Abstract

Big box retailers such as Walmart need to have accurate models of purchasing patterns at their various stores. Overestimating consumer demand could lead to losses due to excess inventory.

On the other hand, underestimating it could result in lower customer service level and lost sales.

Our dataset contains historical sales data for 45 Walmart stores located in different regions. Each store contains a number of departments.

For the purpose of this lab, we are analyzing a randomly chosen store 4 and department 12 (based on a random number generator)

- Is there a relationship between temperature and sales?
- Is there a relationship between CPI and sales?
- Is there a relationship between fuel price and sales?
- Is there a relationship between unemployment and sales?

Finally, we are going to forecast sales for this store which will help them maximize sales and optimize inventory.

Loading the source datasets and the required libraries

```
train<-read.csv('train.csv')
features<-read.csv('features.csv')
train <- train[train['Dept']==12,]
train<-train[,c(1,2,3,4)]
```

Function to perform EDA on time series

```
tseda<-function(timeseries,xlabel,ylabel,mainlabel){
  plot(timeseries,xlab=xlabel,ylab=ylabel,main=mainlabel)
  #par(mfrow=c(1,2))
  Acf(timeseries,main = paste(mainlabel,' ACF'))
  Pacf(timeseries,main =
  paste(mainlabel,' Pacf'))
  adf.test(timeseries)
  }
```

Function to perform out of sample tests on forecasts using RMSE

```
rmse<-function(timeseries1,timeseries2){
return(sqrt(mean((timeseries1-timeseries2)^2))) }
```

Merging and filtering the data

```
input<-merge(train,features,by  =c("Store","Date"))
input<-input[,c(1,2,3,4,5,6,12,13)]
input<-input[(input["Store"]==4)|(input["Store"]==14)|(input["Store"]==15),]
store4_data<-input[input["Store"]==4,]
```

Describe the time series in the dataset

```
str(store4_data)
```

```
## 'data.frame':    143 obs. of  8 variables:
##  $ Store       : int  4 4 4 4 4 4 4 4 4 4 ...
##  $ Date        : Factor w/ 143 levels "2010-02-05","2010-02-12",..: 1 2 3 4 5 6 7 8 9
 10 ...
##  $ Dept        : int  12 12 12 12 12 12 12 12 12 12 ...
##  $ Weekly_Sales: num  8245 6689 6791 8821 8800 ...
##  $ Temperature : num  43.8 28.8 36.5 41.4 43.5 ...
##  $ Fuel_Price  : num  2.6 2.57 2.54 2.59 2.65 ...
##  $ CPI         : num  126 126 127 127 127 ...
##  $ Unemployment: num  8.62 8.62 8.62 8.62 8.62 ...
```

```
summary(store4_data)
```

```
##      Store            Date           Dept     Weekly_Sales    Temperature
##  Min.   :4   2010-02-05:  1   Min.   :12   Min.   : 4209   Min.   :28.84
##  1st Qu.:4   2010-02-12:  1   1st Qu.:12   1st Qu.: 7277   1st Qu.:48.47
##  Median :4   2010-02-19:  1   Median :12   Median : 8336   Median :64.22
##  Mean   :4   2010-02-26:  1   Mean   :12   Mean   : 8142   Mean   :62.25
##  3rd Qu.:4   2010-03-05:  1   3rd Qu.:12   3rd Qu.: 9037   3rd Qu.:77.44
##  Max.   :4   2010-03-12:  1   Max.   :12   Max.   :10870   Max.   :86.09
##              (Other)   :137
##    Fuel_Price         CPI          Unemployment
##  Min.   :2.540   Min.   :126.1   Min.   :3.879
##  1st Qu.:2.764   1st Qu.:126.6   1st Qu.:4.607
##  Median :3.290   Median :129.1   Median :5.946
##  Mean   :3.217   Mean   :128.7   Mean   :5.965
##  3rd Qu.:3.587   3rd Qu.:130.5   3rd Qu.:7.127
##  Max.   :3.881   Max.   :131.2   Max.   :8.623
##
```

Based on an analysis of the structure of the data, it doesn't appear as though there are missing values in the data set.

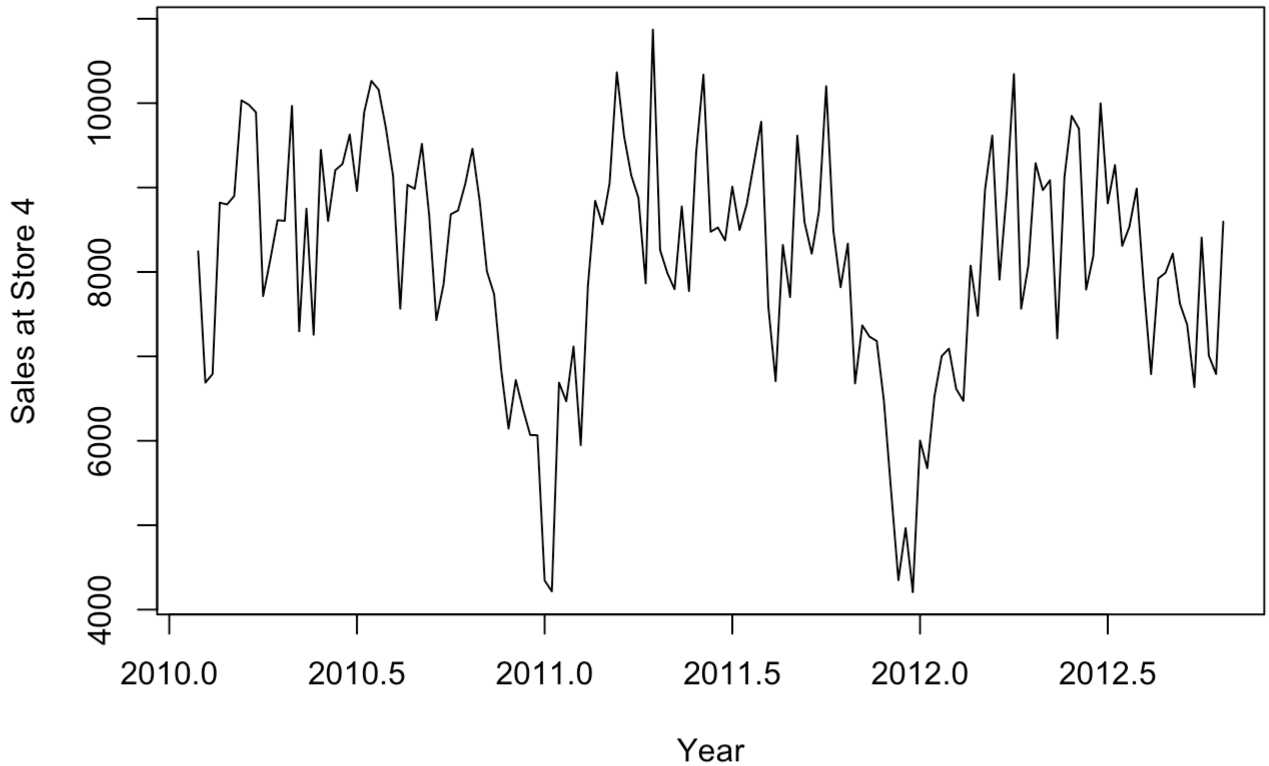
Examine the various series for stationarity

- The ACF and PACF plots for all the time series show the classic signature for an AR model of root 1. The ACF drops down slowly while the PACF drops suddenly at lag 1. This series' will need to be differenced at least once to get to stationarity

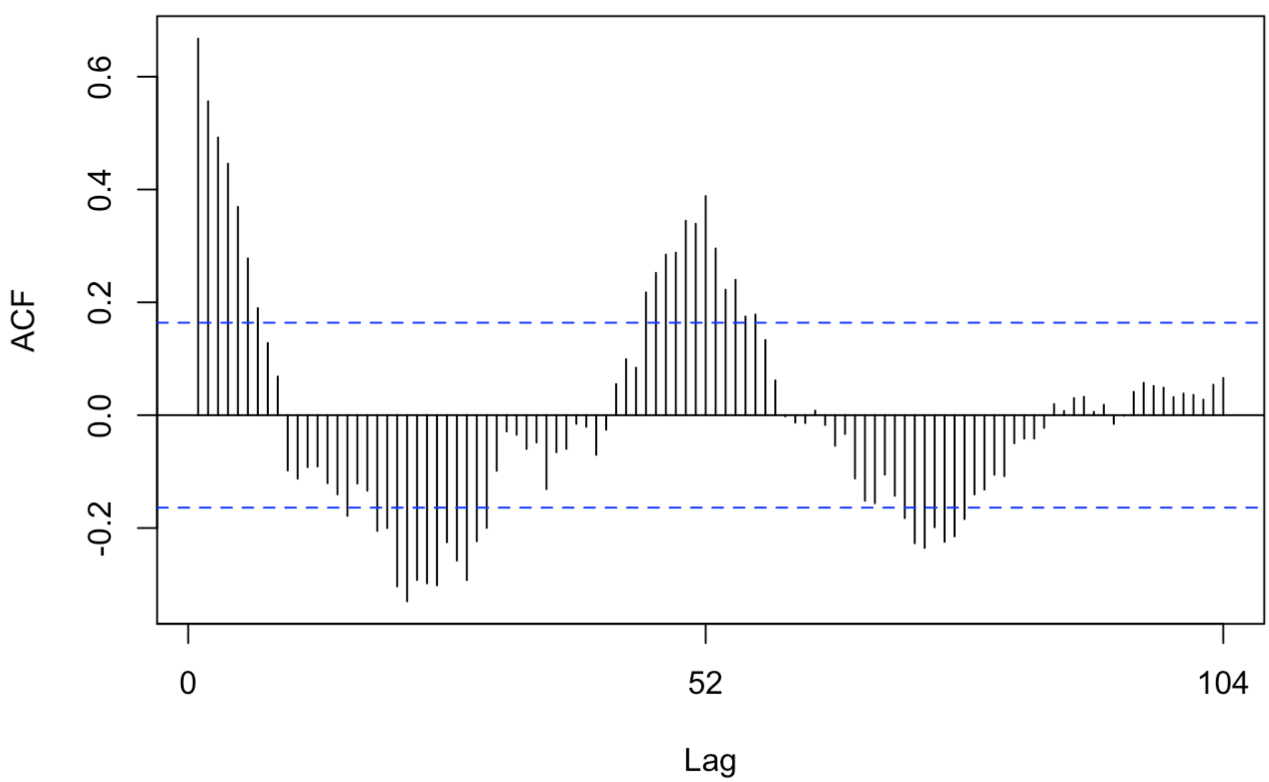
Time Series 1: Sales

```
store4_sales.ts<-ts(store4_data$Weekly_Sales,start =c(2010,5),freq=52)
tseda(store4_sales.ts,"Year","Sales at Store 4","Store 4
Sales by Year")
```

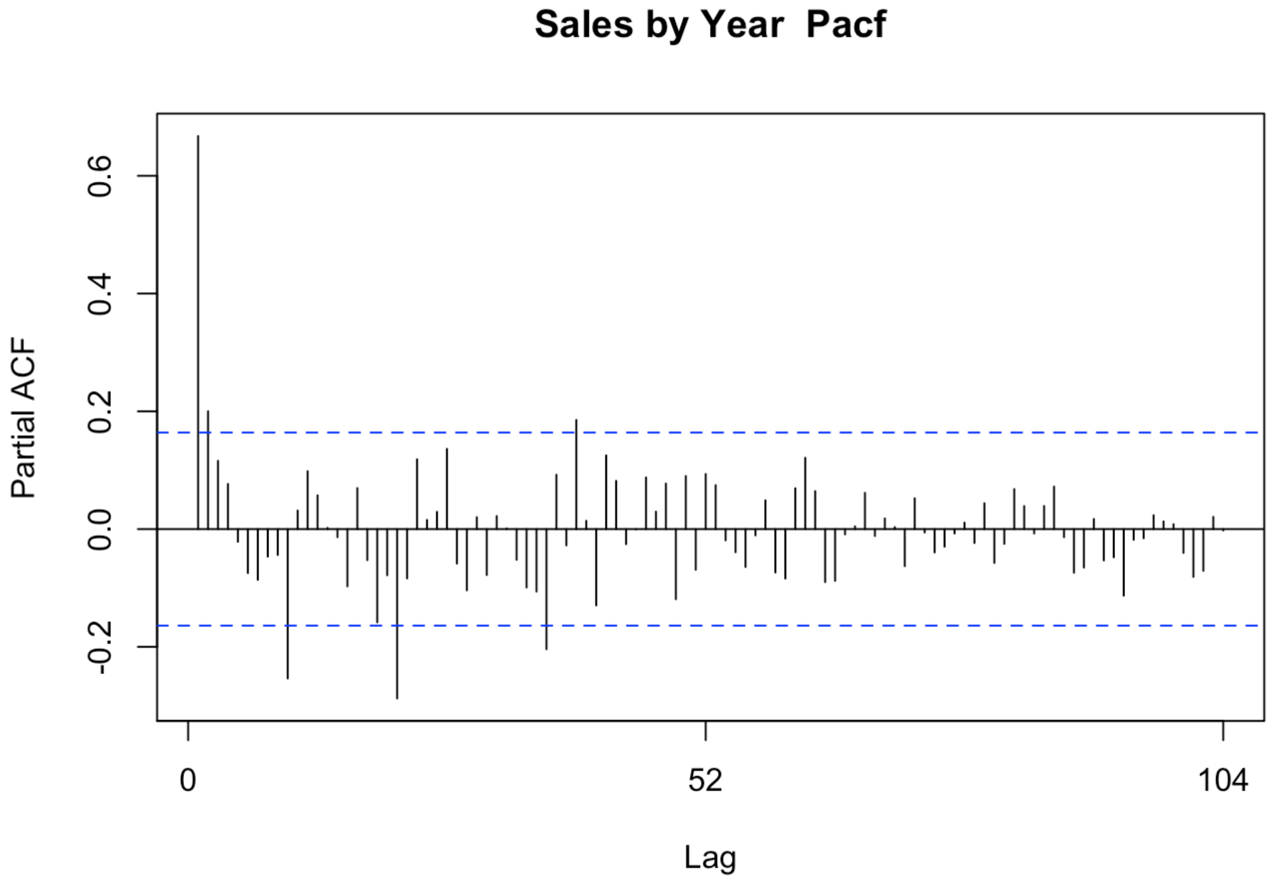
Store 4
Sales by Year



Store 4
Sales by Year ACF



Store 4

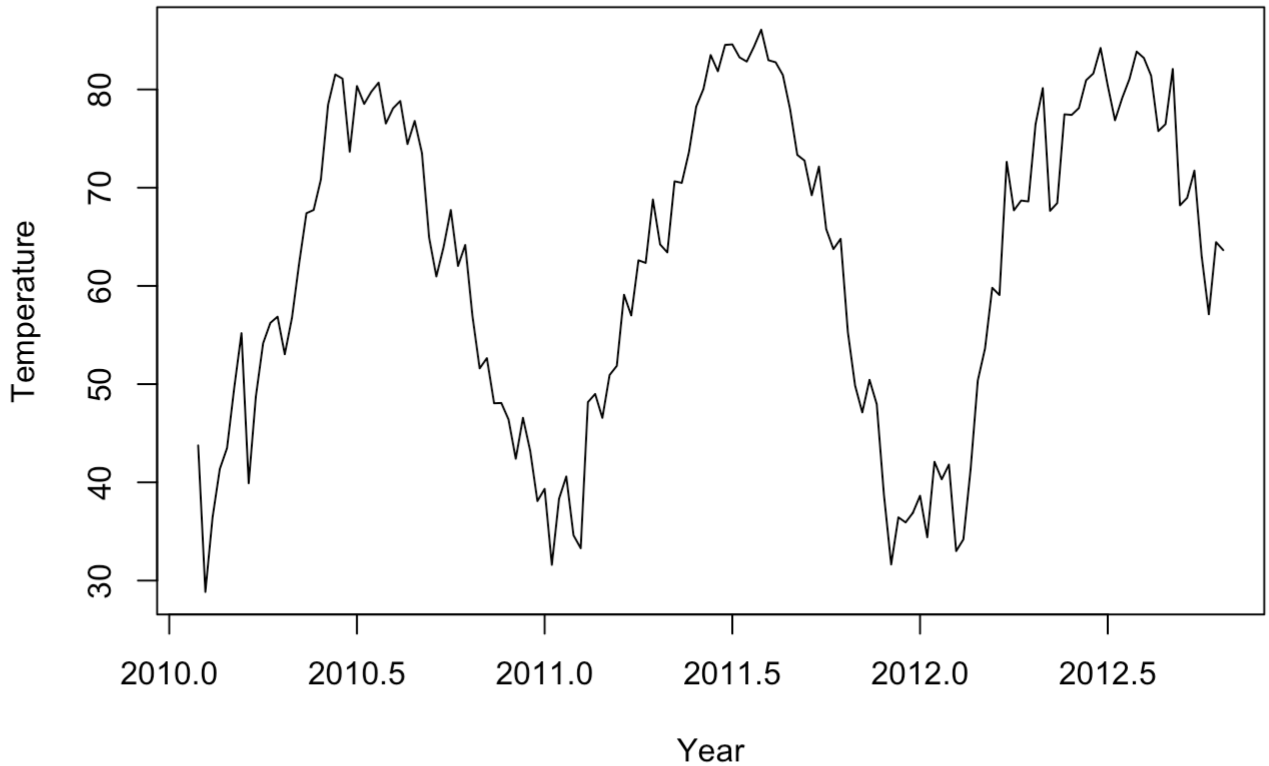



```
##
##  Augmented Dickey-Fuller Test
##
## data:  timeseries
## Dickey-Fuller = -3.0062, Lag order = 5, p-value = 0.1576
## alternative hypothesis: stationary
```

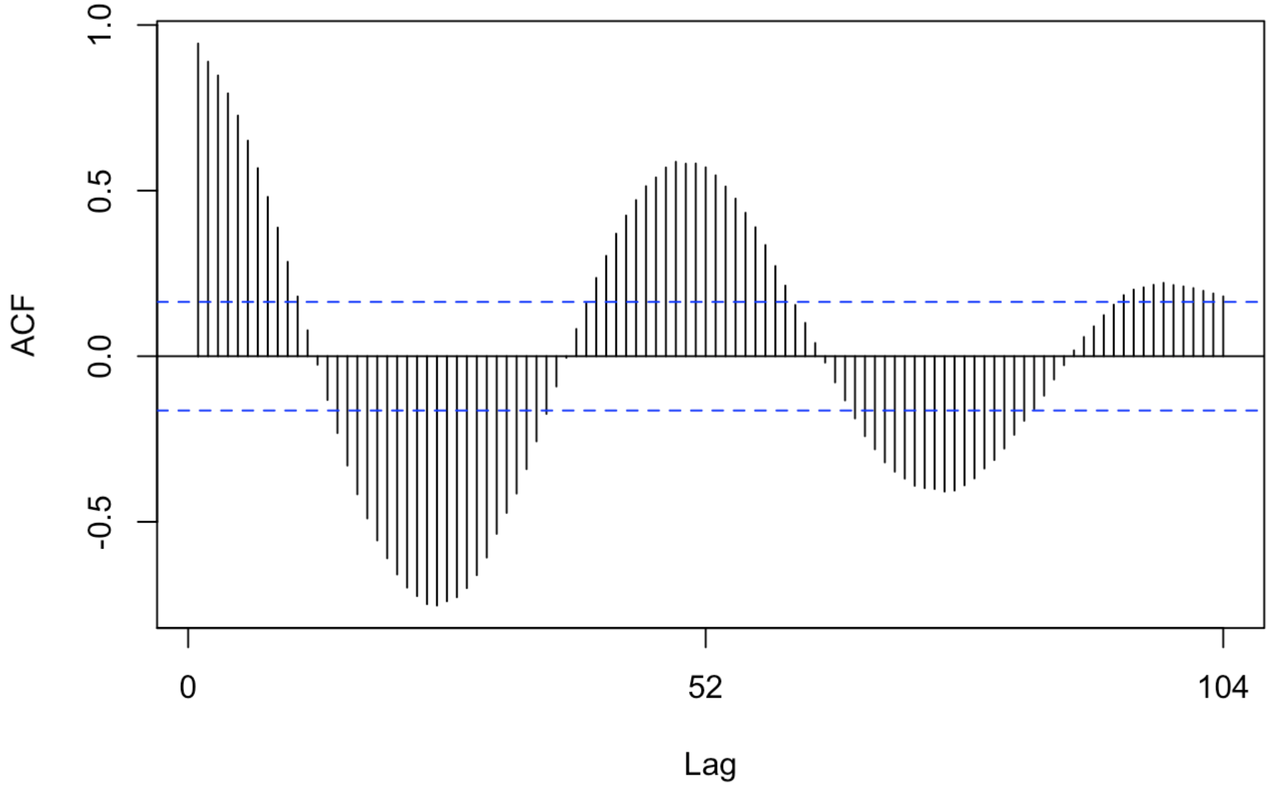
Time Series 2: Temperature

```
store4_temp.ts<-ts(store4_data$Temperature,start =c(2010,5),freq=52)
tseda(store4_temp.ts,"Year","Temperature","Temperature by
Year")
```

