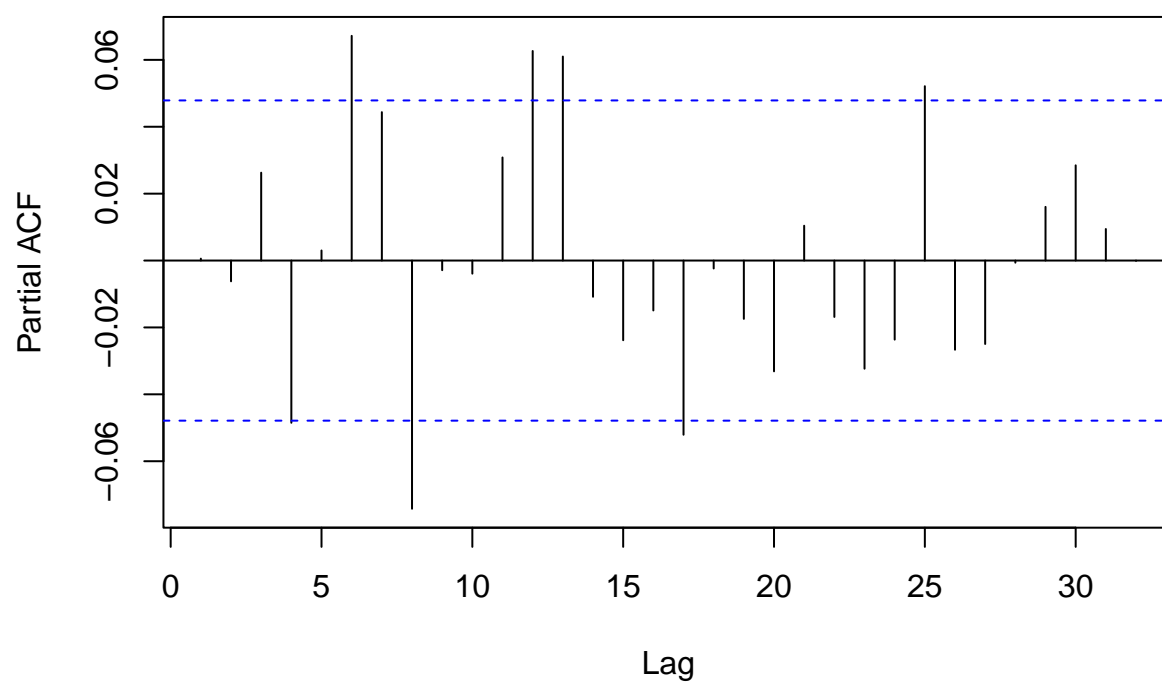
```
acf(resid(store_1.arima))
```


Series resid(store_1.arma)



