TimeSeries Walmart Project

Kenya Roy and Lani Lewis

2024-12-08

We invite you to watch our powerpoint presentation on our predictive modeling work online: https://youtu.be/eyyFiHjTxng

EDA & CLEAN DATA

Merge Data

Validation Dataset

Clean up merge duplicate columns

Clean up date types

Training Dataset

Validation Dataset

NA Values

```
##        Store         Date         Dept Weekly_Sales    IsHoliday  Temperature
##      0.00000      0.00000      0.00000      0.00000      0.00000      0.00000
##   Fuel_Price    MarkDown1    MarkDown2    MarkDown3    MarkDown4    MarkDown5
##      0.00000     64.25718     73.61103     67.48085     67.98468     64.07904
##          CPI Unemployment         Type         Size
##      0.00000      0.00000      0.00000      0.00000
```

Remove NA >50%

SPLIT

Filtered Training Dataset

Aggregate Weekly_Sales and retain other columns

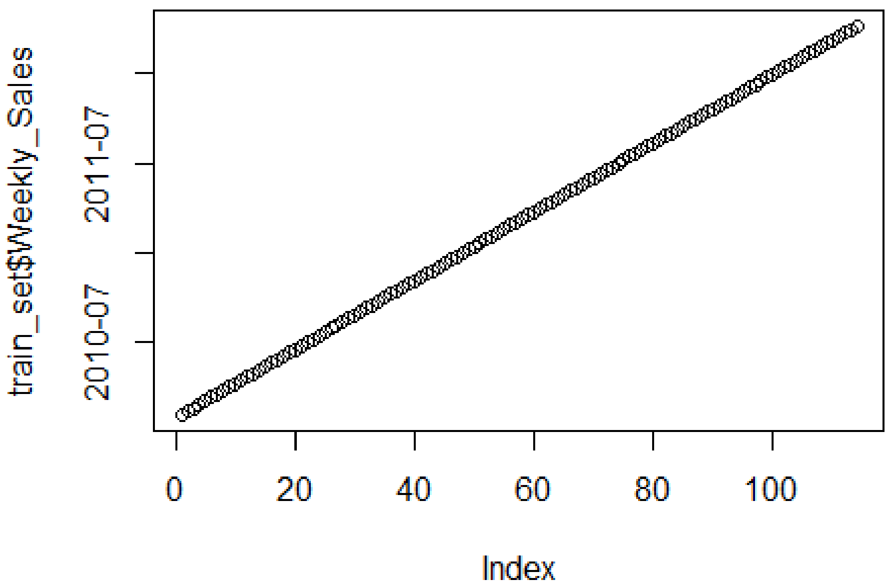
80/20 Train Test Split

```
## Training set dimensions: 114 8

## Test set dimensions: 29 8

## Training set percentage: 79.72%

## Test set percentage: 20.28%
```

TEST SPLIT OF DATA

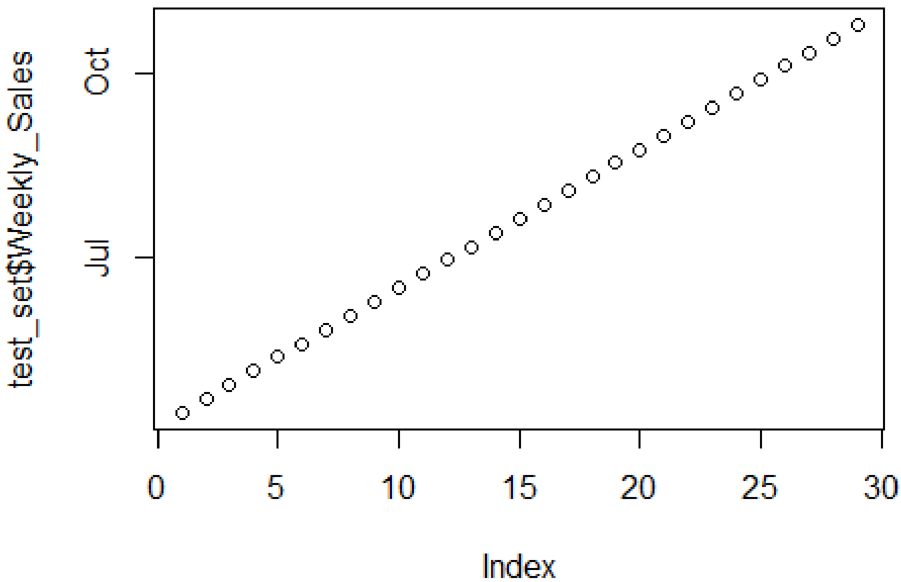
Aggregated Date

```
##
## 2010 2011 2012
##   48   52   43
```



