Project 2

Alvin Ng, Austin Lee, Yi Ran Wang, Ruixi Zhou

Introduction

This report looks at the relation between monthly attendence at a non-profit children museum and monthly Irvine unemployment data. Both datasets contain data from Sept 2009 to Dec 2018. The monthly unemployment data was obtained from State of California Employment Development Department and The monthly attendence data was provided by the non-profit with adjustments by Alvin Ng. We will attempt to forecast the monthly attendence using information from the Irvine unemployment data.

We are using the unemployment data because if unemployment increase, we should see a decrease in spending thus indirectly affect the monthly attendence.

Results

a.) Time Series Plot

```r
library(readxl)
```

```
## Warning: package 'readxl' was built under R version 3.5.2
```

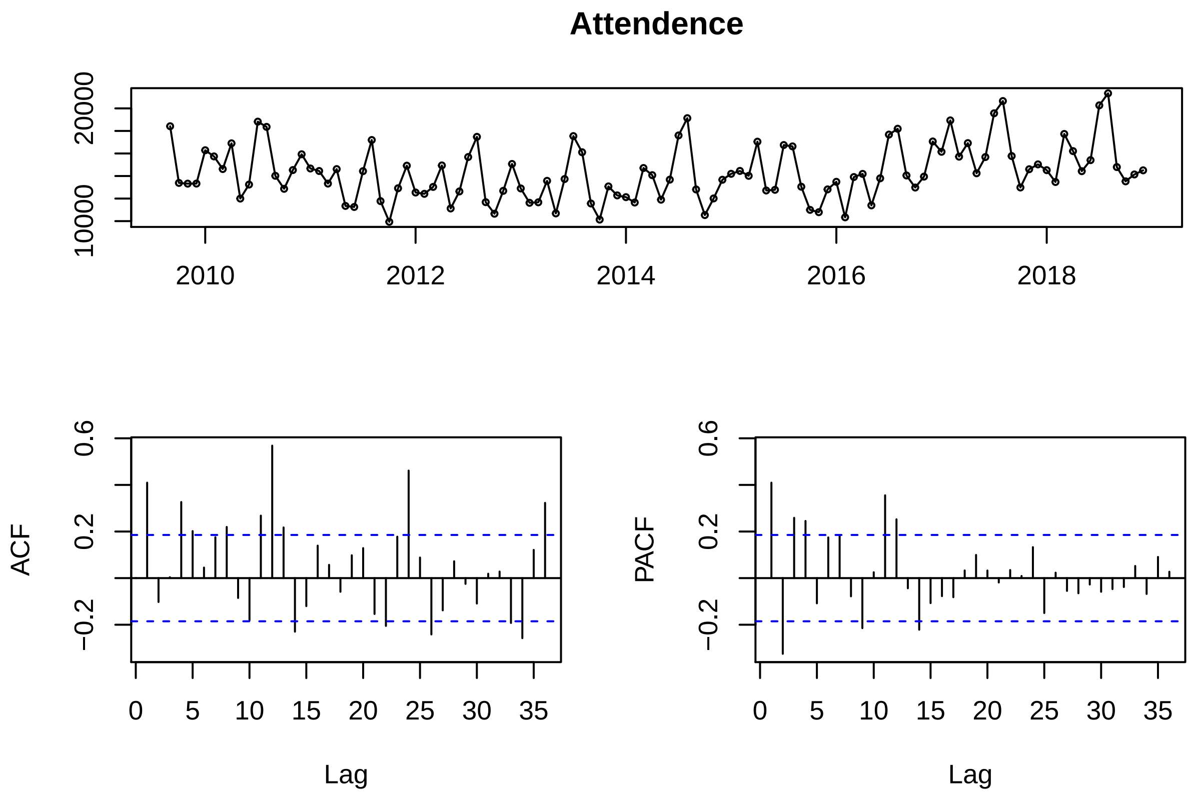
```r
library(tseries)
```

```
## Warning: package 'tseries' was built under R version 3.5.2
```

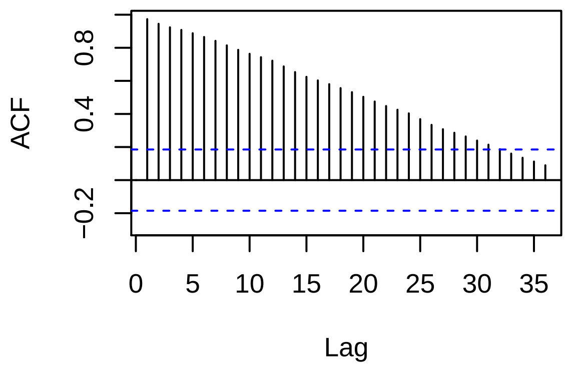
```r
library(forecast)
```

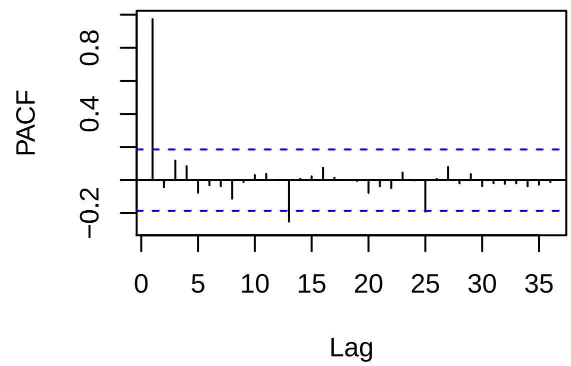
```
## Warning: package 'forecast' was built under R version 3.5.2
```

```r
Attendence_data <- read_excel("PretendCity Daily Attendence.xlsx",
    sheet = "Monthly")
unemployment_data <- read_excel("Local-Area-Unemployment-Statistics.xlsx",
    sheet = "Data")

Attendence <- ts(Attendence_data$Attendence, start = c(2009, 9), freq = 12)
Unemployment <- ts(unemployment_data$Unemployment, start = c(2009, 9), freq = 12)

tsdisplay(Attendence)
```




```r
tsdisplay(Unemployment)
```





b.) Trend, Seasonality, Cyclical

Quaduratic Trend with seasonal but no ARIMA

```
t <- seq(2009+(8/12), 2018.96,length=length(Attendence))
t2 <- t^2
Attendence_half_model <- tslm(Attendence~ t + I(t^2) + season)
summary(Attendence_half_model)
```

```
##
## Call:
## tslm(formula = Attendence ~ t + I(t^2) + season)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3190.6  -824.9  -122.4   862.5  4580.2
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  5.706e+08  7.663e+07   7.447 3.77e-11 ***
## t           -5.667e+05  7.608e+04  -7.449 3.73e-11 ***
## I(t^2)       1.407e+02  1.889e+01   7.451 3.69e-11 ***
## season2     -4.307e+02  6.133e+02  -0.702  0.48418
## season3      1.889e+02  6.133e+02   0.308  0.75876
## season4      1.257e+03  6.134e+02   2.049  0.04315 *
## season5     -1.937e+03  6.135e+02  -3.158  0.00211 **
```

```
## season6     -5.810e+02  6.136e+02  -0.947  0.34599
## season7      3.415e+03  6.137e+02   5.565 2.28e-07 ***
## season8      4.143e+03  6.139e+02   6.750 1.05e-09 ***
## season9     -5.473e+02  5.985e+02  -0.914  0.36274
## season10    -2.552e+03  5.985e+02  -4.264 4.62e-05 ***
## season11    -1.166e+03  5.986e+02  -1.948  0.05430 .
## season12     7.673e+01  5.987e+02   0.128  0.89829
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1301 on 98 degrees of freedom
## Multiple R-squared:  0.7507, Adjusted R-squared:  0.7176
## F-statistic:  22.7 on 13 and 98 DF,  p-value: < 2.2e-16
```

```
AIC(Attendence_half_model)
```

```
## [1] 1939.156
```

```
BIC(Attendence_half_model)
```