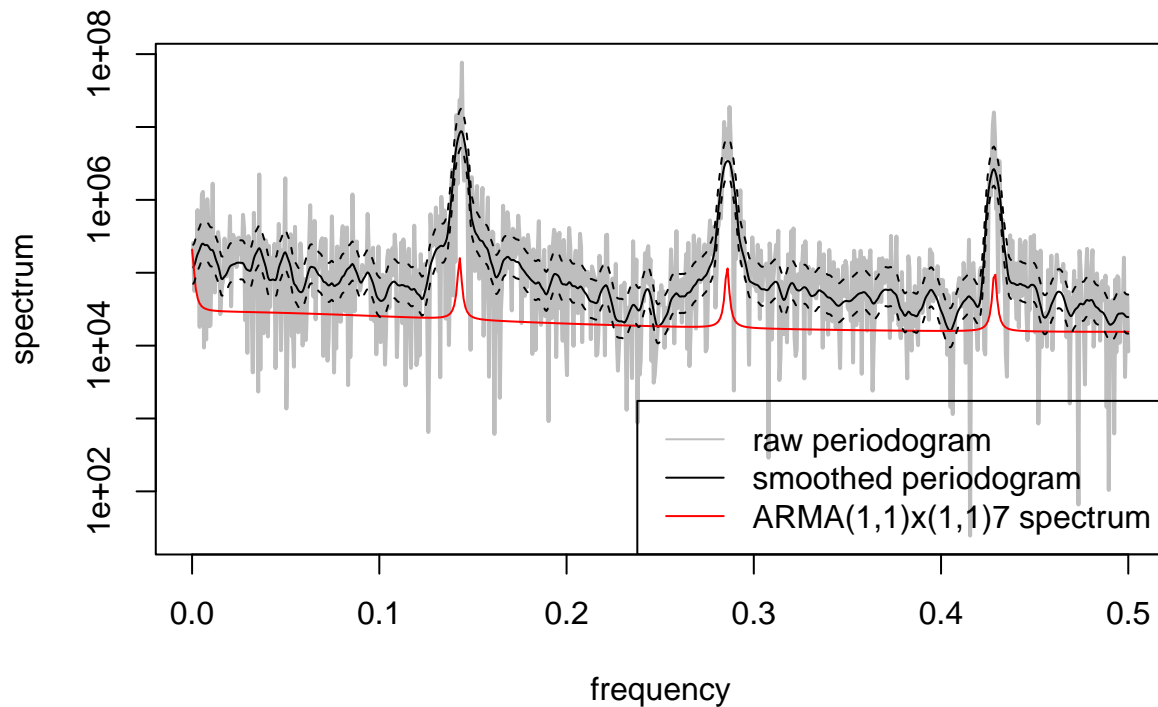
```
pacf(resid(store_1.arma))
```

Series resid(store_1.arma)



```
#LjungBoxPlot(store_1.arma)
```

Plotting the periodogram against the fitted spectrum



It's worth noting that the fitted spectrum is incorrect when we knit the appendix; the fitted spectrum that's in the final report is the correct spectrum.

Store 28

Initial plot of store 28 transactions

```
store_28 <- transactions %>%
  filter(store_nbr == 28) %>%
  select(date, transactions)
```

Adding factor variables

```
christmas_week <- store_28 %>%
  filter(date == "2013-12-18" | date == "2013-12-19" | date == "2013-12-20" |
         date == "2013-12-21" | date == "2013-12-22" | date == "2013-12-23" |
         date == "2014-12-18" | date == "2014-12-19" | date == "2014-12-20" |
         date == "2014-12-21" | date == "2014-12-22" | date == "2014-12-23" |
         date == "2015-12-18" | date == "2015-12-19" | date == "2015-12-20" |
         date == "2015-12-21" | date == "2015-12-22" | date == "2015-12-23" |
         date == "2016-12-18" | date == "2016-12-19" | date == "2016-12-20" |
         date == "2016-12-21" | date == "2016-12-22" | date == "2016-12-23")

dec_24 <- store_28 %>%
  filter(date == "2013-12-24" | date == "2014-12-24" | date == "2015-12-24" | date == "2016-12-24")

dec_31 <- store_28 %>%
  filter(date == "2013-12-31" | date == "2014-12-31" | date == "2015-12-31" | date == "2016-12-31")

store_28 <- transactions %>%
  filter(store_nbr == 28) %>%
  select(date, transactions) %>%
  mutate(isSaturday = ifelse(wday(date) %in% 7, "Yes", "No")) %>%
```

```

mutate(isSunday = ifelse(wday(date) %in% 1, "Yes", "No")) %>%
mutate(isHoliday = ifelse(date %in% store_1_holidays$date, "Yes", "No")) %>%
mutate(isChristmasWeek = ifelse(date %in% christmas_week$date, "Yes", "No")) %>%
mutate(isDec24 = ifelse(date %in% dec_24$date, "Yes", "No")) %>%
mutate(isDec31 = ifelse(date %in% dec_31$date, "Yes", "No"))

store_28 <- store_28 %>% #Turning the variables into factor variables
mutate(decimal_date = decimal_date(date)) %>%
mutate(isSaturday = ifelse(isSaturday == "Yes", 1, 0)) %>%
mutate(isSunday = ifelse(isSunday == "Yes", 1, 0)) %>%
mutate(isHoliday = ifelse(isHoliday == "Yes", 1, 0)) %>%
mutate(isChristmasWeek = ifelse(isChristmasWeek == "Yes", 1, 0)) %>%
mutate(isDec24 = ifelse(isDec24 == "Yes", 1, 0)) %>%
mutate(isDec31 = ifelse(isDec31 == "Yes", 1, 0))