Train Test Set

Test Set - This is different then the validation set



SCALE VARIABLES

AGGREGATED

TRAINED

TEST

LOG VARIABLES

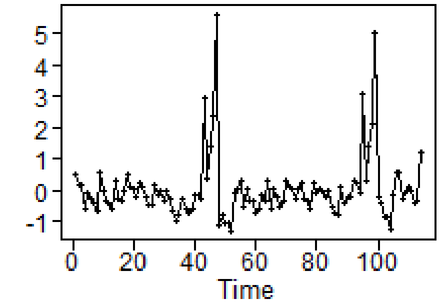
AGGREGATED

TRAINED

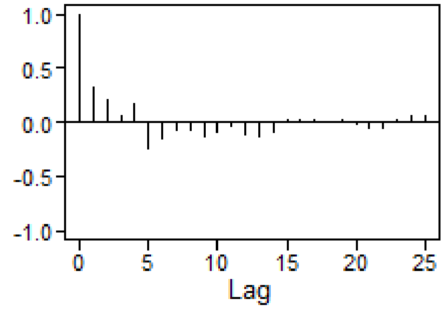
TEST

COVERT DATASETS TS

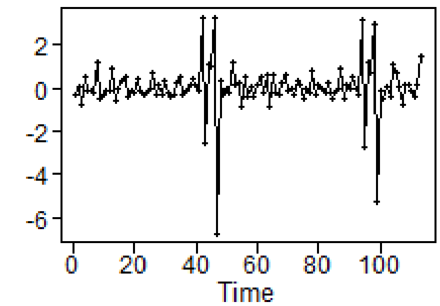
DIFFERENCE DATA



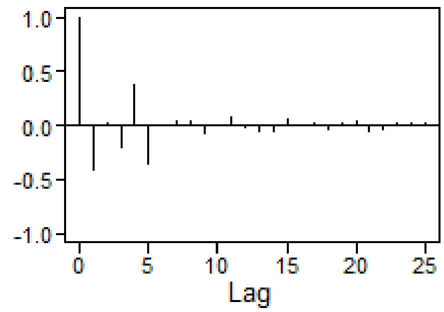
Realization



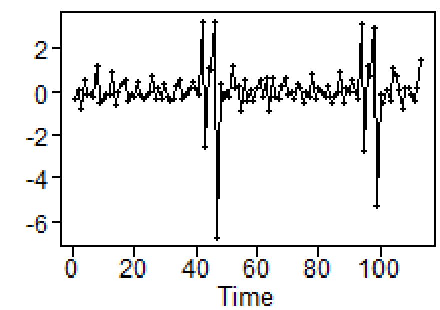
Sample Autocorrelations



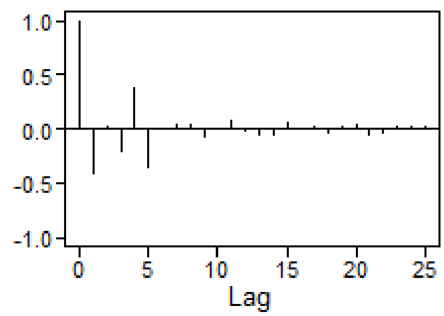
Transformed Realization



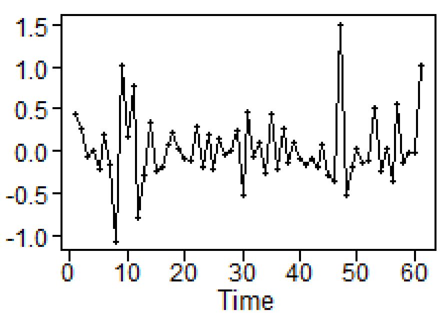
Sample Autocorrelations



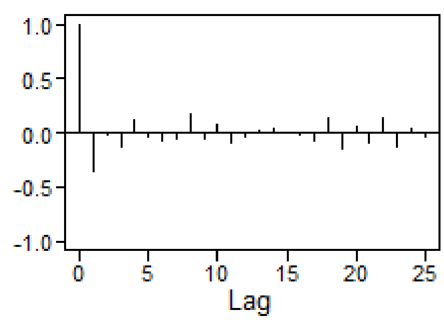
Realization



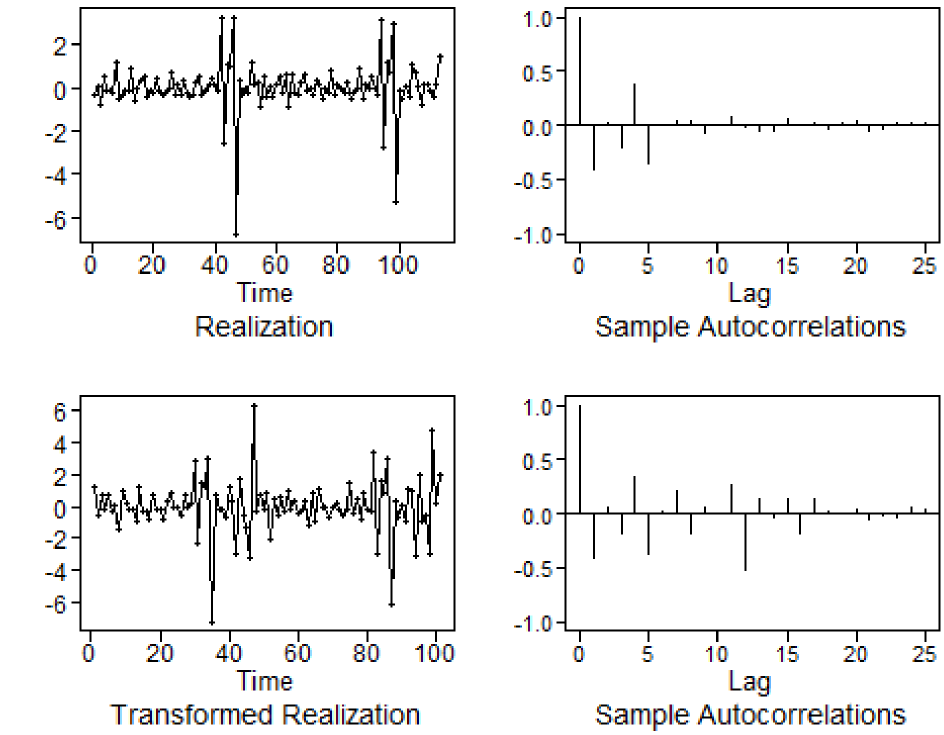
Sample Autocorrelations



Transformed Realization

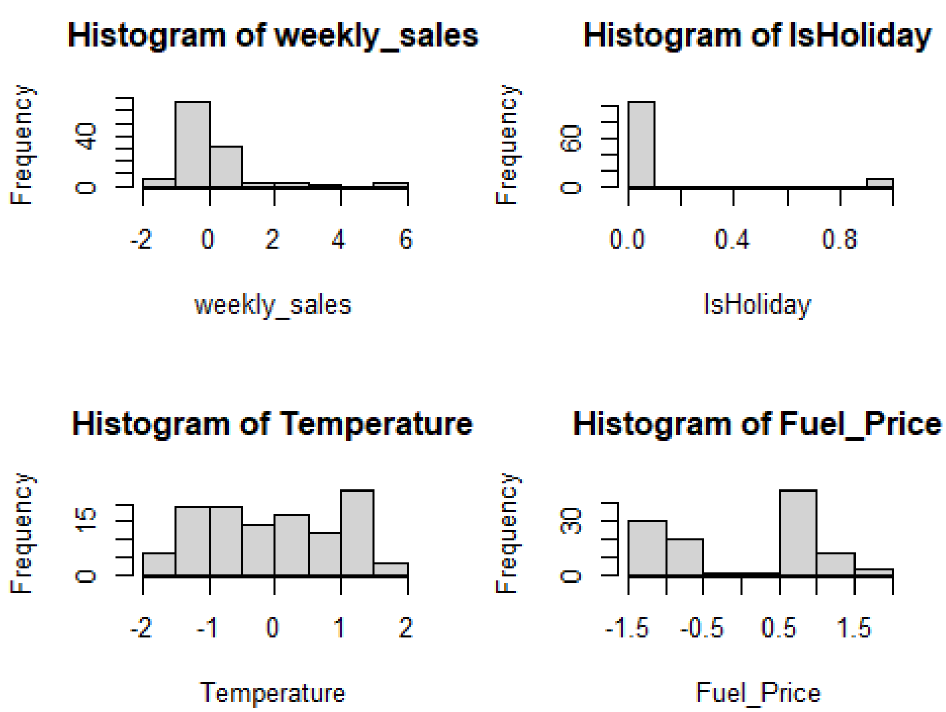


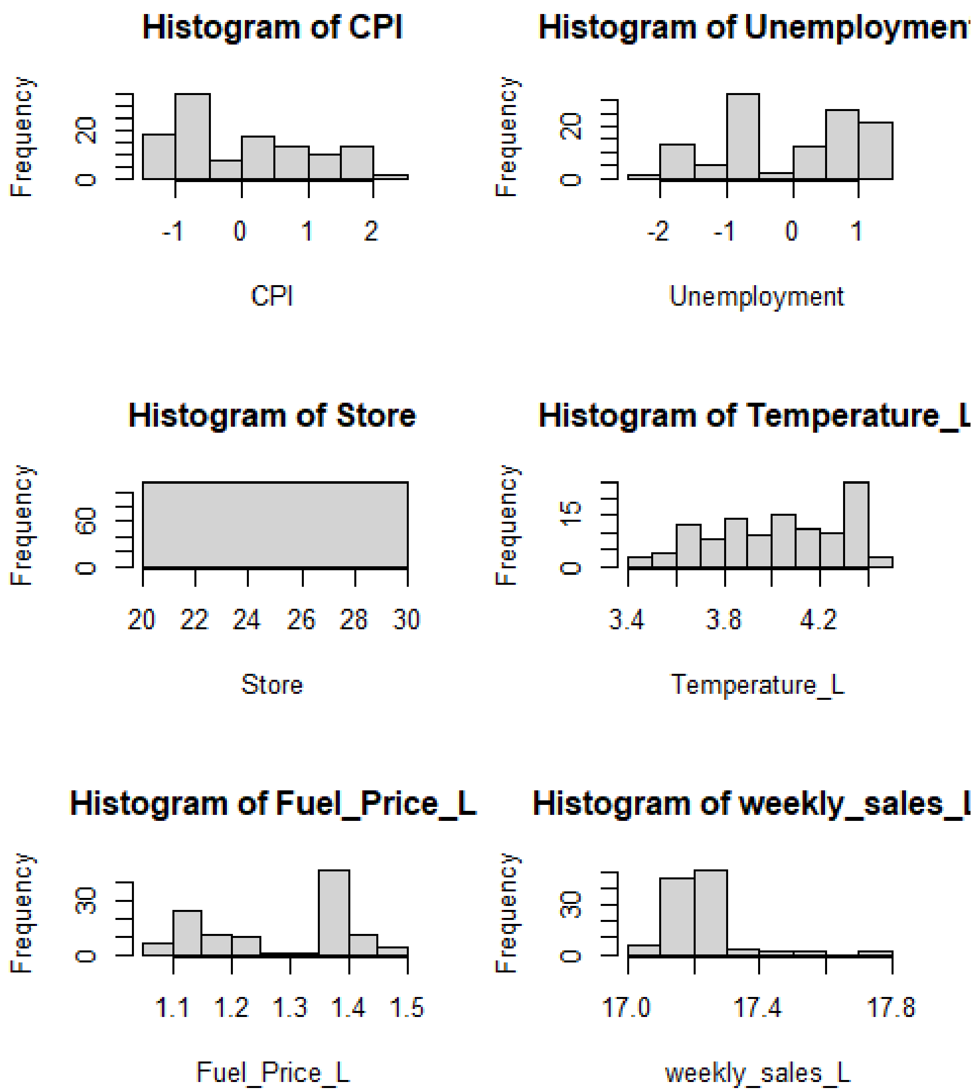
Sample Autocorrelations



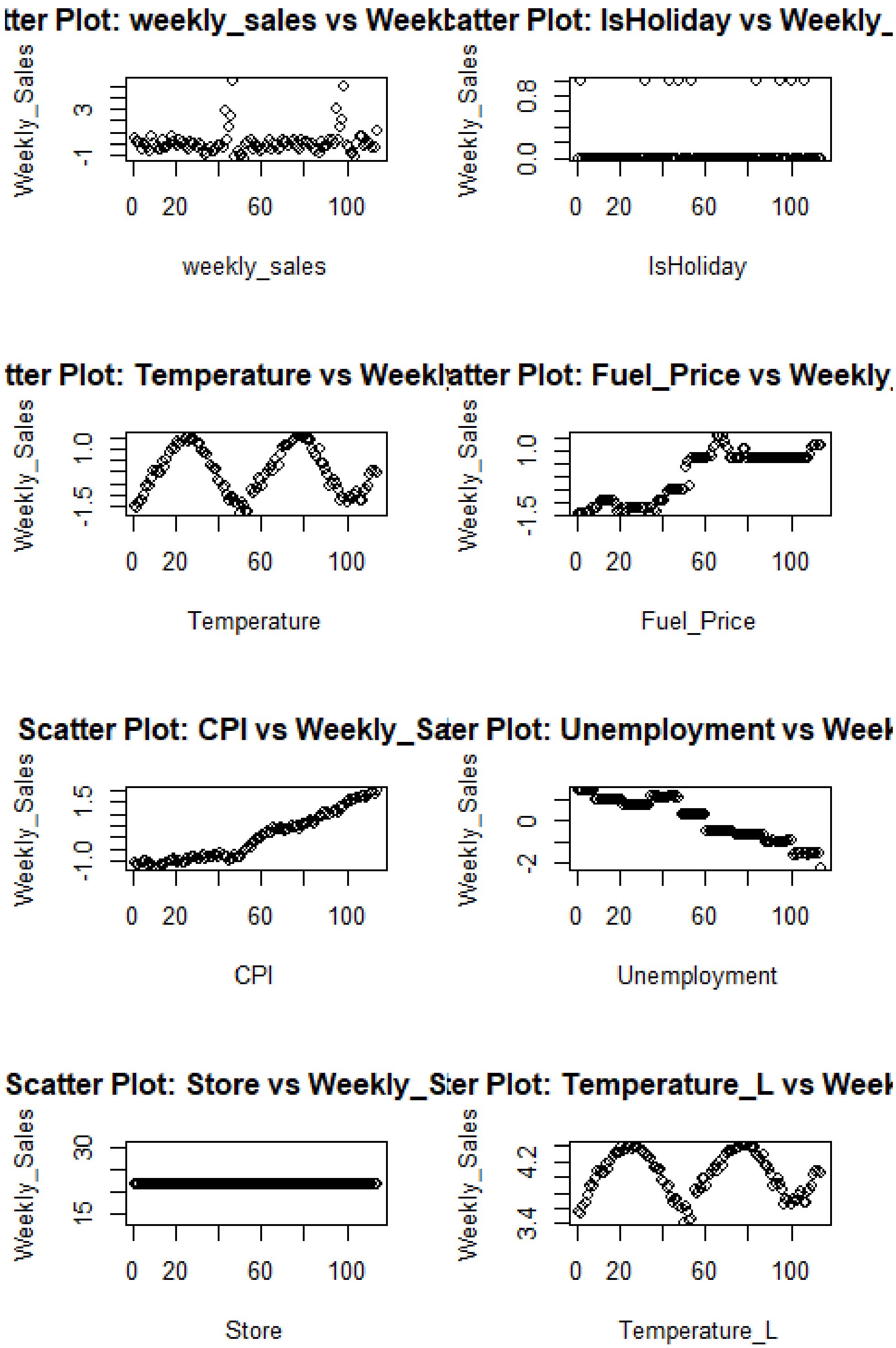
PLOTS

Variables that are highly skewed often benefit from a log transformation.

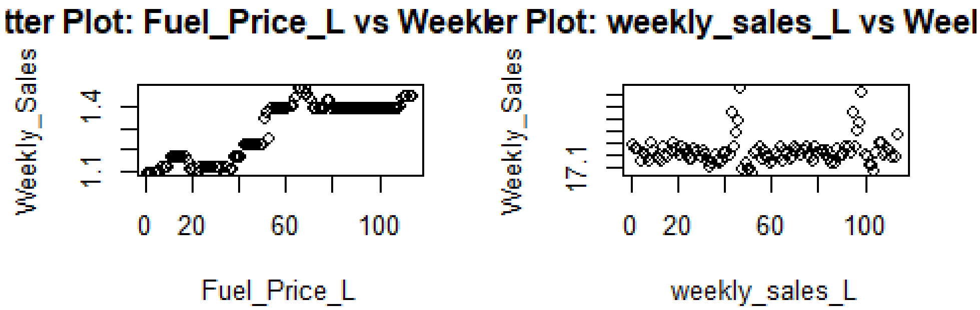




Look for non-linear relationships where a log transformation could help linearize the data. - Possibly log variables



tter Plot: weekly_sales vs Week
atter Plot: IsHoliday vs Weekly_
Weekly_Sales
-1
3
0.0
0.8
0
20
60
100
weekly_sales
IsHoliday
tter Plot: Temperature vs Weekl
atter Plot: Fuel_Price vs Weekly
-1.5
1.0
Temperature
Fuel_Price
Scatter Plot: CPI vs Weekly_Sa
er Plot: Unemployment vs Week
-1.0
1.5
-2
0
CPI
Unemployment
Scatter Plot: Store vs Weekly_S
er Plot: Temperature_L vs Week
15
30
3.4
4.2
Store
Temperature_L

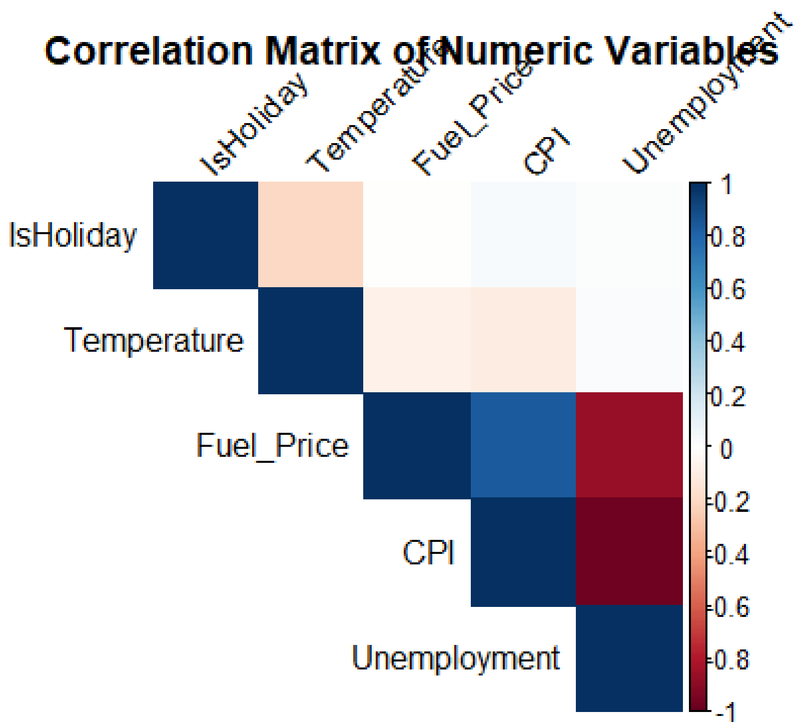


May not need to use Highly correlated data - Keep all variables

##Correlation Matrix of Numeric Variables Unemployment has high correlation with Fuel_Price and CPI variables.

```
numeric_vars <- train_set %>%
  select_if(is.numeric) # Select all numeric columns

# drop weekly_sales
numeric_vars <- numeric_vars[, !colnames(numeric_vars) %in% "weekly_sales"]
numeric_vars <- numeric_vars[, !colnames(numeric_vars) %in% "Store"]
numeric_vars <- numeric_vars[, !colnames(numeric_vars) %in% "Dept"]
numeric_vars <- numeric_vars[, !colnames(numeric_vars) %in% "Size"]

# Compute the correlation matrix
cor_matrix <- cor(numeric_vars, use = "complete.obs")  # Use only complete cases

# Print the correlation matrix
print(cor_matrix)

##                IsHoliday Temperature   Fuel_Price         CPI Unemployment
## IsHoliday     1.000000000 -0.20963274 -0.002213931  0.03250266   0.01133062
## Temperature  -0.209632737  1.00000000 -0.071108692 -0.10476660   0.02227830
## Fuel_Price   -0.002213931 -0.07110869  1.000000000  0.83926661  -0.86517633
## CPI           0.032502658 -0.10476660  0.839266607  1.00000000  -0.97189949
## Unemployment  0.011330622  0.02227830 -0.865176331 -0.97189949   1.00000000
```

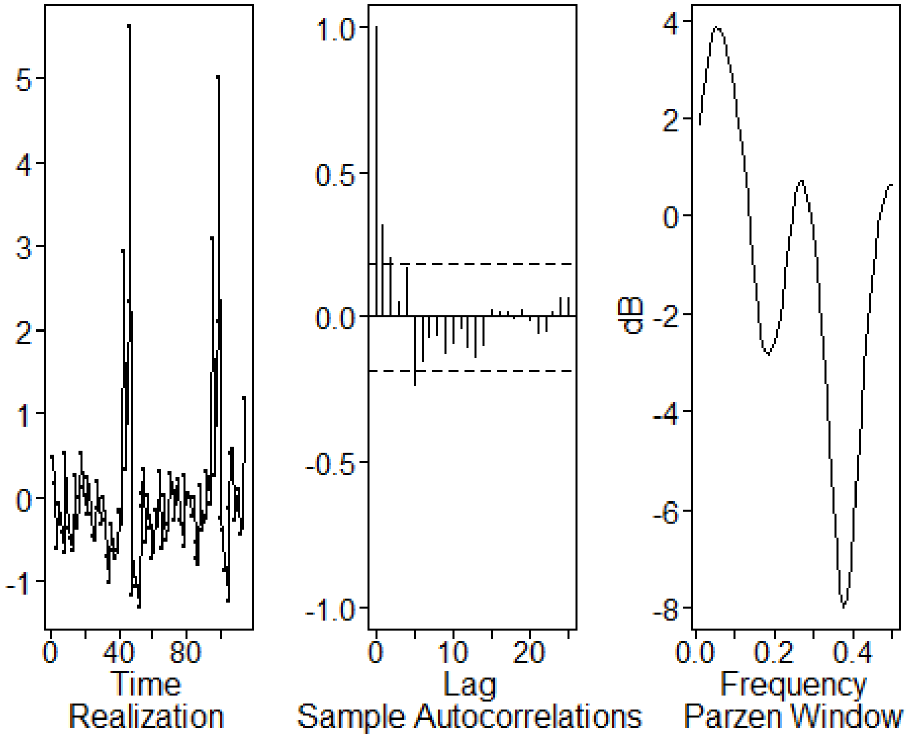
```r
# Adjust margins to create space for the title
par(mar = c(5, 5, 7, 5))  # Increase the top margin (3rd value)
corrplot(cor_matrix,
         method = "color",
         type = "upper",
         tl.col = "black",
         tl.srt = 45)
title("Correlation Matrix of Numeric Variables", line = 5)  # Add title with proper spacing
```



TS Plots

- Realization: The sharp spikes and upward or downward trends suggest that the series is non-stationary (mean and variance are not constant over time).
- ACF: Strong correlations at lag 1 and lag 2, gradually decreasing, also indicating non-stationarity.
- Spectral Density: looks like we have three high peaks the first is the strongest and we see high peaks around period 12 (13/14)
- The next import periods are around 3 then 2…

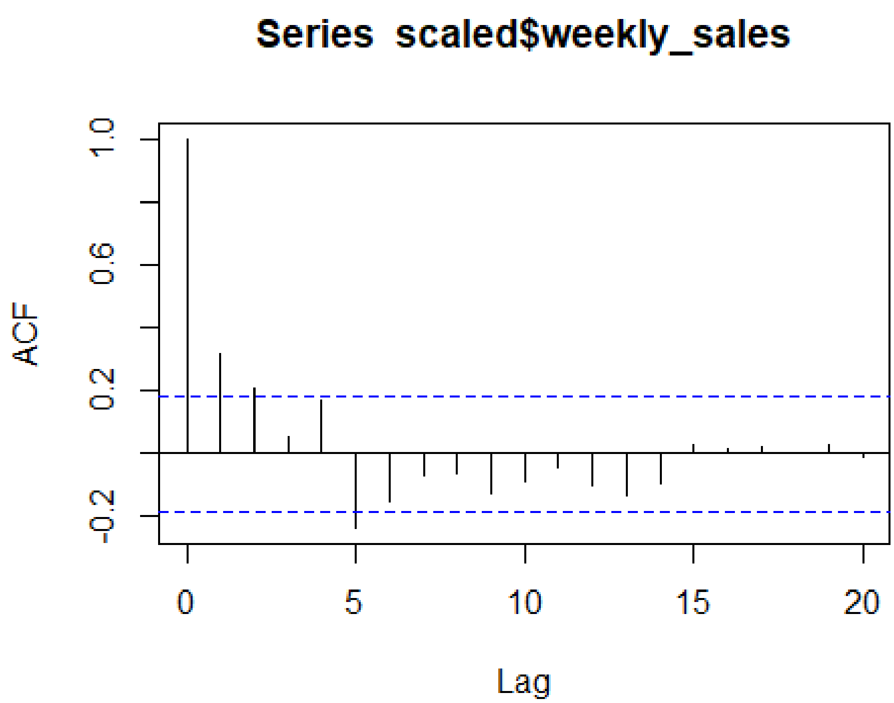
- The overall behavior indicates that differencing might be required to stabilize the



series.