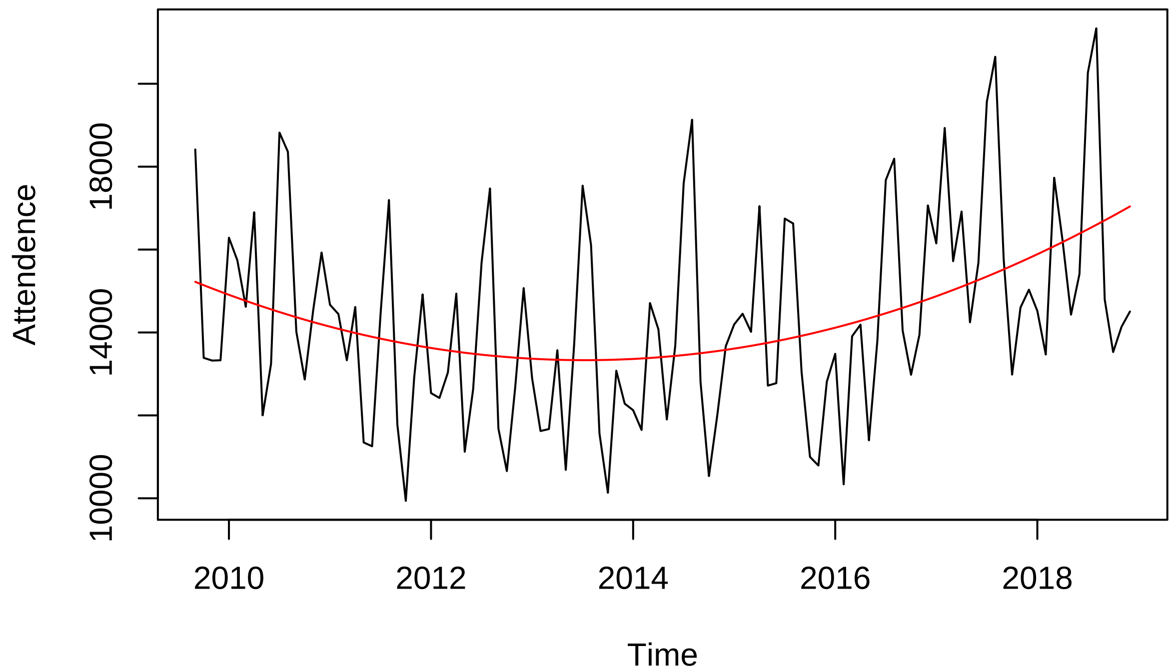
```
## [1] 1979.933
```

Quaduratic Trend with ARIMA seasonal

```
Attendence_model <- tslm(Attendence~ t + t2)
summary(Attendence_model)
```

```
##
## Call:
## tslm(formula = Attendence ~ t + t2)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -3832  -1679   -537   1338   5657
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  5.113e+08  1.323e+08   3.864 0.000190 ***
## t           -5.078e+05  1.314e+05  -3.866 0.000189 ***
## t2           1.261e+02  3.261e+01   3.867 0.000188 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2261 on 109 degrees of freedom
## Multiple R-squared:  0.1622, Adjusted R-squared:  0.1468
## F-statistic: 10.55 on 2 and 109 DF,  p-value: 6.485e-05
```

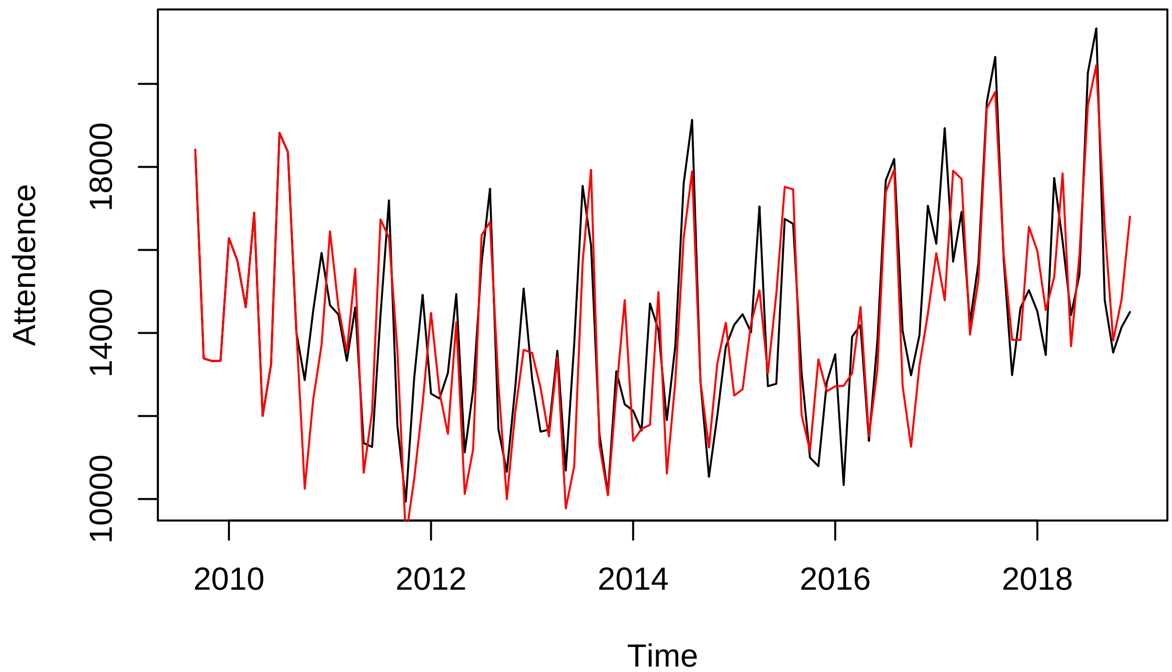
```
plot(Attendence)
lines(Attendence_model$fitted.values, col="red")
```




```r
Attendence_full_model <- auto.arima(Attendence)
summary(Attendence_full_model)
```

```
## Series: Attendence
## ARIMA(1,1,1)(2,1,0)[12]
##
## Coefficients:
##          ar1      ma1     sar1     sar2
##       0.3087  -0.8252  -0.7451  -0.2053
## s.e.  0.1488   0.0927   0.1134   0.1227
##
## sigma^2 estimated as 2061027:  log likelihood=-861.9
## AIC=1733.81   AICc=1734.45   BIC=1746.78
##
## Training set error measures:
##                    ME     RMSE      MAE       MPE     MAPE      MASE
## Training set 168.2067 1322.192 1001.043 0.8683353 7.121259 0.6743661
##                      ACF1
## Training set -0.003525128
```

```r
plot(Attendence)
lines(Attendence_full_model$fitted, col="red")
```



Since modeling without ARIMA gave us a higher AIC and BIC, we selected the model with trend capture by the AR process - AR(3), with S-AR(1) 12 period, drift enabled, and MA(1) to capture the short term effect. The fitted values from the model fit the actual values well with exception from time period before 2011.