```

Creating the OLS model for store 28

```

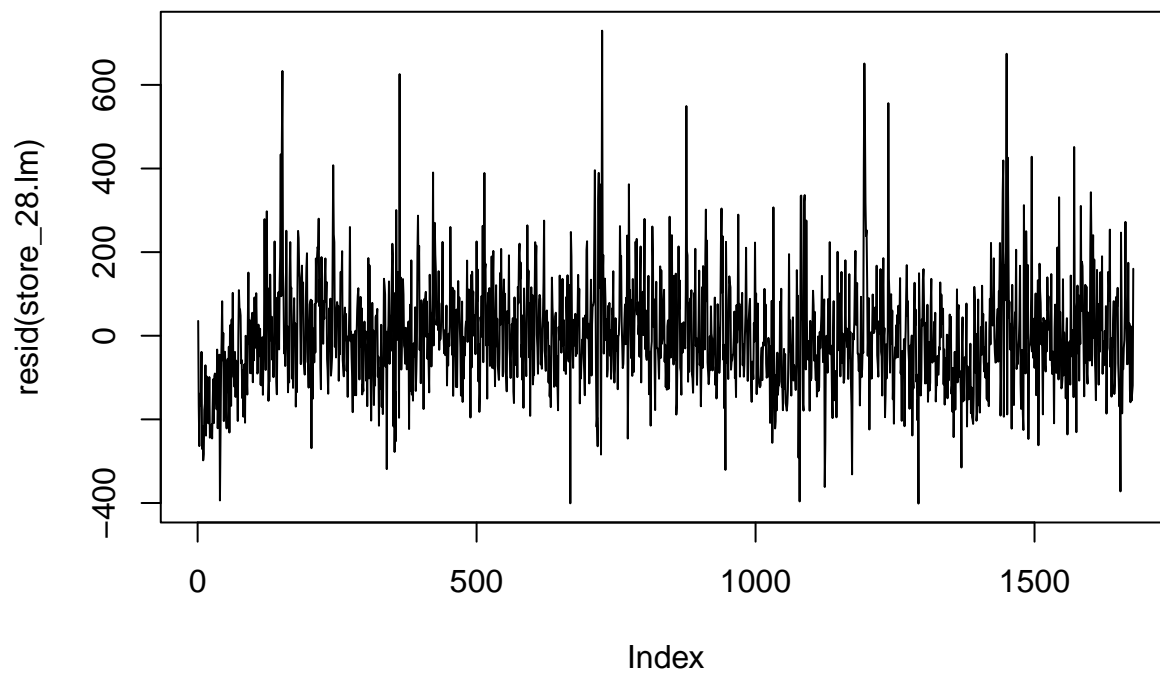
store_28.lm <- lm(transactions ~ decimal_date + isSaturday + isSunday +
                  isHoliday + isChristmasWeek + isDec24 + isDec31, data = store_28)
summary(store_28.lm)

```

```

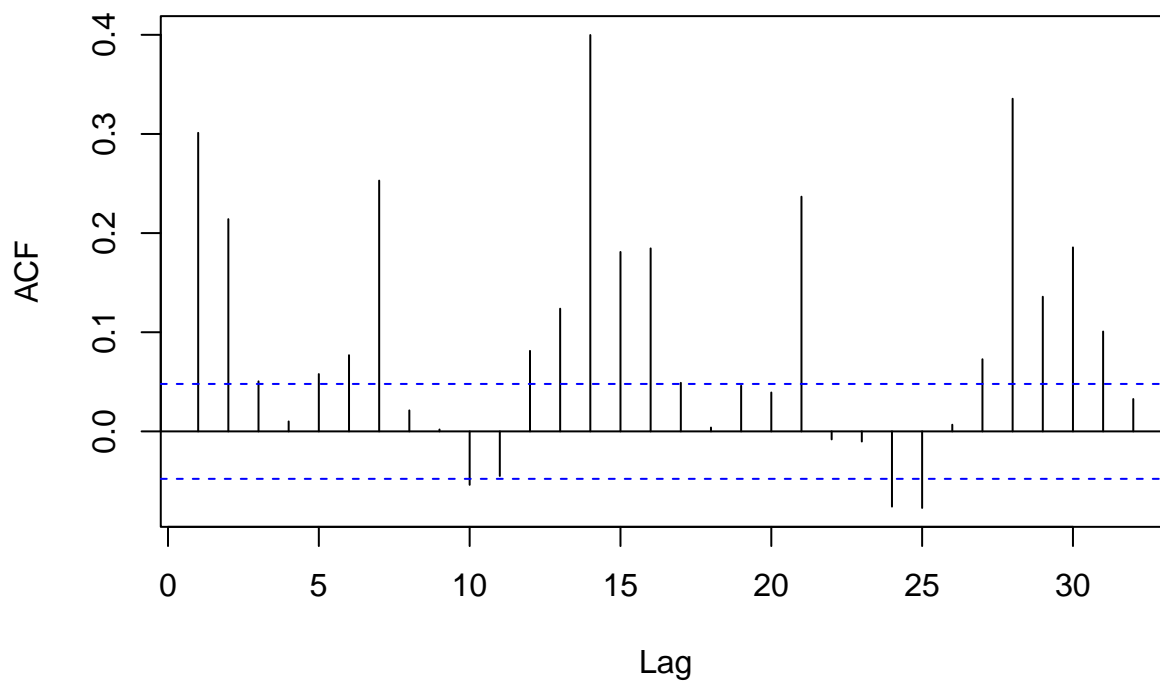
##
## Call:
## lm(formula = transactions ~ decimal_date + isSaturday + isSunday +
##      isHoliday + isChristmasWeek + isDec24 + isDec31, data = store_28)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -401.39  -85.62  -13.05   73.15  729.68
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -1.161e+05  4.795e+03  -24.203  <2e-16 ***
## decimal_date    5.810e+01  2.379e+00   24.421  <2e-16 ***
## isSaturday     2.567e+02  9.189e+00   27.931  <2e-16 ***
## isSunday       4.581e+02  9.234e+00   49.611  <2e-16 ***
## isHoliday      1.856e+02  1.967e+01    9.432  <2e-16 ***
## isChristmasWeek 4.223e+02  2.675e+01   15.788  <2e-16 ***
## isDec24        1.289e+03  6.512e+01   19.801  <2e-16 ***
## isDec31        1.120e+03  6.512e+01   17.196  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 130 on 1669 degrees of freedom
## Multiple R-squared:  0.7255, Adjusted R-squared:  0.7243
## F-statistic:  630 on 7 and 1669 DF,  p-value: < 2.2e-16
plot(resid(store_28.lm), type = 'l')