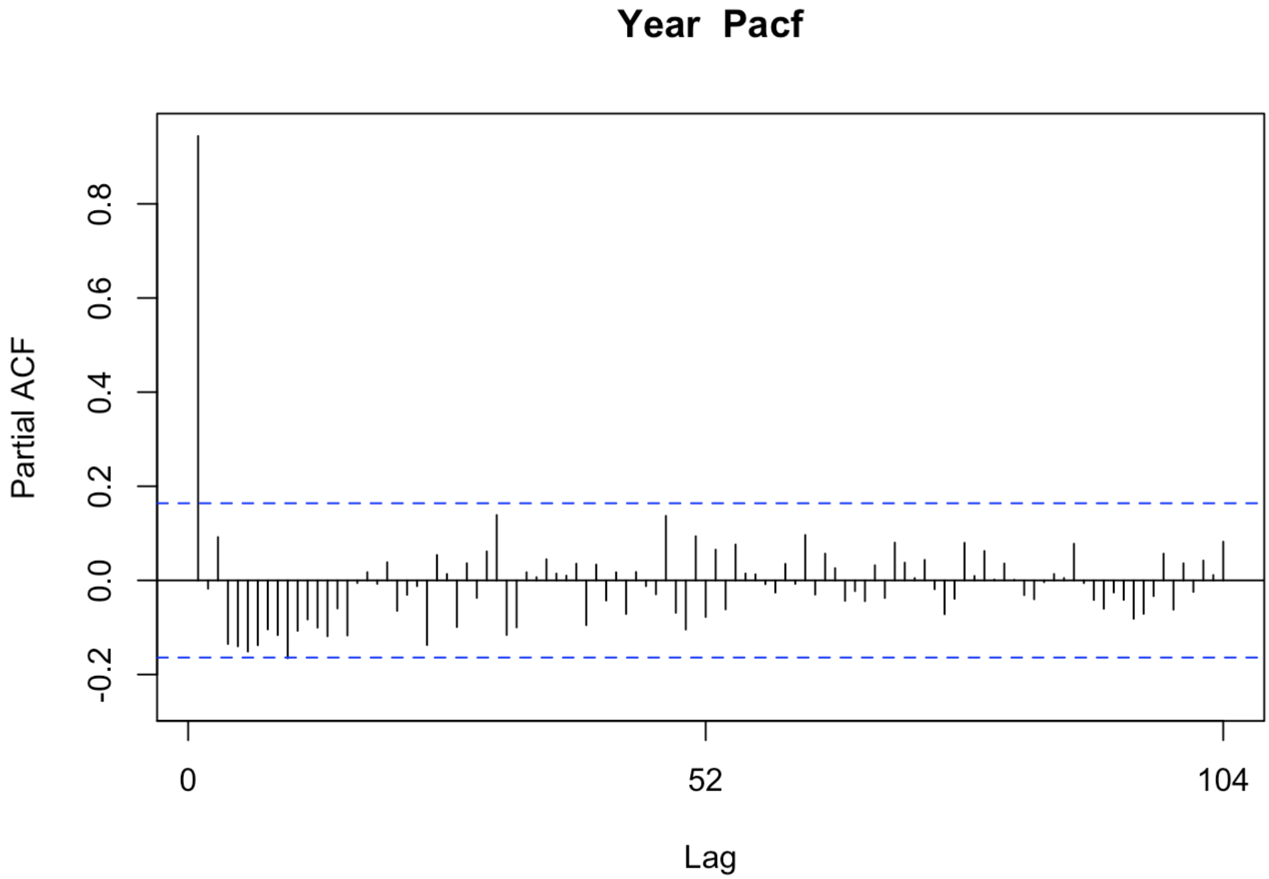
Temperature by Year



Temperature by Year ACF



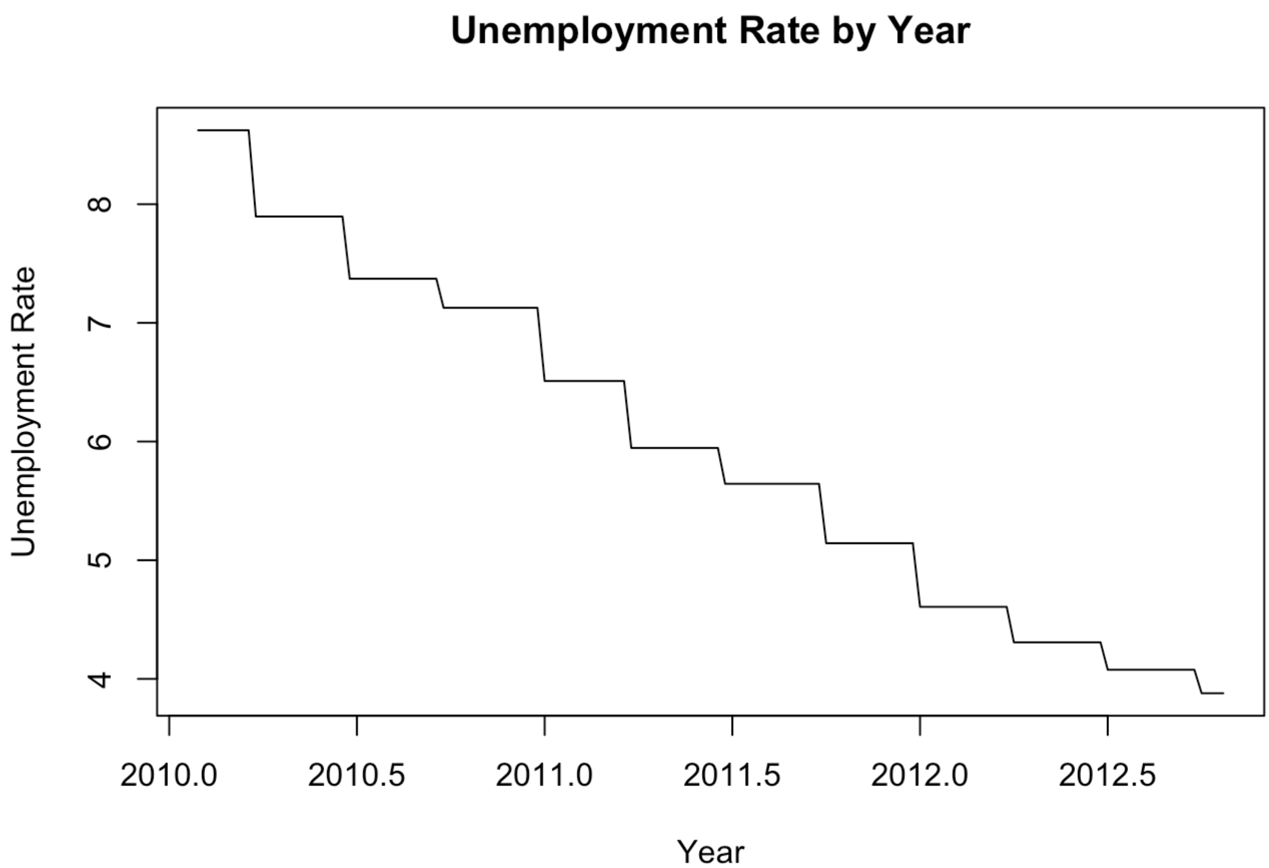
Temperature by

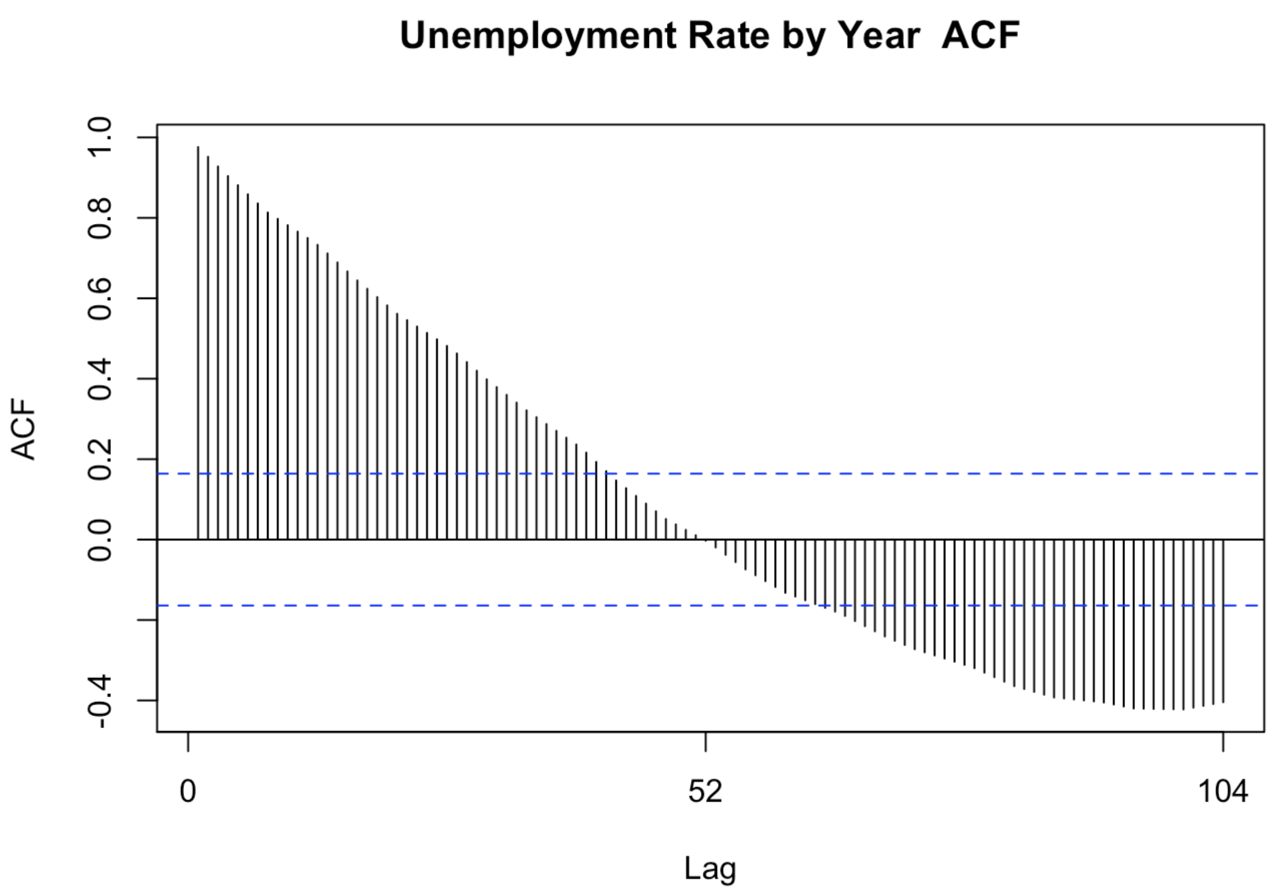


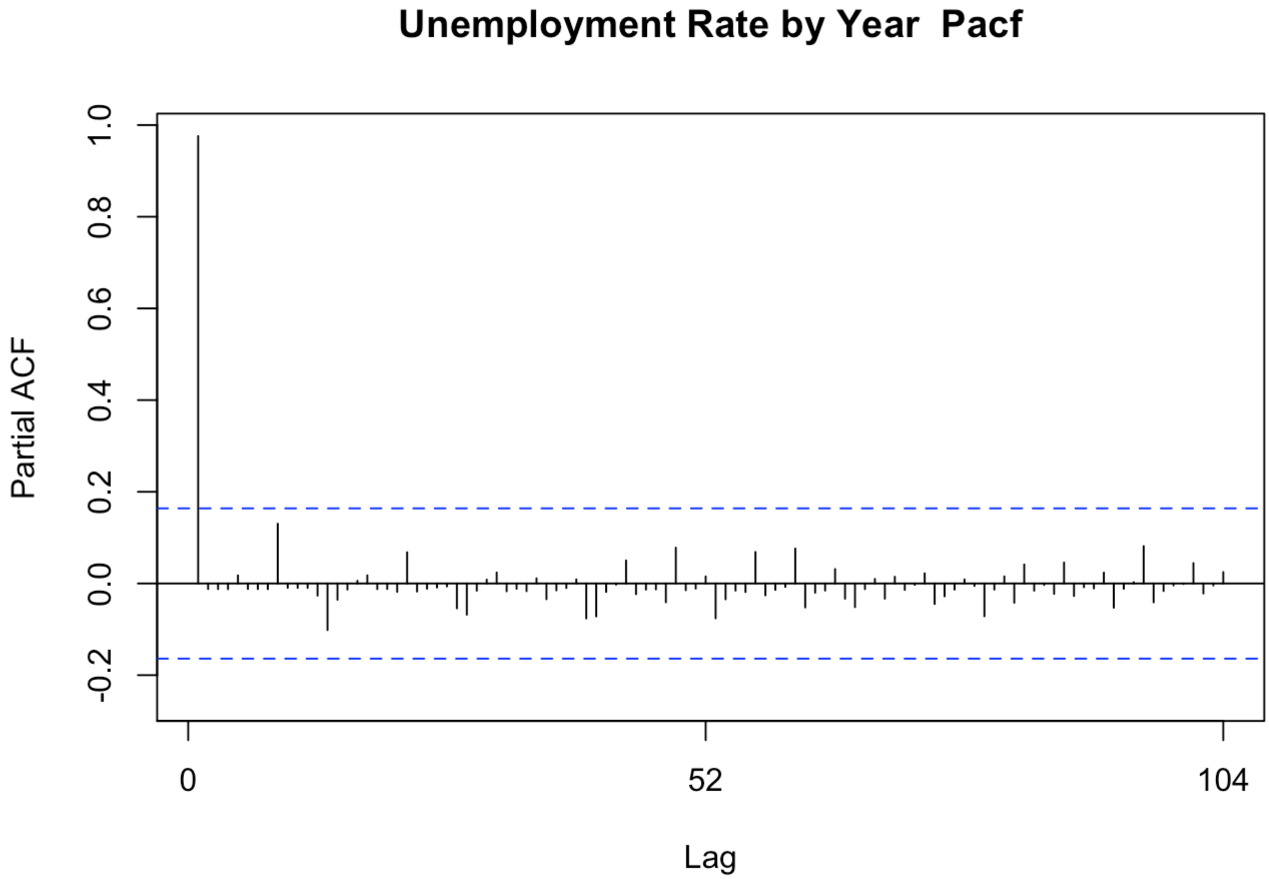

```
##
##  Augmented Dickey-Fuller Test
##
## data:  timeseries
## Dickey-Fuller = -2.7561, Lag order = 5, p-value = 0.2617
## alternative hypothesis: stationary
```

Time Series 3: Unemployment Rate

```
store4_ue.ts<-ts(store4_data$Unemployment,start = c(2010,5),freq=52)
tseda(store4_ue.ts,"Year","Unemployment Rate","Unemployment Rate by Year")
```


Unemployment Rate by Year
Unemployment Rate
8
7
6
5
4
2010.0
2010.5
2011.0
2011.5
2012.0
2012.5
Year

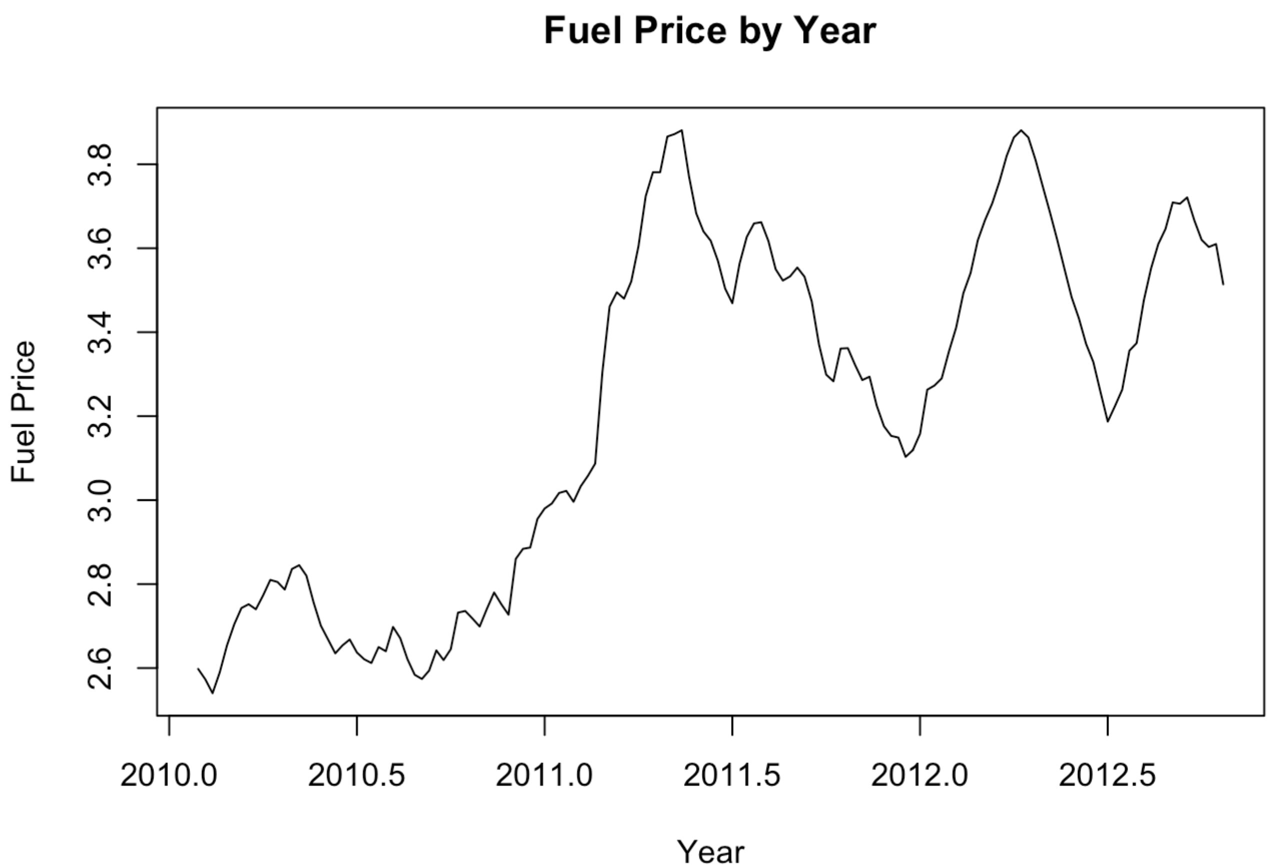

Unemployment Rate by Year ACF
ACF
1.0
0.8
0.6
0.4
0.2
0.0
-0.4
0
52
104
Lag

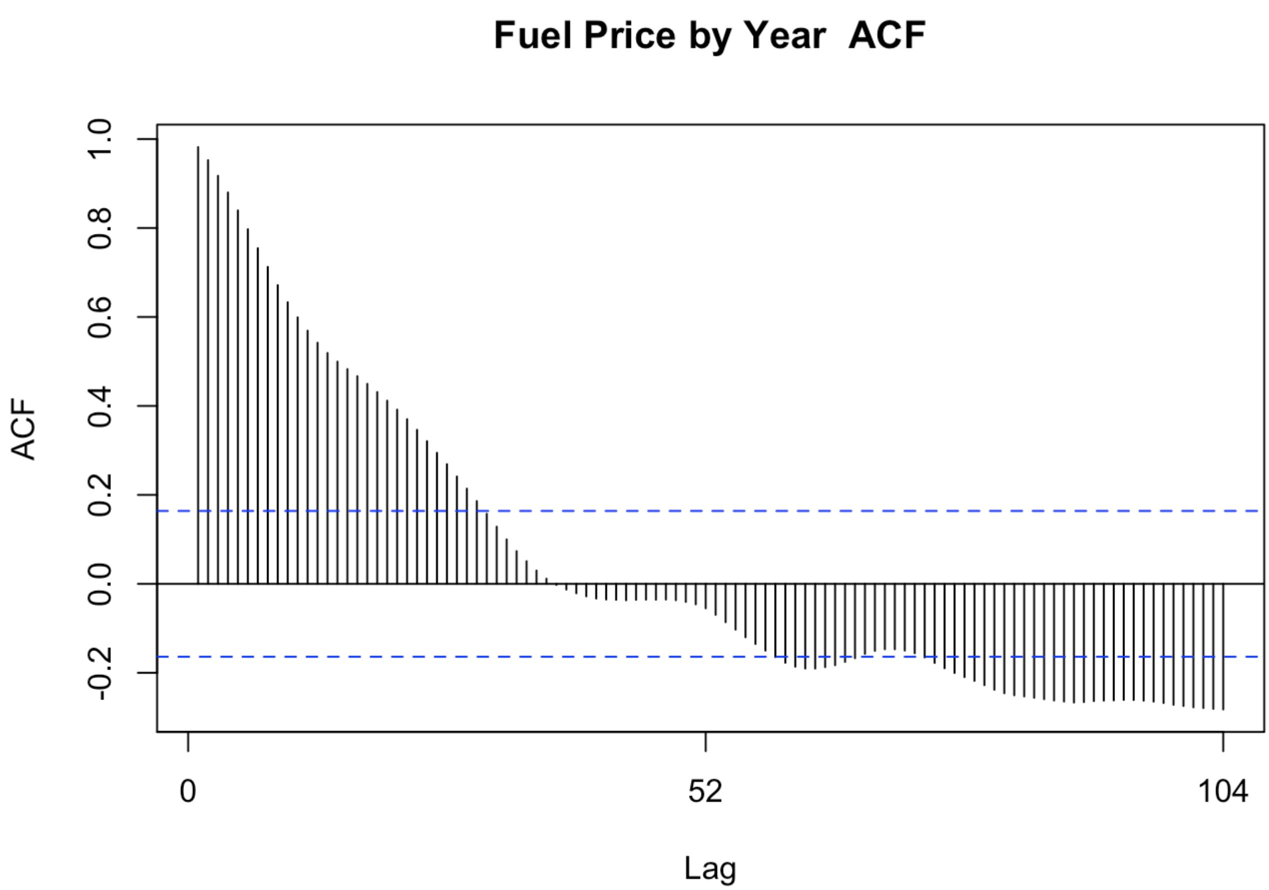


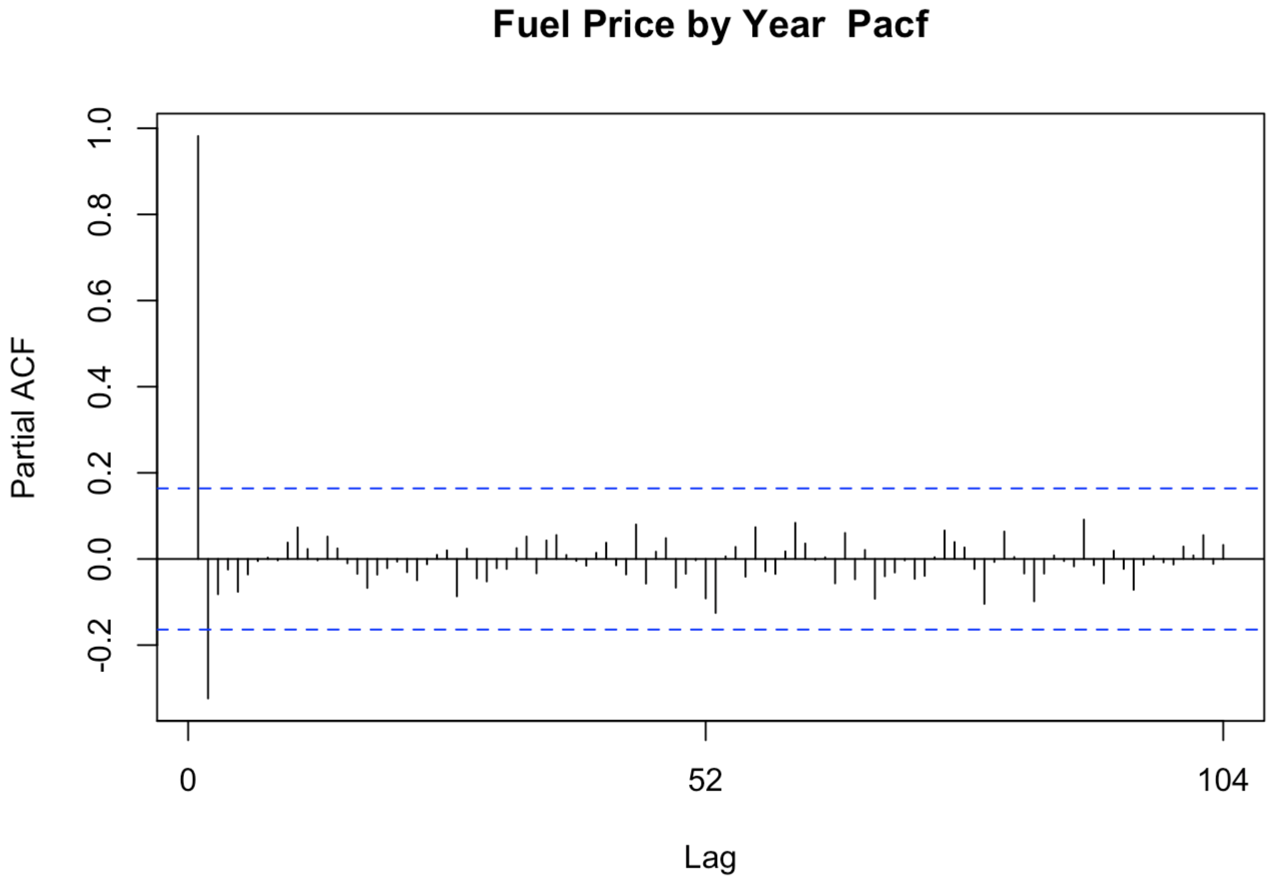

```
##
##  Augmented Dickey-Fuller Test
##
## data:  timeseries
## Dickey-Fuller = -3.6827, Lag order = 5, p-value = 0.02829
## alternative hypothesis: stationary
```

Time Series 4: Fuel Price

```
store4_fp.ts<-ts(store4_data$Fuel_Price,start = c(2010,5),freq=52)
tseda(store4_fp.ts,"Year","Fuel Price","Fuel Price by Year")
```


Fuel Price by Year
Fuel Price
3.8
3.6
3.4
3.2
3.0
2.8
2.6
2010.0
2010.5
2011.0
2011.5
2012.0
2012.5
Year

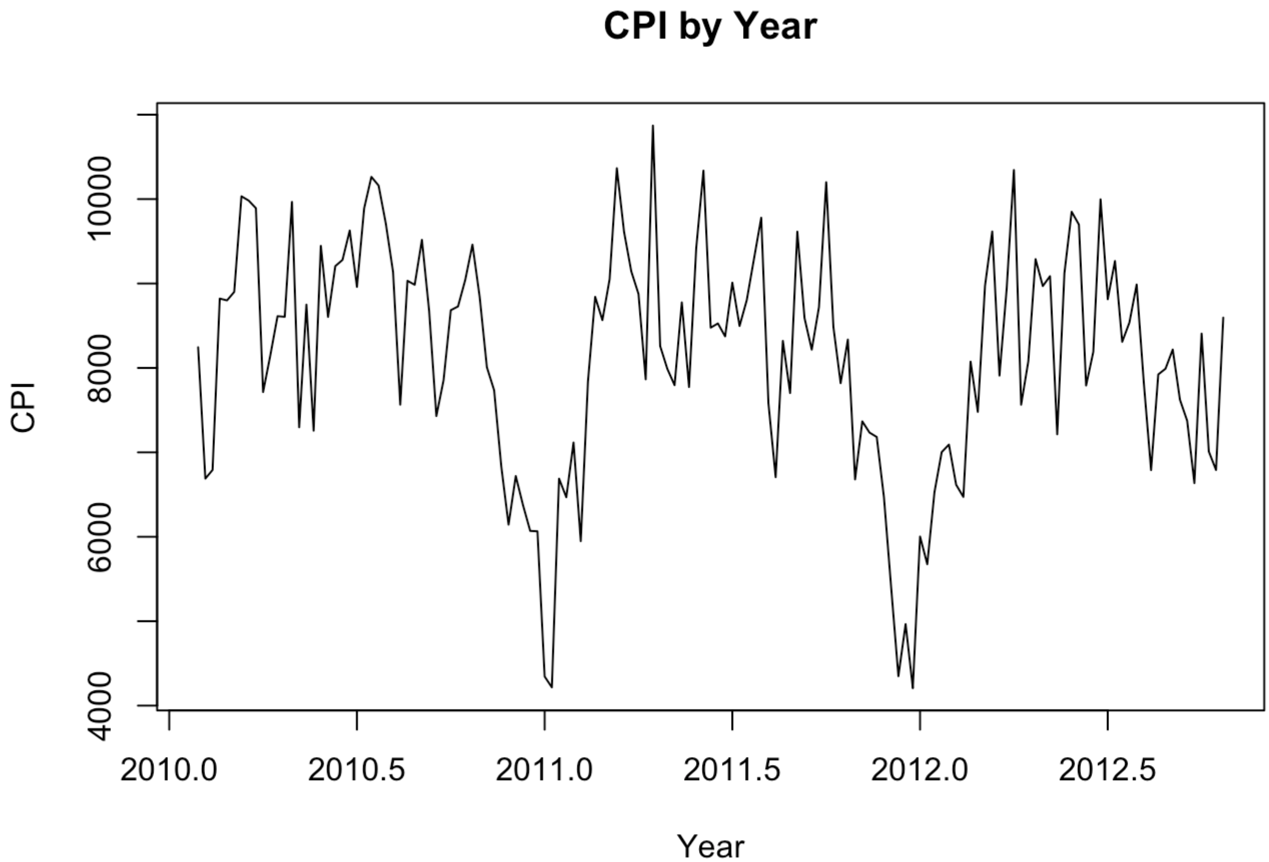

Fuel Price by Year ACF
ACF
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
0
52
104
Lag