Unemployment data

Linear trend with seasonal

```r
Unemployment_half_model <- tslm(Unemployment~trend + season)
summary(Unemployment_half_model)
```

```
##
## Call:
## tslm(formula = Unemployment ~ trend + season)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -11920  -5464   1439   4828  15671
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  38697.97    2407.53  16.074  < 2e-16 ***
## trend         1102.97      19.17  57.530  < 2e-16 ***
## season2       2574.80    3086.71   0.834 0.406200
## season3      -2194.84    3086.89  -0.711 0.478744
## season4     -11820.03    3087.19  -3.829 0.000226 ***
```

```
## season5     -11711.89    3087.60  -3.793 0.000256 ***
## season6      -5781.53    3088.14  -1.872 0.064134 .
## season7      -5806.73    3088.79  -1.880 0.063057 .
## season8      -4876.37    3089.57  -1.578 0.117679
## season9      -9571.50    3008.74  -3.181 0.001959 **
## season10     -6474.48    3009.04  -2.152 0.033855 *
## season11     -4877.45    3009.47  -1.621 0.108263
## season12     -6360.42    3010.02  -2.113 0.037108 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6548 on 99 degrees of freedom
## Multiple R-squared:  0.9713, Adjusted R-squared:  0.9678
## F-statistic:   279 on 12 and 99 DF,  p-value: < 2.2e-16
```

```r
AIC(Unemployment_half_model)
```

```
## [1] 2300.285
```

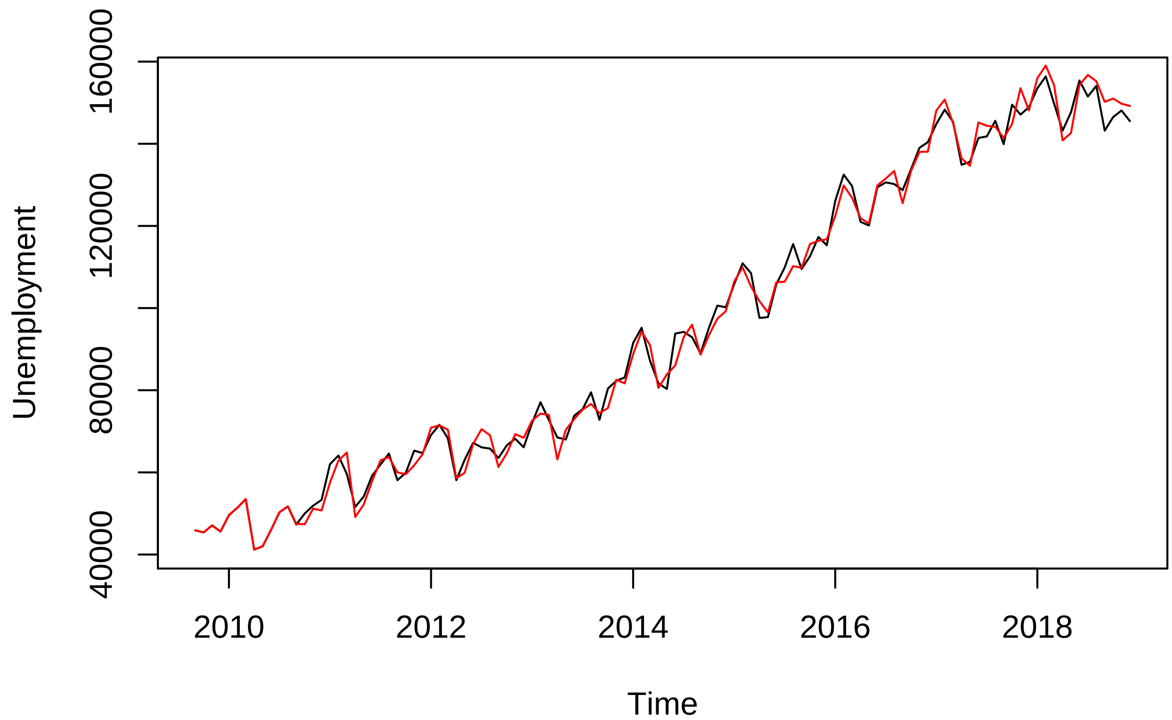
```r
BIC(Unemployment_half_model)
```

```
## [1] 2338.344
```

```r
Unemployment_full_model <- auto.arima(Unemployment)
summary(Unemployment_full_model)
```

```
## Series: Unemployment
## ARIMA(0,1,1)(2,1,0)[12]
##
## Coefficients:
##           ma1     sar1     sar2
##       -0.2168  -0.6309  -0.2369
## s.e.   0.1058   0.1120   0.1165
##
## sigma^2 estimated as 8231278:  log likelihood=-929.69
## AIC=1867.37   AICc=1867.8   BIC=1877.75
##
## Training set error measures:
##                    ME     RMSE      MAE       MPE     MAPE      MASE
## Training set 32.16916 2656.197 2028.982 0.1990982 2.171159 0.1639051
##                     ACF1
## Training set 0.005383186
```

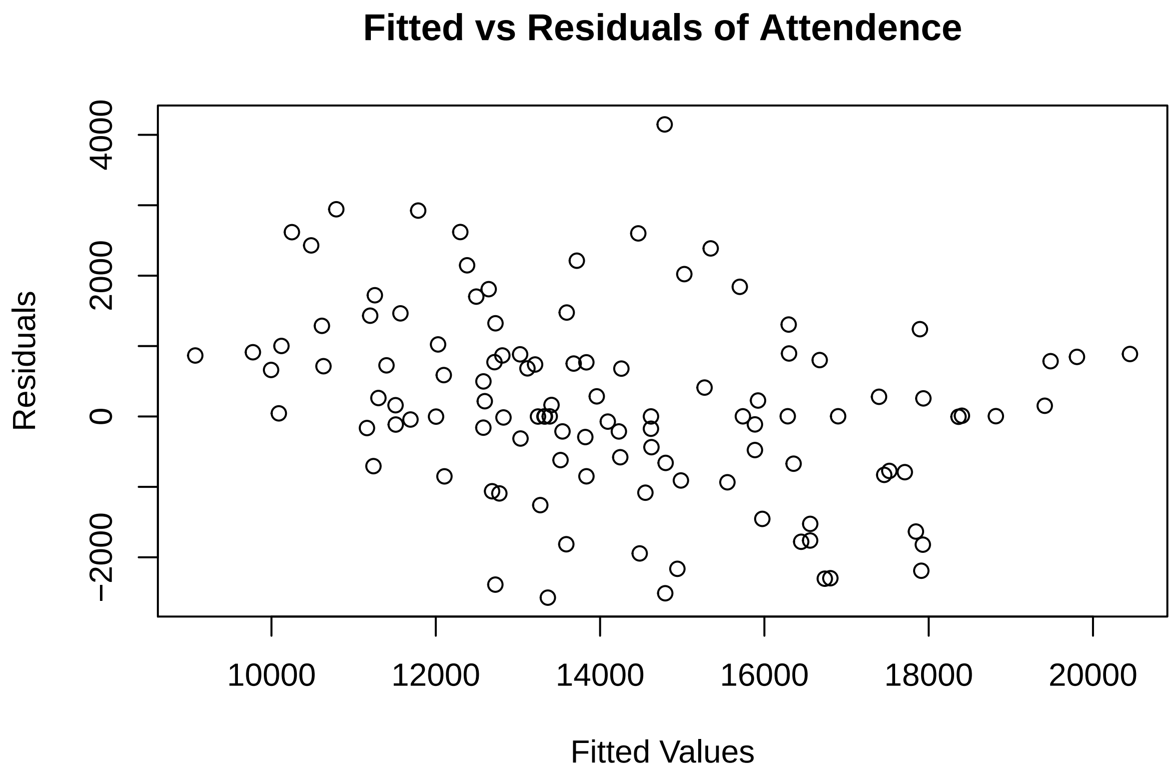
```r
plot(Unemployment)
lines(Unemployment_full_model$fitted, col="red")
```