```



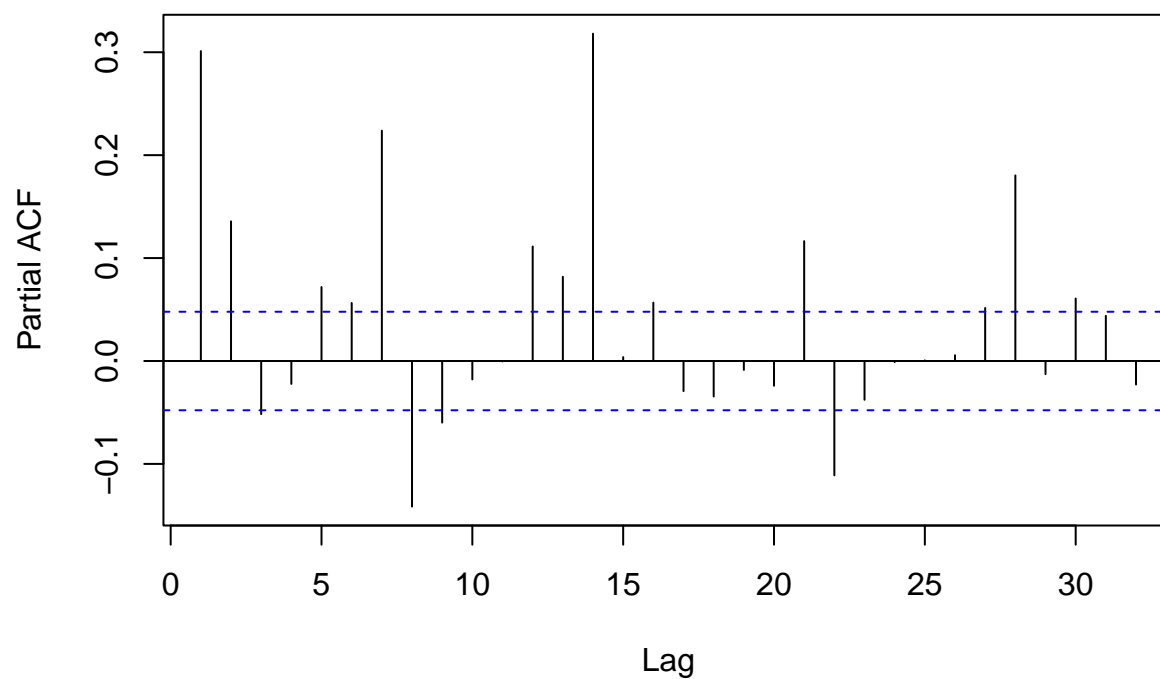
```
acf(resid(store_28.lm))
```

Series resid(store_28.lm)



```
pacf(resid(store_28.lm))
```

Series resid(store_28.lm)