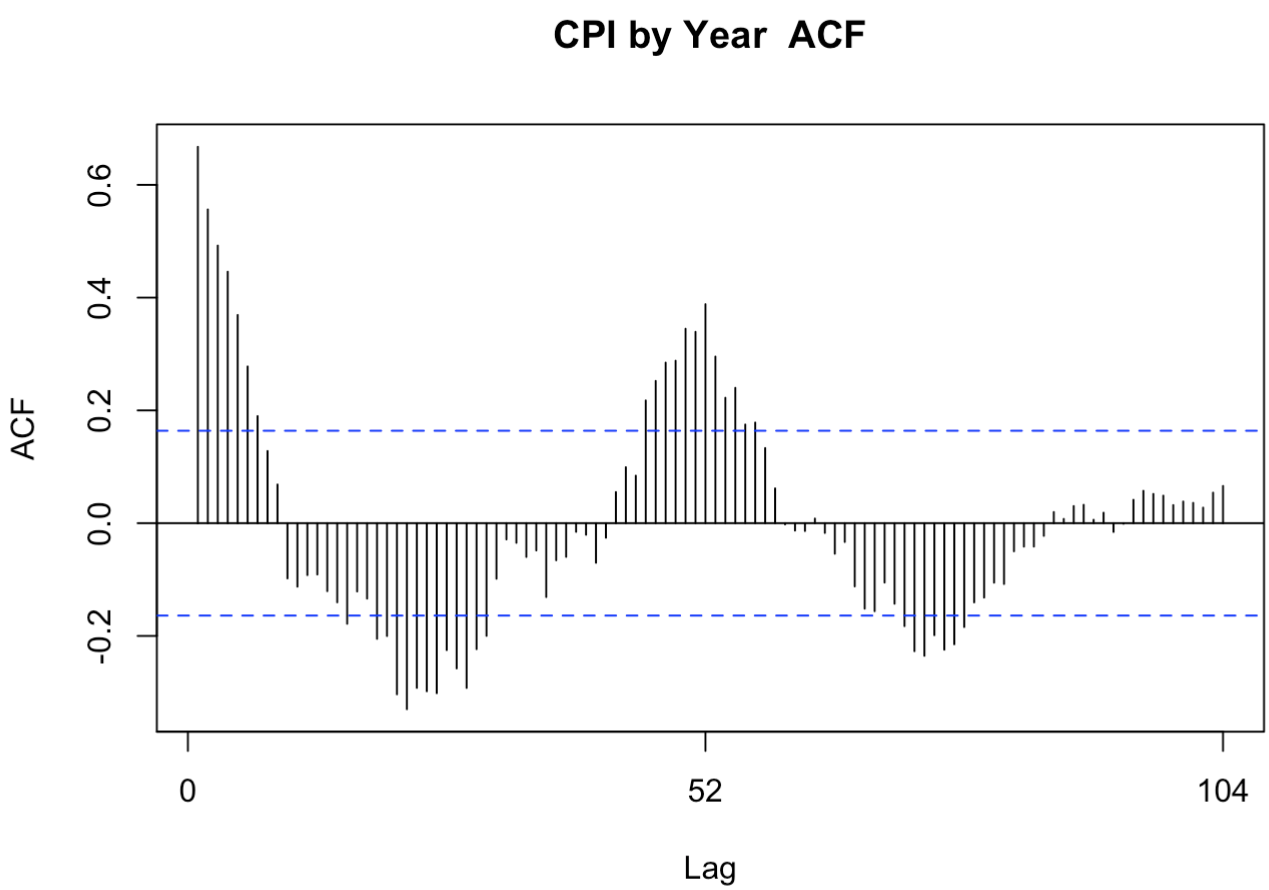


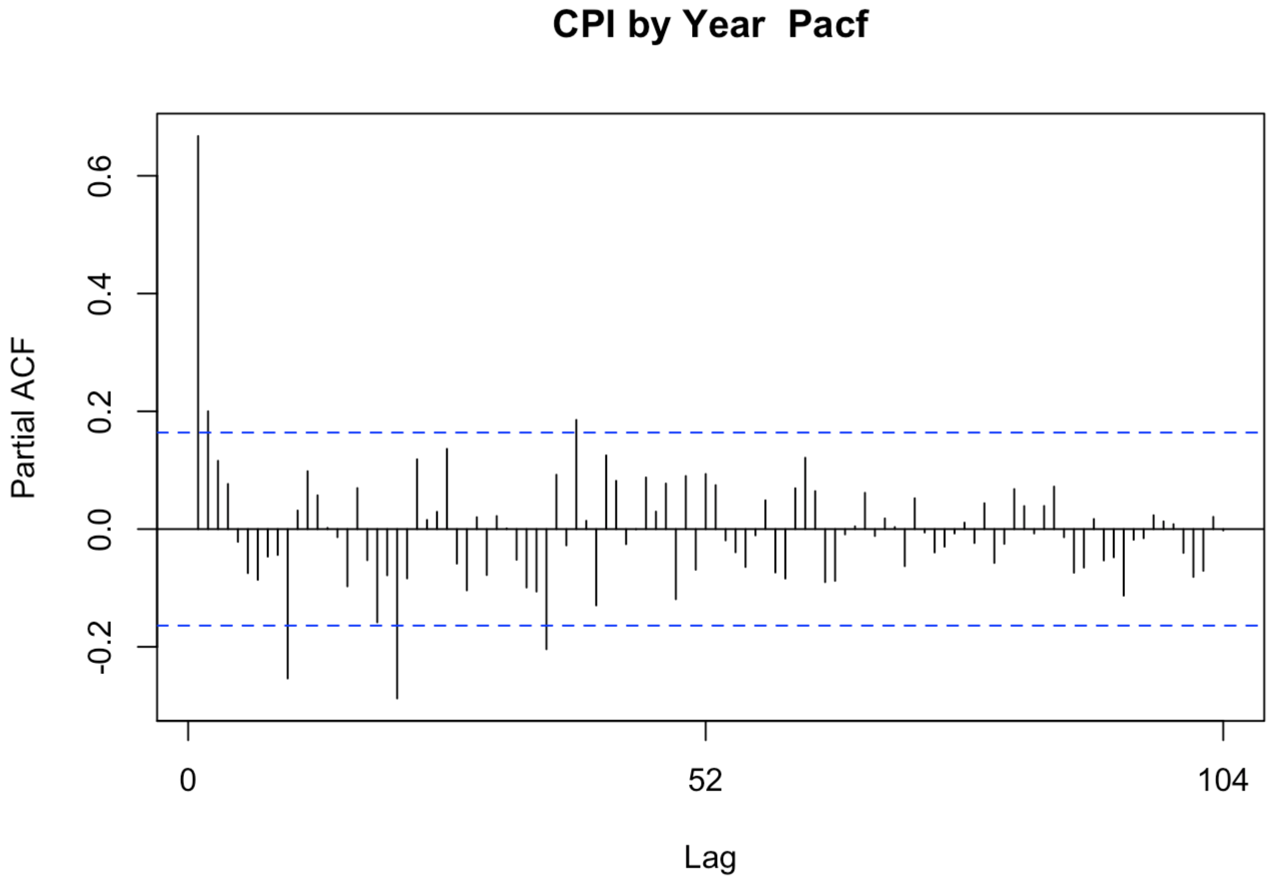

```
##
##  Augmented Dickey-Fuller Test
##
## data:  timeseries
## Dickey-Fuller = -2.5205, Lag order = 5, p-value = 0.3597
## alternative hypothesis: stationary
```

Time Series 5: CPI

```
store4_cpi.ts<-ts(store4_data$CPI,start =c(2010,5),freq=52)
tseda(store4_sales.ts,"Year","CPI","CPI by Year")
```


CPI by Year
CPI
Year
4000
6000
8000
10000
2010.0
2010.5
2011.0
2011.5
2012.0
2012.5


CPI by Year ACF
ACF
Lag
-0.2
0.0
0.2
0.4
0.6
0
52
104

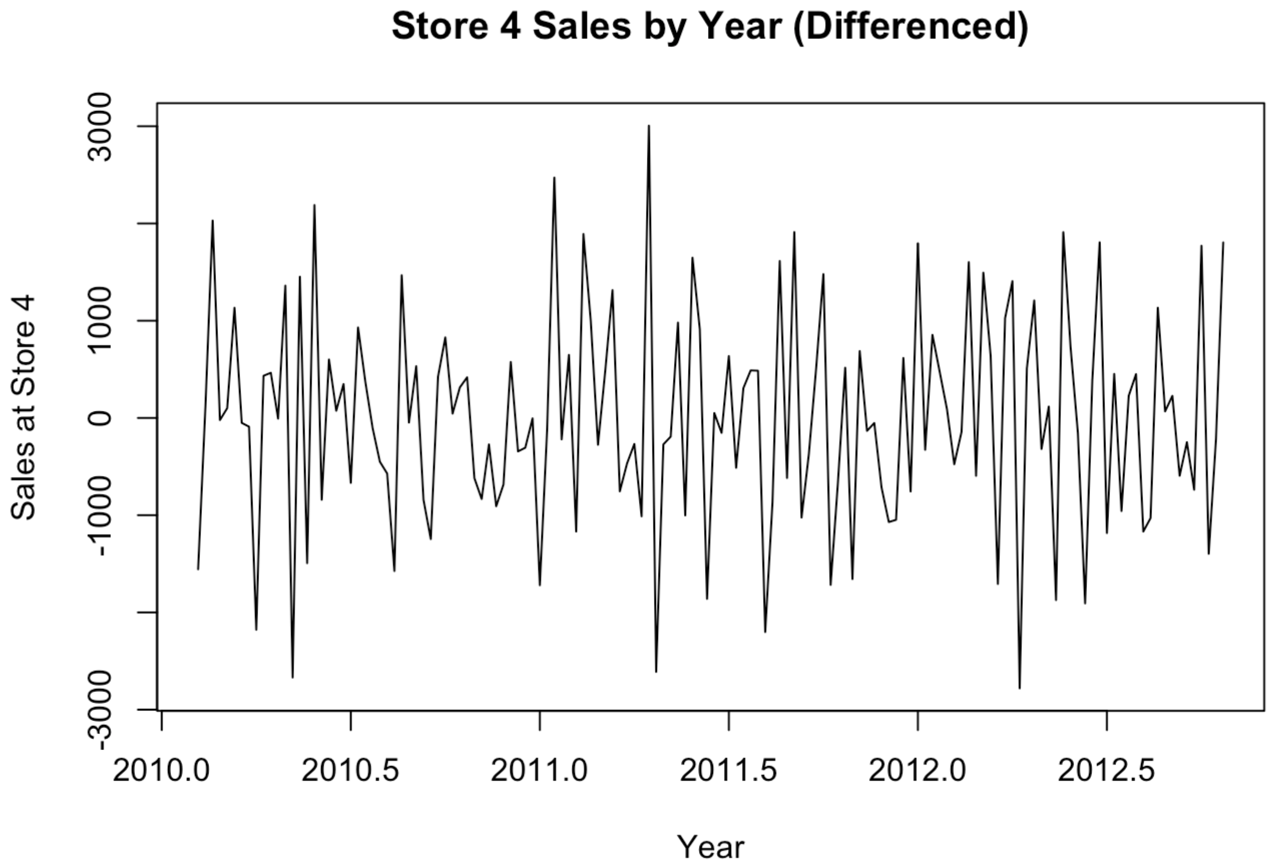


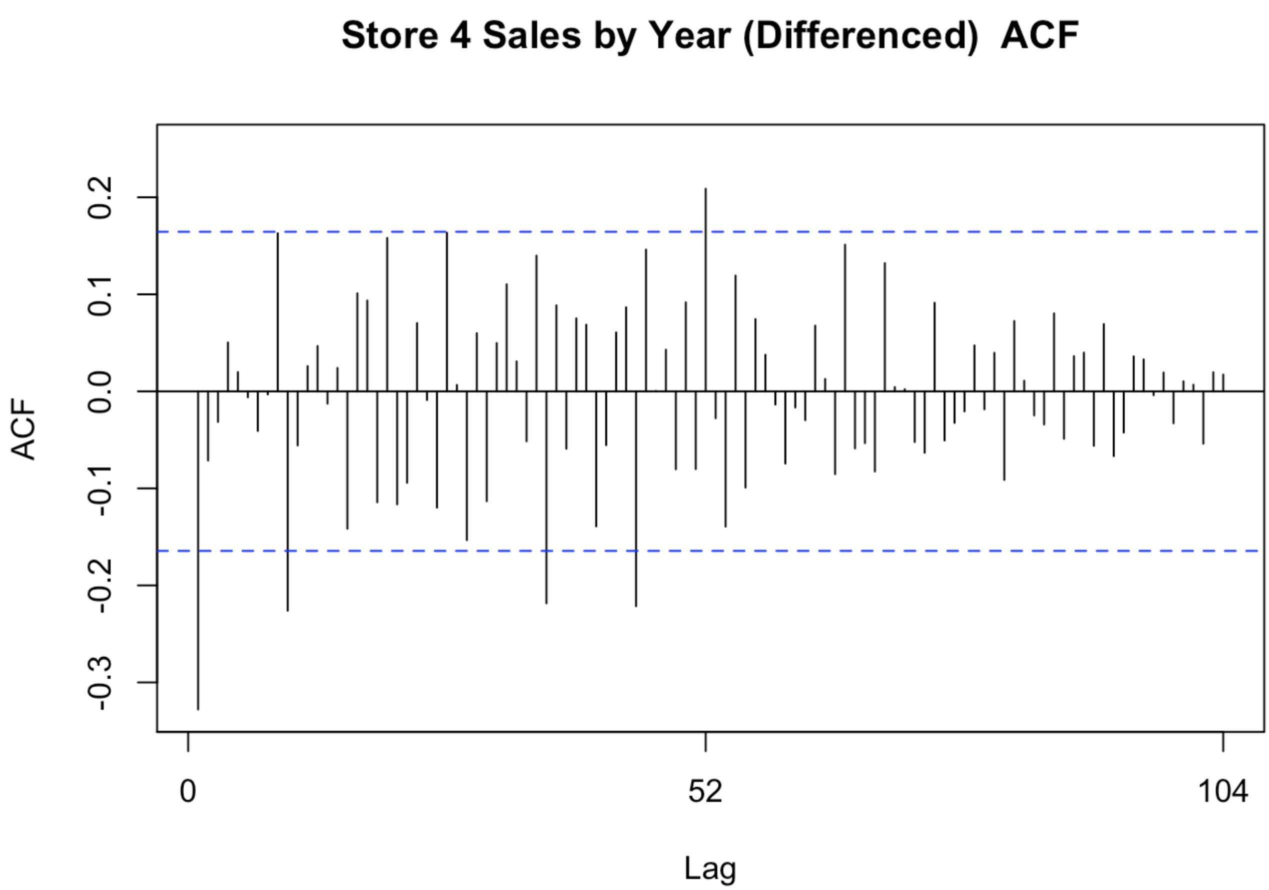

```
##
##  Augmented Dickey-Fuller Test
##
## data:  timeseries
## Dickey-Fuller = -3.0062, Lag order = 5, p-value = 0.1576
## alternative hypothesis: stationary
```

Converting non-stationary time series to stationary series by differencing

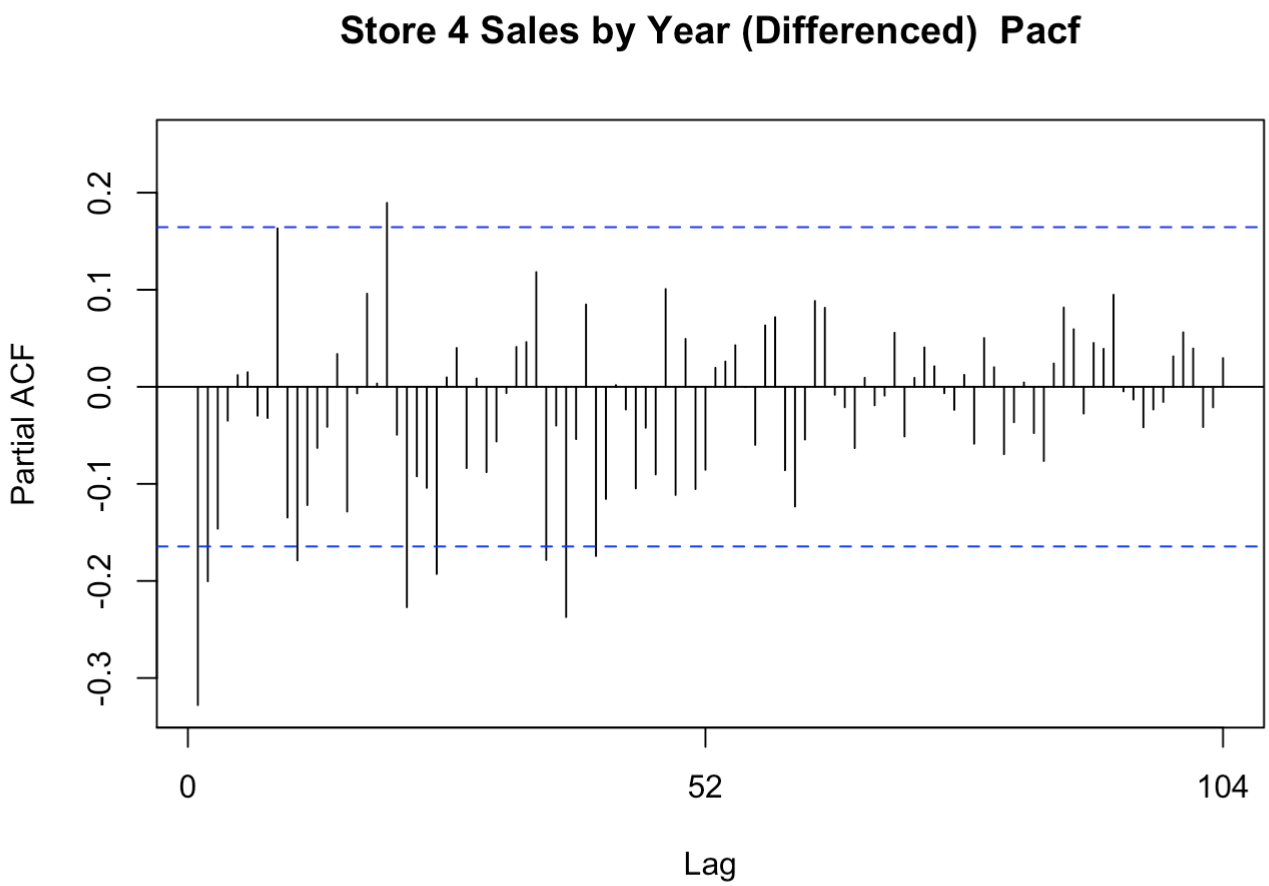
Time Series 1: Sales

```
store4_sales.d.ts<-diff(store4_sales.ts,differences=1,lag=1)
tseda(store4_sales.d.ts,"Year","Sales at Store 4","Store 4 Sales by Year (Differenced)")
```


Store 4 Sales by Year (Differenced)
Sales at Store 4
3000
1000
0
-1000
-3000
2010.0
2010.5
2011.0
2011.5
2012.0
2012.5
Year


Store 4 Sales by Year (Differenced) ACF
ACF
0.2
0.1
0.0
-0.1
-0.2
-0.3
0
52
104
Lag

```
## Warning in adf.test(timeseries): p-value smaller than printed p-value
```

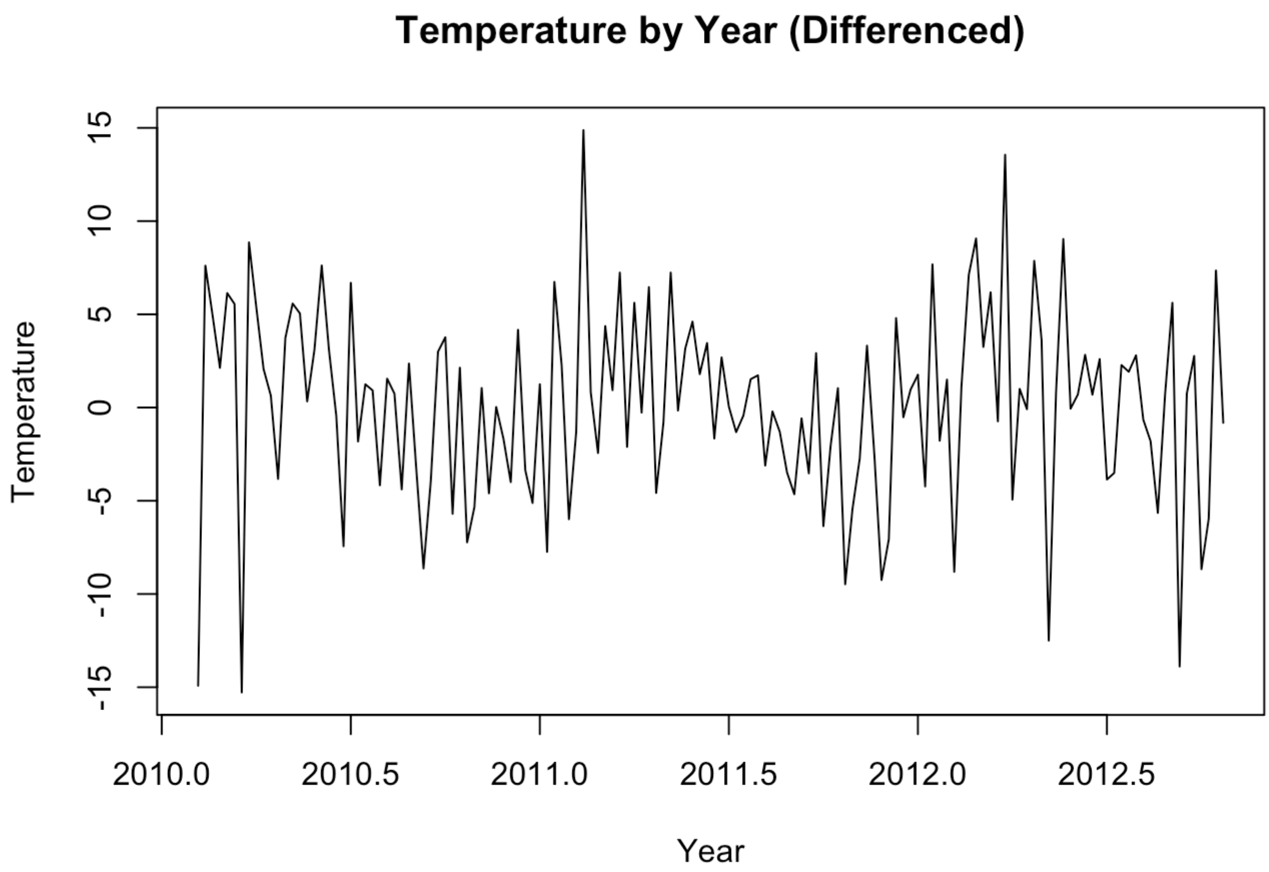


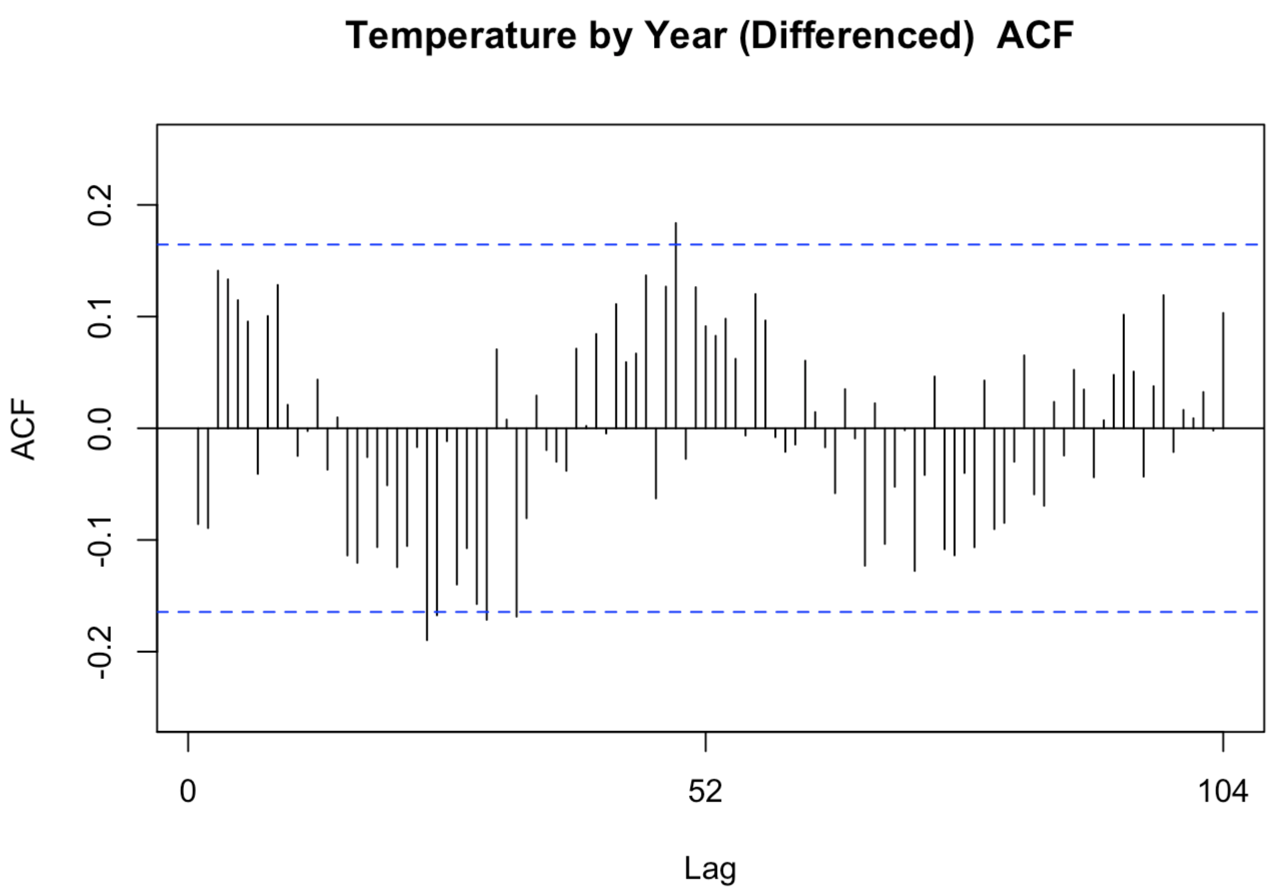

```
##
##  Augmented Dickey-Fuller Test
##
## data:  timeseries
## Dickey-Fuller = -5.3952, Lag order = 5, p-value = 0.01
## alternative hypothesis: stationary
```