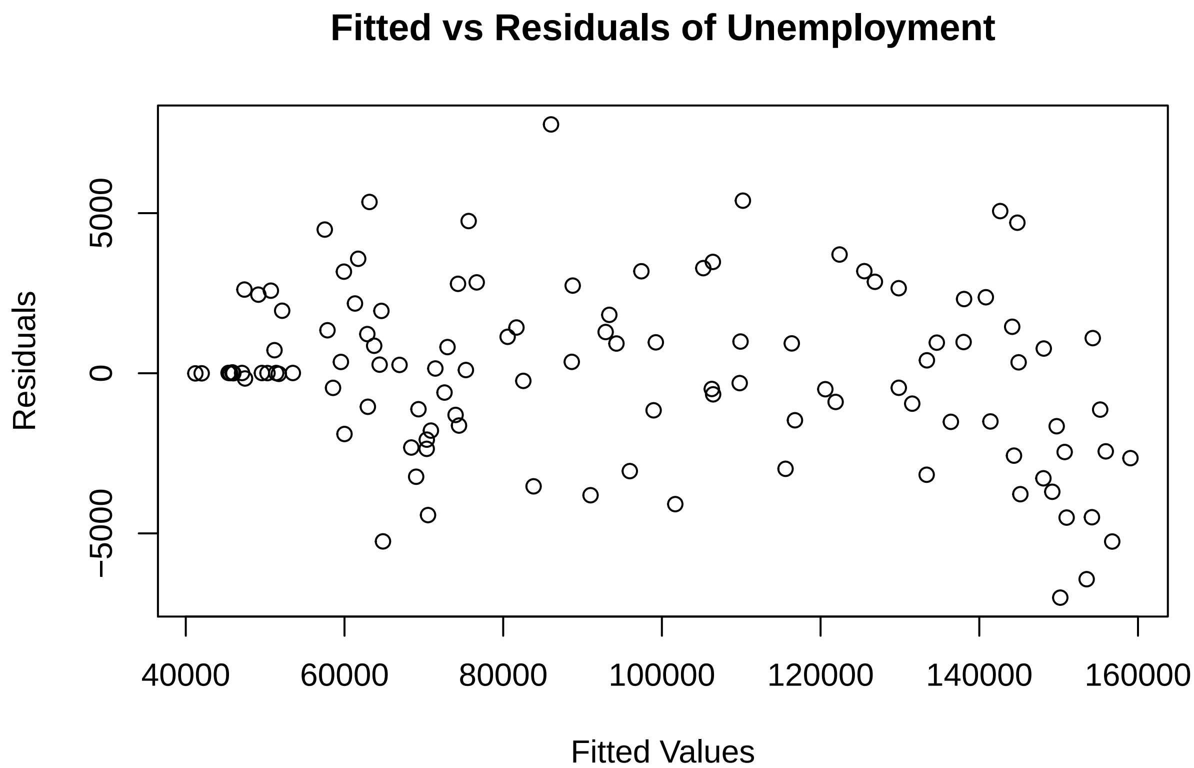
For the unemployment data, our model contain a linear trend, an AR(1) process with drift and an seasonal-AR(2) process with a 12 month period. The fitted values from the model seems to be fitting the actual value nicely.

c.) Residual vs fitted values

```
plot(as.vector(Attendence_full_model$fitted),
     as.vector(Attendence_full_model$residuals),
     ylab="Residuals",
     xlab="Fitted Values", main = "Fitted vs Residuals of Attendence")
```



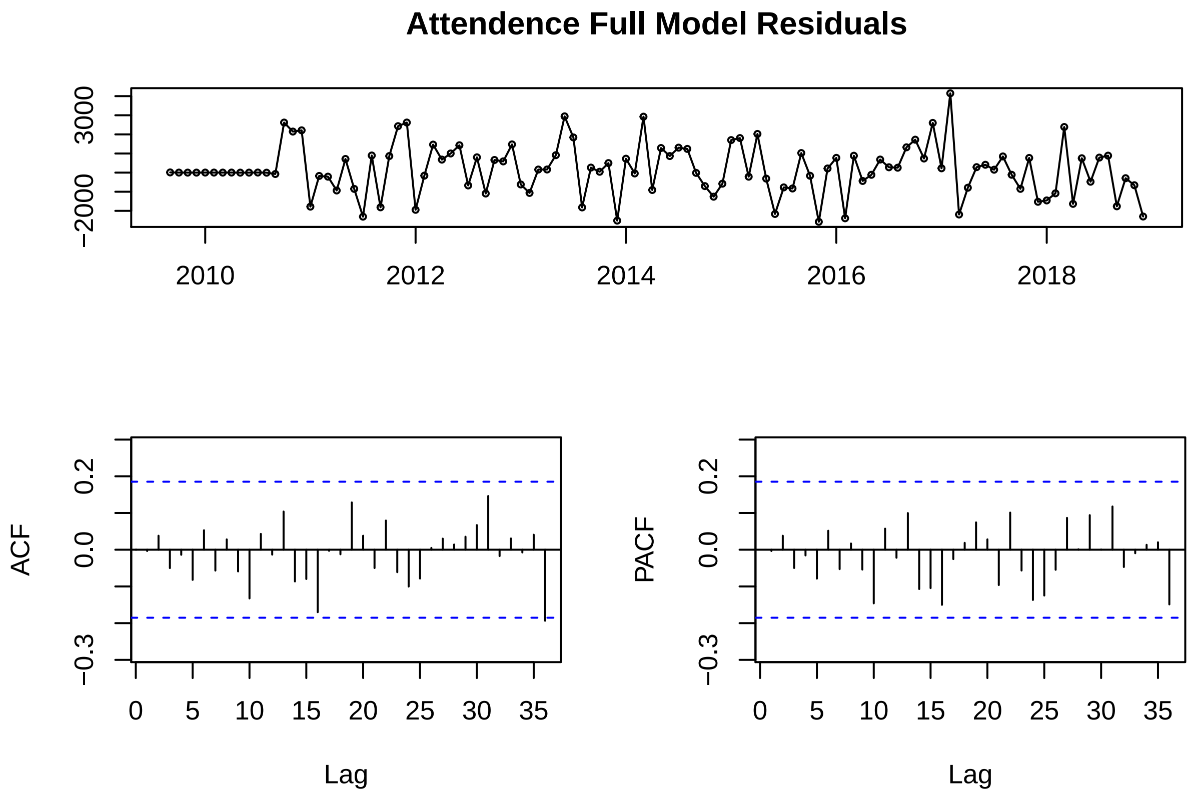

```
plot(as.vector(Unemployment_full_model$fitted),
     as.vector(Unemployment_full_model$residuals),
     ylab="Residuals",
     xlab="Fitted Values", main = "Fitted vs Residuals of Unemployment")
```