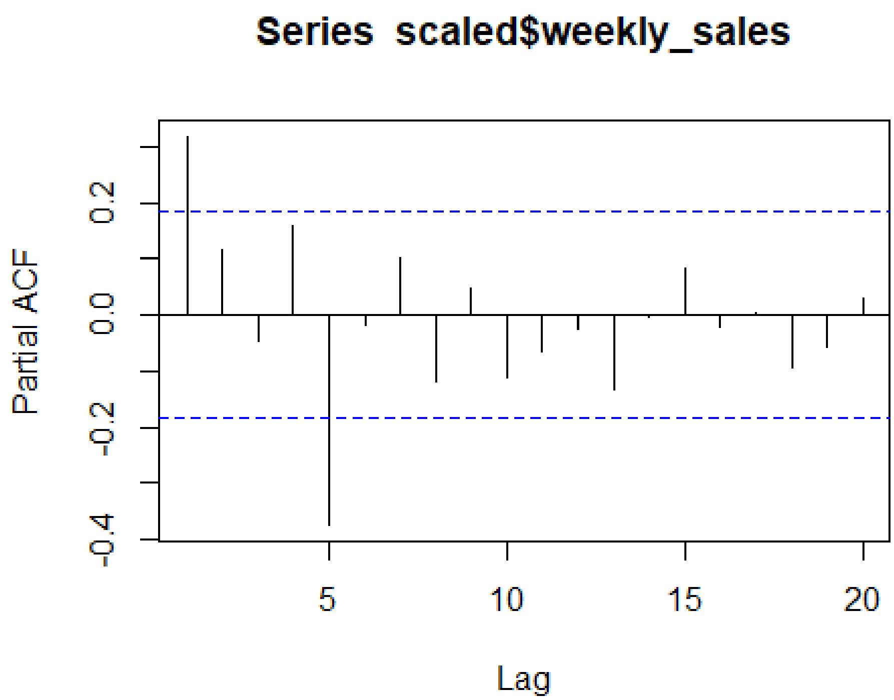
ACF | MA(q) Review

- Maybe MA(4)



PACF | AR(p) Review

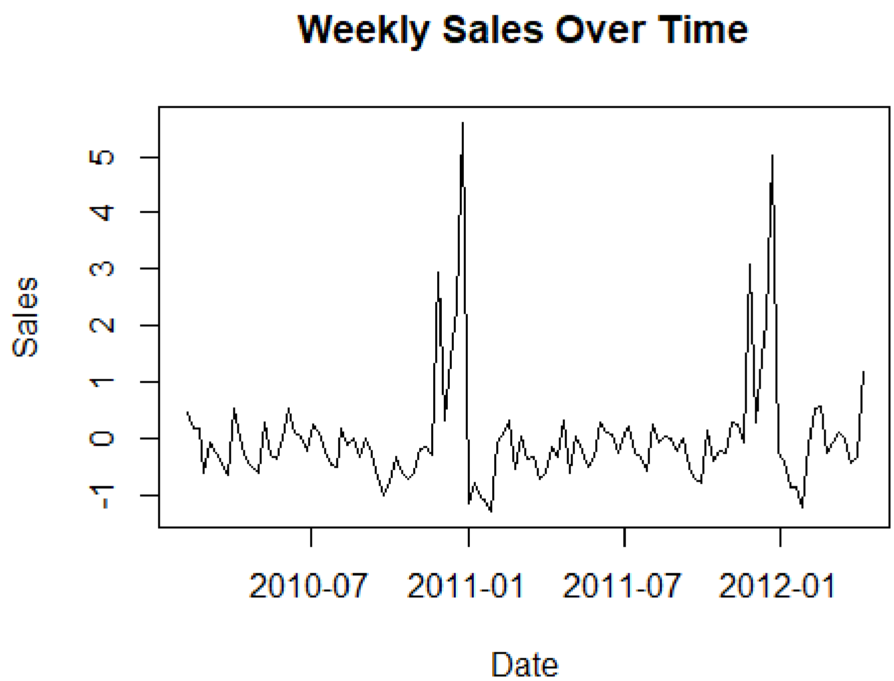
- maybe a AR(4)
- scaled it looks like an AR(2)



Plot for Trend

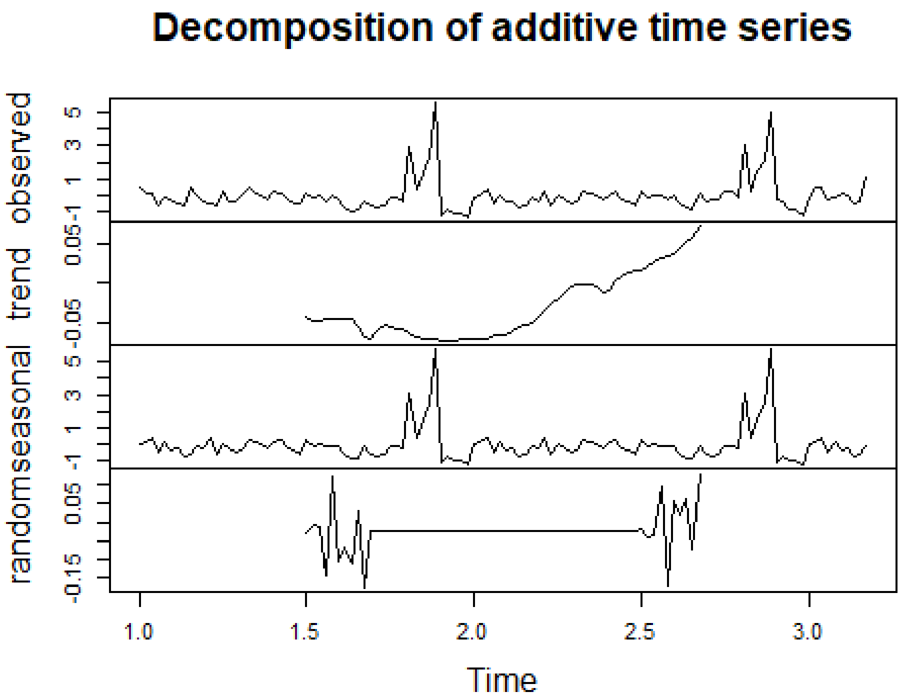
- It looks like there are spike around 13(11 - Nov) and 14(12 - Dec) period.

- Maybe a frequency peak at around .071 - .077



Decompose Time Series

Summary of Insights - **Trend**: A clear upward trend is present between time 2 and 3, indicating long-term growth. - **Seasonality**: The data has a strong seasonal component, which should be explicitly modeled. - **Stationarity**: The presence of both trend and seasonality suggests the series is non-stationary. First-order differencing ($d=1$) and/or seasonal differencing ($D=1$) may be required to stabilize the series for ARIMA modeling. - **Outliers**: The sharp spikes in the residuals suggest potential outliers or irregular events that may require further investigation or adjustments. —



Dickey-Fuller Test for Stationarity

- Reject the null hypothesis of Non-Stationarity
- This test was not as helpful as I had hopped cause we know that we need to difference the model from the plots above.
- This is actually a non-stationary model based off Kaggle as well. On the site they already mention this is a non-stationary model

ARMA MODEL

AIC Model Determination

- I will test the following three models
    - ARMA(5,1) increase the values for more options since we are at the max
        - ARMA(5,1) is still in the top 5 so I will try this option
        - ARMA(5,3) and ARMA(6,1) will be the next I play with.
- After testing I found that ARMA(5,3) performs the best

```
## ---------WORKING... PLEASE WAIT...
##
##
## Five Smallest Values of  aic
##    p   q        aic
##    5   3 -0.2302411
##    5   0 -0.2106635
```

```
##     6    1 -0.1933963
##     6    0 -0.1932953
##     5    1 -0.1932337
```

BIC Model Determination

- Not going to use this model cause we know this is not white noise. So I increased the model
- Still not valid options

```
## ---------WORKING... PLEASE WAIT...

## Error in aic calculation at 0 0

## Error in aic calculation at 0 1

## Error in aic calculation at 0 2

## Error in aic calculation at 0 3

## Error in aic calculation at 1 0

## Error in aic calculation at 1 1

## Error in aic calculation at 1 2

## Error in aic calculation at 1 3

## Error in aic calculation at 2 0

## Error in aic calculation at 2 1

## Error in aic calculation at 2 2

## Error in aic calculation at 2 3

## Error in aic calculation at 3 0

## Error in aic calculation at 3 1

## Error in aic calculation at 3 2

## Error in aic calculation at 3 3

## Error in aic calculation at 4 0

## Error in aic calculation at 4 1

## Error in aic calculation at 4 2

## Error in aic calculation at 4 3

## Error in aic calculation at 5 0

## Error in aic calculation at 5 1
```

```
## Error in aic calculation at 5 2

## Error in aic calculation at 5 3

## Error in aic calculation at 6 0

## Error in aic calculation at 6 1

## Error in aic calculation at 6 2

## Error in aic calculation at 6 3

## Error in aic calculation at 7 0

## Error in aic calculation at 7 1

## Error in aic calculation at 7 2

## Error in aic calculation at 7 3
## Five Smallest Values of  bic
##     p    q        bic
##     0    0     999999
##     0    1     999999
##     0    2     999999
##     0    3     999999
##     1    0     999999
```

ARMA(5,3) Factor Table Check | Model Estimates

- Here the 1-B is very apparent. This might be the best model to use since I know I need to difference my model
- The frequency also seems to match up almost exactly with what I am seeing in the original realization

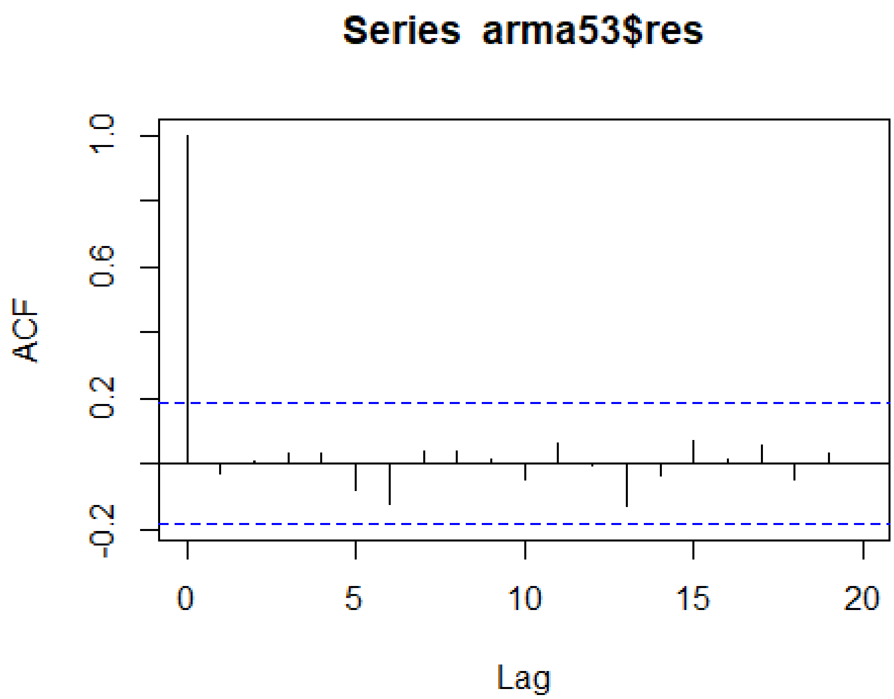
```
## 
## 
## Coefficients of AR polynomial: 
## 0.1845 0.5098 0.3896 0.0227 -0.3018
## 
##                        AR Factor Table
## Factor                 Roots                Abs Recip    System Freq
## 1-1.6012B+0.6501B^2    1.2315+-0.1471i      0.8063       0.0189
## 1+0.6551B+0.6095B^2   -0.5374+-1.1627i      0.7807       0.3189
## 1+0.7616B             -1.3131               0.7616       0.5000
## 
## 
## 
## 
## Coefficients of MA polynomial: 
## -0.1845 0.4290 0.7555
## 
##                        MA FACTOR TABLE
## Factor                 Roots                Abs Recip    System Freq
```

```
## 1-1.0000B            1.0000            1.0000      0.0000
## 1+1.1845B+0.7555B^2  -0.7839+-0.8421i    0.8692      0.3693
##
##
```

RESIDUAL CHECKS ARMA(5,3)

ACF Residual Check

- Passed the ACF White Noise Residual Check

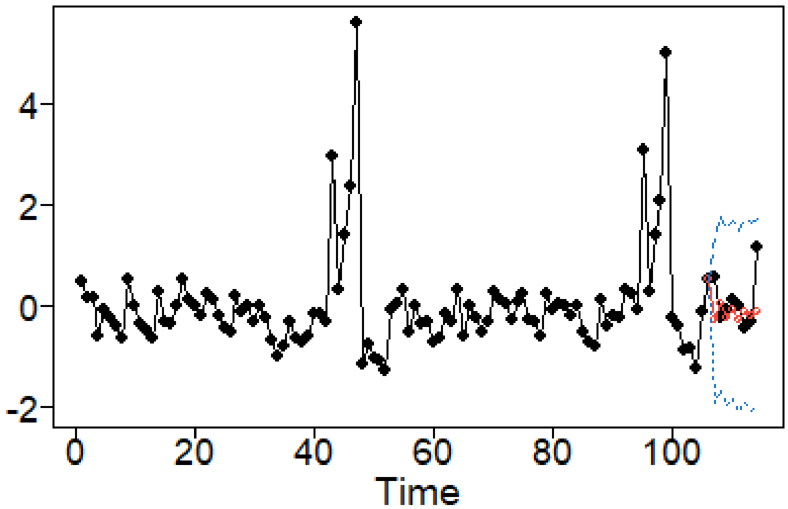


Ljung Residual Check

- Fail to reject the Null Hypothesis of White Noise as the p-value is greater then .05 for both tests
- This is another pass for this model

```
## p-value 0.8608067

## p-value 0.9770861
```

Forecast ARMA Model



ASE

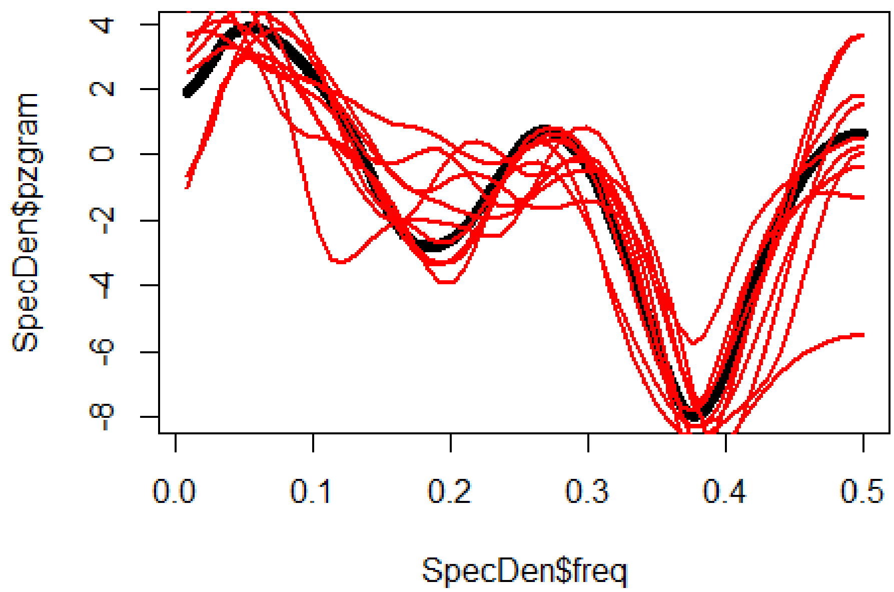
```
## [1] 1.014748
```

WMAE | Kaggle

```
## [1] 0.8192932
```

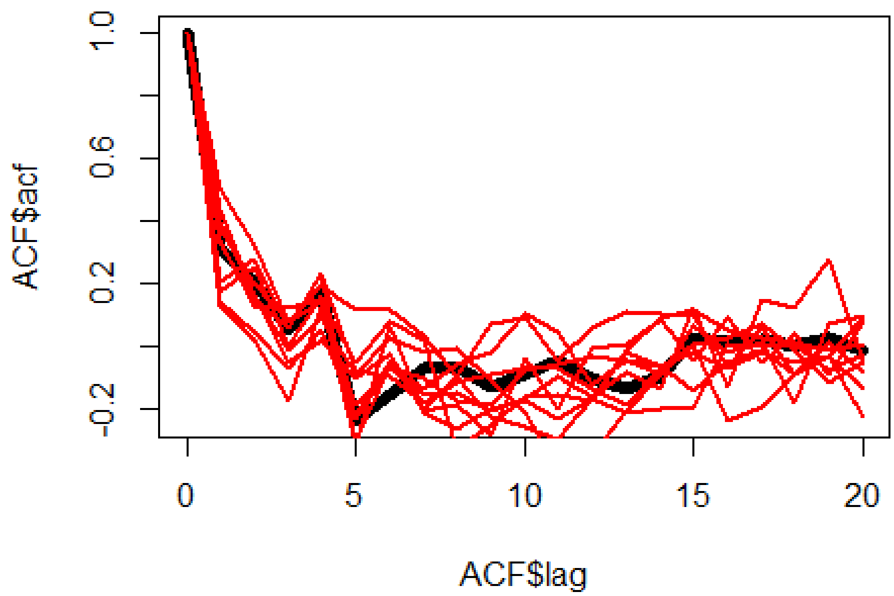
Compare Multiple Spectral Densities

- This model appears to perform well with generating the spectral densities



Compare Multiple ACFs

- Does fairly well modeling the ACFs



FINAL ARMA(5,3) MODEL

$$(1 - 0.1845B - 0.5098B^2 - 0.3896B^3 - 0.0227B^4 + 0.3018B^5)(X_t + 4.470451e{-16}) = (1 + 0.1845B - 0.4290B^2 - 0.7555B^3)a_t,\ \hat{\sigma}_a^2 = 0.6783202$$