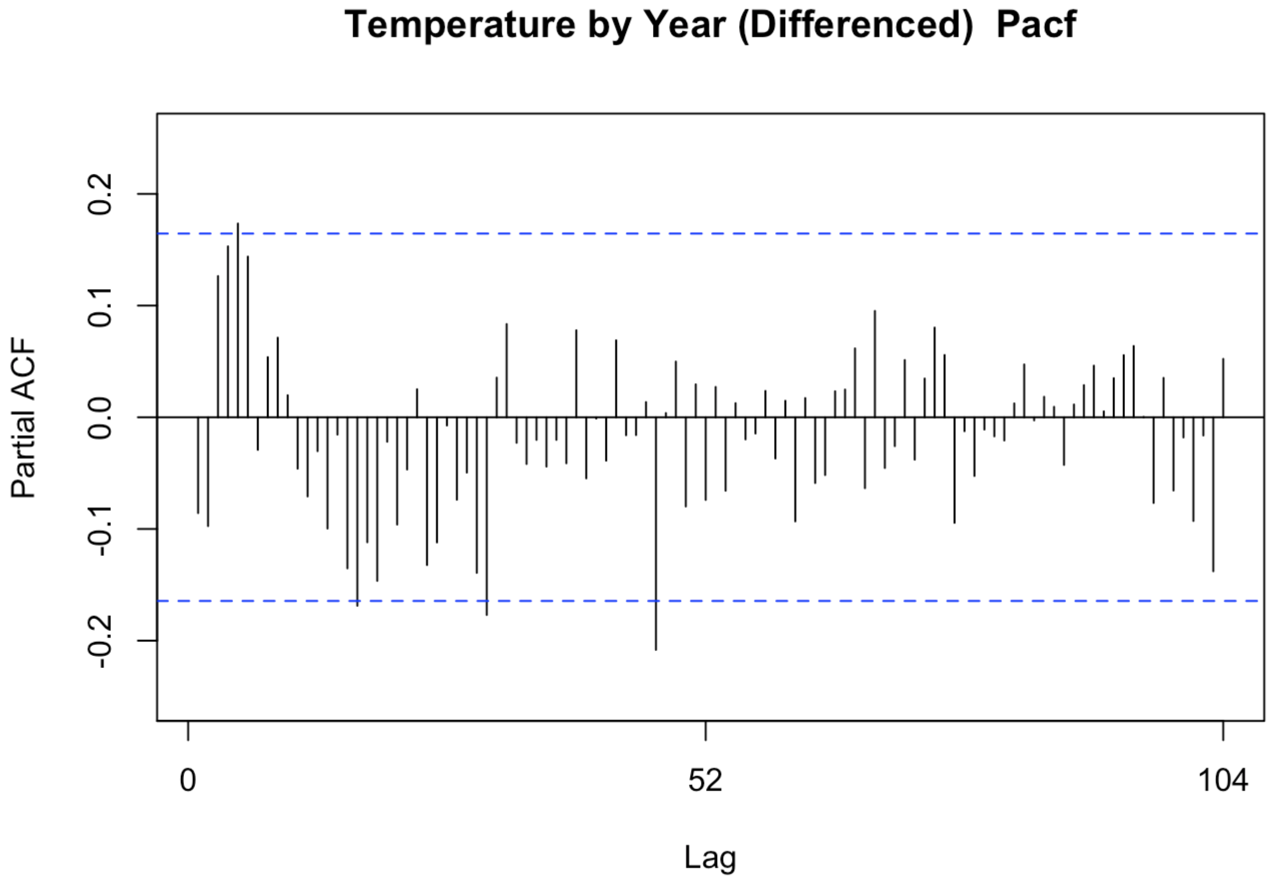
By differencing the “sales” time series once, we’re able to see via the plot and the ADF test that the differenced series is stationary. This means that the “sales” series was integrated with order 1 (I(1)). The differenced series has been stored in a new variable (store4_sales.d.ts)

Time Series 2: Temperature

```
store4_temp.d.ts<-diff(store4_temp.ts,differences=1,lag=1)
tseda(store4_temp.d.ts,"Year","Temperature","Temperature by Year (Differenced)")
```


Temperature by Year (Differenced)
Temperature
15
10
5
0
-5
-10
-15
2010.0
2010.5
2011.0
2011.5
2012.0
2012.5
Year


Temperature by Year (Differenced) ACF
ACF
0.2
0.1
0.0
-0.1
-0.2
0
52
104
Lag

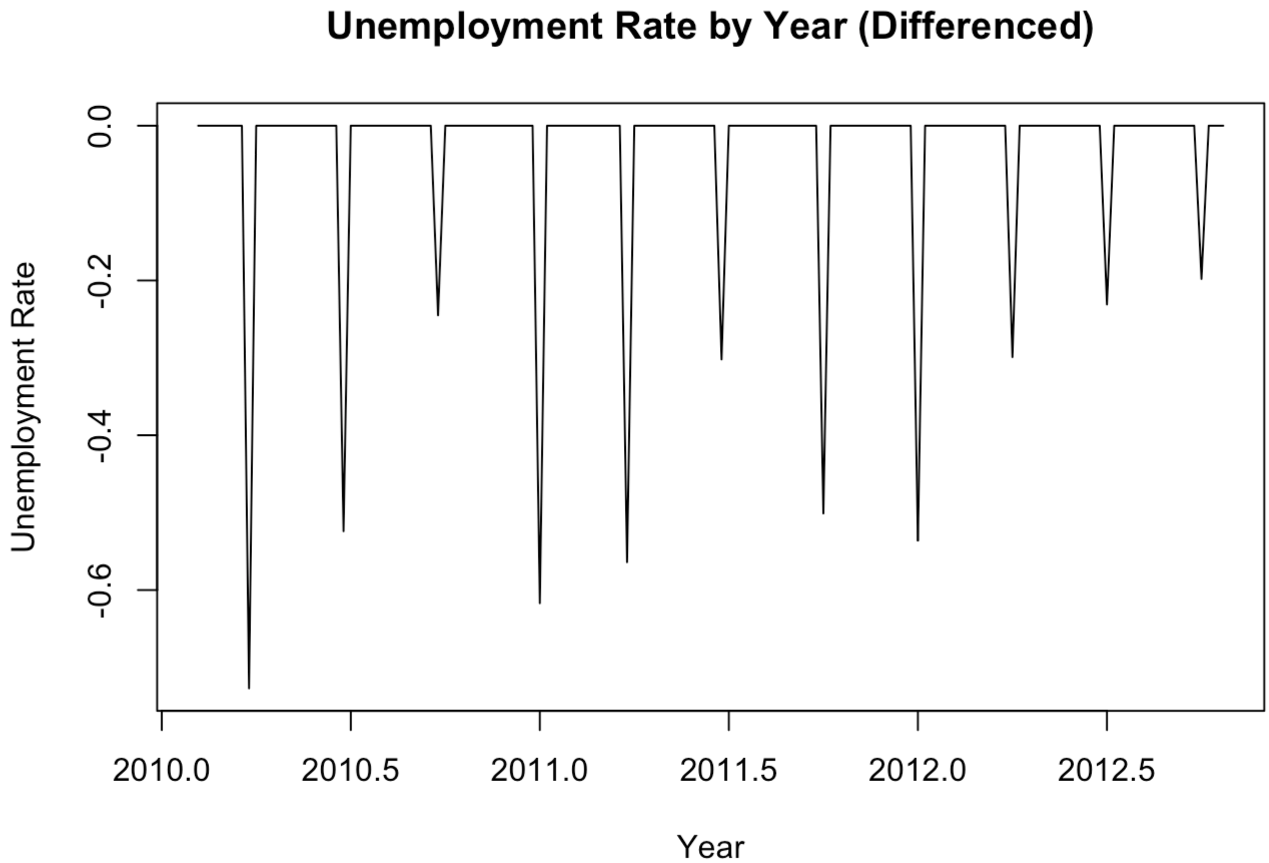



```
##
##  Augmented Dickey-Fuller Test
##
## data:  timeseries
## Dickey-Fuller = -2.9376, Lag order = 5, p-value = 0.1861
## alternative hypothesis: stationary
```

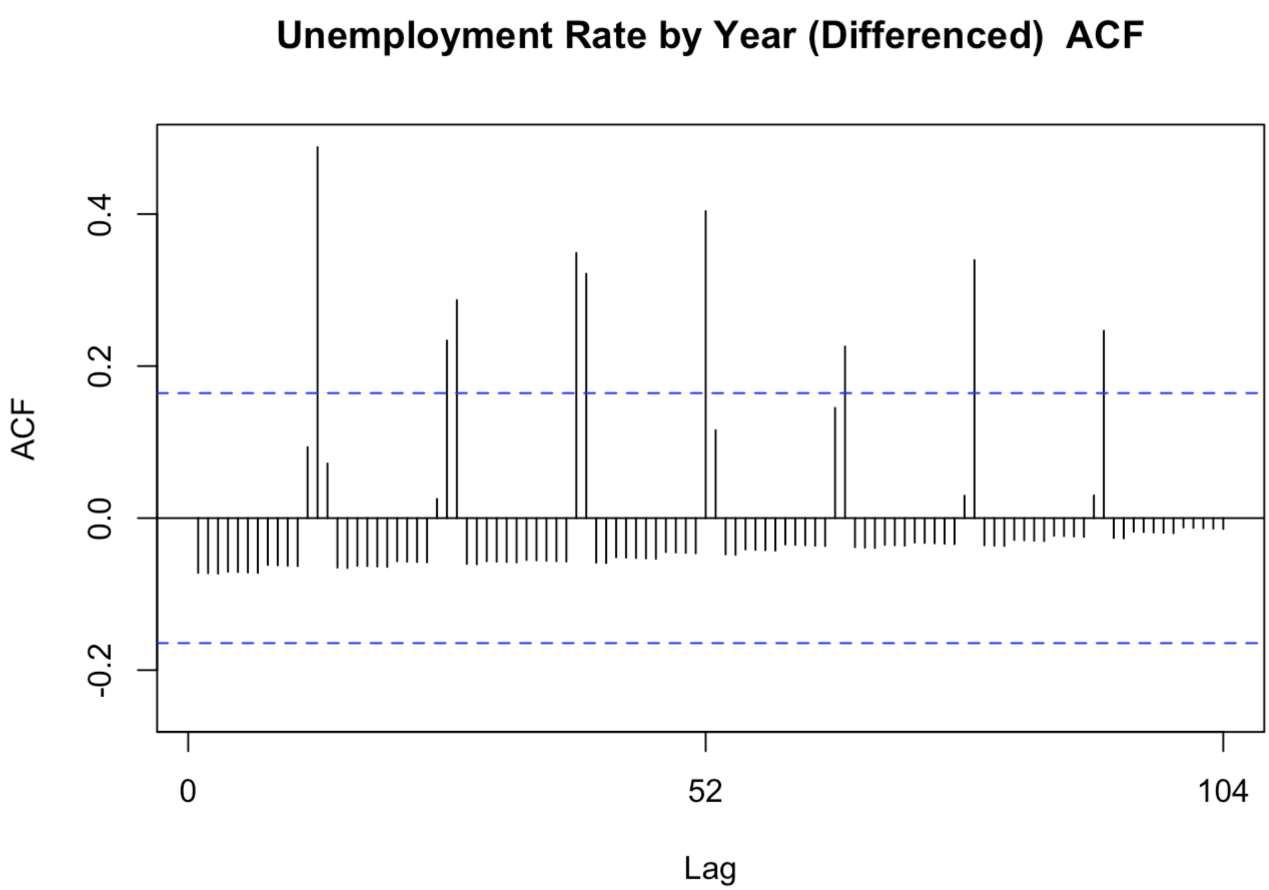
Differencing the “temperature” time series once has created a time series that “looks” stationary with the first order difference. However, the ADT test says that the null hypothesis that the process is still unit root can’t be rejected, and this can be seen in the damping of the ACF and the pacfs which are slightly significant. However, we want to avoid over-differencing and for our purposes here, we’re dont think our analysis would be impacted if we assumed that temperature is an I(1) series.

Time Series 3: Unemployment

```
store4_ue.d.ts<- diff(store4_ue.ts,differences=1,lag=1)
tseda(store4_ue.d.ts,"Year","Unemployment Rate","Unemployment Rate by Year (Difference
d)")
```

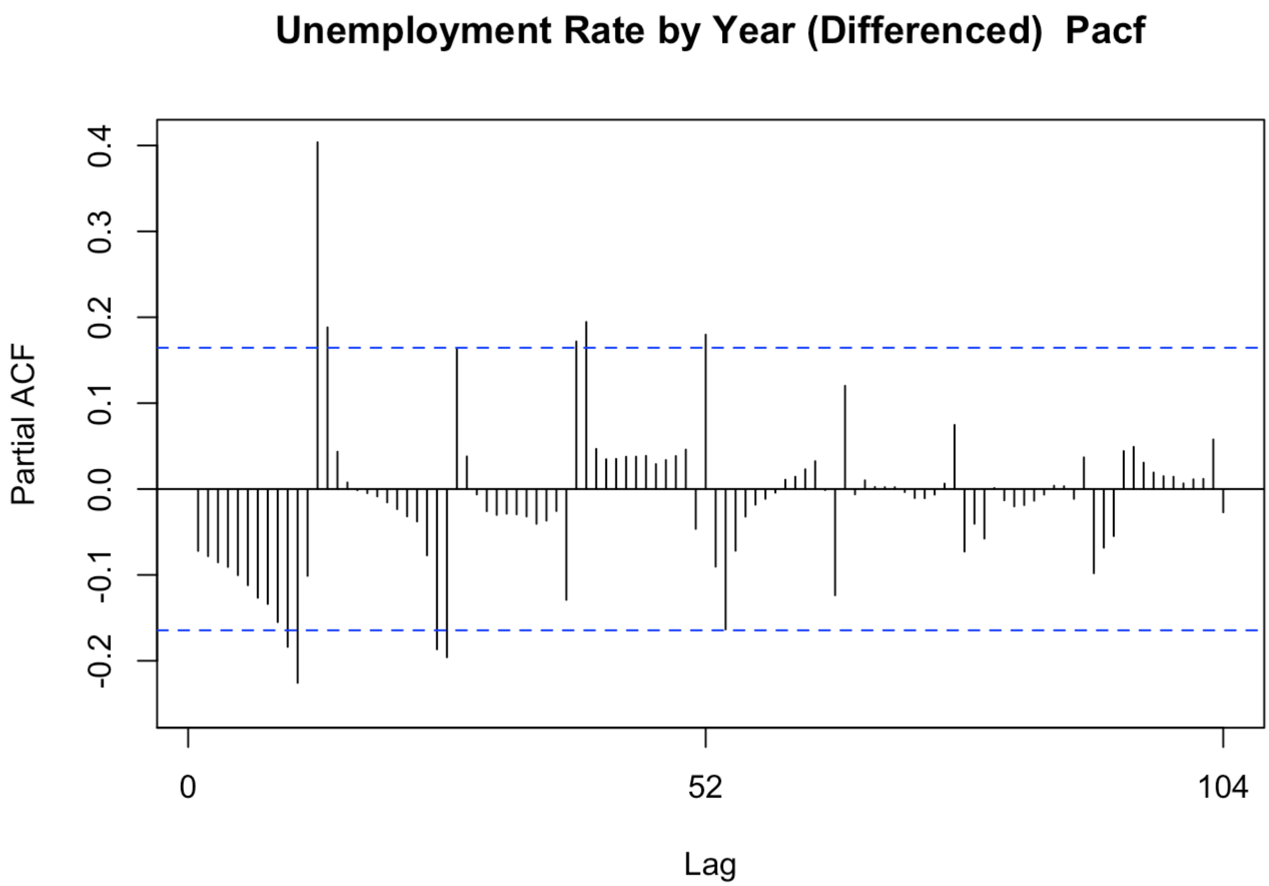


Unemployment Rate by Year (Differenced)
Unemployment Rate
0.0
-0.2
-0.4
-0.6
2010.0
2010.5
2011.0
2011.5
2012.0
2012.5
Year



Unemployment Rate by Year (Differenced) ACF
ACF
0.4
0.2
0.0
-0.2
0
52
104
Lag

```
## Warning in adf.test(timeseries): p-value smaller than printed p-value
```



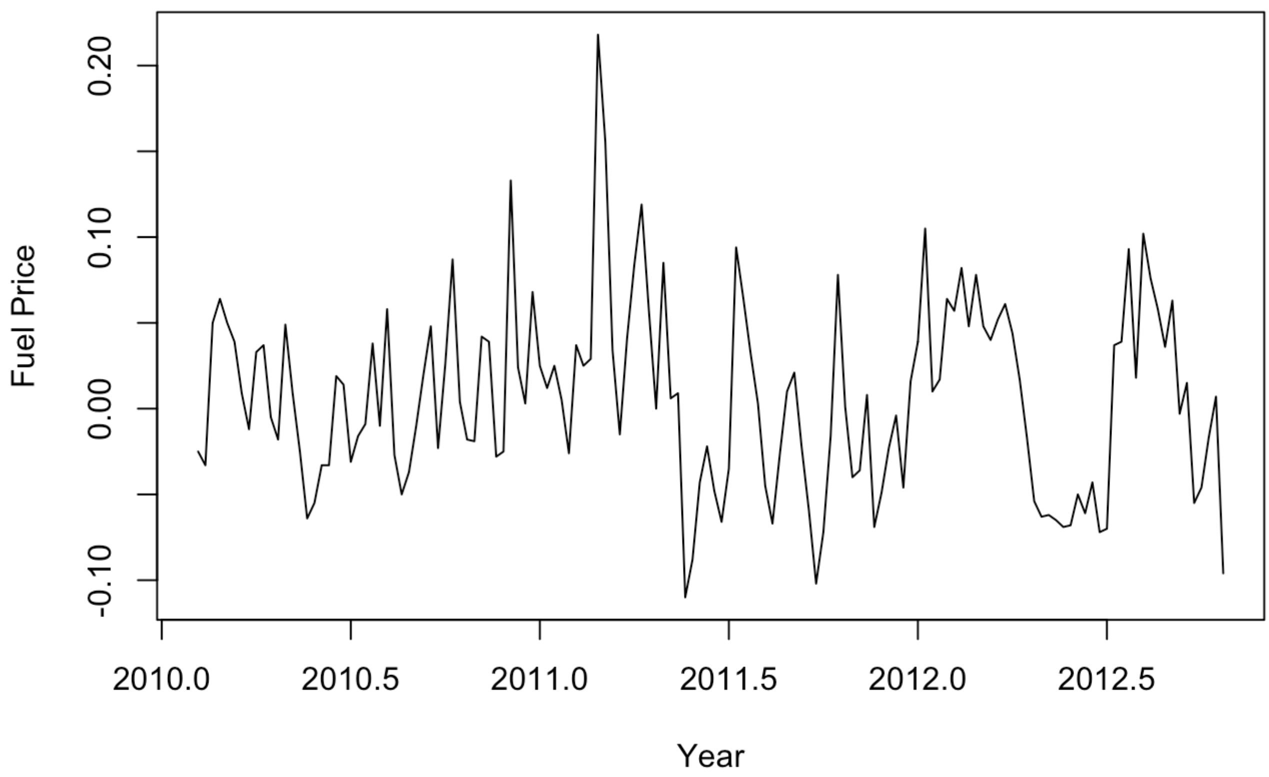

```
##
##  Augmented Dickey-Fuller Test
##
## data:  timeseries
## Dickey-Fuller = -6.6314, Lag order = 5, p-value = 0.01
## alternative hypothesis: stationary
```

Analyzing the unemployment variable after the first order difference, the ADF function returns a p value of 0.01 rejecting the null hypothesis of unit root stationarity