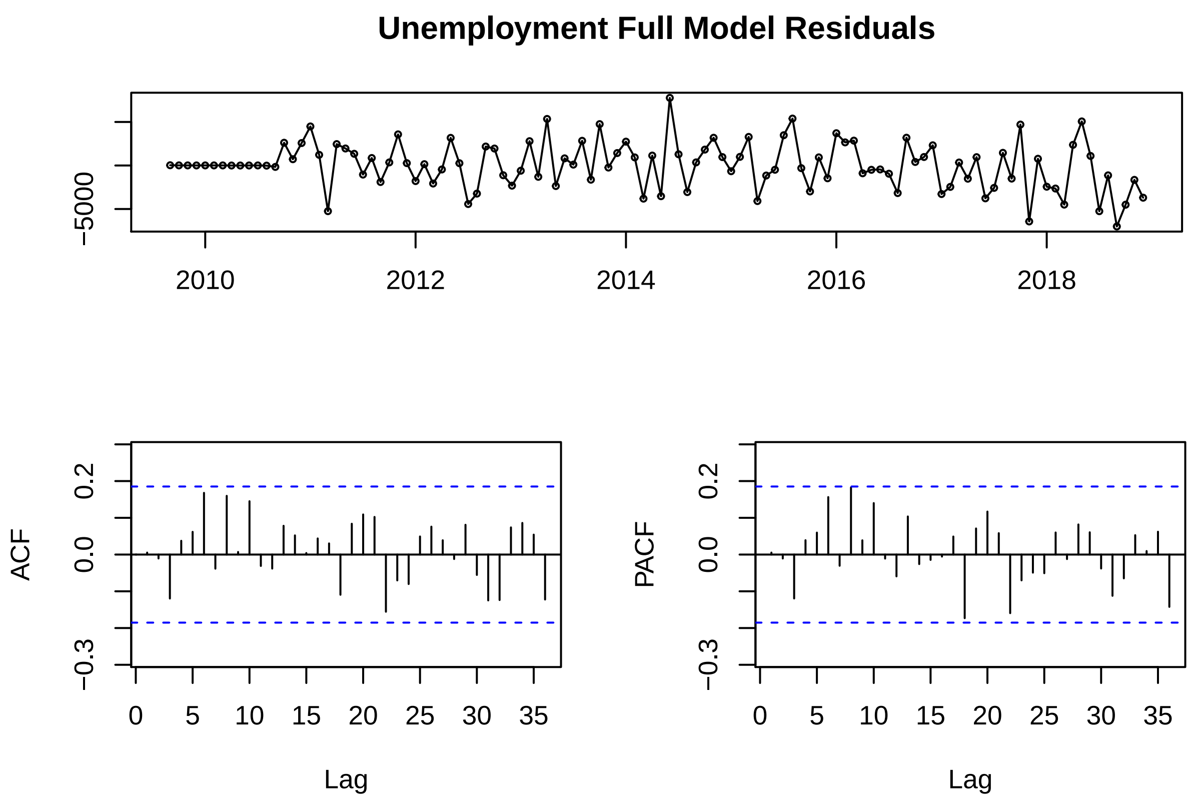
Both fitted vs residuals plots seem to be evenly spread horizontally around zero, implying no bias is observed.

e.) ACF and PACF of Residuals

```r
# Computing ACF and PACF of the residuals of Daily Attendance
tsdisplay(Attendence_full_model$residuals, main = "Attendence Full Model Residuals")
```




```r
# Computing ACF and PACF of the residuals of Unemployment
tsdisplay(Unemployment_full_model$residuals, main = "Unemployment Full Model Residuals")
```



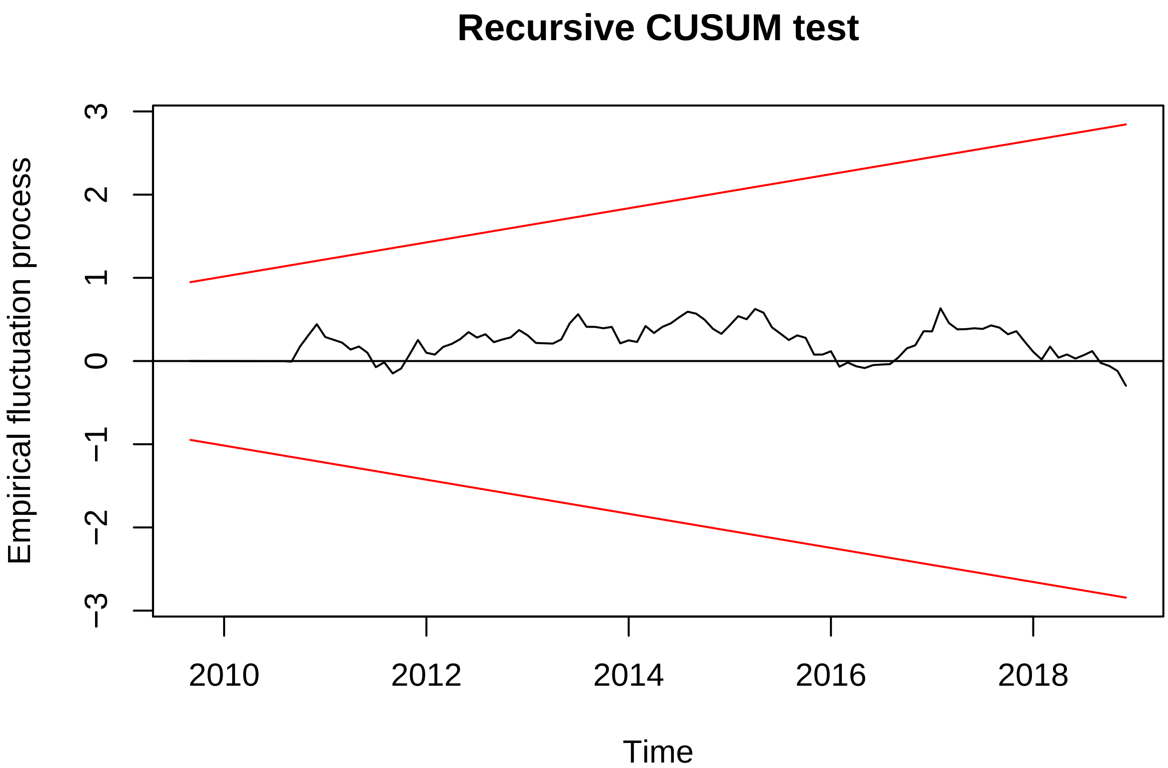
For both models, the PACF and ACF indicate a white noise process because none of the lags spike past the area of significance.

f.) Plot CUSUM

```r
library("strucchange")
```

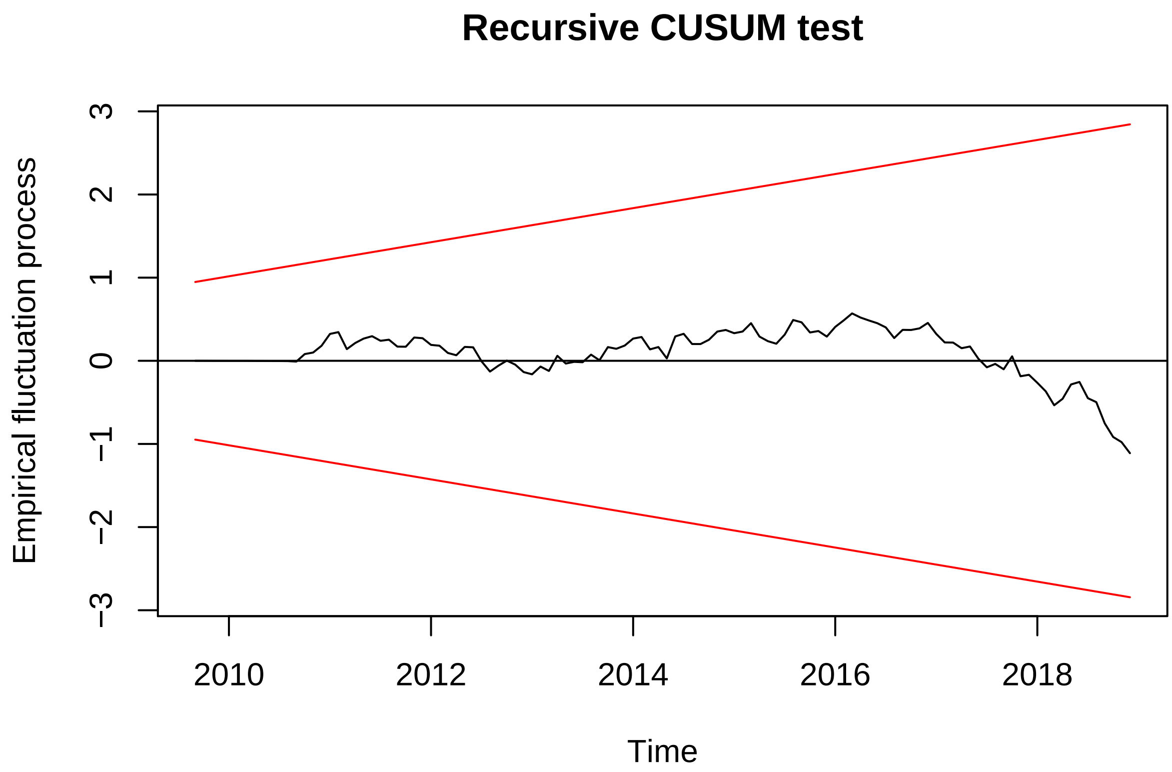
```
## Warning: package 'strucchange' was built under R version 3.5.2

## Loading required package: zoo

## Warning: package 'zoo' was built under R version 3.5.2

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric

## Loading required package: sandwich

## Warning: package 'sandwich' was built under R version 3.5.2
```

```r
plot(efp(Attendence_full_model$residuals ~ 1, type = "Rec-CUSUM"))
```