Modeling testing for store 28

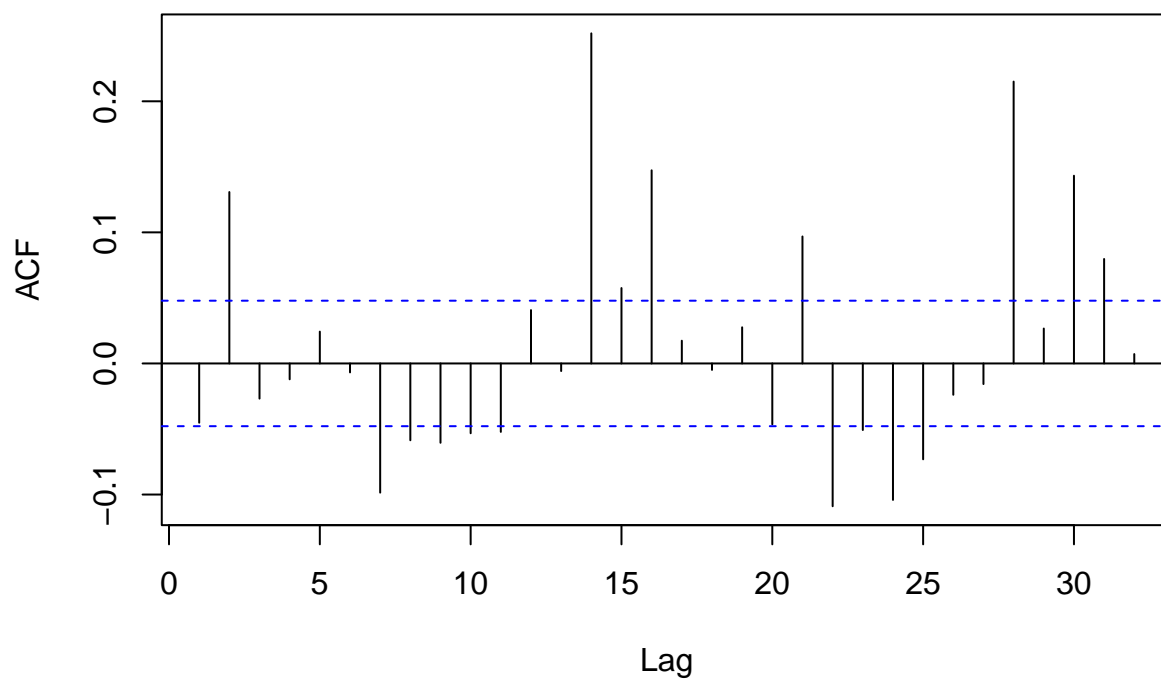
```
store_28.arima <- arima(store_28$transactions, order = c(1, 0, 0),  
                        seasonal = list(order = c(1, 0, 0), period = 7),  
                        xreg = cbind(store_28$isSaturday,  
                                     store_28$isSunday,  
                                     store_28$isHoliday,  
                                     store_28$isChristmasWeek,  
                                     store_28$isDec24,  
                                     store_28$isDec31,  
                                     store_28$decimal_date))
```

```
AIC(store_28.arima)
```

```
## [1] 20745.13
```