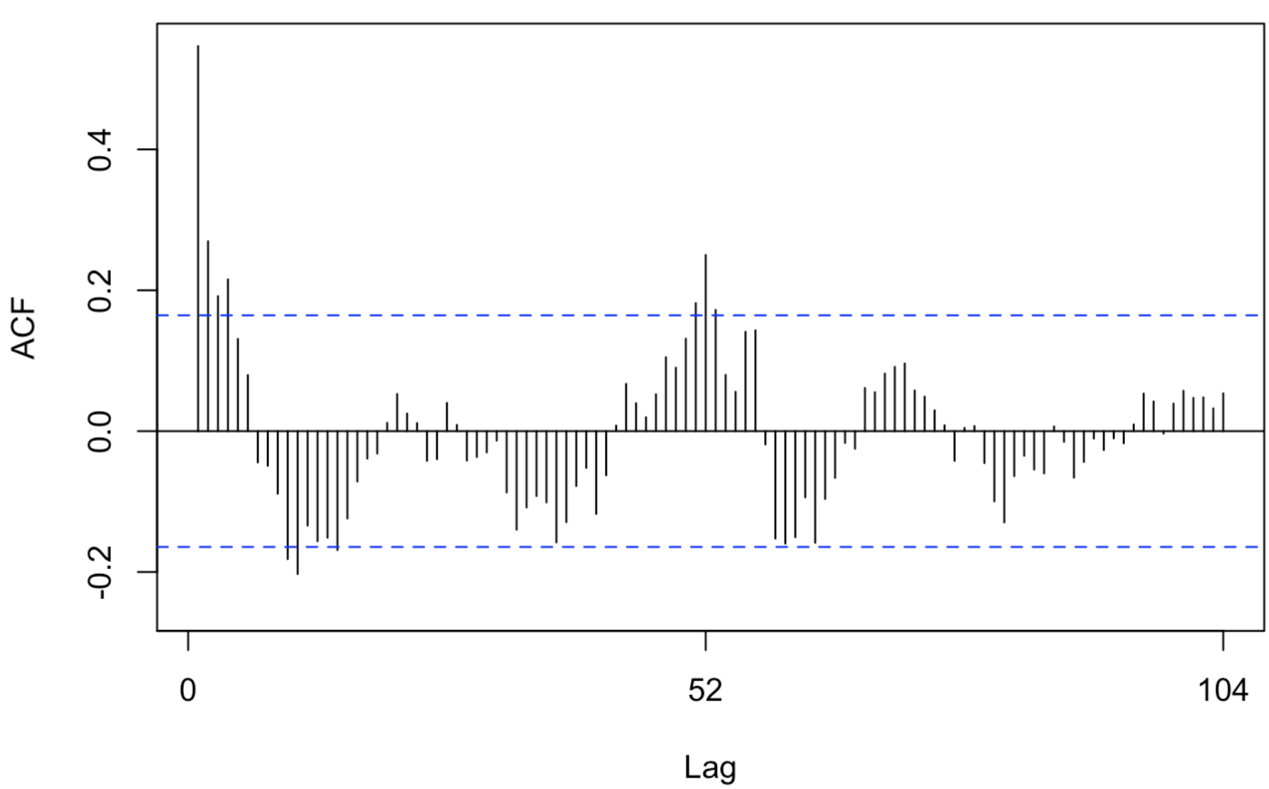
Time Series 4: Fuel Price

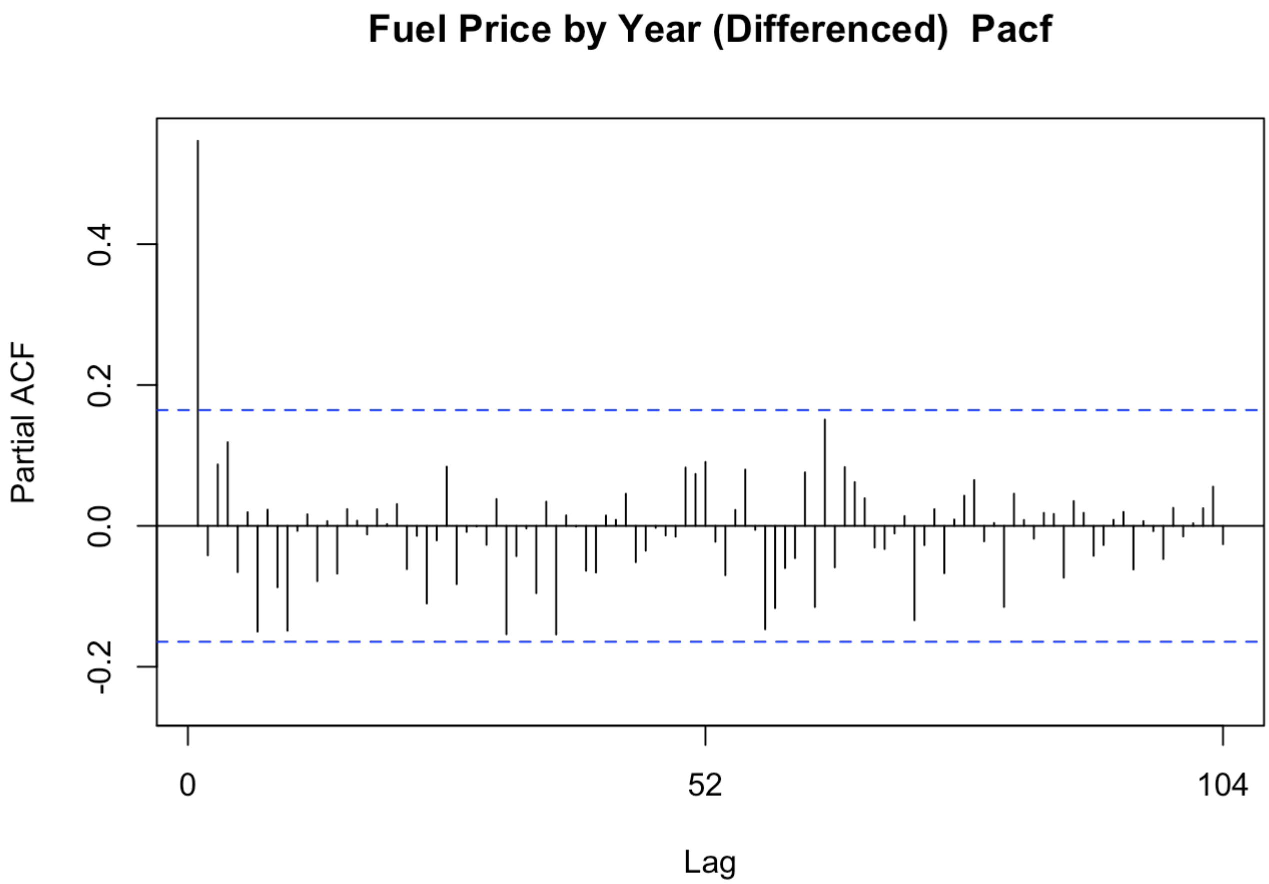
```
store4_fp.d.ts<-diff(store4_fp.ts,differences=1,lag=1)
tseda(store4_fp.d.ts,"Year","Fuel Price","Fuel Price by Year (Differenced)")
```

Fuel Price by Year (Differenced)



Fuel Price by Year (Differenced) ACF



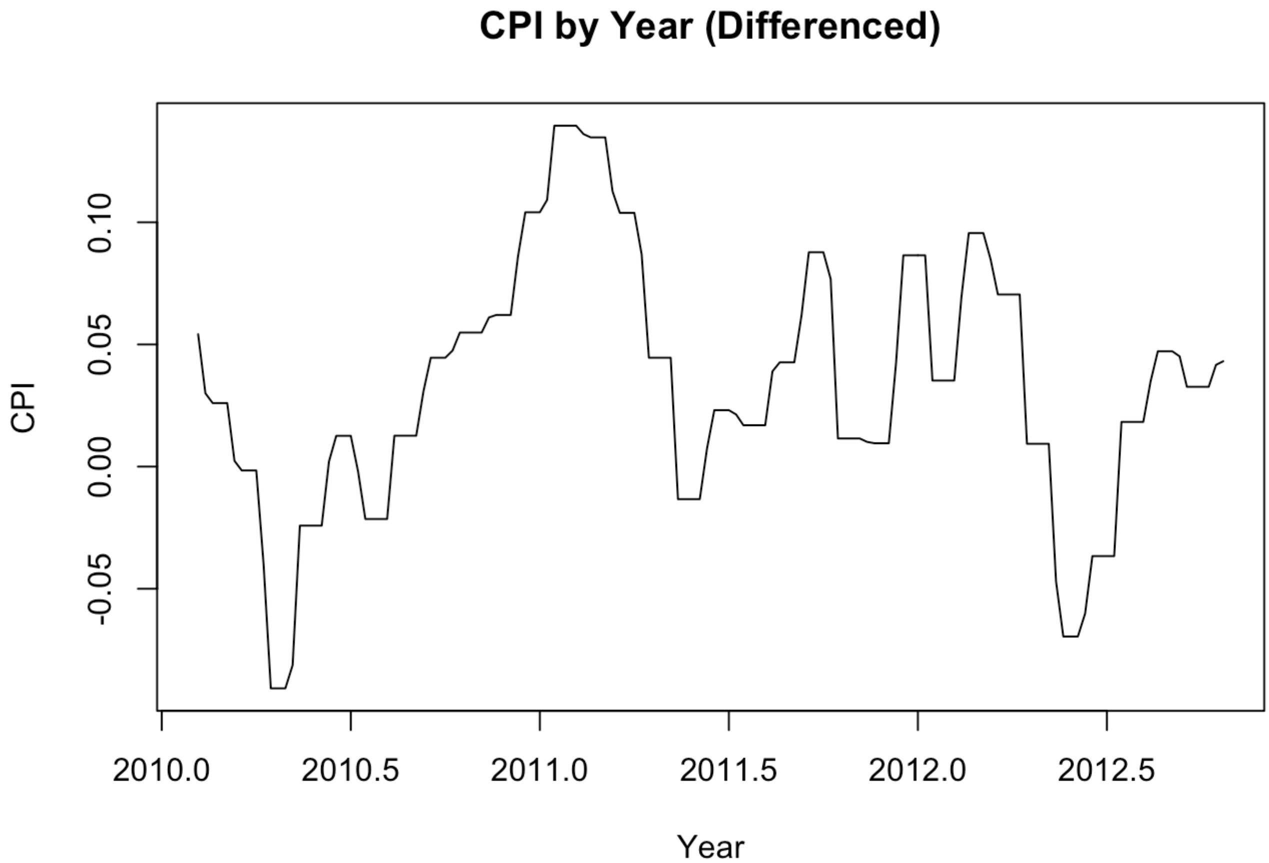


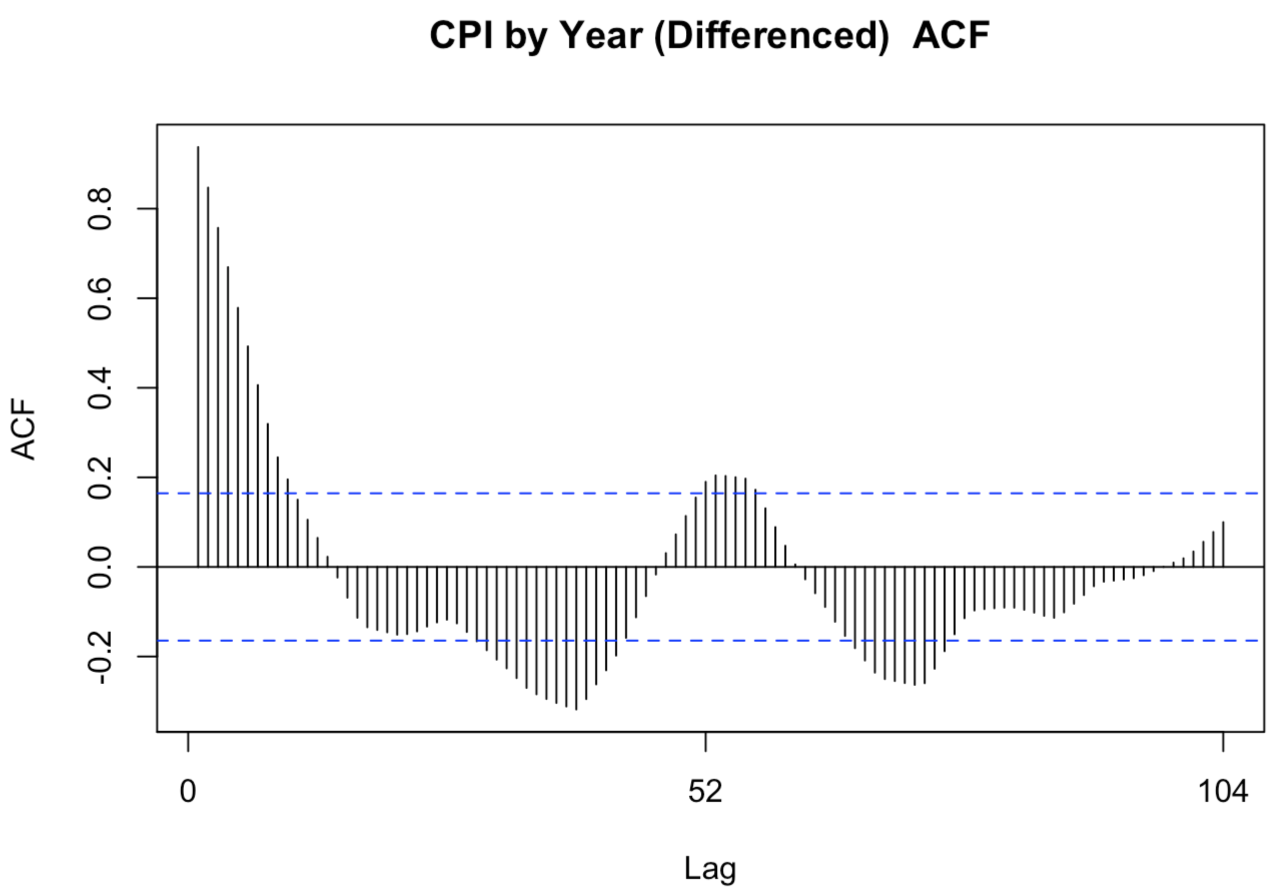

```
##
##  Augmented Dickey-Fuller Test
##
## data:  timeseries
## Dickey-Fuller = -3.4417, Lag order = 5, p-value = 0.05046
## alternative hypothesis: stationary
```

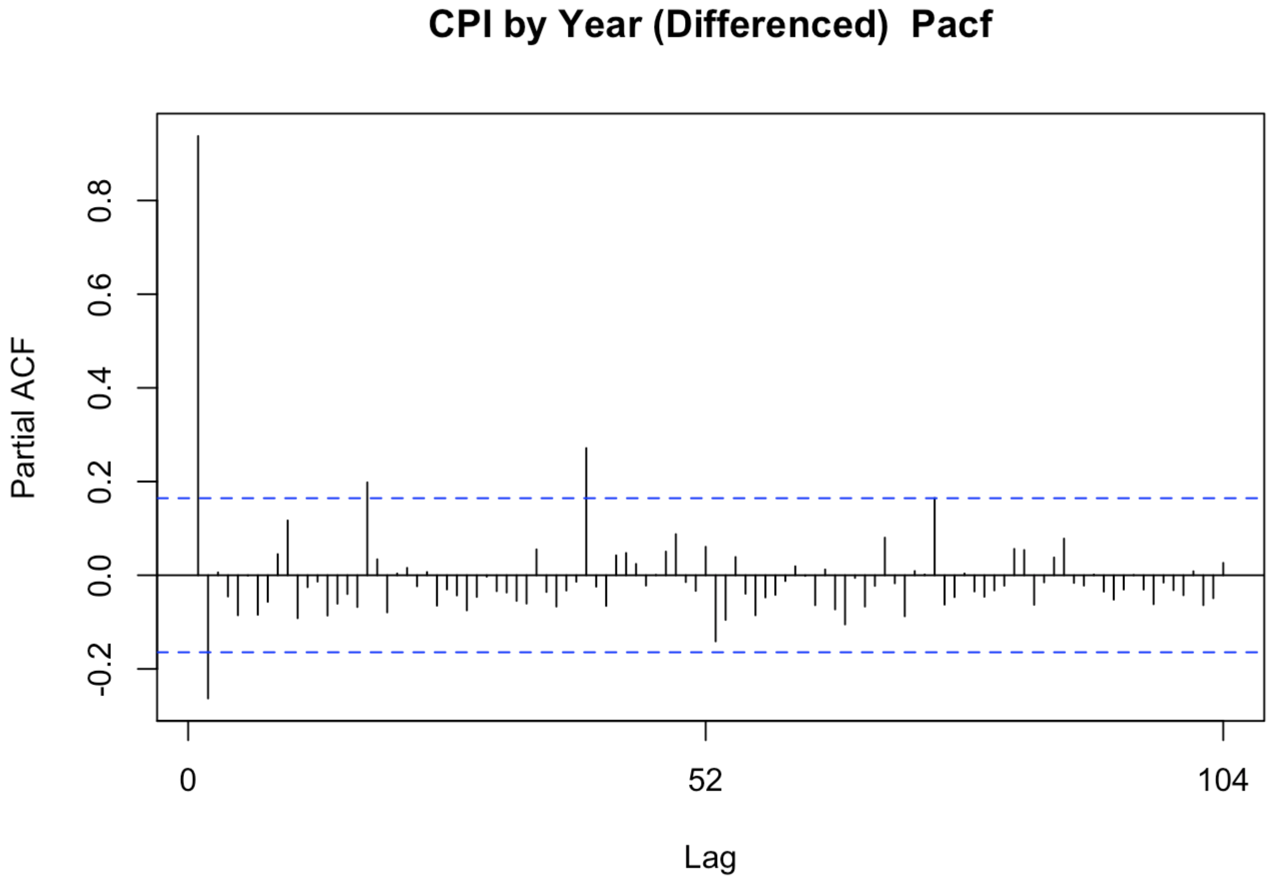
The first order difference of the “fuel price” time series satisfies the conditions of stationarity from an analysis of the acf,pacf and the ADF test.

Time Series 5: CPI

```
store4_cpi.d.ts<-diff(store4_cpi.ts,differences=1,lag=1)
tseda(store4_cpi.d.ts,"Year","CPI","CPI by Year (Differenced)")
```

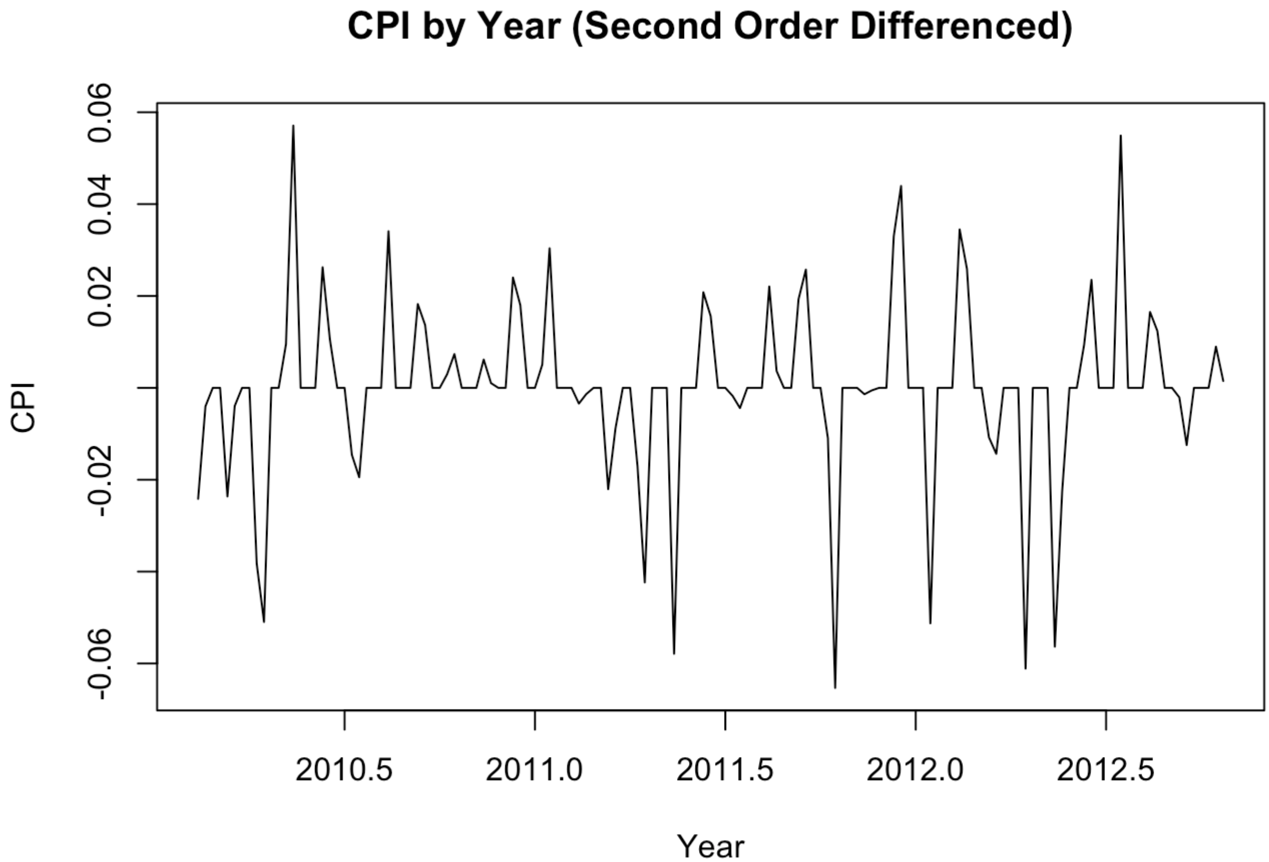

CPI by Year (Differenced)
CPI
0.10
0.05
0.00
-0.05
2010.0
2010.5
2011.0
2011.5
2012.0
2012.5
Year

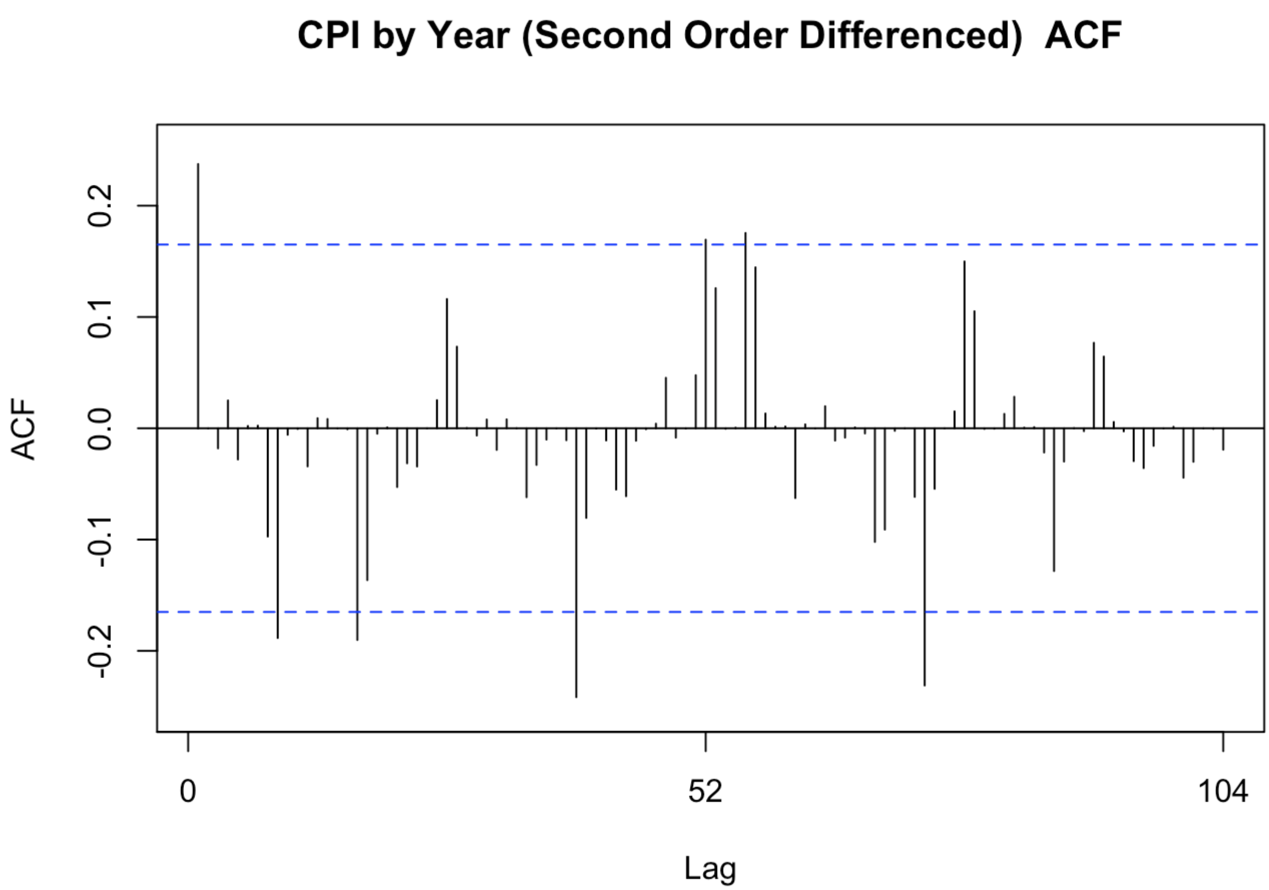

CPI by Year (Differenced) ACF
ACF
0.8
0.6
0.4
0.2
0.0
-0.2
0
52
104
Lag



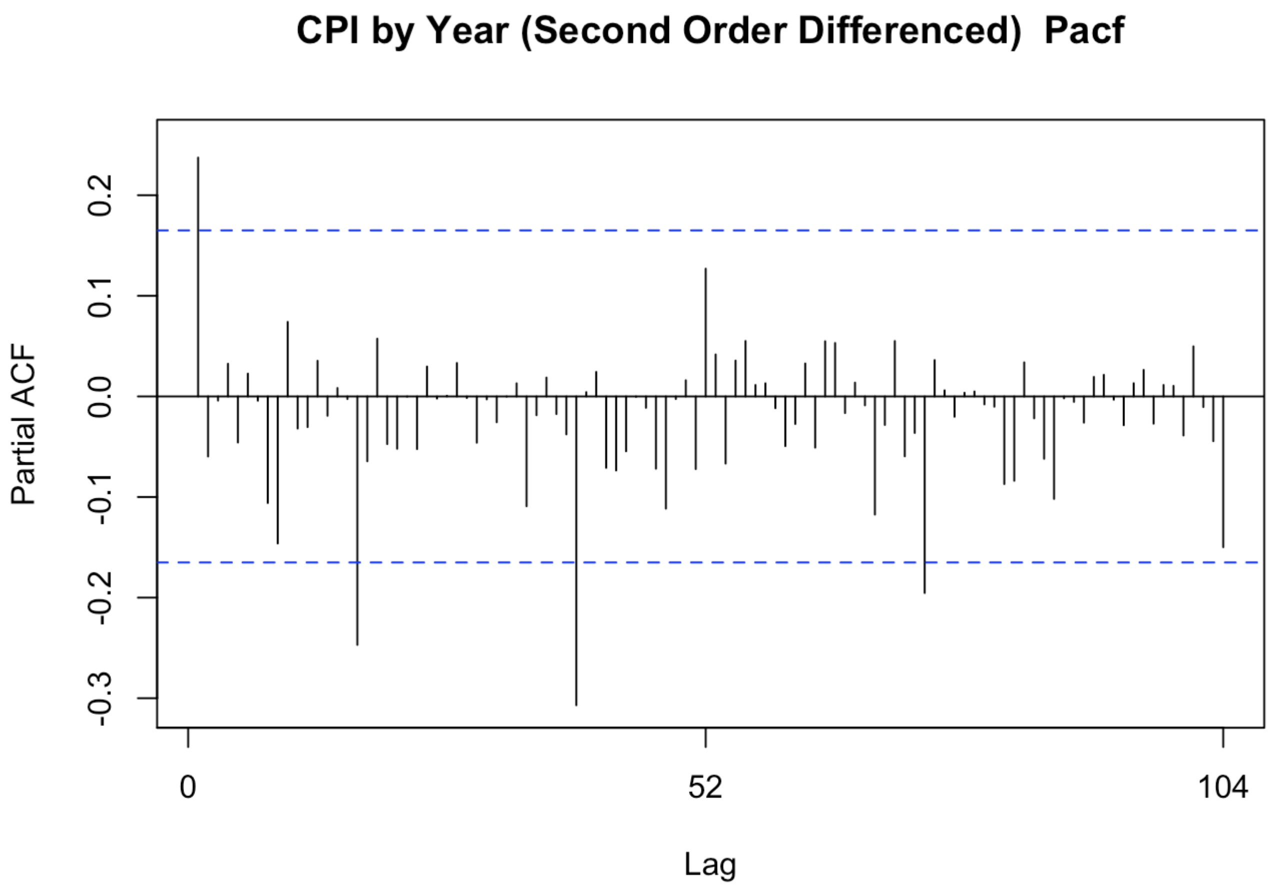

```
##
##  Augmented Dickey-Fuller Test
##
## data:  timeseries
## Dickey-Fuller = -2.7002, Lag order = 5, p-value = 0.285
## alternative hypothesis: stationary
```

```
store4_cpi.d.ts<-diff(store4_cpi.d.ts,differences=1,lag=1)
tseda(store4_cpi.d.ts,"Year","CPI","CPI by Year (Second Order Differenced)")
```