ARIMA (5,1,3)

Dickey-Fuller Test for Stationarity

- Reject the null hypothesis of Non-Stationarity

AIC Model Determineation

- We don't see ARMA(5,3) in the top options but decided to move forward with this option do the results from the first model

```
## ---------WORKING... PLEASE WAIT...
##
##
## Five Smallest Values of  aic
##     p    q        aic
##     5    2 -0.1757346
##     5    1 -0.1632874
##     6    3 -0.1589239
```

```
##    6   2 -0.1587582
##    7   2 -0.1518339
```

BIC Model Determination

```
## ---------WORKING... PLEASE WAIT...
##
##
## Five Smallest Values of  bic
##    p   q          bic
##    1   3 -0.003345037
##    5   1  0.005665782
##    5   2  0.017354819
##    6   2  0.058467360
##    5   3  0.077225616
```

ARIMA (5,1,3) Factor Table Check | Model Estimates

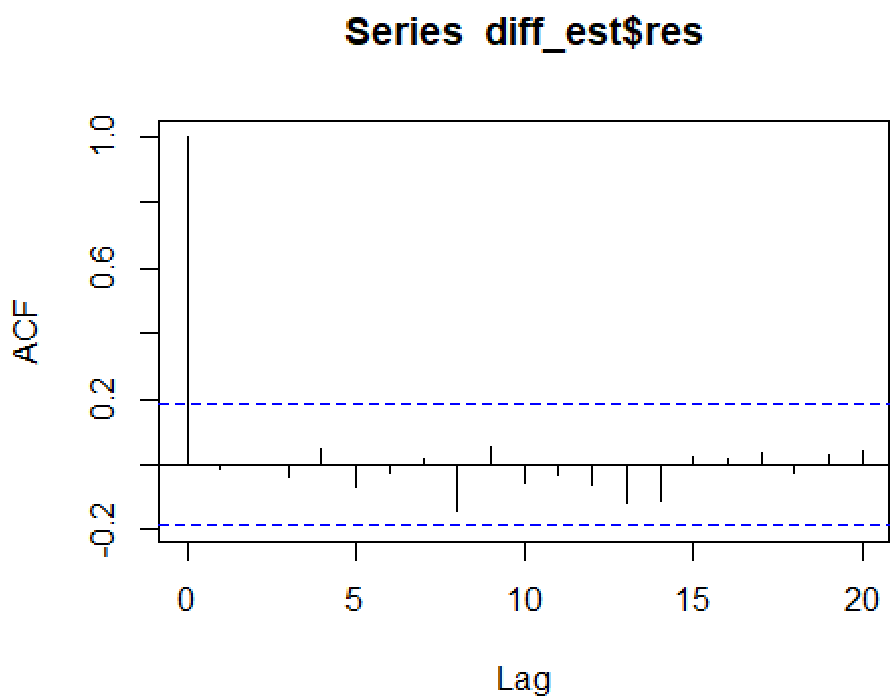
- Frequencies don't match exactly with the original model as the dominance has change
- We still see the 1-B in the Moving Average Model

```
##
##
## Coefficients of AR polynomial:
## 0.2202 -0.0766 0.0163 0.2829 -0.3569
##
##                           AR Factor Table
## Factor                 Roots                Abs Recip    System Freq
## 1+0.2570B+0.7446B^2   -0.1726+-1.1459i      0.8629       0.2738
## 1+0.8433B             -1.1858               0.8433       0.5000
## 1-1.3206B+0.5683B^2    1.1619+-0.6401i      0.7539       0.0801
##
##
##
##
## Coefficients of MA polynomial:
## 0.8525 -0.0815 0.2290
##
##                              MA FACTOR TABLE
## Factor                 Roots                Abs Recip    System Freq
## 1-1.0000B              1.0000               1.0000       0.0000
## 1+0.1475B+0.2290B^2   -0.3221+-2.0647i      0.4786       0.2746
##
##
```

RESIDUAL CHECKS ARIMA (5,1,3)

ACF Residual Check

- Passed the ACF White Noise Residual Check

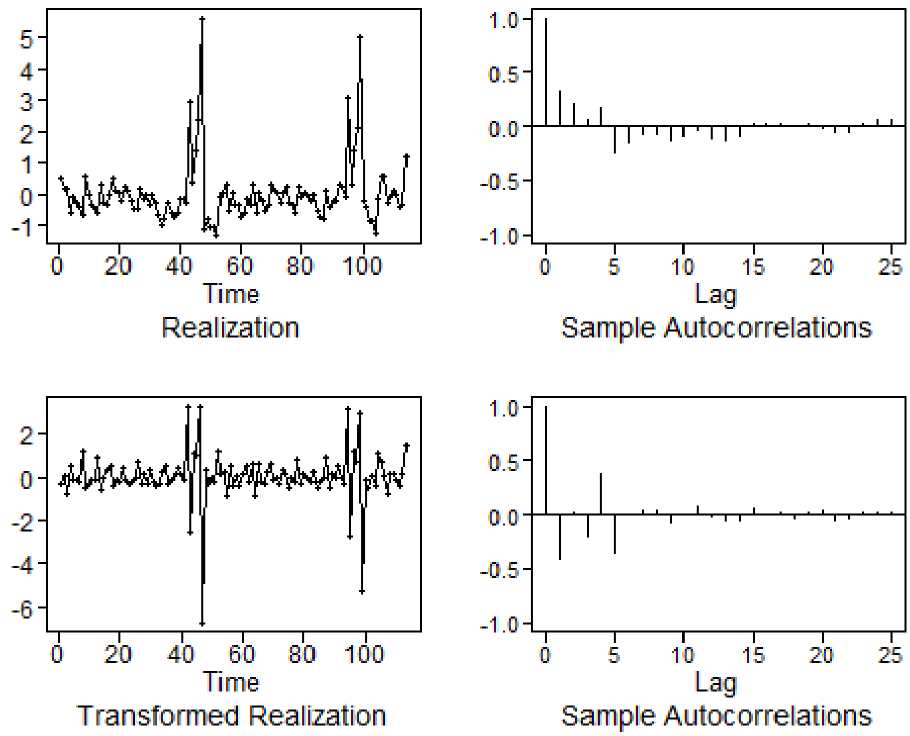


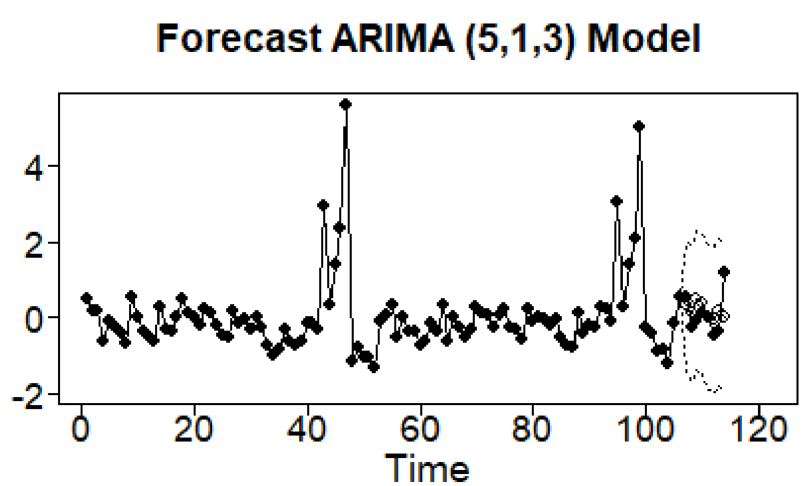
Ljung Residual Check

- Fail to reject the Null Hypothesis of White Noise as the p-value is greater then .05 for both tests
- This is another pass for this model

```
## p-value 0.7744939

## p-value 0.9933498
```

Forecast ARIMA (5,1,3) Model





ASE

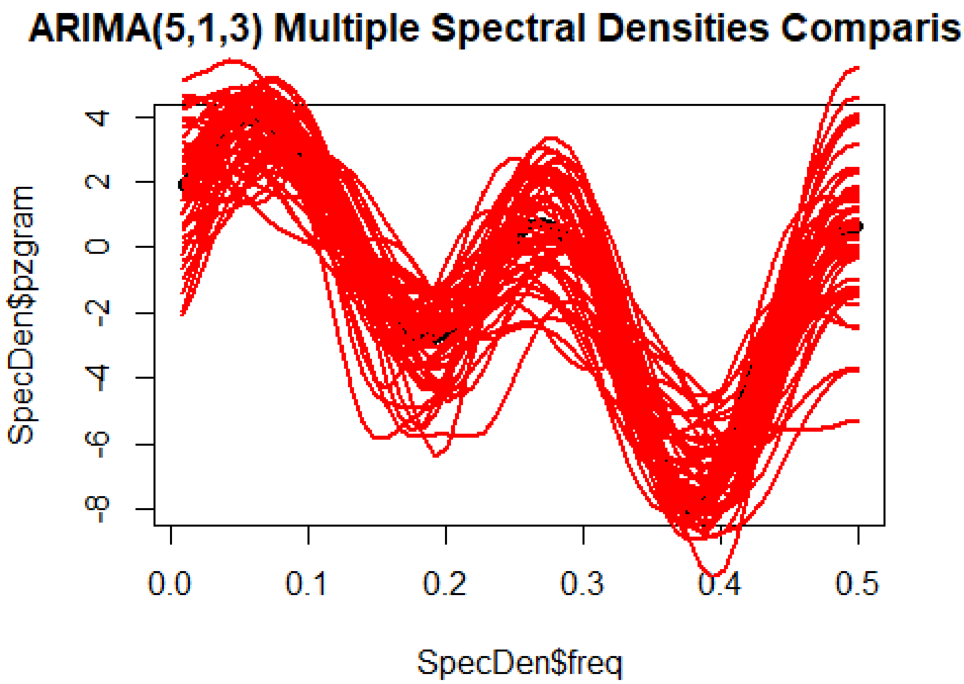
```
## [1] 1.077798
```

WMAE | Kaggle

```
## [1] 0.8133012
```

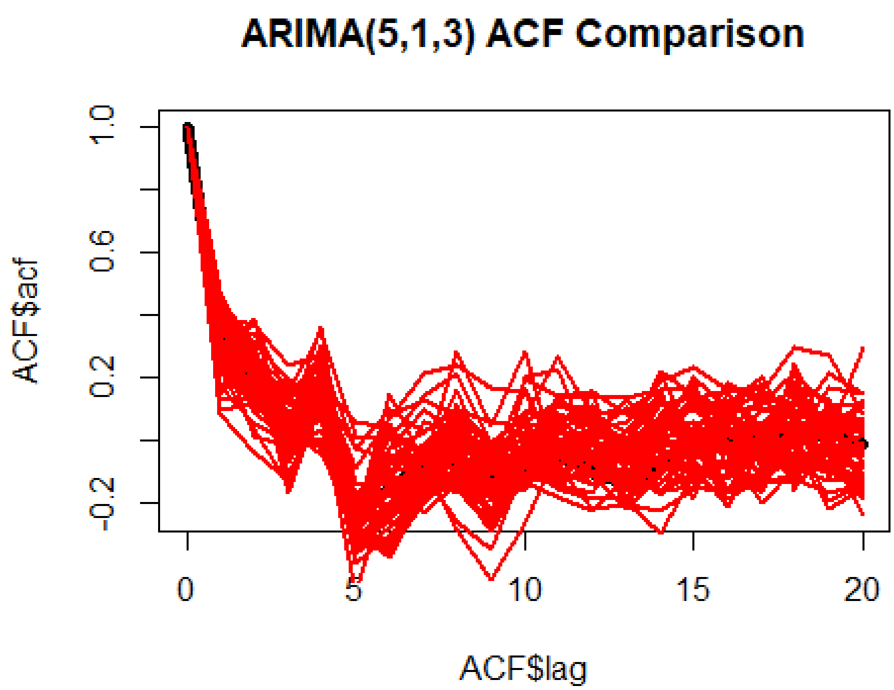
Compare Multiple Spectral Densities

- This model appears to perform well with generating the spectral densities



Compare Multiple ACFs

- Does fairly well modeling the ACFs



FINAL ARIMA (5,1,3) MODEL

$$(1-0.2202B + 0.0766B^2 - 0.0163B^3 - 0.2829B^4 + 0.3569B^5)(1-B)(X_t - 0.006105923)$$
$$= (1-0.8525B + 0.0815B^2 - 0.2290B^3)a_t,\ \hat{\sigma}_a^2 = 0.7413429$$

ARUMA(3,1,0) w/ s=52

Dickey-Fuller Test for Stationarity

- Reject the null hypothesis of Non-Stationarity

AIC Model Determineation

```
## ---------WORKING... PLEASE WAIT...
##
##
## Five Smallest Values of  aic
##     p    q        aic
##     3    0  -1.984716
##     7    0  -1.972271
##     4    0  -1.963414
##     6    0  -1.943017
##     5    0  -1.941745
```

BIC Model Determineation

```
## ---------WORKING... PLEASE WAIT...
##
##
## Five Smallest Values of  bic
##     p    q        bic
##     3    0  -1.846299
##     1    0  -1.831823
##     2    0  -1.819269
##     4    0  -1.790391
##     0    0  -1.737236
```

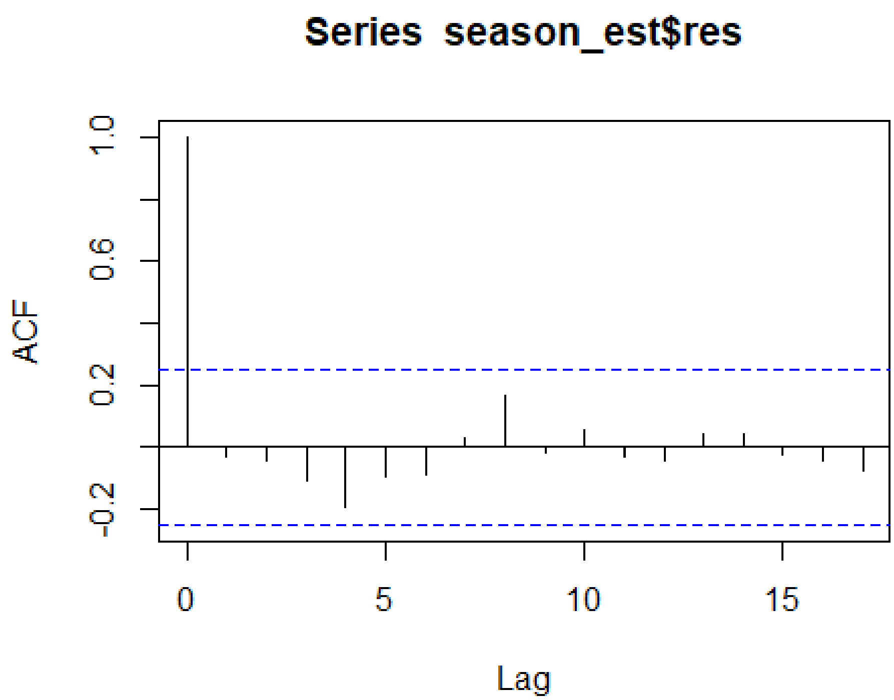
ARUMA(3,1,0) w/ s = 52 Factor Table Check | Model Estimates

```
##
##
## Coefficients of AR polynomial:
## -0.5804 -0.4008 -0.3096
##
##                           AR Factor Table
## Factor                    Roots                Abs Recip    System Freq
## 1-0.0904B+0.4615B^2    0.0980+-1.4688i      0.6793       0.2394
## 1+0.6708B            -1.4907               0.6708      0.5000
##
##
```