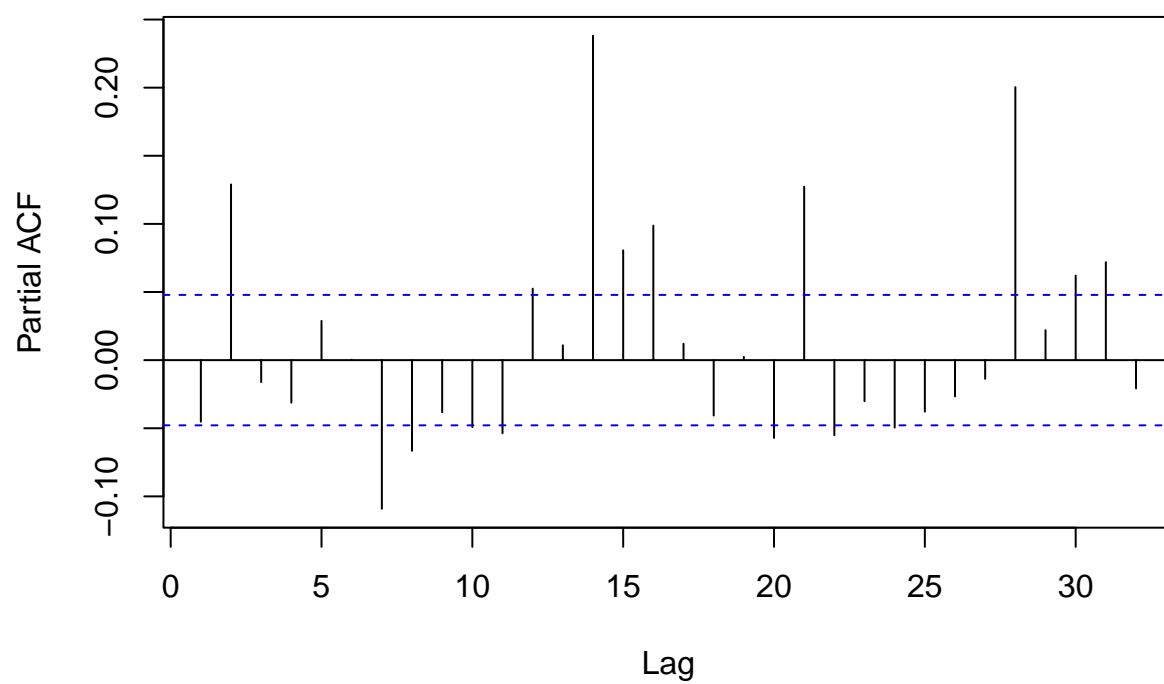
```
acf(resid(store_28.arima))
```

Series resid(store_28.arima)



```
pacf(resid(store_28.arima))
```

Series resid(store_28.arima)



```
#LjungBoxPlot(store_28.arima)
```

```
store_28.arima <- arima(store_28$transactions, order = c(1, 0, 1),
  seasonal = list(order = c(1, 0, 1), period = 7),
  xreg = cbind(store_28$isSaturday,
    store_28$isSunday,
    store_28$isHoliday,
    store_28$isChristmasWeek,
    store_28$isDec24,
    store_28$isDec31,
    store_28$decimal_date))
```

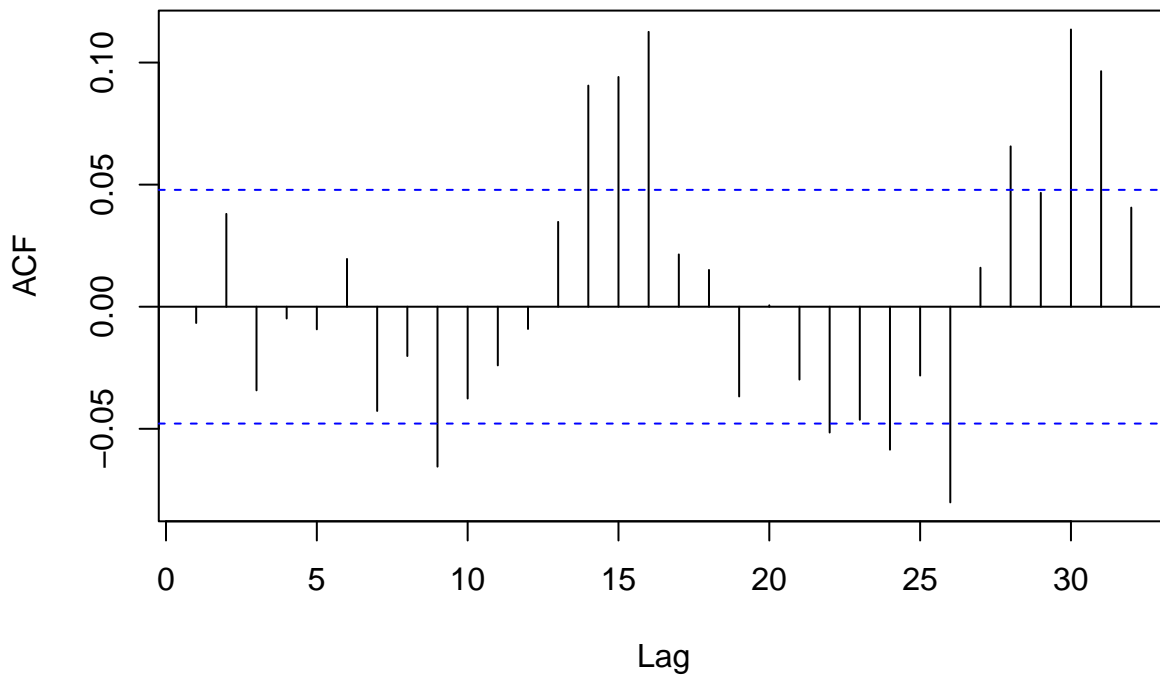
```
## Warning in stats::arima(x = x, order = order, seasonal = seasonal, xreg =
## xreg, : possible convergence problem: optim gave code = 1
```

```
AIC(store_28.arima)
```

```
## [1] 20460
```