CPI by Year (Second Order Differenced)
0.06
0.04
0.02
-0.02
-0.06
CPI
2010.5
2011.0
2011.5
2012.0
2012.5
Year


CPI by Year (Second Order Differenced) ACF
0.2
0.1
0.0
-0.1
-0.2
ACF
0
52
104
Lag

```
## Warning in adf.test(timeseries): p-value smaller than printed p-value
```

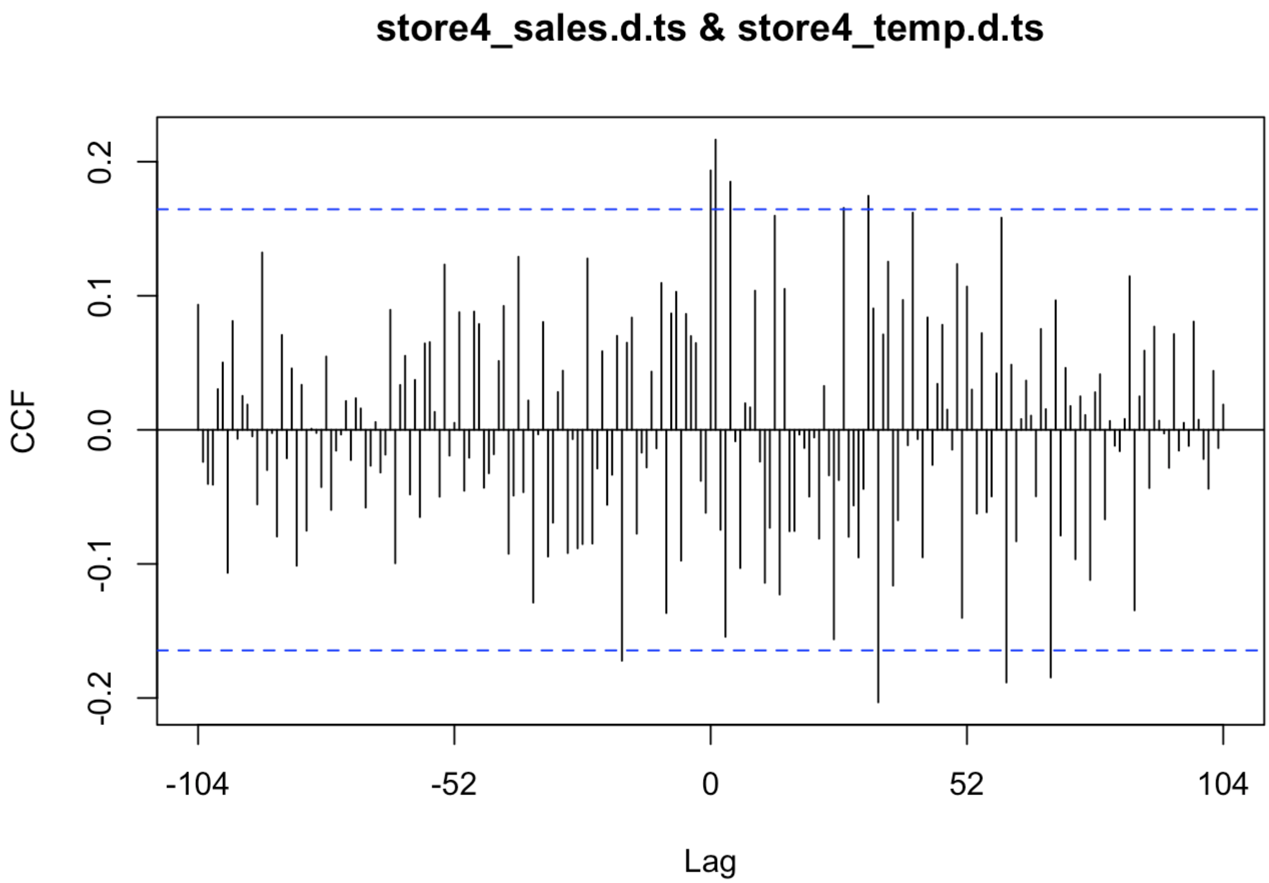



```
##
##  Augmented Dickey-Fuller Test
##
## data:  timeseries
## Dickey-Fuller = -4.5331, Lag order = 5, p-value = 0.01
## alternative hypothesis: stationary
```

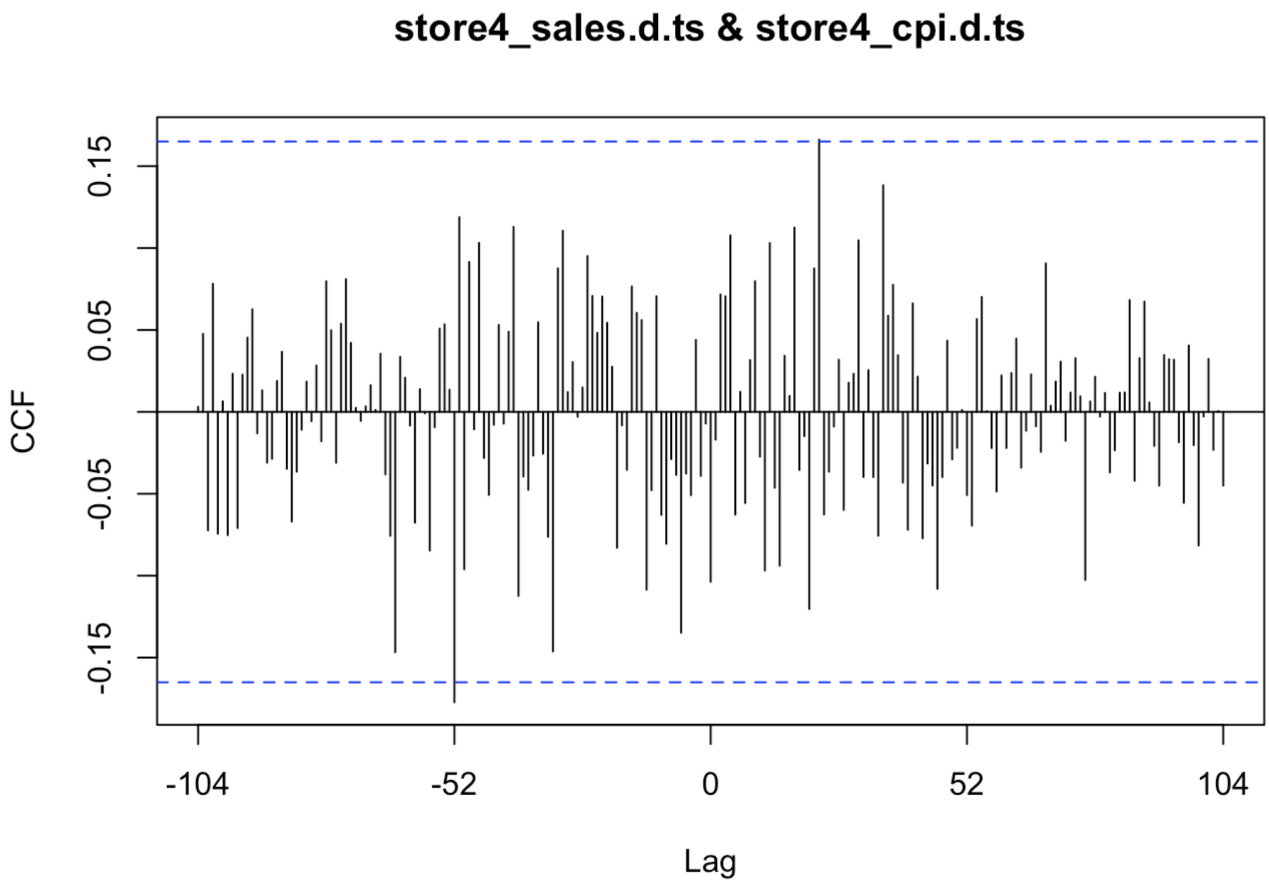
After differencing once, the store4_cpi.ts time series continues to show strong evidence that it is a unit root series (through the ACF, PACF and ADF test). We have differenced the series once more to make it stationary.

Analyze cross-correlations

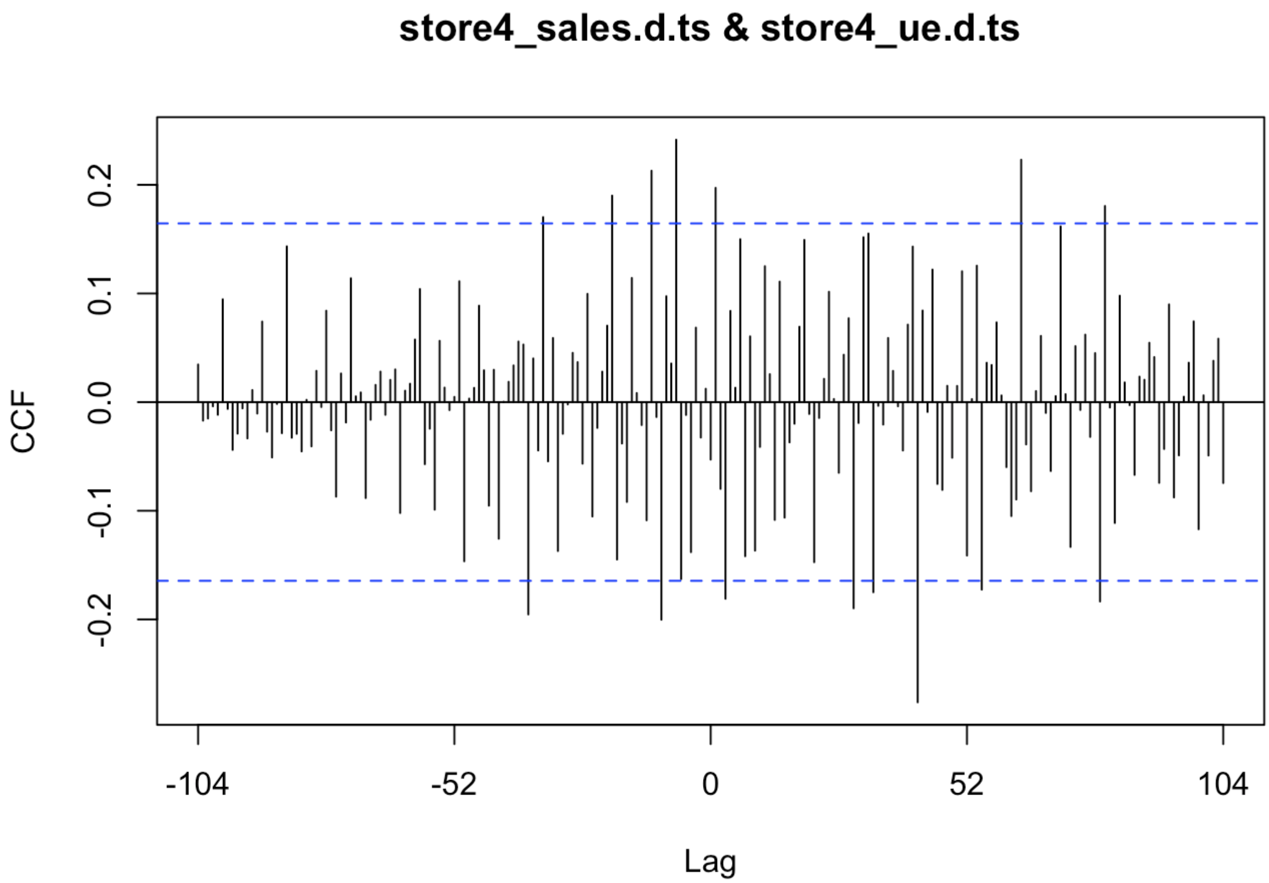
```
Ccf(store4_sales.d.ts,store4_temp.d.ts)
```



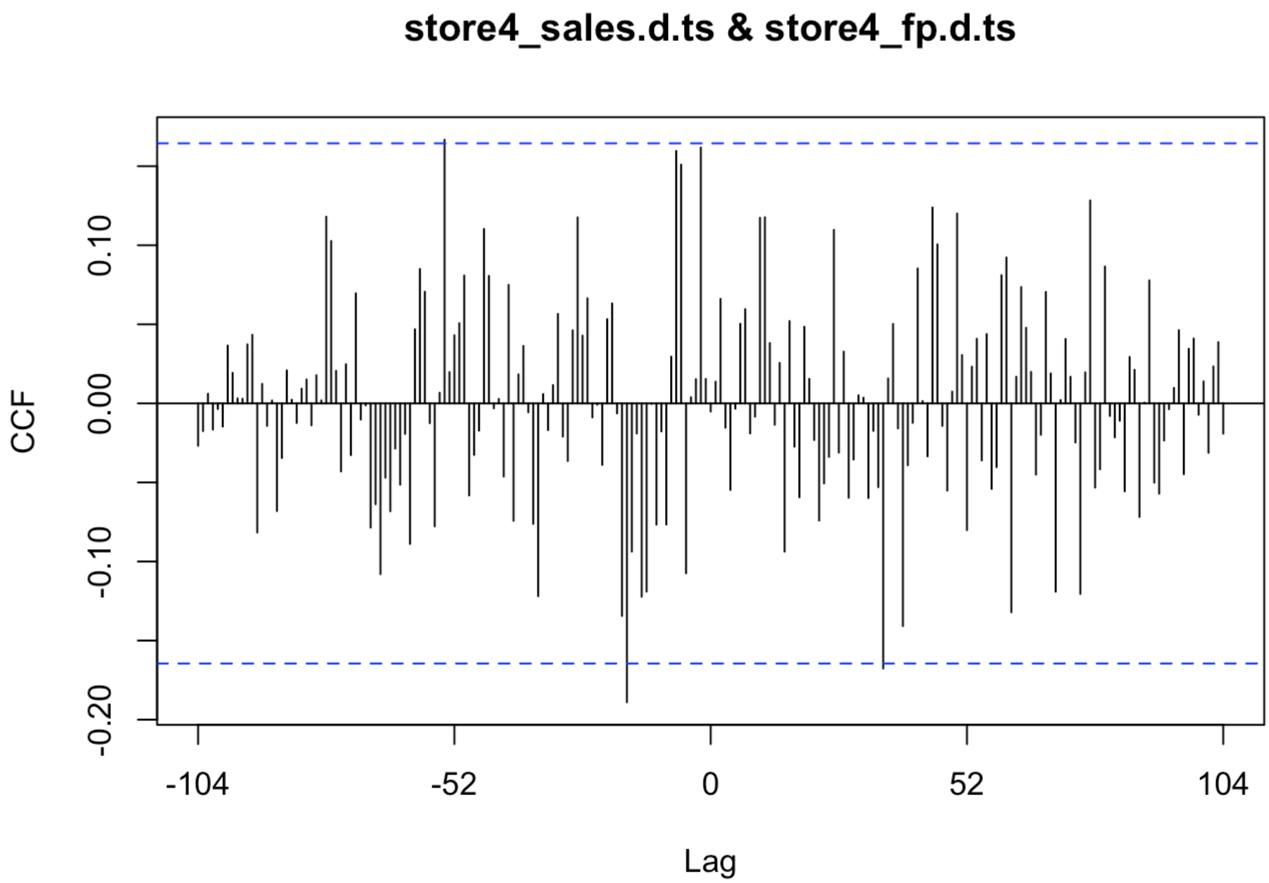

```
Ccf(store4_sales.d.ts,store4_cpi.d.ts)
```




```
Ccf(store4_sales.d.ts,store4_ue.d.ts)
```




```
Ccf(store4_sales.d.ts,store4_fp.d.ts)
```



Cross correlations are evaluated against differenced time series below.

- Based on the cross-correlation betwen temperature and sales, it appears as though there may not be a significant relationship between these two time series after the first 2-3 lags, although the first lag seems to be somewhat strongly correlated

*CPI and Sales do not seem to have any cross correlations of any importance at all

*Unemployment and Sales seem to have some small cross-correlations of interest although many seem coincidental (for example the 0.2 cross correlation at around lag 60)

*Fuel price and sales do not seem to have any cross correlations of any importance at all

We will model all these relationships during estimation, but based on these cross-correlations, the relationship between Sales and Temperature seems to be the most important, while unemployment may be of secondary importance.

Analyze cointegrations

```
po.test(cbind(store4_temp.ts,store4_sales.ts))
```

```
##
##  Phillips-Ouliaris Cointegration Test
##
## data:  cbind(store4_temp.ts, store4_sales.ts)
## Phillips-Ouliaris demeaned = -22.979, Truncation lag parameter =
## 1, p-value = 0.03126
```

```
po.test(cbind(diff(store4_cpi.ts),store4_sales.ts))
```

```
##
##  Phillips-Ouliaris Cointegration Test
##
## data:  cbind(diff(store4_cpi.ts), store4_sales.ts)
## Phillips-Ouliaris demeaned = -15.12, Truncation lag parameter = 1,
## p-value = 0.1449
```

```
po.test(cbind(store4_ue.ts,store4_sales.ts))
```

```
## Warning in po.test(cbind(store4_ue.ts, store4_sales.ts)): p-value greater
## than printed p-value
```

```
##
##  Phillips-Ouliaris Cointegration Test
##
## data:  cbind(store4_ue.ts, store4_sales.ts)
## Phillips-Ouliaris demeaned = -1.9539, Truncation lag parameter =
## 1, p-value = 0.15
```

```
po.test(cbind(store4_fp.ts,store4_sales.ts))
```

```
## Warning in po.test(cbind(store4_fp.ts, store4_sales.ts)): p-value greater
## than printed p-value
```

```
##
##  Phillips-Ouliaris Cointegration Test
##
## data:  cbind(store4_fp.ts, store4_sales.ts)
## Phillips-Ouliaris demeaned = -2.7698, Truncation lag parameter =
## 1, p-value = 0.15
```

- Based on this analysis, it appears as if sales are cointegrated with temperature (p-value of 0.03). In order to correctly model cointegrated series, we would need to use VECM (Vector Error Correcton Models) which have not been covered in this course so far. For the purpose of this lab, we will proceed by building standard VAR models after making a note of this detail.

Estimation

Break data into in-sample and out of Sample Data Sets

We are breaking the time-series' up into in-sample and out-of-sample components. The in-sample set has 129 rows while the out of sample data contains 13 rows (10%).

```
istartindex=2
iendindex=130
ostartindex=iendindex+1
oendindex=dim(as.matrix(store4_sales.d.ts))[c(1)]

modelinput = cbind(store4_sales.d.ts,store4_temp.d.ts,store4_ue.d.ts,store4_fp.d.ts,stor
e4_cpi.d.ts)[istartindex:iendindex,]
modeloos=cbind(store4_sales.d.ts,store4_temp.d.ts,store4_ue.d.ts,store4_fp.d.ts,store4_c
pi.d.ts)[ostartindex:oendindex,]
```

Estimate VAR model

This function will be used to calculate out of sample RSME's for our models

```
oosrmse<-function(model){
  p<-predict(model,n.ahead=oendindex-iendindex+1)
  rmse(p$fcst$store4_sales.d.ts[,c(1)],modeloos[,c(1)])
}
```

AR Model with Sales alone

```
modelar<-ar(modelinput[,c(1)],method = 'ols',dmean=T,intercept=F)
summary(modelar)
```

```
##              Length Class  Mode
## order          1    -none- numeric
## ar             3    -none- numeric
## var.pred       1    -none- numeric
## x.mean         1    -none- numeric
## x.intercept    0    -none- NULL
## aic           22    -none- numeric
## n.used         1    -none- numeric
## order.max      1    -none- numeric
## partialacf     0    -none- NULL
## resid        129    -none- numeric
## method         1    -none- character
## series         1    -none- character
## frequency      1    -none- numeric
## call           5    -none- call
## asy.se.coef    2    -none- list
```

```
modelar$ar
```

```
## , , 1
##
##            [,1]
## [1,] -0.4252121
## [2,] -0.2242020
## [3,] -0.1420759
```

```
modelar$aic
```

```
##          0          1          2          3          4          5
## 17.0175335  4.7734886  0.2639013  0.0000000  2.6019130  3.2694219
##          6          7          8          9         10         11
##  6.0285055  9.0581929  7.0816763  7.3688286  6.3327024  4.8976876
##         12         13         14         15         16         17
##  1.3652162  1.5722340  3.2096832  5.0598598  3.6140768  6.7368885
##         18         19         20         21
##  7.5433646 10.6961916  5.2510607  6.6547486
```

For a simple AR model, the fit only considers the first three lag terms. This gives us a hint which is further validated by the Acf and Pacf of sales that higher order lag terms may not be significant when it pertains to the sales variable.