RESIDUAL CHECKS ARUMA(3,1,0) w/ s=52

ACF Residual Check

- Passed the ACF White Noise Residual Check

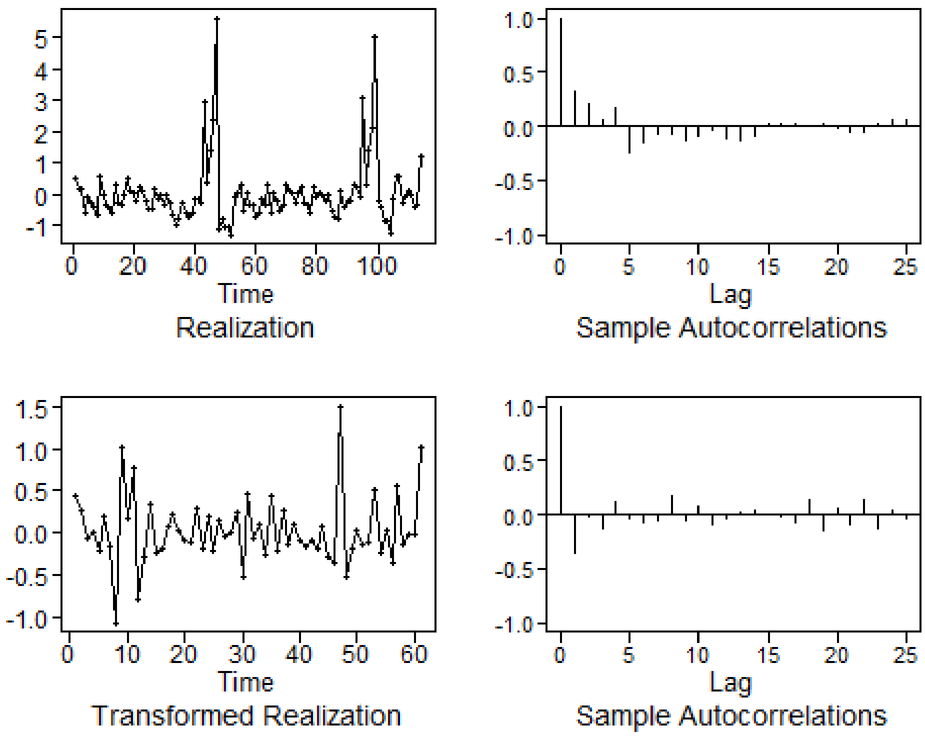


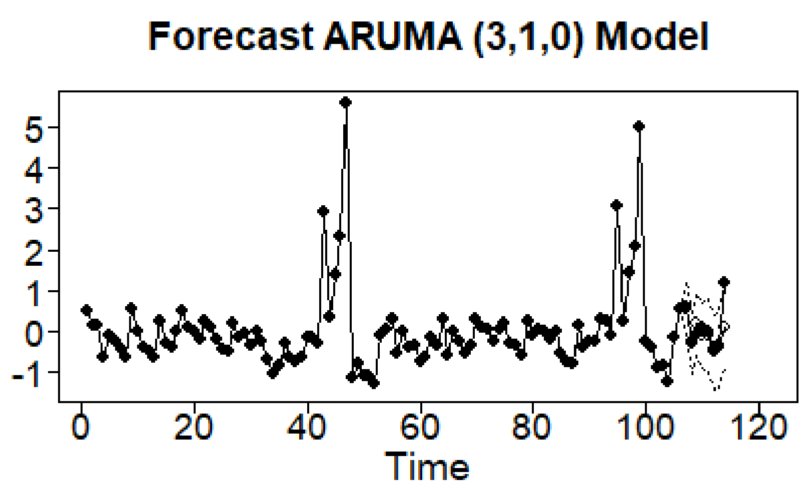
Ljung Residual Check

- Fail to reject the Null Hypothesis of White Noise as the p-value is greater then .05 for both tests
- This is another pass for this model

```
## p-value 0.9764235

## p-value 0.7185762
```

Forecast ARUMA(3,1,0) w/ s=52 Model





ASE

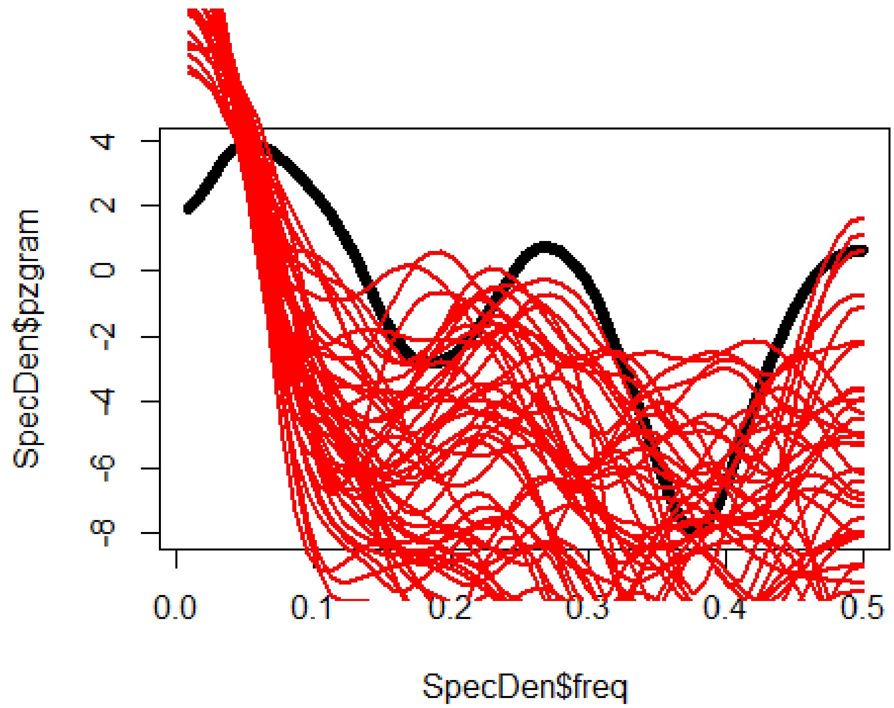
```
## [1] 1.099792
```

WMAE | Kaggle

```
## [1] 0.8255756
```

Compare Multiple Spectral Densities

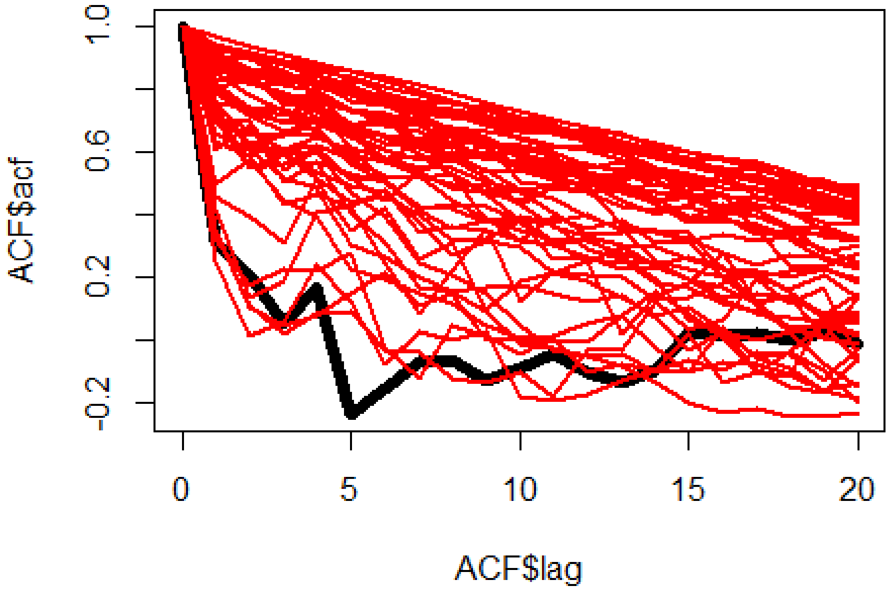
- This model does not appear to perform very well with generating the spectral



densities

Compare Multiple ACFs

- Doesn't do very well modeling the ACFs

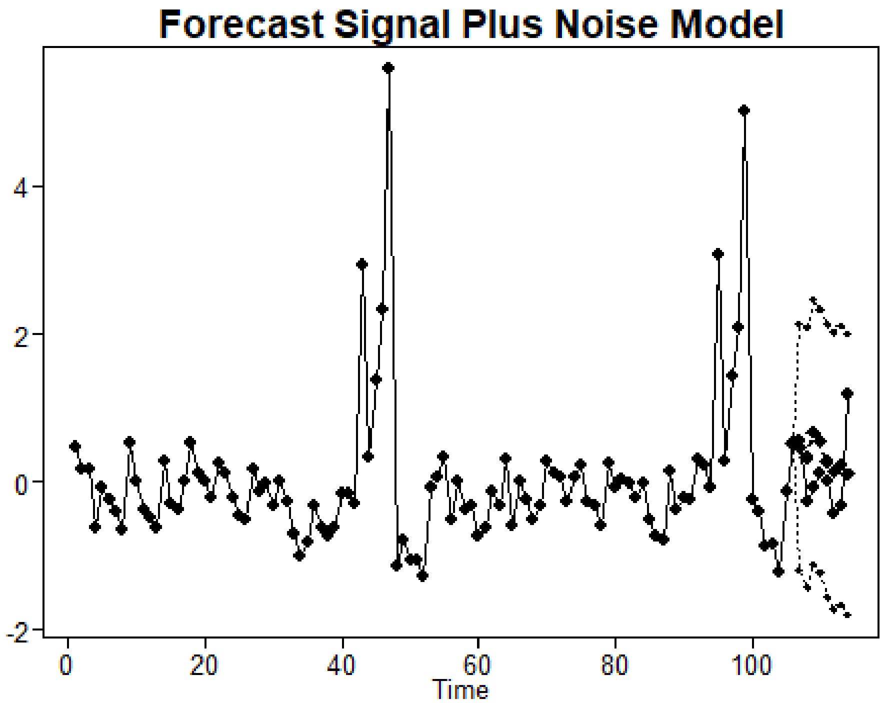


FINAL ARUMA (3,1,0) MODEL with s=52

$(1 + 0.5804B + 0.4008B^2 + 0.3096B^3)(1-B)(X_t-0.1205291) = a_t$, $\hat{\sigma}_a^2 = 0.1205291$

Signal Plus Noise

Forecast Scaled Signal Plus Noise as Best Model



ASE

```
## [1] 1.139449
```

WMAE | Kaggle

```
## [1] 0.8257315
```

FINAL SIGNAL PLUS NOISE REGRESSION EQUATION

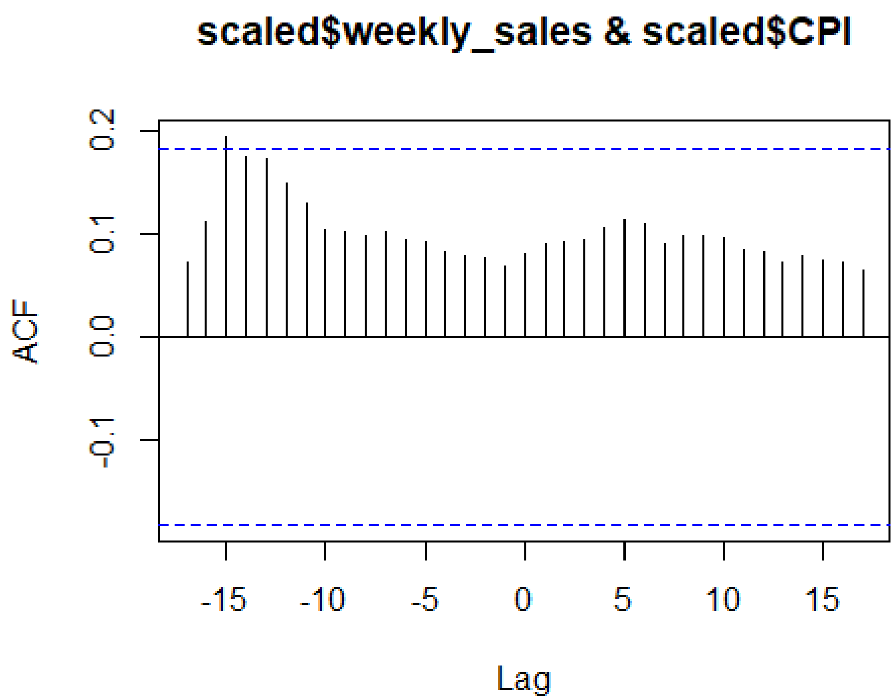
- Sales tend to decrease overtime

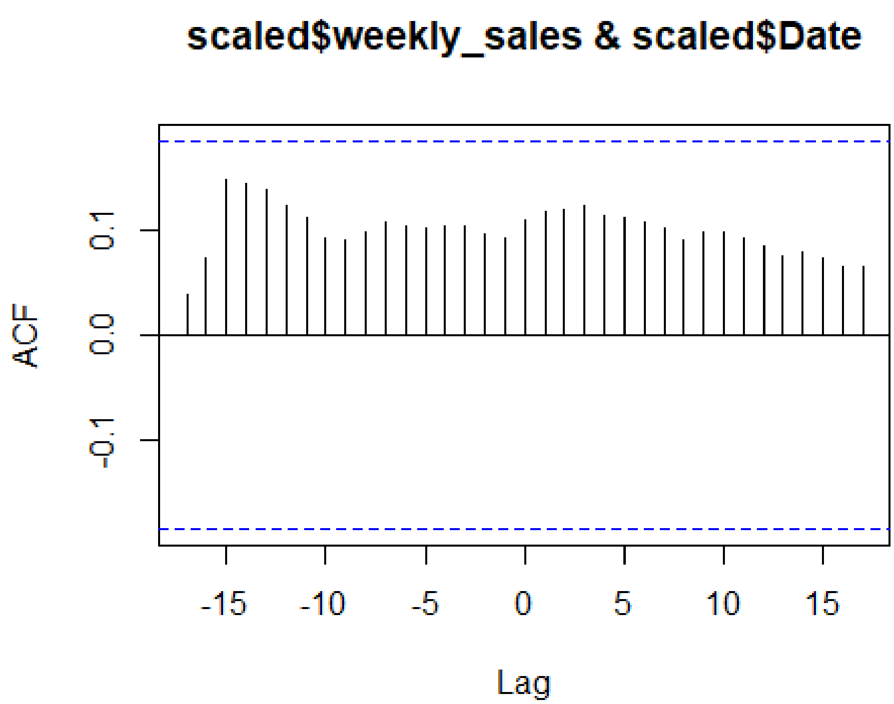
$$\text{sales} = -0.1896494 + .003298251(time)$$