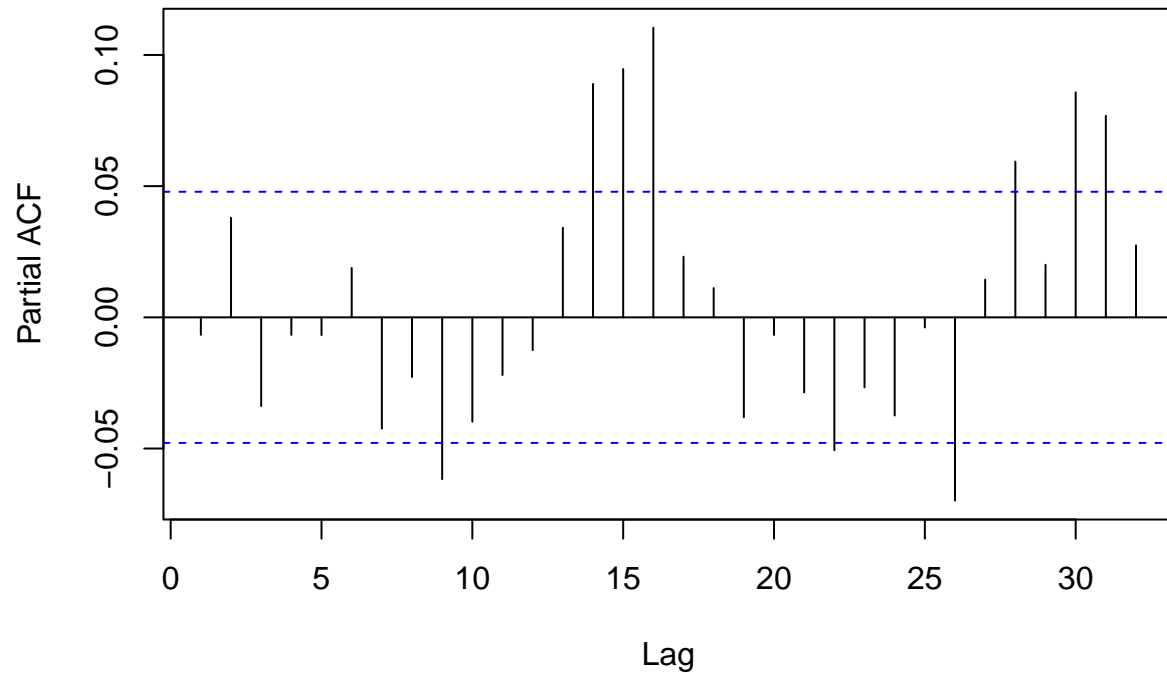
```
acf(resid(store_28.arima))
```

Series resid(store_28.arima)



```
pacf(resid(store_28.arima))
```

Series resid(store_28.arima)



```
#LjungBoxPlot(store_28.arima)
```

Final ARMA Modeling for store 28, and the ACF and PACF of the model

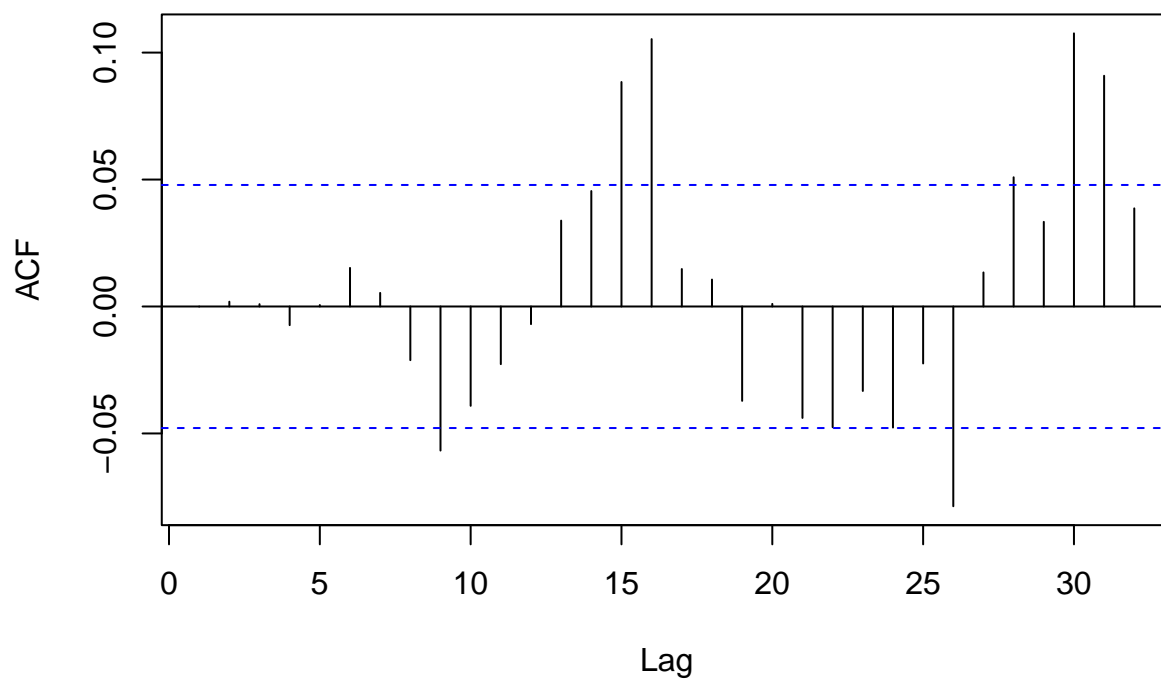
```
store_28.arima <- arima(store_28$transactions, order = c(2, 0, 1),  
                        seasonal = list(order = c(2, 0, 1), period = 7),  
                        xreg = cbind(store_28$isSaturday,  
                                     store_28$isSunday,  