Model1: Sales+Temperature

```
VARselect(modelinput[,c(1,2)], lag.max = 6)
```

```
## $selection
## AIC(n)  HQ(n)  SC(n) FPE(n)
##      5      1      1      5
##
## $criteria
##                   1            2            3            4            5
## AIC(n) 1.698624e+01 1.699175e+01 1.700132e+01 1.698986e+01 1.695441e+01
## HQ(n)  1.704197e+01 1.708462e+01 1.713134e+01 1.715702e+01 1.715872e+01
## SC(n)  1.712342e+01 1.722038e+01 1.732140e+01 1.740140e+01 1.745740e+01
## FPE(n) 2.382544e+07 2.395859e+07 2.419277e+07 2.392370e+07 2.310047e+07
##                   6
## AIC(n) 1.698216e+01
## HQ(n)  1.722362e+01
## SC(n)  1.757661e+01
## FPE(n) 2.376553e+07
```

```
model1<-VAR(modelinput[,c(1,2)],p=1,ic="AIC")
summary(model1)
```

```
##
## VAR Estimation Results:
## =========================
## Endogenous variables: store4_sales.d.ts, store4_temp.d.ts
## Deterministic variables: const
## Sample size: 128
## Log Likelihood: -1447.808
## Roots of the characteristic polynomial:
## 0.3599 0.06589
## Call:
## VAR(y = modelinput[, c(1, 2)], p = 1, ic = "AIC")
##
##
## Estimation results for equation store4_sales.d.ts:
## ==================================================
## store4_sales.d.ts = store4_sales.d.ts.l1 + store4_temp.d.ts.l1 + const
##
##                      Estimate Std. Error t value Pr(>|t|)
## store4_sales.d.ts.l1  -0.3916     0.0820  -4.776 4.92e-06 ***
## store4_temp.d.ts.l1   65.5567    18.4252   3.558 0.000529 ***
## const                 -3.9275    89.6887  -0.044 0.965141
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Residual standard error: 1011 on 125 degrees of freedom
## Multiple R-Squared: 0.1932,  Adjusted R-squared: 0.1802
## F-statistic: 14.96 on 2 and 125 DF,  p-value: 1.494e-06
##
##
## Estimation results for equation store4_temp.d.ts:
## =================================================
## store4_temp.d.ts = store4_sales.d.ts.l1 + store4_temp.d.ts.l1 + const
##
##                        Estimate Std. Error t value Pr(>|t|)
## store4_sales.d.ts.l1 -0.0001577  0.0004025  -0.392    0.696
## store4_temp.d.ts.l1  -0.0341590  0.0904294  -0.378    0.706
## const                 0.3866036  0.4401846   0.878    0.381
##
##
## Residual standard error: 4.963 on 125 degrees of freedom
## Multiple R-Squared: 0.002943,    Adjusted R-squared: -0.01301
## F-statistic: 0.1845 on 2 and 125 DF,  p-value: 0.8317
##
##
##
## Covariance matrix of residuals:
##                   store4_sales.d.ts store4_temp.d.ts
## store4_sales.d.ts           1022678          1090.69
## store4_temp.d.ts               1091            24.63
##
## Correlation matrix of residuals:
##                   store4_sales.d.ts store4_temp.d.ts
```

```
## store4_sales.d.ts            1.0000           0.2173
## store4_temp.d.ts             0.2173           1.0000
```

```
oosrmse(model1)
```

```
## Warning in timeseries1 - timeseries2: longer object length is not a
## multiple of shorter object length
```

```
## [1] 1055.817
```

After adding temperature, VARSelect recommends an order of 5 based on AIC. However, there isn't a significant difference between VAR(1) and VAR(5) in terms of AIC and so in the interest of parsimony, we are choosing a VAR(1) model

Model2: Sales+Temperature+Unemployment

```
VARselect(modelinput[,c(1,2,3)], lag.max = 6)
```

```
## $selection
## AIC(n)  HQ(n)  SC(n) FPE(n) 
##      6      1      1      6 
## 
## $criteria
##                  1            2            3            4            5
## AIC(n)     12.98761     13.06180     13.05077     13.03633     13.01513
## HQ(n)      13.09905     13.25682     13.32938     13.39853     13.46091
## SC(n)      13.26197     13.54193     13.73666     13.92800     14.11257
## FPE(n) 436994.00327 470789.29483 465955.32818 459875.87305 451160.00615
##                  6
## AIC(n)     12.92464
## HQ(n)      13.45400
## SC(n)      14.22785
## FPE(n) 413379.12443
```

```
model2<-VAR(modelinput[,c(1,2,3)],p=1,ic="AIC")
summary(model2)
```

```
##
## VAR Estimation Results:
## =========================
## Endogenous variables: store4_sales.d.ts, store4_temp.d.ts, store4_ue.d.ts
## Deterministic variables: const
## Sample size: 128
## Log Likelihood: -1364.818
## Roots of the characteristic polynomial:
## 0.3528 0.09082 0.09082
## Call:
## VAR(y = modelinput[, c(1, 2, 3)], p = 1, ic = "AIC")
##
##
## Estimation results for equation store4_sales.d.ts:
## ==================================================
## store4_sales.d.ts = store4_sales.d.ts.l1 + store4_temp.d.ts.l1 + store4_ue.d.ts.l1 +
 const
##
##                         Estimate Std. Error t value Pr(>|t|)
## store4_sales.d.ts.l1   -0.38433    0.08089  -4.751 5.49e-06 ***
## store4_temp.d.ts.l1    63.92182   18.17471   3.517  0.00061 ***
## store4_ue.d.ts.l1    1466.65193  677.13678   2.166  0.03223 *
## const                  48.72289   91.67446   0.531  0.59604
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Residual standard error: 996.7 on 124 degrees of freedom
## Multiple R-Squared: 0.2226,  Adjusted R-squared: 0.2038
## F-statistic: 11.83 on 3 and 124 DF,  p-value: 7.204e-07
##
##
## Estimation results for equation store4_temp.d.ts:
## =================================================
## store4_temp.d.ts = store4_sales.d.ts.l1 + store4_temp.d.ts.l1 + store4_ue.d.ts.l1 + c
onst
##
##                         Estimate Std. Error t value Pr(>|t|)
## store4_sales.d.ts.l1 -0.0001606  0.0004044  -0.397    0.692
## store4_temp.d.ts.l1  -0.0335023  0.0908606  -0.369    0.713
## store4_ue.d.ts.l1    -0.5891476  3.3852013  -0.174    0.862
## const                 0.3654542  0.4583070   0.797    0.427
##
##
## Residual standard error: 4.983 on 124 degrees of freedom
## Multiple R-Squared: 0.003187,    Adjusted R-squared: -0.02093
## F-statistic: 0.1321 on 3 and 124 DF,  p-value: 0.9408
##
##
## Estimation results for equation store4_ue.d.ts:
## ===============================================
## store4_ue.d.ts = store4_sales.d.ts.l1 + store4_temp.d.ts.l1 + store4_ue.d.ts.l1 + con
st
```

```
##
##                         Estimate Std. Error t value Pr(>|t|)
## store4_sales.d.ts.l1 -1.080e-06  1.067e-05  -0.101  0.91956
## store4_temp.d.ts.l1   2.381e-03  2.398e-03   0.993  0.32263
## store4_ue.d.ts.l1    -7.769e-02  8.933e-02  -0.870  0.38611
## const                -3.923e-02  1.209e-02  -3.244  0.00152 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Residual standard error: 0.1315 on 124 degrees of freedom
## Multiple R-Squared: 0.01344, Adjusted R-squared: -0.01042
## F-statistic: 0.5633 on 3 and 124 DF,  p-value: 0.6402
##
##
##
## Covariance matrix of residuals:
##                    store4_sales.d.ts store4_temp.d.ts store4_ue.d.ts
## store4_sales.d.ts         993343.158        1.115e+03       -6.11371
## store4_temp.d.ts            1114.587        2.483e+01        0.02263
## store4_ue.d.ts                -6.114        2.263e-02        0.01729
##
## Correlation matrix of residuals:
##                    store4_sales.d.ts store4_temp.d.ts store4_ue.d.ts
## store4_sales.d.ts            1.00000          0.22444       -0.04666
## store4_temp.d.ts             0.22444          1.00000        0.03454
## store4_ue.d.ts              -0.04666          0.03454        1.00000
```

```
oosrmse(model2)
```

```
## Warning in timeseries1 - timeseries2: longer object length is not a
## multiple of shorter object length
```

```
## [1] 1058.585
```

After adding Unemployment, VARSelect recommends an order of 6 based on AIC. However, there isn’t a significant difference between VAR(1) and VAR(6) in terms of AIC and so in the interest of parsimony, we are choosing a VAR(1) model

Model3: Sales+Temperature+Unemployment+FuelPrice

```
VARselect(modelinput[,c(1,2,3,4)], lag.max = 6)
```

```
## $selection
## AIC(n)  HQ(n)  SC(n) FPE(n)
##      1      1      1      1
##
## $criteria
##                 1           2           3           4           5
## AIC(n)   6.866215    6.959583    7.012296    7.019004    6.980806
## HQ(n)    7.051955    7.293915    7.495221    7.650521    7.760914
## SC(n)    7.323481    7.782661    8.201187    8.573708    8.901322
## FPE(n) 959.482893 1054.298616 1113.725653 1125.661705 1090.415310
##                  6
## AIC(n)    6.968931
## HQ(n)     7.897632
## SC(n)     9.255260
## FPE(n) 1087.779988
```

```
model3<-VAR(modelinput[,c(1,2,3,4)],p=1,ic="AIC")
summary(model3)
```

```
##
## VAR Estimation Results:
## =========================
## Endogenous variables: store4_sales.d.ts, store4_temp.d.ts, store4_ue.d.ts, store4_fp.
d.ts
## Deterministic variables: const
## Sample size: 128
## Log Likelihood: -1145.954
## Roots of the characteristic polynomial:
## 0.5493 0.3504 0.07766 0.07766
## Call:
## VAR(y = modelinput[, c(1, 2, 3, 4)], p = 1, ic = "AIC")
##
##
## Estimation results for equation store4_sales.d.ts:
## ==================================================
## store4_sales.d.ts = store4_sales.d.ts.l1 + store4_temp.d.ts.l1 + store4_ue.d.ts.l1 +
 store4_fp.d.ts.l1 + const
##
##                         Estimate Std. Error t value Pr(>|t|)
## store4_sales.d.ts.l1   -0.38477    0.08121  -4.738 5.86e-06 ***
## store4_temp.d.ts.l1    63.77225   18.25310   3.494 0.000663 ***
## store4_ue.d.ts.l1    1460.15047  680.18460   2.147 0.033780 *
## store4_fp.d.ts.l1     414.71646 1629.34060   0.255 0.799511
## const                  46.02248   92.63171   0.497 0.620194
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Residual standard error: 1000 on 123 degrees of freedom
## Multiple R-Squared: 0.223,   Adjusted R-squared: 0.1977
## F-statistic: 8.824 on 4 and 123 DF,  p-value: 2.687e-06
##
##
## Estimation results for equation store4_temp.d.ts:
## =================================================
## store4_temp.d.ts = store4_sales.d.ts.l1 + store4_temp.d.ts.l1 + store4_ue.d.ts.l1 + s
tore4_fp.d.ts.l1 + const
##
##                         Estimate Std. Error t value Pr(>|t|)
## store4_sales.d.ts.l1 -0.0001775  0.0003998  -0.444   0.6578
## store4_temp.d.ts.l1  -0.0392399  0.0898507  -0.437   0.6631
## store4_ue.d.ts.l1    -0.8385523  3.3482034  -0.250   0.8027
## store4_fp.d.ts.l1    15.9090916  8.0204165   1.984   0.0495 *
## const                 0.2618627  0.4559789   0.574   0.5668
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Residual standard error: 4.925 on 123 degrees of freedom
## Multiple R-Squared: 0.03408, Adjusted R-squared: 0.002673
## F-statistic: 1.085 on 4 and 123 DF,  p-value: 0.3669
##
```

```
##
## Estimation results for equation store4_ue.d.ts:
## ===============================================
## store4_ue.d.ts = store4_sales.d.ts.l1 + store4_temp.d.ts.l1 + store4_ue.d.ts.l1 + sto
re4_fp.d.ts.l1 + const
##
##                       Estimate Std. Error t value Pr(>|t|)
## store4_sales.d.ts.l1 -1.140e-06  1.071e-05  -0.106  0.91546
## store4_temp.d.ts.l1   2.361e-03  2.408e-03   0.980  0.32884
## store4_ue.d.ts.l1    -7.858e-02  8.973e-02  -0.876  0.38289
## store4_fp.d.ts.l1     5.628e-02  2.149e-01   0.262  0.79388
## const                -3.960e-02  1.222e-02  -3.240  0.00154 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Residual standard error: 0.132 on 123 degrees of freedom
## Multiple R-Squared: 0.01399, Adjusted R-squared: -0.01807
## F-statistic: 0.4364 on 4 and 123 DF,  p-value: 0.7821
##
##
## Estimation results for equation store4_fp.d.ts:
## ===============================================
## store4_fp.d.ts = store4_sales.d.ts.l1 + store4_temp.d.ts.l1 + store4_ue.d.ts.l1 + sto
re4_fp.d.ts.l1 + const
##
##                       Estimate Std. Error t value Pr(>|t|)
## store4_sales.d.ts.l1 -3.666e-07  3.699e-06  -0.099    0.921
## store4_temp.d.ts.l1   4.874e-05  8.313e-04   0.059    0.953
## store4_ue.d.ts.l1    -5.071e-02  3.098e-02  -1.637    0.104
## store4_fp.d.ts.l1     5.583e-01  7.421e-02   7.523 9.75e-12 ***
## const                 1.285e-03  4.219e-03   0.305    0.761
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Residual standard error: 0.04556 on 123 degrees of freedom
## Multiple R-Squared: 0.3224,  Adjusted R-squared: 0.3004
## F-statistic: 14.63 on 4 and 123 DF,  p-value: 8.383e-10
##
##
##
## Covariance matrix of residuals:
##                  store4_sales.d.ts store4_temp.d.ts store4_ue.d.ts
## store4_sales.d.ts         1.001e+06       1103.42546     -6.2349501
## store4_temp.d.ts          1.103e+03         24.25254      0.0200676
## store4_ue.d.ts           -6.235e+00          0.02007      0.0174172
## store4_fp.d.ts            2.077e+00         -0.01405      0.0001282
##                  store4_fp.d.ts
## store4_sales.d.ts      2.0770486
## store4_temp.d.ts      -0.0140516
## store4_ue.d.ts         0.0001282
## store4_fp.d.ts         0.0020760
##
```

```
## Correlation matrix of residuals:
##                  store4_sales.d.ts store4_temp.d.ts store4_ue.d.ts
## store4_sales.d.ts           1.00000          0.22396       -0.04722
## store4_temp.d.ts            0.22396          1.00000        0.03088
## store4_ue.d.ts             -0.04722          0.03088        1.00000
## store4_fp.d.ts              0.04557         -0.06262        0.02132
##                  store4_fp.d.ts
## store4_sales.d.ts       0.04557
## store4_temp.d.ts       -0.06262
## store4_ue.d.ts          0.02132
## store4_fp.d.ts          1.00000
```

```
oosrmse(model3)
```

```
## Warning in timeseries1 - timeseries2: longer object length is not a
## multiple of shorter object length
```

```
## [1] 1059.666
```

After adding Fuel Price, VARSelect continues to recommend an order of 1. We use this order to build the model.

Model4: Sales+Temperature+Unemployment+FuelPrice+CPI

```
VARselect(modelinput[,c(1,2,3,4,5)], lag.max = 6)
```

```
## $selection
## AIC(n)  HQ(n)  SC(n) FPE(n)
##      1      1      1      1
##
## $criteria
##                 1          2          3          4          5          6
## AIC(n) -1.0024140 -0.9372290 -0.9238915 -0.8855573 -0.8265214 -0.7935867
## HQ(n)  -0.7238037 -0.4264434 -0.1809307  0.0895787  0.3807899  0.6458999
## SC(n)  -0.3165154  0.3202518  0.9051715  1.5150878  2.1457060  2.7502229
## FPE(n)  0.3671347  0.3926513  0.3999247  0.4195083  0.4519581  0.4780252
```

```
model4<-VAR(modelinput[,c(1,2,3,4,5)],p=1,ic="AIC")
summary(model4)
```

```
##
## VAR Estimation Results:
## =========================
## Endogenous variables: store4_sales.d.ts, store4_temp.d.ts, store4_ue.d.ts, store4_fp.
d.ts, store4_cpi.d.ts
## Deterministic variables: const
## Sample size: 128
## Log Likelihood: -811.817
## Roots of the characteristic polynomial:
## 0.5437 0.3523 0.244 0.07825 0.07825
## Call:
## VAR(y = modelinput[, c(1, 2, 3, 4, 5)], p = 1, ic = "AIC")
##
##
## Estimation results for equation store4_sales.d.ts:
## ==================================================
## store4_sales.d.ts = store4_sales.d.ts.l1 + store4_temp.d.ts.l1 + store4_ue.d.ts.l1 +
 store4_fp.d.ts.l1 + store4_cpi.d.ts.l1 + const
##
##                        Estimate Std. Error t value Pr(>|t|)
## store4_sales.d.ts.l1 -3.909e-01  8.177e-02  -4.781 4.92e-06 ***
## store4_temp.d.ts.l1   6.336e+01  1.829e+01   3.464 0.000736 ***
## store4_ue.d.ts.l1     1.458e+03  6.814e+02   2.140 0.034350 *
## store4_fp.d.ts.l1     3.290e+02  1.636e+03   0.201 0.840996
## store4_cpi.d.ts.l1   -3.586e+03  4.760e+03  -0.753 0.452722
## const                 4.573e+01  9.280e+01   0.493 0.623062
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Residual standard error: 1002 on 122 degrees of freedom
## Multiple R-Squared: 0.2266,  Adjusted R-squared: 0.1949
## F-statistic: 7.148 on 5 and 122 DF,  p-value: 6.748e-06
##
##
## Estimation results for equation store4_temp.d.ts:
## =================================================
## store4_temp.d.ts = store4_sales.d.ts.l1 + store4_temp.d.ts.l1 + store4_ue.d.ts.l1 + s
tore4_fp.d.ts.l1 + store4_cpi.d.ts.l1 + const
##
##                        Estimate Std. Error t value Pr(>|t|)
## store4_sales.d.ts.l1 -0.0001760  0.0004034  -0.436   0.6634
## store4_temp.d.ts.l1  -0.0391399  0.0902580  -0.434   0.6653
## store4_ue.d.ts.l1    -0.8380628  3.3619045  -0.249   0.8036
## store4_fp.d.ts.l1    15.9299309  8.0727425   1.973   0.0507 .
## store4_cpi.d.ts.l1    0.8712465 23.4851864   0.037   0.9705
## const                 0.2619346  0.4578453   0.572   0.5683
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Residual standard error: 4.945 on 122 degrees of freedom
## Multiple R-Squared: 0.0341,  Adjusted R-squared: -0.005491
```

```
## F-statistic: 0.8613 on 5 and 122 DF,  p-value: 0.5093
##
##
## Estimation results for equation store4_ue.d.ts:
## ================================================
## store4_ue.d.ts = store4_sales.d.ts.l1 + store4_temp.d.ts.l1 + store4_ue.d.ts.l1 + sto
re4_fp.d.ts.l1 + store4_cpi.d.ts.l1 + const
##
##                         Estimate Std. Error t value Pr(>|t|)
## store4_sales.d.ts.l1 -1.349e-06  1.081e-05  -0.125   0.9009
## store4_temp.d.ts.l1   2.347e-03  2.418e-03   0.970   0.3338
## store4_ue.d.ts.l1    -7.864e-02  9.008e-02  -0.873   0.3844
## store4_fp.d.ts.l1     5.336e-02  2.163e-01   0.247   0.8056
## store4_cpi.d.ts.l1   -1.220e-01  6.293e-01  -0.194   0.8466
## const                -3.961e-02  1.227e-02  -3.229   0.0016 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Residual standard error: 0.1325 on 122 degrees of freedom
## Multiple R-Squared: 0.0143,  Adjusted R-squared: -0.0261
## F-statistic: 0.3539 on 5 and 122 DF,  p-value: 0.8789
##
##
## Estimation results for equation store4_fp.d.ts:
## ================================================
## store4_fp.d.ts = store4_sales.d.ts.l1 + store4_temp.d.ts.l1 + store4_ue.d.ts.l1 + sto
re4_fp.d.ts.l1 + store4_cpi.d.ts.l1 + const
##
##                         Estimate Std. Error t value Pr(>|t|)
## store4_sales.d.ts.l1  6.115e-08  3.712e-06   0.016    0.987
## store4_temp.d.ts.l1   7.732e-05  8.306e-04   0.093    0.926
## store4_ue.d.ts.l1    -5.057e-02  3.094e-02  -1.635    0.105
## store4_fp.d.ts.l1     5.642e-01  7.429e-02   7.595 6.92e-12 ***
## store4_cpi.d.ts.l1    2.489e-01  2.161e-01   1.152    0.252
## const                 1.306e-03  4.213e-03   0.310    0.757
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Residual standard error: 0.0455 on 122 degrees of freedom
## Multiple R-Squared: 0.3297,  Adjusted R-squared: 0.3022
## F-statistic:    12 on 5 and 122 DF,  p-value: 1.857e-09
##
##
## Estimation results for equation store4_cpi.d.ts:
## =================================================
## store4_cpi.d.ts = store4_sales.d.ts.l1 + store4_temp.d.ts.l1 + store4_ue.d.ts.l1 + st
ore4_fp.d.ts.l1 + store4_cpi.d.ts.l1 + const
##
##                         Estimate Std. Error t value Pr(>|t|)
## store4_sales.d.ts.l1  7.201e-07  1.511e-06   0.477  0.63453
## store4_temp.d.ts.l1  -1.639e-04  3.380e-04  -0.485  0.62871
## store4_ue.d.ts.l1     1.665e-03  1.259e-02   0.132  0.89500
```

```
## store4_fp.d.ts.l1    -1.425e-02  3.024e-02  -0.471  0.63837   
## store4_cpi.d.ts.l1    2.349e-01  8.796e-02   2.670  0.00861 **
## const                 1.769e-04  1.715e-03   0.103  0.91799   
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## 
## Residual standard error: 0.01852 on 122 degrees of freedom
## Multiple R-Squared: 0.06096, Adjusted R-squared: 0.02247 
## F-statistic: 1.584 on 5 and 122 DF,  p-value: 0.1696 
## 
## 
## 
## Covariance matrix of residuals:
##                    store4_sales.d.ts store4_temp.d.ts store4_ue.d.ts
## store4_sales.d.ts          1.004e+06        1.114e+03     -6.445e+00
## store4_temp.d.ts           1.114e+03        2.445e+01      2.027e-02
## store4_ue.d.ts            -6.445e+00        2.027e-02      1.755e-02
## store4_fp.d.ts             2.418e+00       -1.425e-02      1.403e-04
## store4_cpi.d.ts           -1.712e+00       -2.981e-03      2.307e-05
##                    store4_fp.d.ts store4_cpi.d.ts
## store4_sales.d.ts      2.418e+00      -1.712e+00
## store4_temp.d.ts      -1.425e-02      -2.981e-03
## store4_ue.d.ts         1.403e-04       2.307e-05
## store4_fp.d.ts         2.071e-03      -7.138e-05
## store4_cpi.d.ts       -7.138e-05       3.430e-04
## 
## Correlation matrix of residuals:
##                    store4_sales.d.ts store4_temp.d.ts store4_ue.d.ts
## store4_sales.d.ts            1.00000          0.22471      -0.048536
## store4_temp.d.ts             0.22471          1.00000       0.030940
## store4_ue.d.ts              -0.04854          0.03094       1.000000
## store4_fp.d.ts               0.05303         -0.06331       0.023266
## store4_cpi.d.ts             -0.09224         -0.03256       0.009403
##                    store4_fp.d.ts store4_cpi.d.ts
## store4_sales.d.ts        0.05303       -0.092238
## store4_temp.d.ts        -0.06331       -0.032557
## store4_ue.d.ts           0.02327        0.009403
## store4_fp.d.ts           1.00000       -0.084702
## store4_cpi.d.ts         -0.08470        1.000000
```

```
oosrmse(model4)
```

```
## Warning in timeseries1 - timeseries2: longer object length is not a
## multiple of shorter object length
```

```
## [1] 1059.279
```

After adding CPI, VARSelect continues to recommend an order of 1. We use this order to build the model.

Model Selection

Model	Time Series' Included	Adjusted R2	RMSE
Model1	Sales+Temperature	0.18	1055.817
Model2	Sales+Temperature+Unemployment	0.20	1058.585
Model3	Sales+Temperature+Unemployment+FuelPrice	0.19	1059.666
Model4	Sales+Temperature+Unemployment+FuelPrice+CPI	0.19	1059.279

Based on the R2 values and the RMSE of the out-of-sample projections vs the actual values, we have chosen Model2 as our final VAR model and the Sales, Temperature and Unemployment as the variables to include in the model.