There are no structural breaks within our attendence model since it does not move above or below the upper and lower bound.

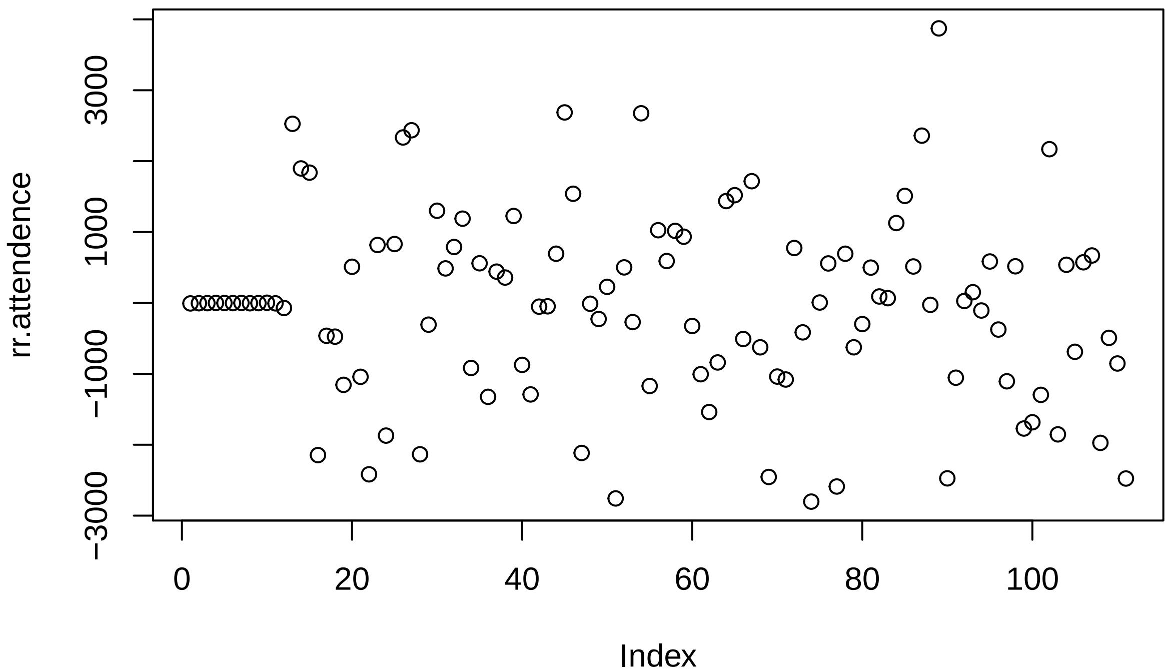
```
plot(efp(Unemployment_full_model$residuals ~ 1, type = "Rec-CUSUM"))
```



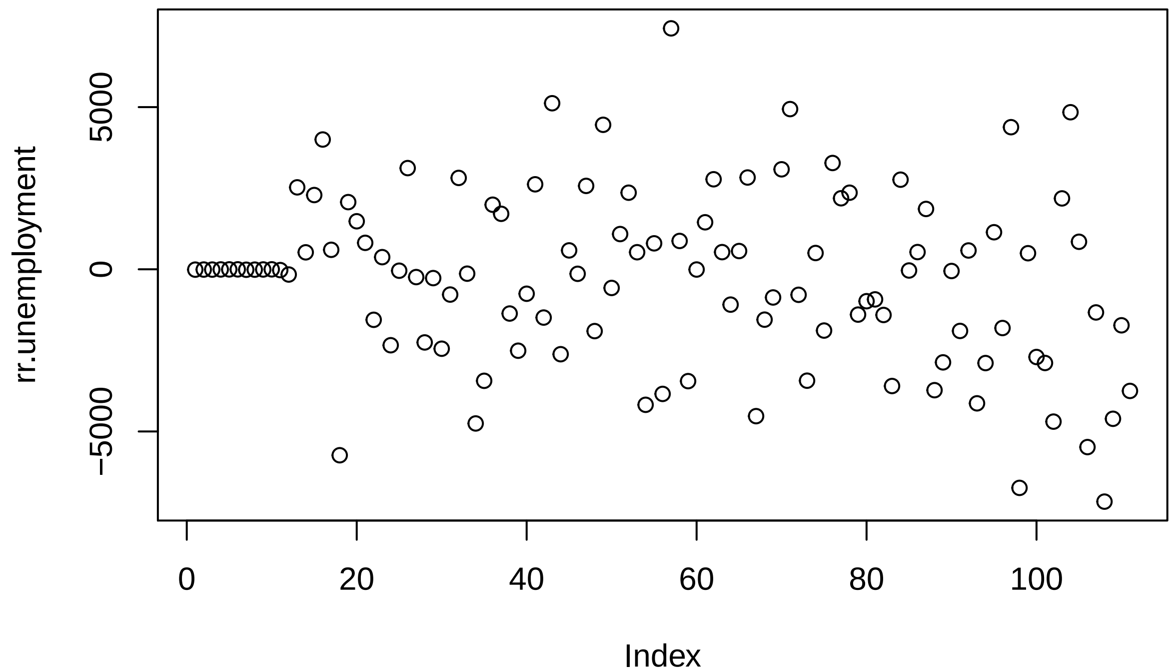
When looking at the recursive cumsum for the unemployment model, there are no structural breaks within our model.

g.) Recursive Residuals

```
library("strucchange")
rr.attendence <- recresid(Attendence_full_model$residuals ~ 1 )
plot(rr.attendence)
```




```r
rr.unemployment <- recresid(Unemployment_full_model$residuals ~ 1 )
plot(rr.unemployment)
```



By looking at the recursive residuals, we can investigate the fit of our model. The plot suggests that our residuals are independently and identically distributed about 0.