                                     store_28$isHoliday,  
                                     store_28$isChristmasWeek,  
                                     store_28$isDec24,  
                                     store_28$isDec31,  
                                     store_28$decimal_date))
```

```
AIC(store_28.arima)
```

```
## [1] 20452.49
```

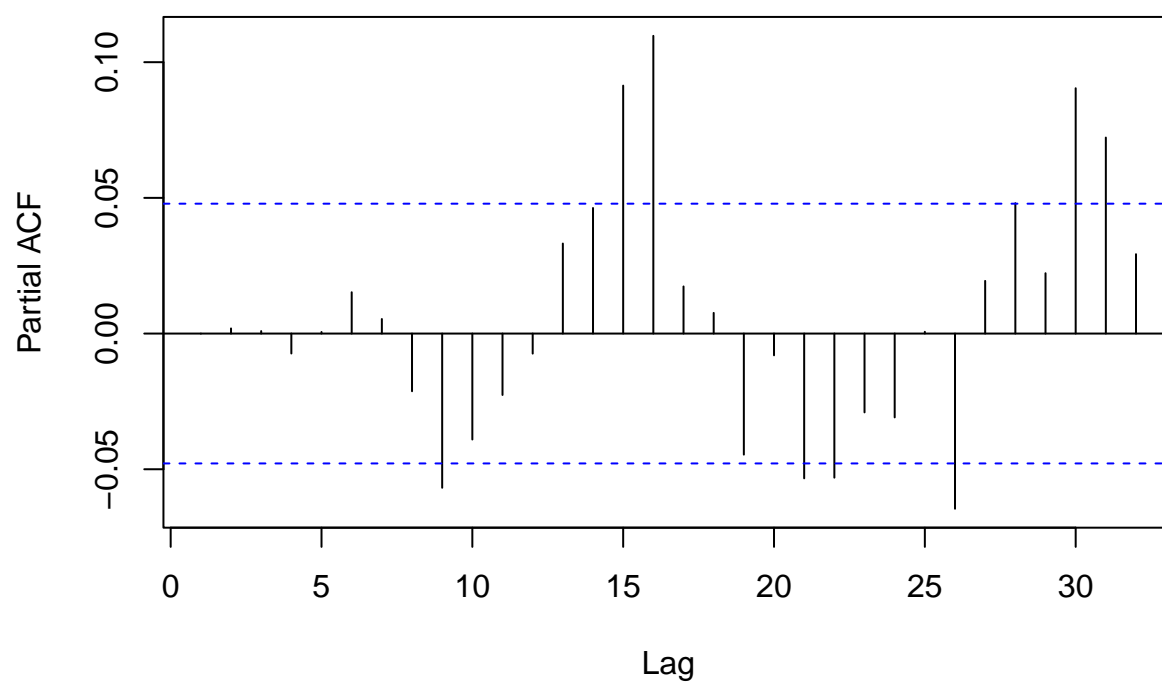
```
acf(resid(store_28.arima))
```


Series resid(store_28.arima)



```
pacf(resid(store_28.arima))
```

Series resid(store_28.arima)



```
#LjungBoxPlot(store_28.arima)
```

Plotting the periodogram against the fitted spectrum