Final VAR model

```
finalmodel<-VAR(modelinput[,c(1,2,3)],p=1,ic="AIC")
summary(finalmodel)
```

```
##
## VAR Estimation Results:
## =========================
## Endogenous variables: store4_sales.d.ts, store4_temp.d.ts, store4_ue.d.ts
## Deterministic variables: const
## Sample size: 128
## Log Likelihood: -1364.818
## Roots of the characteristic polynomial:
## 0.3528 0.09082 0.09082
## Call:
## VAR(y = modelinput[, c(1, 2, 3)], p = 1, ic = "AIC")
##
##
## Estimation results for equation store4_sales.d.ts:
## ==================================================
## store4_sales.d.ts = store4_sales.d.ts.l1 + store4_temp.d.ts.l1 + store4_ue.d.ts.l1 +
 const
##
##                        Estimate Std. Error t value Pr(>|t|)
## store4_sales.d.ts.l1   -0.38433    0.08089  -4.751 5.49e-06 ***
## store4_temp.d.ts.l1    63.92182   18.17471   3.517  0.00061 ***
## store4_ue.d.ts.l1    1466.65193  677.13678   2.166  0.03223 *
## const                  48.72289   91.67446   0.531  0.59604
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Residual standard error: 996.7 on 124 degrees of freedom
## Multiple R-Squared: 0.2226,  Adjusted R-squared: 0.2038
## F-statistic: 11.83 on 3 and 124 DF,  p-value: 7.204e-07
##
##
## Estimation results for equation store4_temp.d.ts:
## =================================================
## store4_temp.d.ts = store4_sales.d.ts.l1 + store4_temp.d.ts.l1 + store4_ue.d.ts.l1 + c
onst
##
##                        Estimate Std. Error t value Pr(>|t|)
## store4_sales.d.ts.l1 -0.0001606  0.0004044  -0.397    0.692
## store4_temp.d.ts.l1  -0.0335023  0.0908606  -0.369    0.713
## store4_ue.d.ts.l1    -0.5891476  3.3852013  -0.174    0.862
## const                 0.3654542  0.4583070   0.797    0.427
##
##
## Residual standard error: 4.983 on 124 degrees of freedom
## Multiple R-Squared: 0.003187,    Adjusted R-squared: -0.02093
## F-statistic: 0.1321 on 3 and 124 DF,  p-value: 0.9408
##
##
## Estimation results for equation store4_ue.d.ts:
## ===============================================
## store4_ue.d.ts = store4_sales.d.ts.l1 + store4_temp.d.ts.l1 + store4_ue.d.ts.l1 + con
st
```

```
##
##                        Estimate Std. Error t value Pr(>|t|)
## store4_sales.d.ts.l1 -1.080e-06  1.067e-05  -0.101  0.91956
## store4_temp.d.ts.l1   2.381e-03  2.398e-03   0.993  0.32263
## store4_ue.d.ts.l1    -7.769e-02  8.933e-02  -0.870  0.38611
## const                -3.923e-02  1.209e-02  -3.244  0.00152 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Residual standard error: 0.1315 on 124 degrees of freedom
## Multiple R-Squared: 0.01344, Adjusted R-squared: -0.01042
## F-statistic: 0.5633 on 3 and 124 DF,  p-value: 0.6402
##
##
##
## Covariance matrix of residuals:
##                  store4_sales.d.ts store4_temp.d.ts store4_ue.d.ts
## store4_sales.d.ts        993343.158        1.115e+03       -6.11371
## store4_temp.d.ts           1114.587        2.483e+01        0.02263
## store4_ue.d.ts               -6.114        2.263e-02        0.01729
##
## Correlation matrix of residuals:
##                  store4_sales.d.ts store4_temp.d.ts store4_ue.d.ts
## store4_sales.d.ts           1.00000          0.22444       -0.04666
## store4_temp.d.ts            0.22444          1.00000        0.03454
## store4_ue.d.ts             -0.04666          0.03454        1.00000
```

```
oosrmse(finalmodel)
```

```
## Warning in timeseries1 - timeseries2: longer object length is not a
## multiple of shorter object length
```

```
## [1] 1058.585
```

Key insights from this model are * The first lag of sales and temperature seem to be significant. This makes sense since these two variables are cointegrated * The first lag of unemployment seems to be slightly significant

Diagnostics

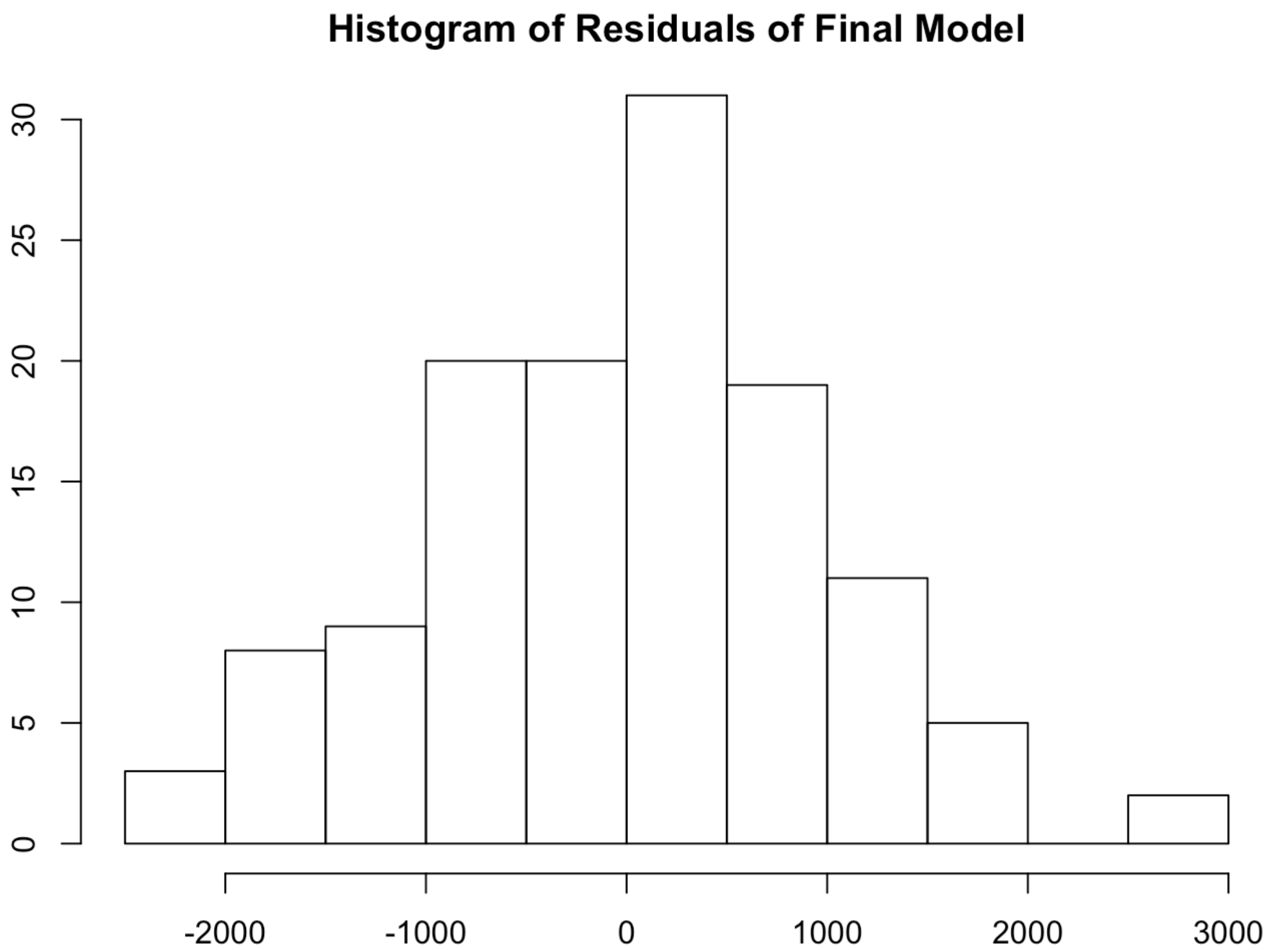
```
par(mar = rep(2, 4))

plot.ts(resid(finalmodel)[,c(1)],main="Plot of Residuals of Sales Time Series")
```

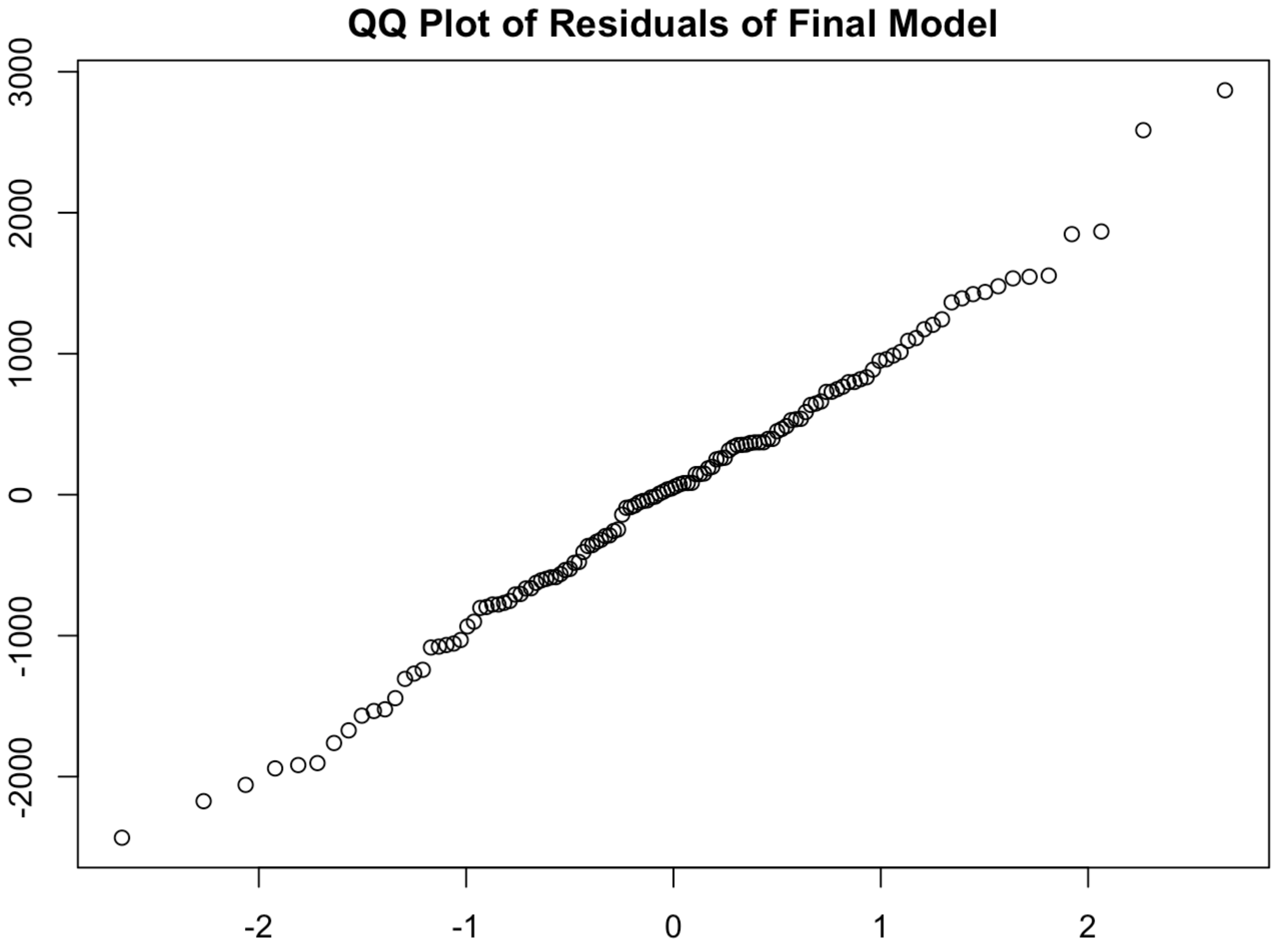

```
Box.test(resid(finalmodel)[,c(1)],type = c("Ljung-Box"))
```

```
##
##  Box-Ljung test
##
## data:  resid(finalmodel)[, c(1)]
## X-squared = 1.1923, df = 1, p-value = 0.2749
```

```
#Tests for normality
hist(resid(finalmodel)[,c(1)],main = "Histogram of Residuals of Final Model",xlab="")
```




```
qqnorm(resid(finalmodel)[,c(1)],main = "QQ Plot of Residuals of Final Model",xlab="")
```



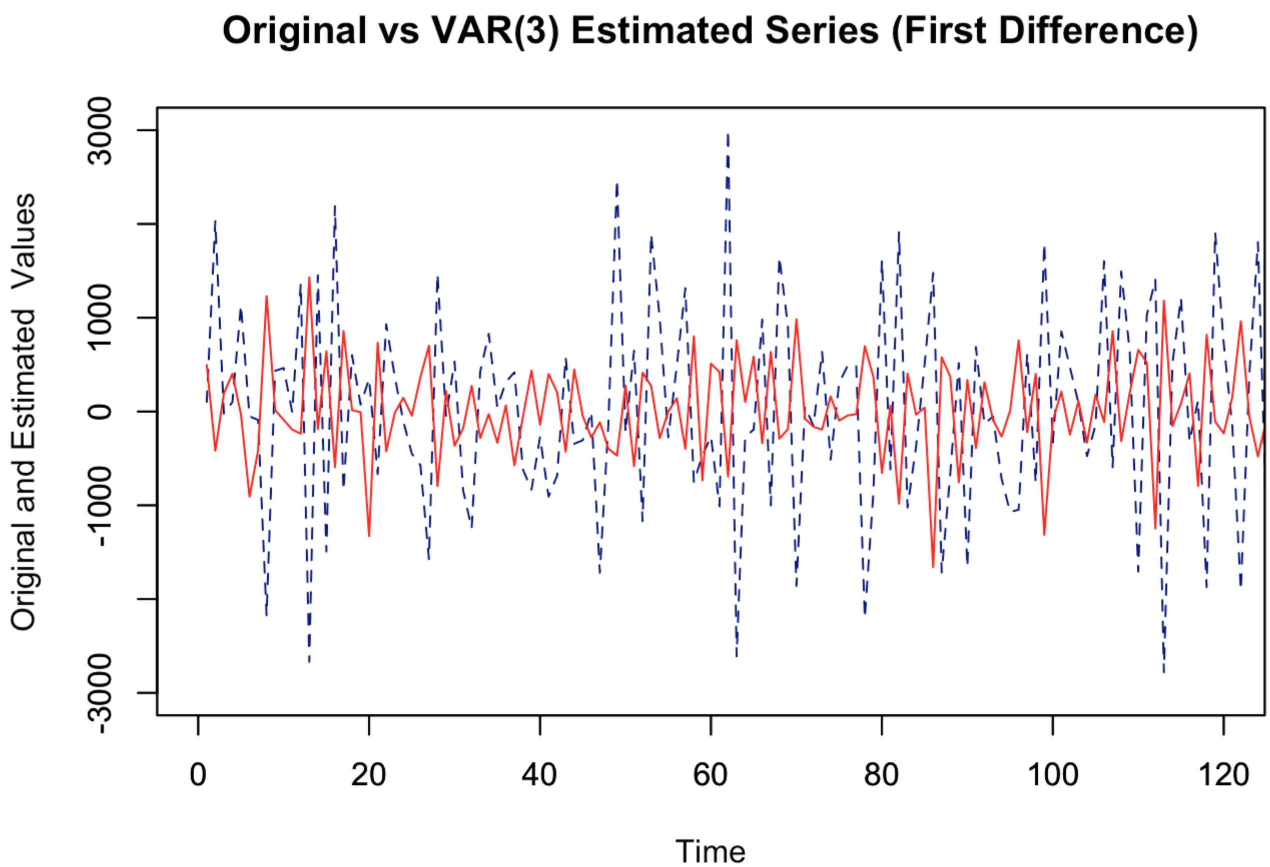
In the plots and tests above, we test the residuals from the chosen model (for sales only)for-

- Independence
- Normality

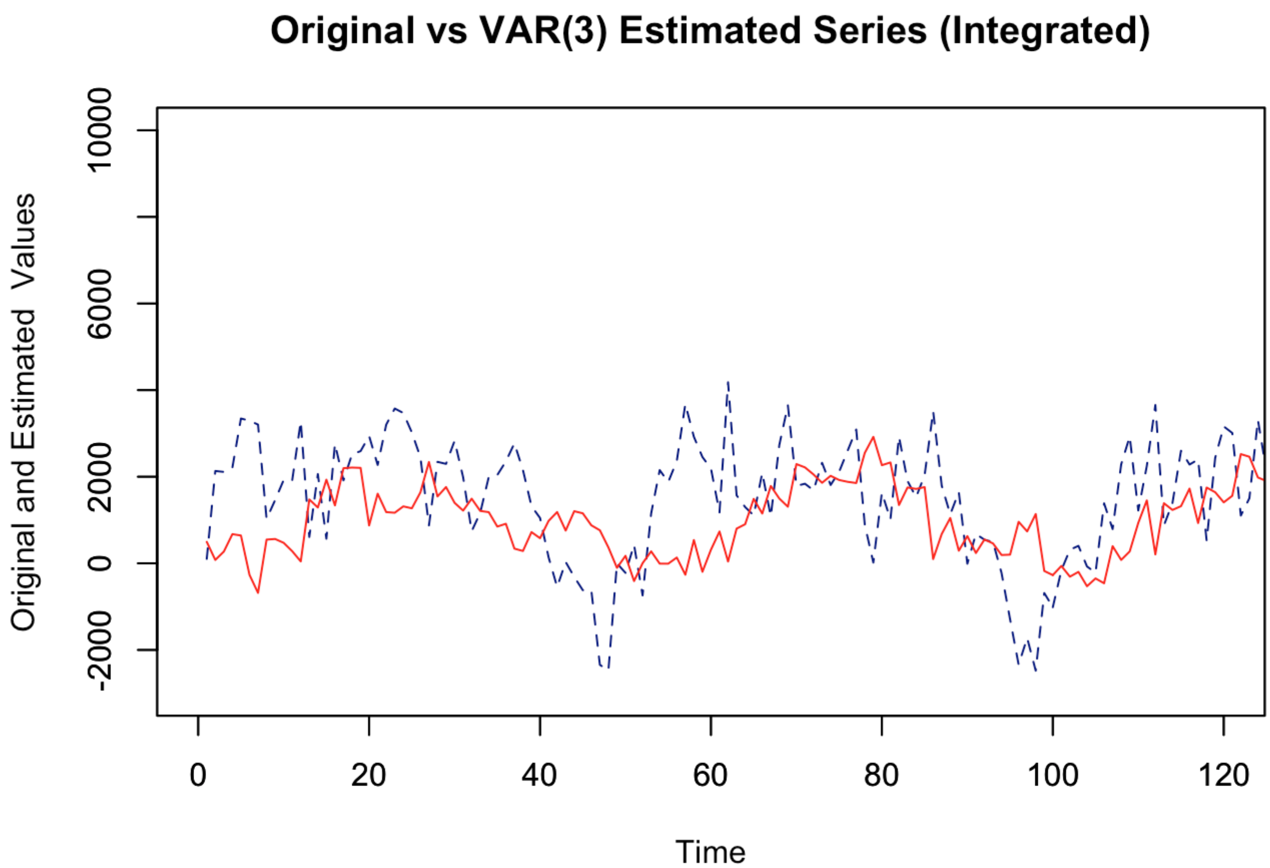
The residuals appear to be independent and identically distributed (iid). Based on the Ljung-Box test, the null hypothesis that the residuals are iid cannot be rejected. The histogram of the residuals and the qqplot further validate the normality of the residuals

Model Performance Evaluation

```
par(mfrow=c(1,1))
 plot.ts(modelinput[,c(1)],col="navy",lty=2,main="Original vs VAR(3) Estimated Series (F
irst Difference)",ylab = "Original and Estimated  Values",xlim = c(0,120),ylim=c(-3000,3
000))
 par(new=T)
 plot.ts(fitted(finalmodel)[,c(1)],col="red",xlim =
c(0,120),ylim=c(-3000,3000),xlab="",ylab="")
```




```
par(mfrow=c(1,1))
 plot.ts(cumsum(modelinput[,c(1)]),col="navy",lty=2,main="Original vs VAR(3) Estimated S
eries (Integrated)",ylab = "Original and Estimated  Values",xlim =
c(0,120),ylim=c(-3000,10000))
 par(new=T)
 plot.ts(cumsum(fitted(finalmodel)[,c(1)]),col="red",xlim =
c(0,120),ylim=c(-3000,10000),xlab="",ylab="")
```



Based on the graphs above, we can conclude that the model provides a reasonably good fit to the sales data

Hypothesis testing

- Is there a relationship between temperature and sales?

There appears to be a strong relationship between sales and the first lag of temperature. Make a note here however that these two time series are cointegrated.

- Is there a relationship between CPI and sales?

There doesn't appear to be a relationship between these two variables

- Is there a relationship between fuel price and sales?

There doesn't appear to be a relationship between these two variables

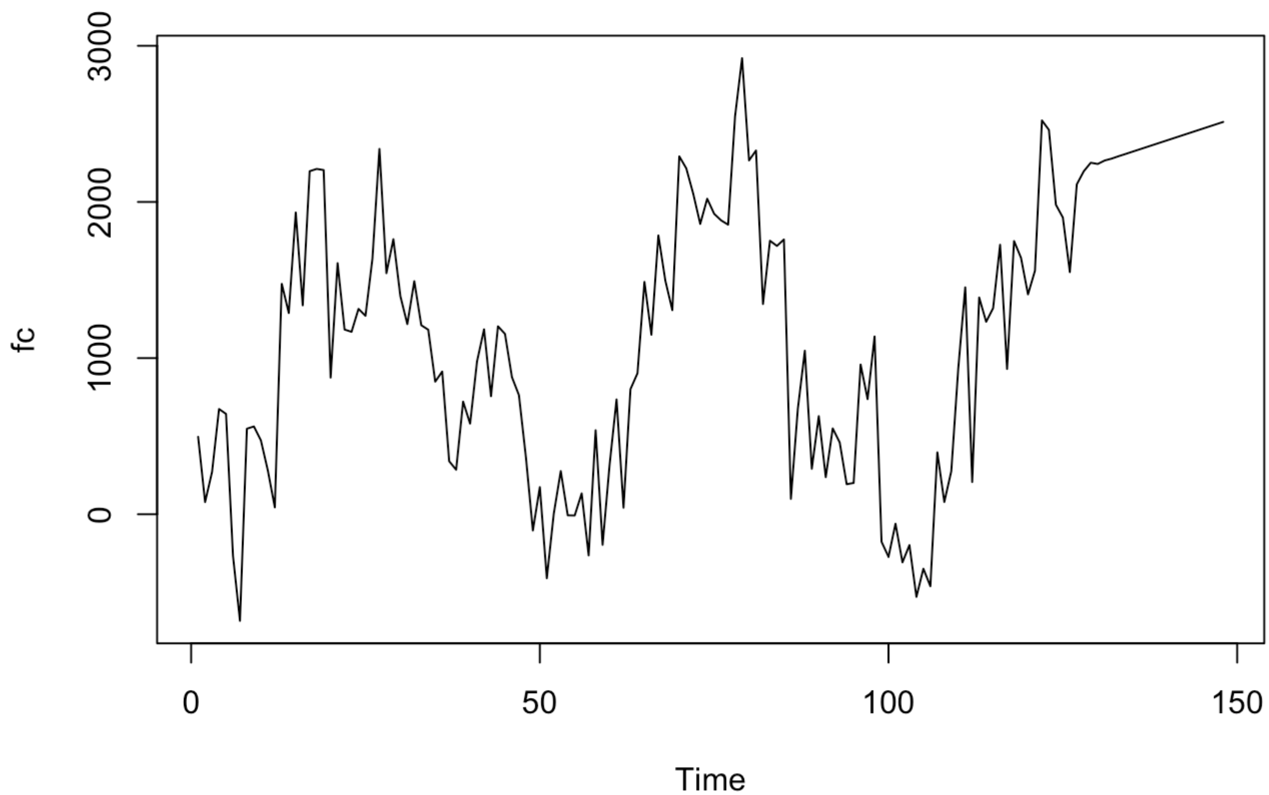
- Is there a relationship between unemployment and sales?

There appears to be a relationship sales and the first lag of unemployment

Forecasting

```
p<-predict(finalmodel,n.ahead=20)

fc<-cumsum(union(fitted(finalmodel)[,c(1)],p$fcst$store4_sales.d.ts[,c(1)]))
plot.ts(fc)
```




```
fc[c(143:148)]
```

```
## [1] 2439.253 2453.896 2468.539 2483.182 2497.825 2512.468
```

We forecast that on an average, the store will sell the units projected above in department 12 for the next 5 time weeks.