Diagnostic Statistics

```r
library('forecast')
print("Attendence Model Diagnostics:")
```

```
## [1] "Attendence Model Diagnostics:"
```

```r
accuracy(Attendence_full_model)
```

```
##                    ME     RMSE      MAE       MPE     MAPE      MASE
## Training set 168.2067 1322.192 1001.043 0.8683353 7.121259 0.6743661
##                      ACF1
## Training set -0.003525128
```

```r
print("Unemployment Model Diagnostics:")
```

```
## [1] "Unemployment Model Diagnostics:"
```

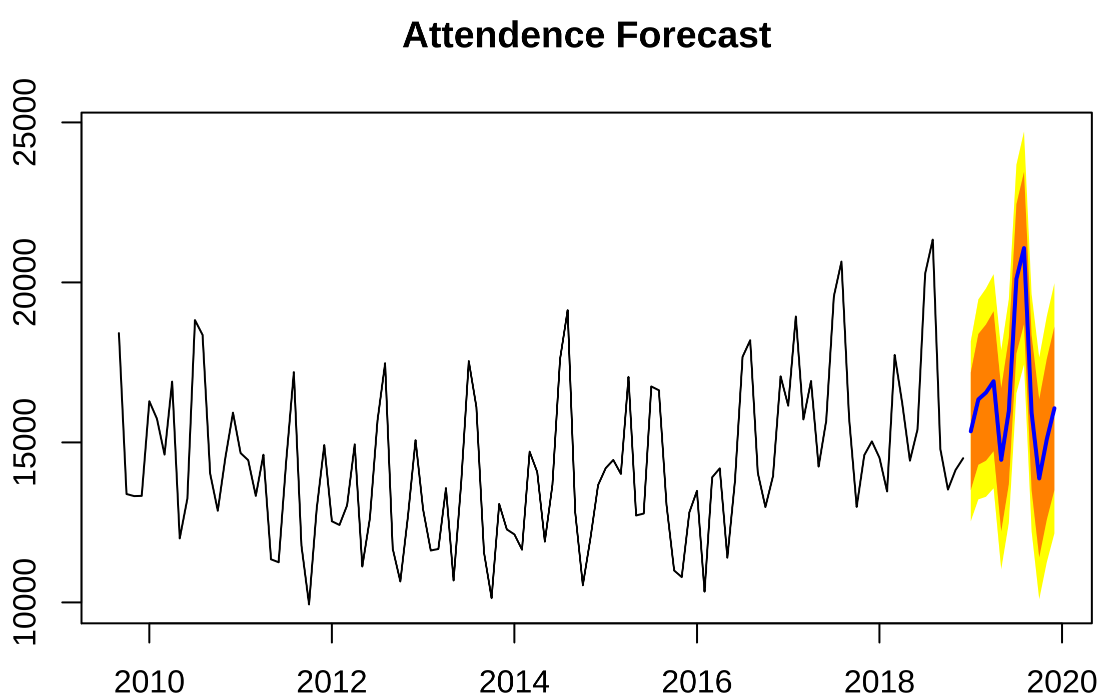
```r
accuracy(Unemployment_full_model)
```

```
##                    ME     RMSE      MAE       MPE     MAPE      MASE
## Training set 32.16916 2656.197 2028.982 0.1990982 2.171159 0.1639051
##                     ACF1
## Training set 0.005383186
```

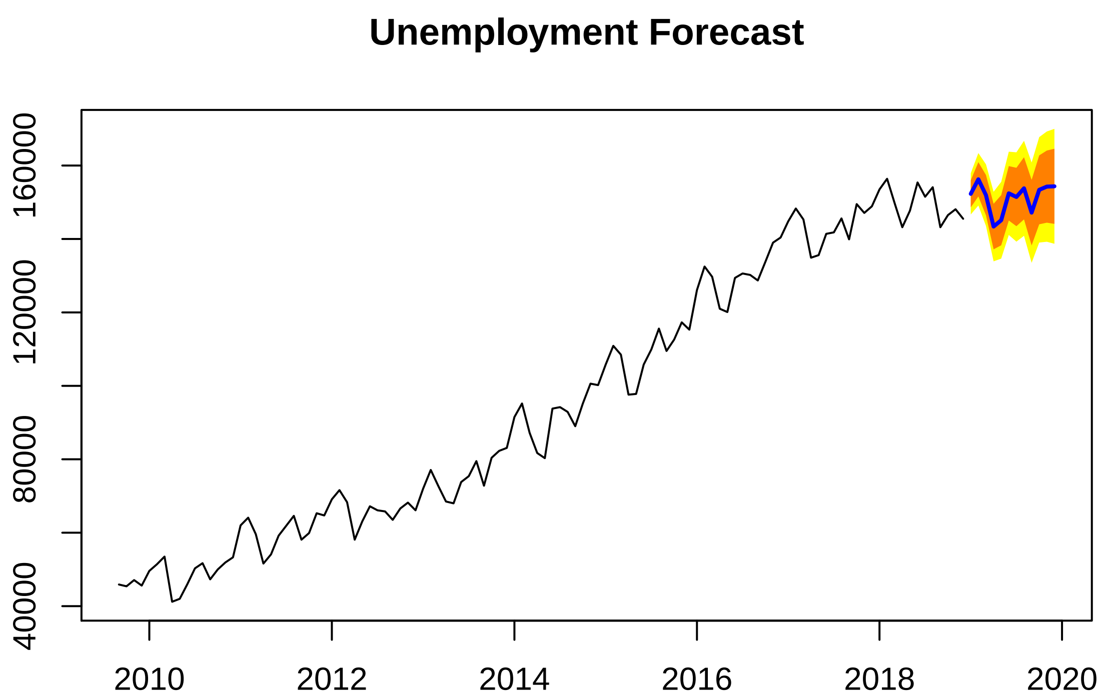
When looking at the Attendence model's diagnostics, we can see that we have a ME(Mean Error) of 168.20,RMSE(Root Mean Squared Error) of 1322.19 ,MAE(Mean Absolute Error) 1001.043,MPE(Mean Percentage Error) 7.121, MAPE(Mean Absolute Percentage Error) 7.121, MASE(Mean Absolute Squared Error)of .674, and ACF1 of -.00352. The MAPE is at 7.12% which is considerably high, meaning the absolute percentage difference between actual data and fitted is about 7%. We can see that this would also mean that the MPE is also high. However, as previously mentioned, we have eliminated ACF and PACF to white noise, as hinted with the ACF1 at an insignificant value.

When looking at the Unemployment model's diagnostics, we can see that we have a ME(Mean Error) of 32.16,RMSE(Root Mean Squared Error) of 2656.19 ,MAE(Mean Absolute Error) 2028.98, MPE(Mean Percentage Error) .199, MAPE(Mean Absolute Percentage Error) 2.1711, MASE(Mean Absolute Squared Error)of .1639051, and ACF1 of .0053831. The MAPE is considerably low, meaning the absolute percentage difference between the actual and fitted data is at 2.17%. The ACF and PACF of the model's residuals have been reduced to white noise as well. When looking at these metrics, its important to compare them to other models, however, which we will compare to our VAR model.

```
plot(forecast(Attendence_full_model, h = 12),
     shadecol = 'oldstyle', main ="Attendence Forecast")
```




```
plot(forecast(Unemployment_full_model, h= 12),
     shadecol = 'oldstyle', main = "Unemployment Forecast")
```



Because our ARIMA attendence model is (1,1,1)(2,1,0), the forecast takes weights recent lags to forecast future events. The model seems to be capturing the short term dynamics within the 12 month period. Our forecast is situated at a 95% confidence interval.

Our ARIMA model for unemployment is (0,1,1)(2,1,0). It looks like our forecast for unemployment is looking at the recent trend as opposed to looking at the entire data set. Otherwise, the trend might have increased over the 12 month period. Our forecast is situated at a 95% confidence interval.