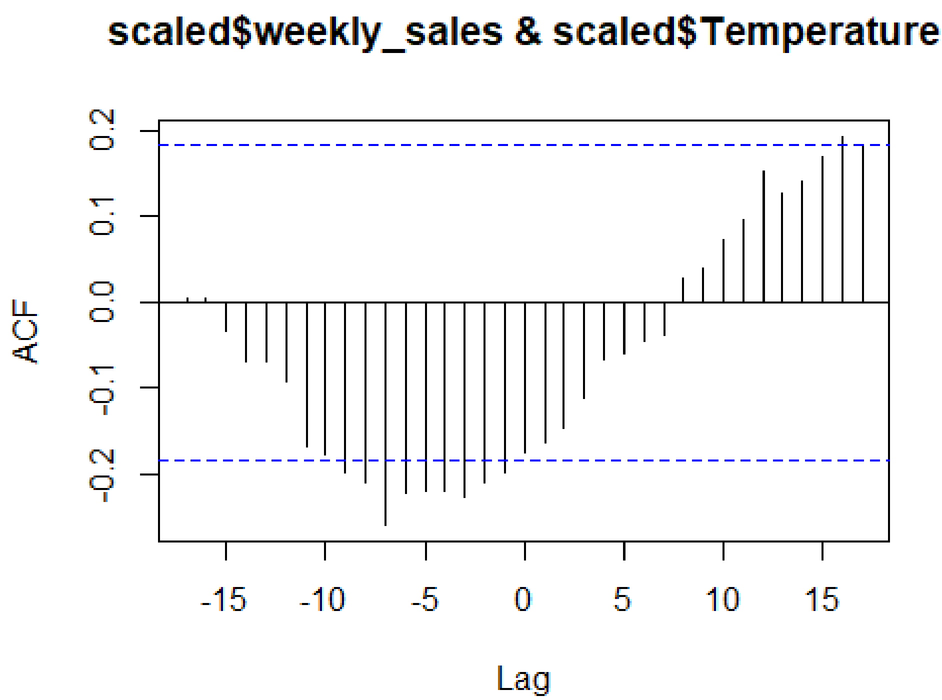
ARUMA MULTIVARIATE MODEL

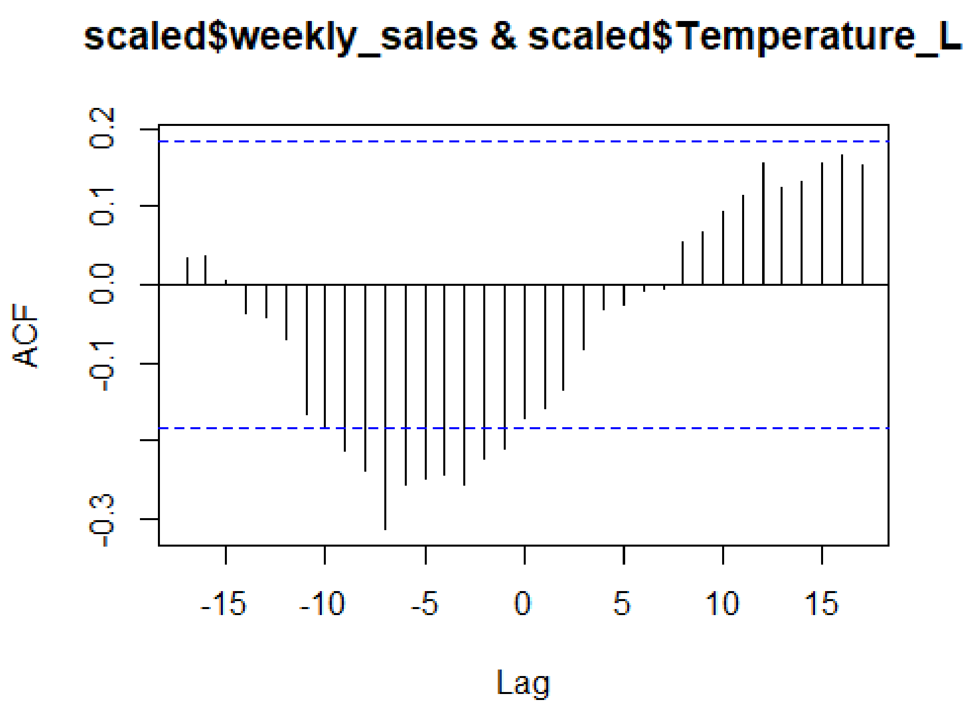
Cross-Correlation Plots

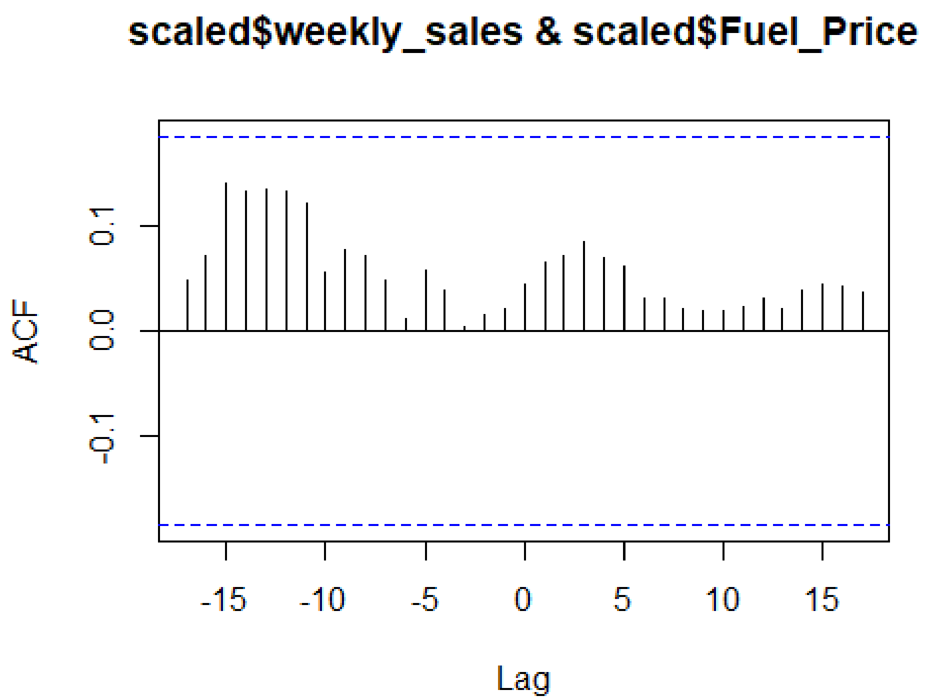
- In this case we are looking for the longest lag to determine how to lag multivariate models
- Since we know Holiday is important and our lags are across the board, we decided to keep it simple and lag everything by 1

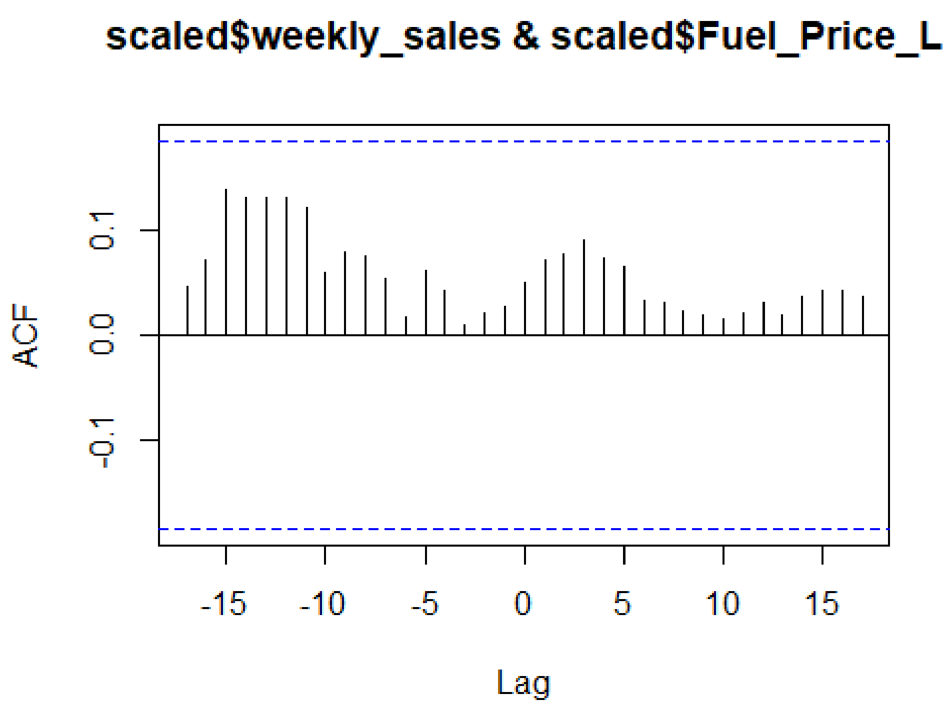

scaled$weekly_sales & scaled$CPI
ACF
0.2
0.1
0.0
-0.1
-15
-10
-5
0
5
10
15
Lag

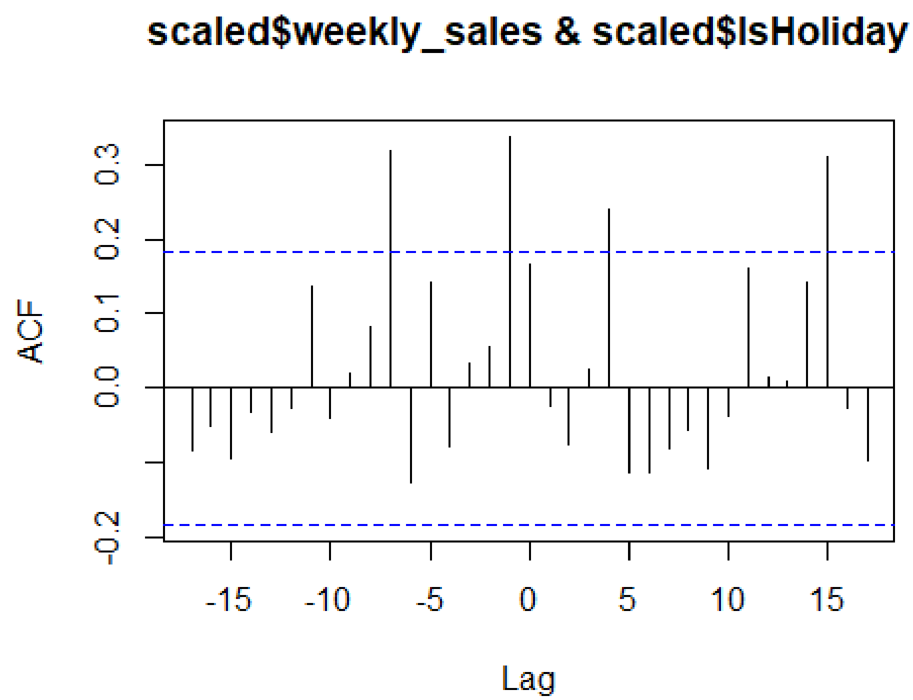

scaled$weekly_sales & scaled$Date
ACF
0.1
0.0
-0.1
-15
-10
-5
0
5
10
15
Lag

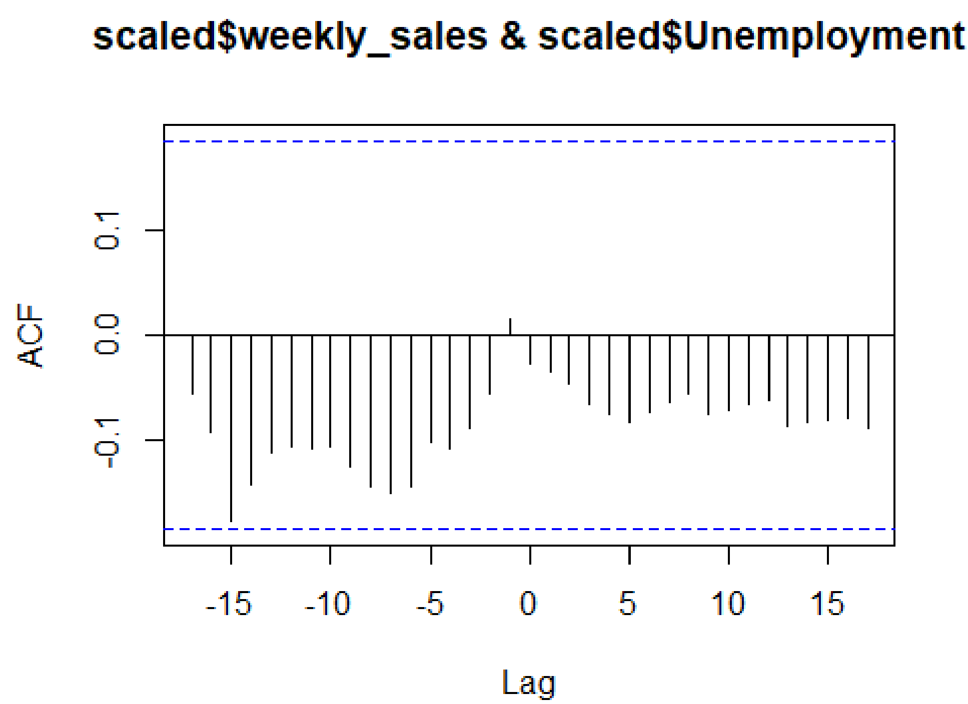

scaled$weekly_sales & scaled$Temperature
ACF
0.2
0.1
0.0
-0.1
-0.2
-15
-10
-5
0
5
10
15
Lag


scaled$weekly_sales & scaled$Temperature_L
ACF
0.2
0.1
0.0
-0.1
-0.3
-15
-10
-5
0
5
10
15
Lag


scaled$weekly_sales & scaled$Fuel_Price
ACF
0.1
0.0
-0.1
-15
-10
-5
0
5
10
15
Lag


scaled$weekly_sales & scaled$Fuel_Price_L
ACF
0.1
0.0
-0.1
-15
-10
-5
0
5
10
15
Lag


scaled$weekly_sales & scaled$IsHoliday
ACF
0.3
0.2
0.1
0.0
-0.2
-15
-10
-5
0
5
10
15
Lag


scaled$weekly_sales & scaled$Unemployment
ACF
0.1
0.0
-0.1
-15
-10
-5
0
5
10
15
Lag

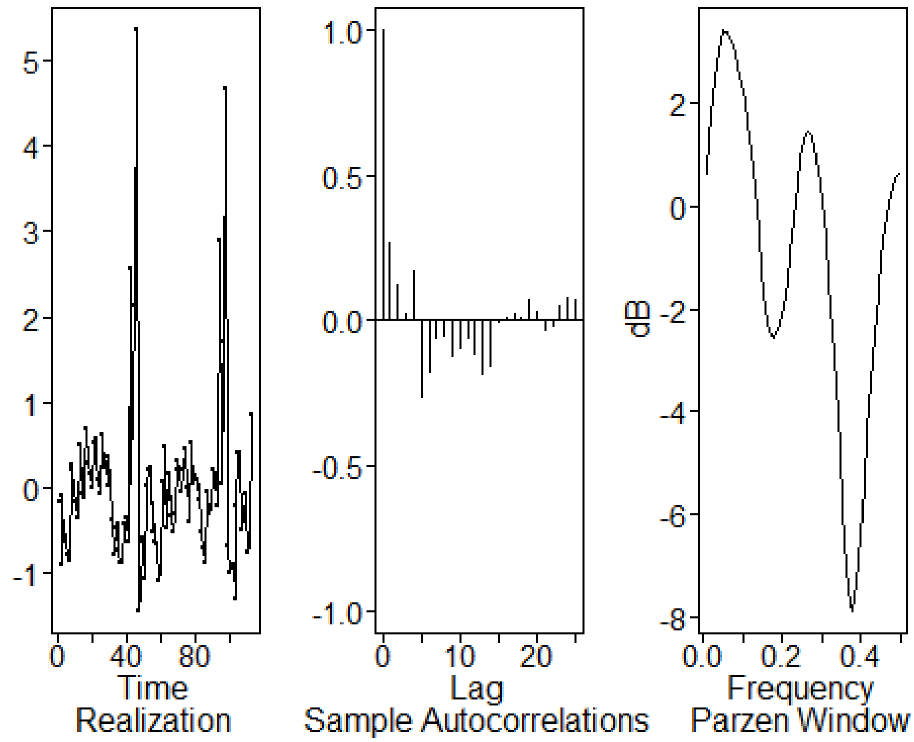
Lag Variables

Fit Regression Model

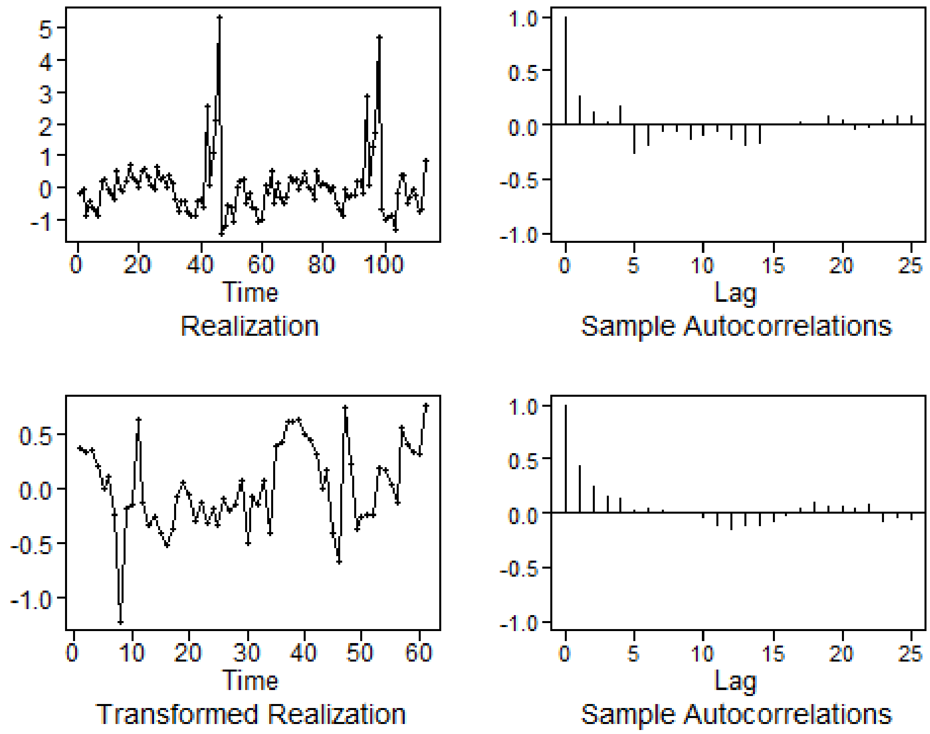
- will use this as the base line for the multivariate model

Check Regression Residual

- Ensure the residuals have been whitened



Difference Regression Residuals

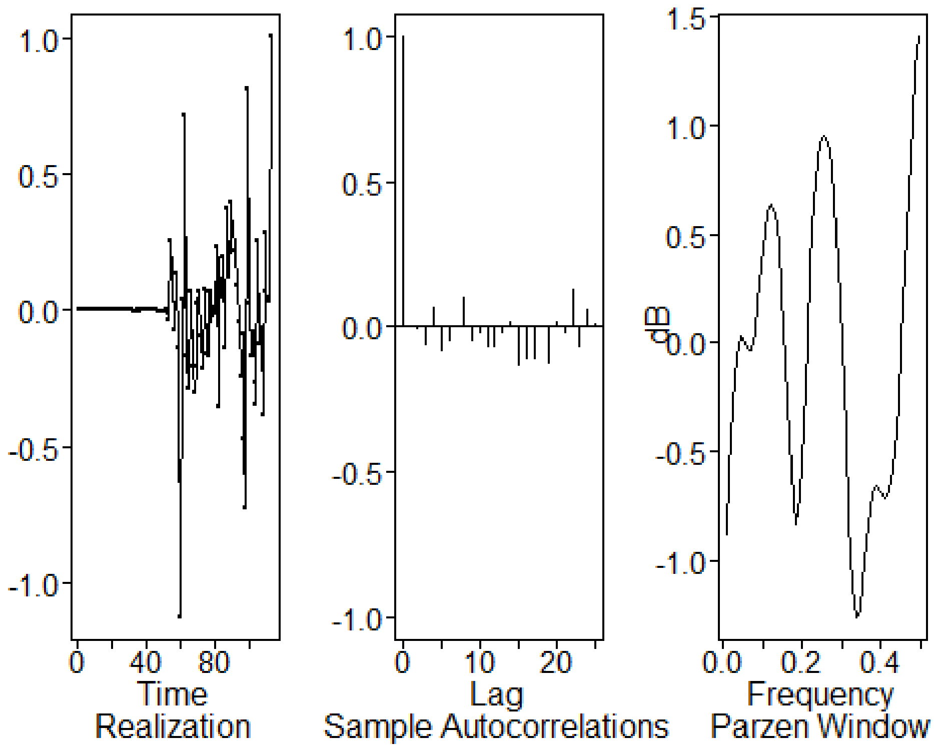


Model Regression Residual

AIC

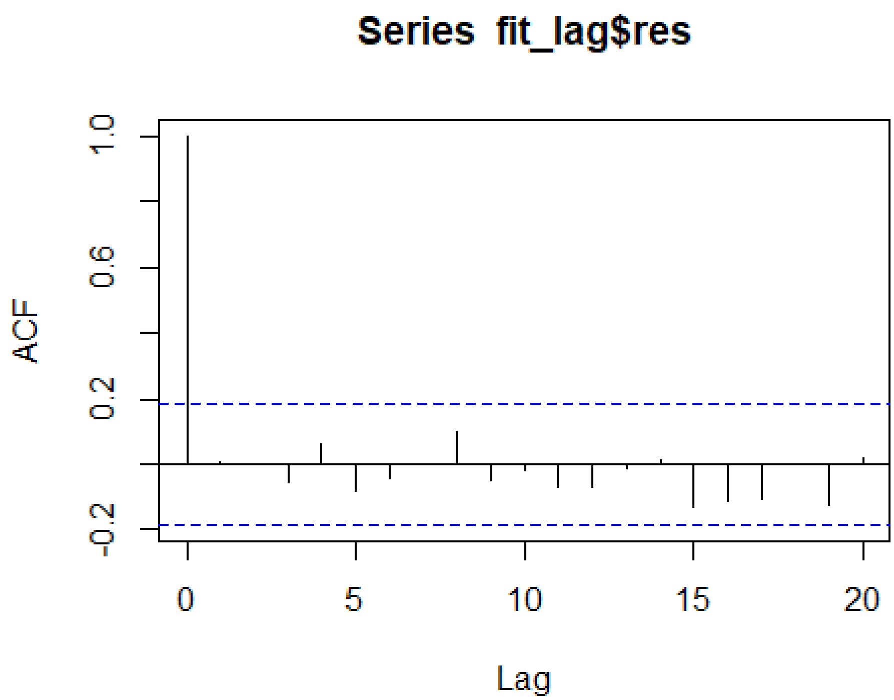
BIC

Forecast ARUMA (1,0,0) w/ s=52



ACF Residual Check

- Passed the ACF White Noise Residual Check

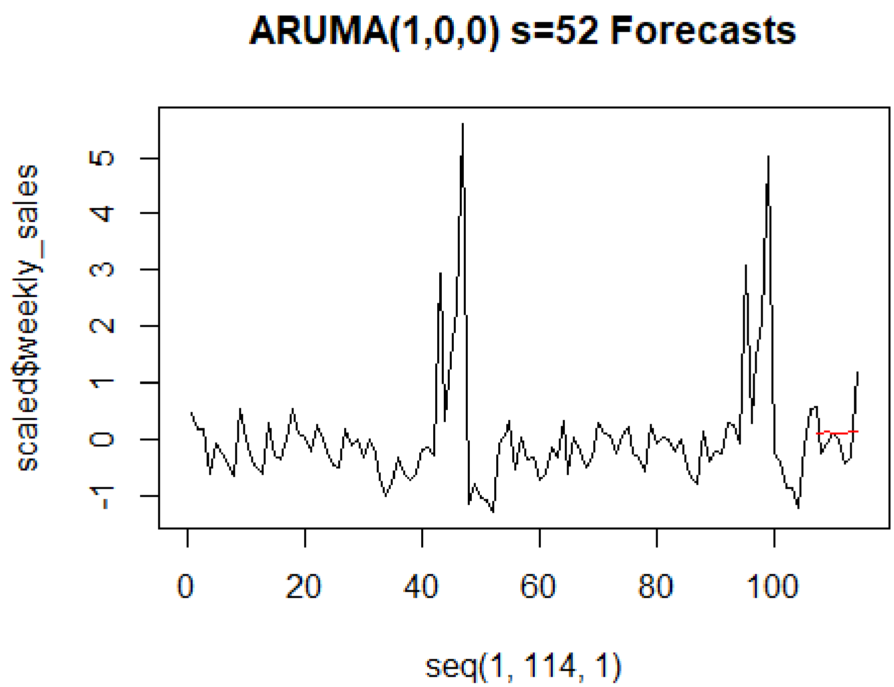


Ljung Residual Check

- Fail to reject the Null Hypothesis of White Noise as the p-value is greater then .05 for both tests
- This is another pass for this model

```
## p-value 0.8664807

## p-value 0.7185762
```

MULTIVARIATE ARUMA(1,0,0) w/ s=52 Predictions



ASE

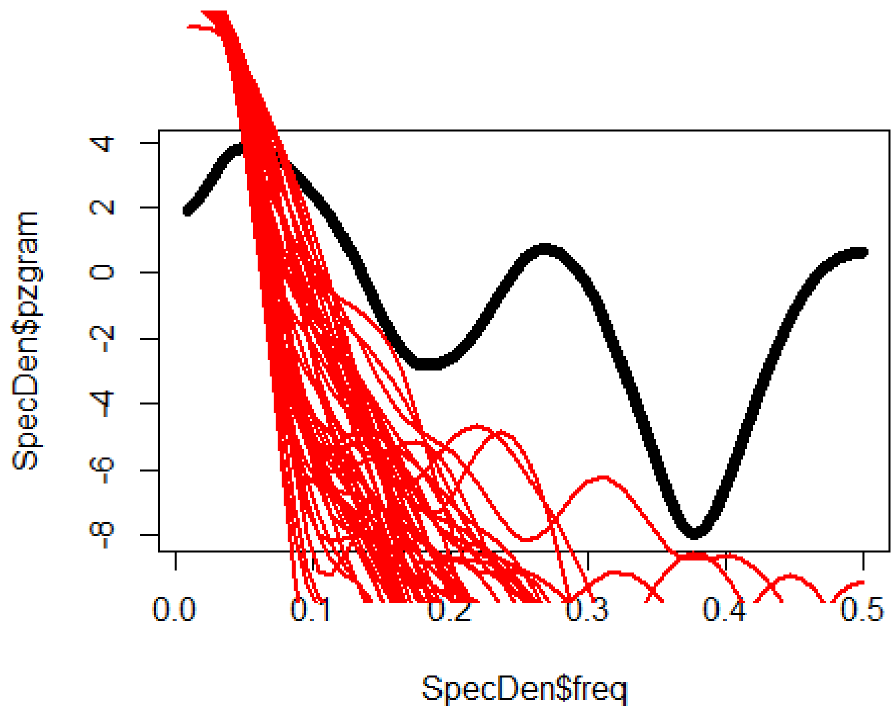
```
## [1] 0.2549112
```

WMAE | Kaggle

```
## [1] 1.032607
```

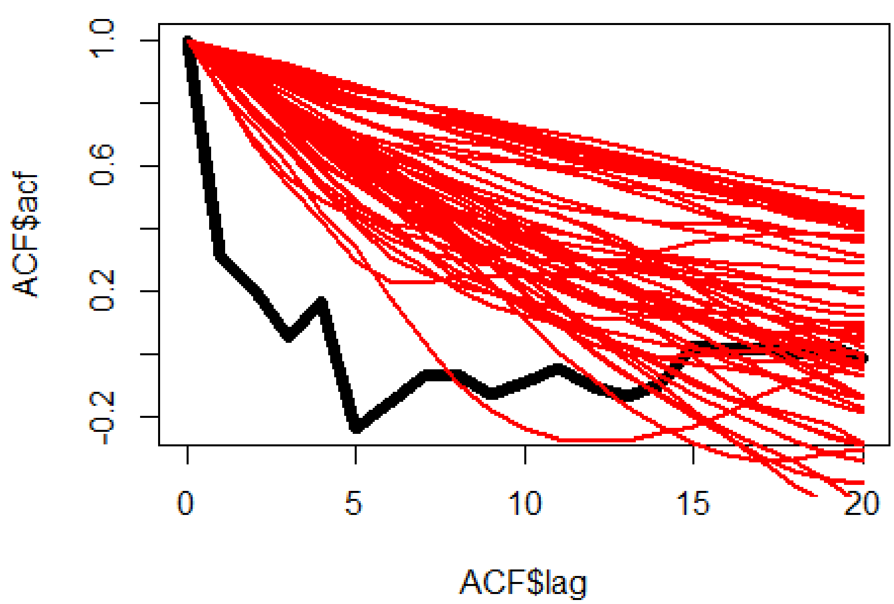
Compare Multiple Spectral Densities

- This model does not appear to perform well with generating the spectral densities



Compare Multiple ACFs

- Does not do fairly well modeling the ACFs



FINAL ARUMA(1,0,0) w/ s=52 MODEL

sales

$$= (1-0.1780B) - 0.0665(lag.fuel) + 0.2753(lag.cpi) + 0.0083(lag.temp) + 0.1352(lag.unemployment)$$

VAR Model

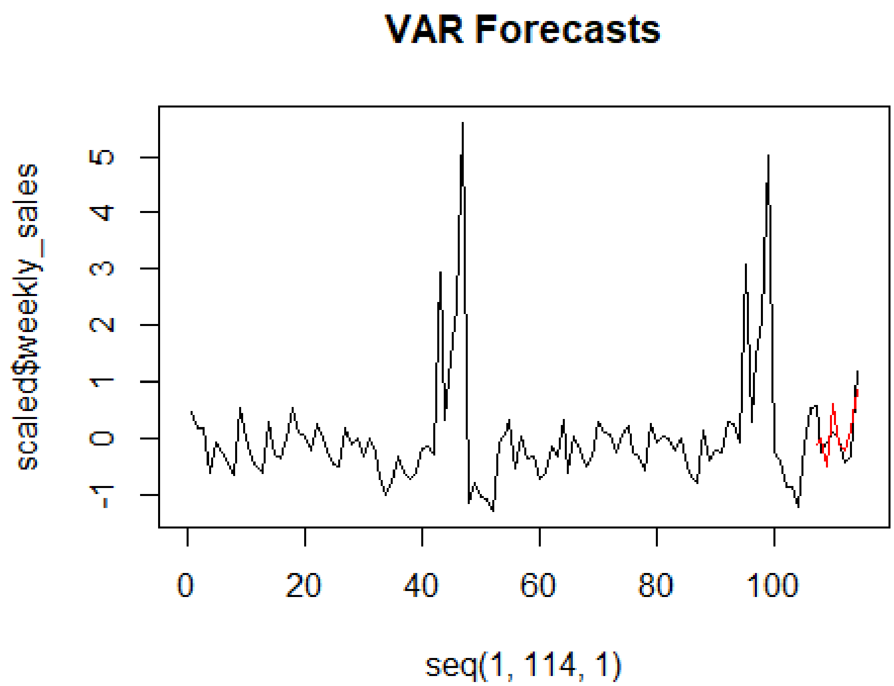
```
## $selection
## AIC(n)  HQ(n)  SC(n) FPE(n)
##      5      5      5      5
##
## $criteria
##                 1          2          3          4           5            6
## AIC(n) 0.01839696 0.03018348 0.04156068 0.04194668 -0.11716576 -0.103400516
## HQ(n)  0.07915561 0.10106857 0.12257221 0.13308466 -0.01590134  0.007990343
## SC(n)  0.16827521 0.20504144 0.24139835 0.26676406  0.13263133  0.171376280
## FPE(n) 1.01868719 1.03083651 1.04272788 1.04325444  0.88992446  0.902423317
##                  7
## AIC(n) -0.08809677
## HQ(n)   0.03342053
## SC(n)   0.21165973
## FPE(n)  0.91654031
```

VAR Model Lag Predictions | Model Estimates

Forecast VAR(5)

```
## $sales
##             fcst      lower      upper        CI
## [1,] -0.12380363 -0.6669861 0.41937884 0.5431825
## [2,] -0.01439203 -0.5728558 0.54407177 0.5584638
## [3,] -0.50304489 -1.0868796 0.08078984 0.5838347
## [4,]  0.59568648 -0.0220197 1.21339266 0.6177062
## [5,] -0.10644980 -0.7358594 0.52295979 0.6294096
## [6,] -0.22854624 -0.8617120 0.40461947 0.6331657
## [7,]  0.13757977 -0.5019809 0.77714042 0.6395606
## [8,]  0.86199150  0.2186140 1.50536898 0.6433775
##
## $fuel_price
##          fcst     lower    upper        CI
## [1,] 1.279460 0.9309083 1.628012 0.3485518
## [2,] 1.681930 1.2779630 2.085898 0.4039673
## [3,] 1.375231 0.9124183 1.838044 0.4628129
## [4,] 1.765363 1.2446244 2.286102 0.5207388
## [5,] 1.970237 1.4171891 2.523285 0.5530480
## [6,] 1.753844 1.1712203 2.336468 0.5826241
## [7,] 1.562466 0.9323247 2.192607 0.6301411
## [8,] 1.508749 0.8604623 2.157036 0.6482867
##
```

```
## $temperature
##            fcst        lower     upper        CI
## [1,] 0.36813877  0.105098082 0.6311795 0.2630407
## [2,] 0.29718752 -0.005453702 0.5998287 0.3026412
## [3,] 0.40027297  0.058712640 0.7418333 0.3415603
## [4,] 0.01047876 -0.336517815 0.3574753 0.3469966
## [5,] 0.75614722  0.380766084 1.1315283 0.3753811
## [6,] 0.63783620  0.236035469 1.0396369 0.4018007
## [7,] 1.04179366  0.625331224 1.4582561 0.4164624
## [8,] 1.21872416  0.791937895 1.6455104 0.4267863
##
## $week
##            fcst      lower     upper       CI
## [1,] 106.149323   62.28988 150.00876 43.85944
## [2,]  31.038505  -17.80365  79.88065 48.84215
## [3,] 122.074539   69.04754 175.10154 53.02700
## [4,] -42.110152 -100.60384  16.38353 58.49368
## [5,] 119.005395   55.41884 182.59195 63.58656
## [6,]  79.944321   14.37431 145.51434 65.57002
## [7,] 118.466261   51.56387 185.36865 66.90239
## [8,]   8.495785  -59.65769  76.64926 68.15348
##
## $cpi
##          fcst    lower    upper        CI
## [1,] 2.134539 1.955909 2.313170 0.1786308
## [2,] 2.143376 1.928143 2.358609 0.2152329
## [3,] 2.088582 1.839356 2.337807 0.2492256
## [4,] 2.377224 2.089825 2.664622 0.2873984
## [5,] 2.387693 2.056794 2.718592 0.3308989
## [6,] 2.403513 2.040551 2.766475 0.3629617
## [7,] 2.462531 2.080257 2.844805 0.3822738
## [8,] 2.537409 2.138100 2.936717 0.3993084
```



Confidence Interval

```
## Average Lower CI: -0.4550076

## Average Upper CI: 0.8153881
```

ASE

```
## [1] 0.1653816
```

WMAE | Kaggle

```
## [1] 0.3707223
```

FINAL VAR Regression Model

Sales Equation:

$$ sales_t = -0.6228 + 0.1681 \cdot sales_{t-1} + 0.1818 \cdot fuel\_price_{t-1} - 0.1168 \cdot temperature_{t-1} + 0.0020 \cdot week_{t-1} + 0.4253 \cdot cpi_{t-1} \\ \quad - 0.1340 \cdot sales_{t-2} - 0.1005 \cdot fuel\_price_{t-2} - 0.1535 \cdot temperature_{t-2} + 0.0028 \cdot week_{t-2} - 0.7972 \cdot cpi_{t-2} + \ldots + \epsilon_t $$

Fuel Price Equation:

$$ fuel\_price_t = -0.9297 + 0.4673 \cdot fuel\_price_{t-1} - 0.0971 \cdot sales_{t-1} - 0.4713 \cdot cpi_{t-1} - 0.5961 \cdot temperature_{t-2} \\ \quad + 0.2138 \cdot fuel\_price_{t-2} + \ldots + \epsilon_t $$

Temperature Equation:

$$\text{temperature}_t = -0.3104 + 0.3282 \cdot \text{temperature}_{t-1} + 0.6624 \cdot \text{cpi}_{t-1} - 0.5386 \cdot \text{fuel\_price}_{t-2} - 0.7805 \cdot \text{cpi}_{t-2} + \ldots + \epsilon_t$$

Week Equation:

$$\text{week}_t = 153.9 - 0.236 \cdot \text{week}_{t-1} - 0.3757 \cdot \text{week}_{t-2} - 43.67 \cdot \text{sales}_{t-4} - 0.6056 \cdot \text{week}_{t-3} + \ldots + \epsilon_t$$

CPI Equation:

$$\text{cpi}_t = -0.2888 + 0.6613 \cdot \text{cpi}_{t-1} + 0.0073 \cdot \text{trend} - 0.0313 \cdot \text{fuel\_price}_{t-4} - 0.1419 \cdot \text{temperature}_{t-4} + \ldots + \epsilon_t$$

Covariance Matrix of Residuals:

$$\begin{bmatrix} 0.0768 & -0.0054 & -0.0035 & -2.3978 & -0.0004 \\ -0.0054 & 0.0316 & 0.0029 & -0.6382 & -0.0010 \\ -0.0035 & 0.0029 & 0.0180 & 0.6042 & -0.0010 \\ -2.3978 & -0.6382 & 0.6042 & 500.7604 & 0.1007 \\ -0.0004 & -0.0010 & -0.0010 & 0.1007 & 0.0083 \end{bmatrix}$$

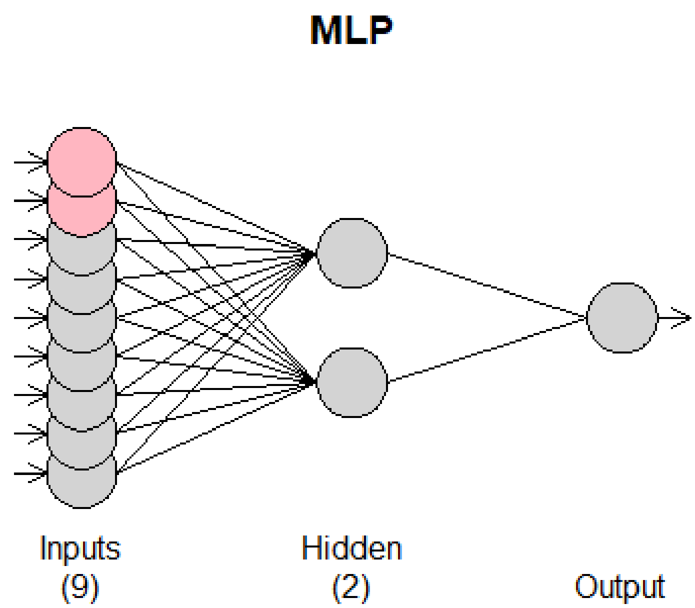
Correlation Matrix of Residuals:

$$\begin{bmatrix} 1.0000 & -0.1097 & -0.0933 & -0.3866 & -0.0153 \\ -0.1097 & 1.0000 & 0.1207 & -0.1604 & -0.0600 \\ -0.0933 & 0.1207 & 1.0000 & 0.2012 & -0.0814 \\ -0.3866 & -0.1604 & 0.2012 & 1.0000 & 0.0494 \\ -0.0153 & -0.0600 & -0.0814 & 0.0494 & 1.0000 \end{bmatrix}$$

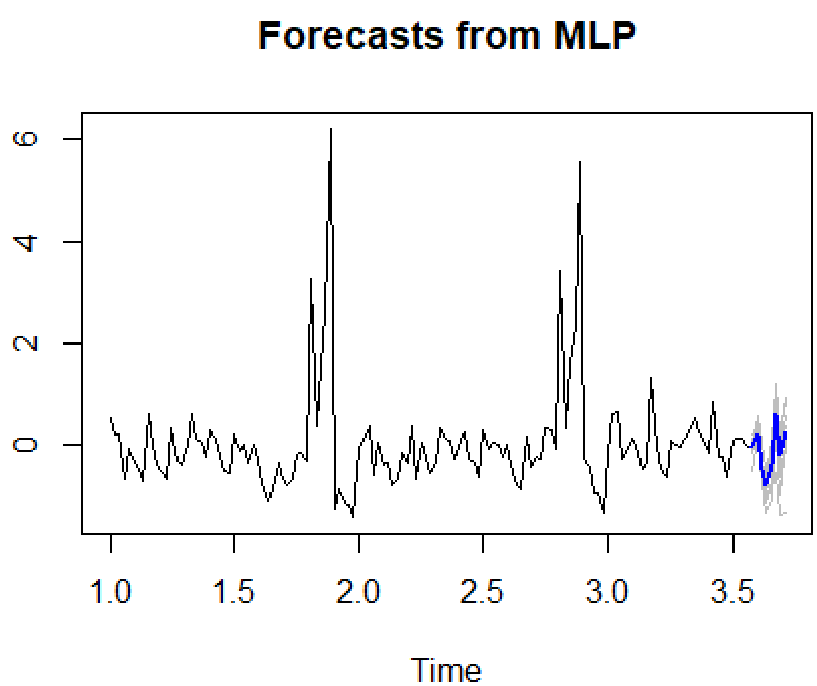
MLP Model

```
## MLP fit with 2 hidden nodes and 10 repetitions.
## Series modelled in differences: D1D12.
## Univariate lags: (11,15,40,47,48,51,52)
## Deterministic seasonal dummies included.
## Forecast combined using the median operator.
## MSE: 0.1938.
```

PLOT MLP Model



MLP Forecast



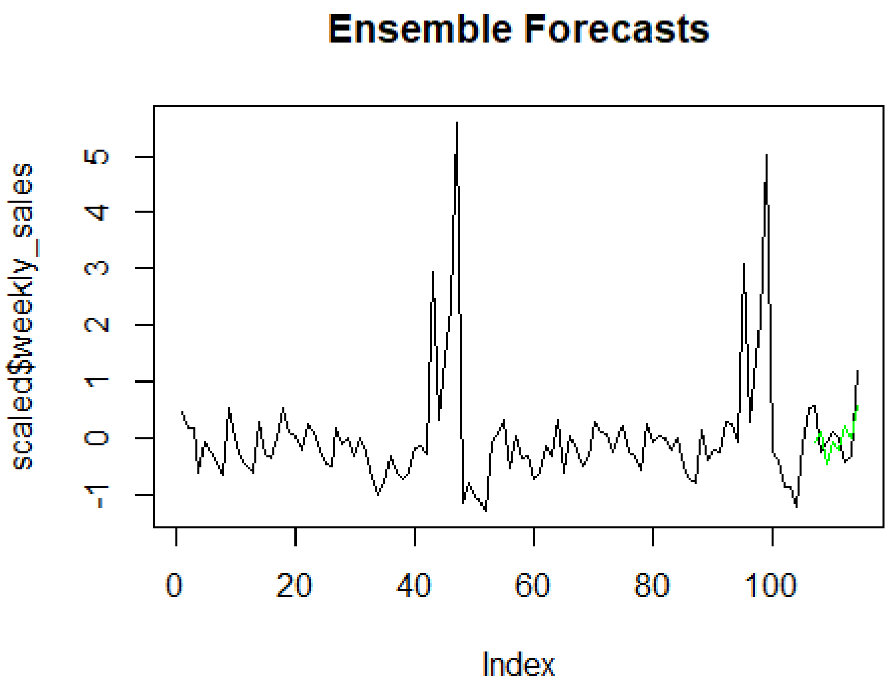
ASE

```
## [1] 0.1026821
```

WMAE | Kaggle

```
## [1] 0.2916268
```

Ensemble - MLP and VAR



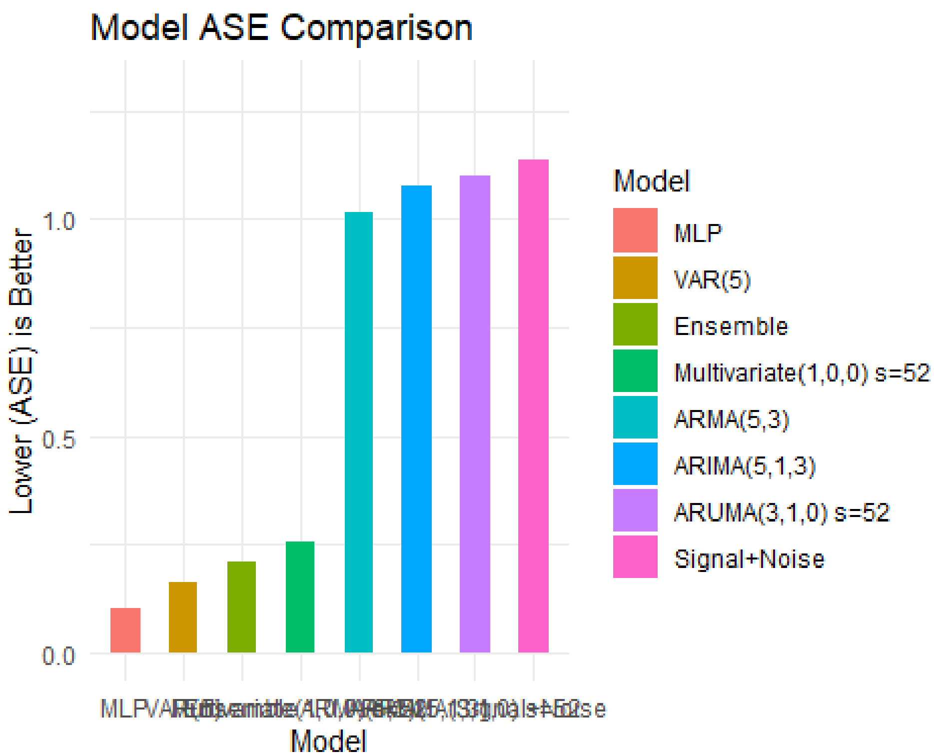
ASE Ensemble

```
## [1] 0.2080437
```

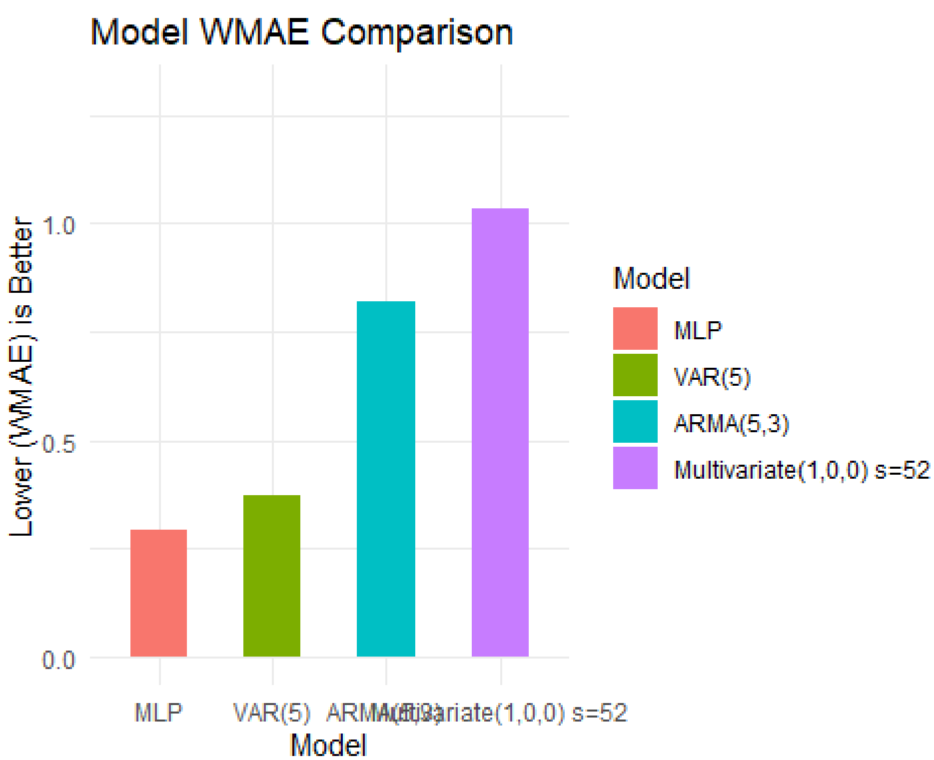
Model Decision - ASE

Model Decision - WMAE

Model Comparison Barchart



Model ASE Comparison
Lower (ASE) is Better
0.0
0.5
1.0
Model
Model
MLP
VAR(5)
Ensemble
Multivariate(1,0,0) s=52
ARMA(5,3)
ARIMA(5,1,3)
ARUMA(3,1,0) s=52
Signal+Noise



Model WMAE Comparison
Lower (WMAE) is Better
0.0
0.5
1.0
MLP
VAR(5)
Model
Model
MLP
VAR(5)
ARMA(5,3)
Multivariate(1,0,0) s=52