Part I: Var Model

```
library('tseries')
library('vars')
```

```
## Warning: package 'vars' was built under R version 3.5.2
## Loading required package: MASS
## Loading required package: urca
## Warning: package 'urca' was built under R version 3.5.2
## Loading required package: lmtest
## Warning: package 'lmtest' was built under R version 3.5.2
```

```
combined_data <- (na.remove(cbind(Attendence, Unemployment)))
tot_combo <- data.frame(combined_data)
VARselect(tot_combo)
```

```
## $selection
```

```
## AIC(n)  HQ(n)  SC(n) FPE(n)
##      9      8      6      9
##
## $criteria
##                   1            2            3            4            5
## AIC(n) 3.244261e+01 3.221602e+01 3.209901e+01 3.191084e+01 3.147968e+01
## HQ(n)  3.250514e+01 3.232023e+01 3.224491e+01 3.209842e+01 3.170894e+01
## SC(n)  3.259702e+01 3.247337e+01 3.245930e+01 3.237407e+01 3.204585e+01
## FPE(n) 1.229314e+14 9.801801e+13 8.721925e+13 7.229412e+13 4.700900e+13
##                   6            7            8            9           10
## AIC(n) 3.127442e+01 3.129110e+01 3.108435e+01 3.107571e+01 3.108931e+01
## HQ(n)  3.154537e+01 3.160373e+01 3.143867e+01 3.147171e+01 3.152699e+01
## SC(n)  3.194353e+01 3.206315e+01 3.195934e+01 3.205364e+01 3.217018e+01
## FPE(n) 3.832823e+13 3.903187e+13 3.180455e+13 3.161057e+13 3.214413e+13
```

For our model, we picked order 6 because AIC tends to overparameterize when compared to BIC.

```r
var_model<-VAR(tot_combo, p = 6)
summary(var_model)
```

```
##
## VAR Estimation Results:
## =========================
## Endogenous variables: Attendence, Unemployment
## Deterministic variables: const
## Sample size: 106
## Log Likelihood: -1935.727
## Roots of the characteristic polynomial:
## 1.003 0.9618 0.9618 0.9455 0.9455 0.8533 0.8533 0.8289 0.8289 0.8136 0.8136 0.2864
## Call:
## VAR(y = tot_combo, p = 6)
##
##
## Estimation results for equation Attendence:
## ===========================================
## Attendence = Attendence.l1 + Unemployment.l1 + Attendence.l2 + Unemployment.l2 + Attendence.l3 + Uner
##
##                   Estimate Std. Error t value Pr(>|t|)
## Attendence.l1    5.168e-01  1.055e-01   4.899 4.06e-06 ***
## Unemployment.l1  5.343e-02  3.571e-02   1.496 0.137953
## Attendence.l2   -7.464e-02  1.002e-01  -0.745 0.458292
## Unemployment.l2 -5.350e-03  4.430e-02  -0.121 0.904150
## Attendence.l3    3.412e-03  9.461e-02   0.036 0.971312
## Unemployment.l3 -1.783e-01  4.377e-02  -4.072 9.77e-05 ***
## Attendence.l4    1.390e-01  9.311e-02   1.493 0.138717
## Unemployment.l4 -5.920e-02  4.853e-02  -1.220 0.225615
## Attendence.l5   -1.667e-01  9.273e-02  -1.798 0.075408 .
## Unemployment.l5  2.710e-01  4.934e-02   5.493 3.42e-07 ***
## Attendence.l6    3.836e-02  8.156e-02   0.470 0.639268
## Unemployment.l6 -6.832e-02  4.689e-02  -1.457 0.148517
## const            6.660e+03  1.787e+03   3.726 0.000333 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
```

```
## Residual standard error: 1484 on 93 degrees of freedom
## Multiple R-Squared: 0.6798,  Adjusted R-squared: 0.6385
## F-statistic: 16.45 on 12 and 93 DF,  p-value: < 2.2e-16
##
##
## Estimation results for equation Unemployment:
## =============================================
## Unemployment = Attendence.l1 + Unemployment.l1 + Attendence.l2 + Unemployment.l2 + Attendence.l3 + Un
##
##                   Estimate Std. Error t value Pr(>|t|)
## Attendence.l1     -0.67788    0.28000  -2.421   0.0174 *
## Unemployment.l1    0.64798    0.09477   6.838 8.33e-10 ***
## Attendence.l2     -0.22914    0.26600  -0.861   0.3912
## Unemployment.l2   -0.03100    0.11759  -0.264   0.7926
## Attendence.l3      0.04341    0.25111   0.173   0.8631
## Unemployment.l3    0.04049    0.11618   0.348   0.7283
## Attendence.l4     -0.29162    0.24711  -1.180   0.2410
## Unemployment.l4    0.06712    0.12880   0.521   0.6035
## Attendence.l5     -0.04353    0.24611  -0.177   0.8600
## Unemployment.l5    0.16846    0.13096   1.286   0.2015
## Attendence.l6      1.08577    0.21647   5.016 2.52e-06 ***
## Unemployment.l6    0.11673    0.12446   0.938   0.3507
## const           3089.17284 4743.58552   0.651   0.5165
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Residual standard error: 3939 on 93 degrees of freedom
## Multiple R-Squared: 0.9892,  Adjusted R-squared: 0.9878
## F-statistic: 707.8 on 12 and 93 DF,  p-value: < 2.2e-16
##
##
##
## Covariance matrix of residuals:
##              Attendence Unemployment
## Attendence      2202283     -1329621
## Unemployment   -1329621     15513103
##
## Correlation matrix of residuals:
##              Attendence Unemployment
## Attendence       1.0000      -0.2275
## Unemployment    -0.2275       1.0000
```

When looking at the diagnostics, for predicting attendence, we have an adjusted R^2 of .6385. Most of the parameters are statistically insignficant except for prediction variables of attendence at lag 1, unemployment at lag 3, and unemployment at lag 5.

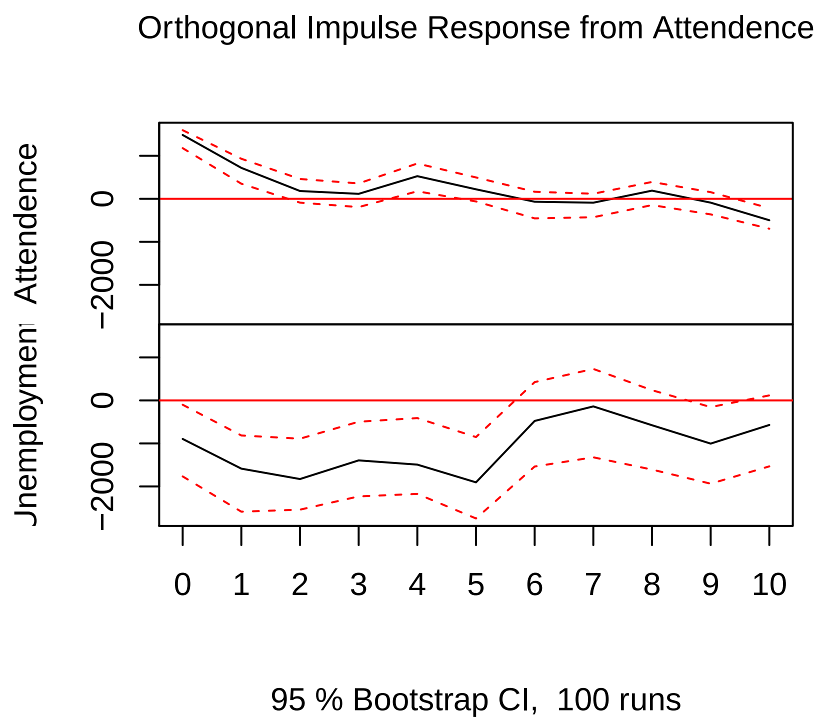
When looking at the diagnostics for predicting unemployment, we have a very high adjusted R^2 of .9878. Three of the parameters, attendence at lag 1, unemployment at lag 1, and attendence at lag 1 , are statistically signficant.

j.) IRF

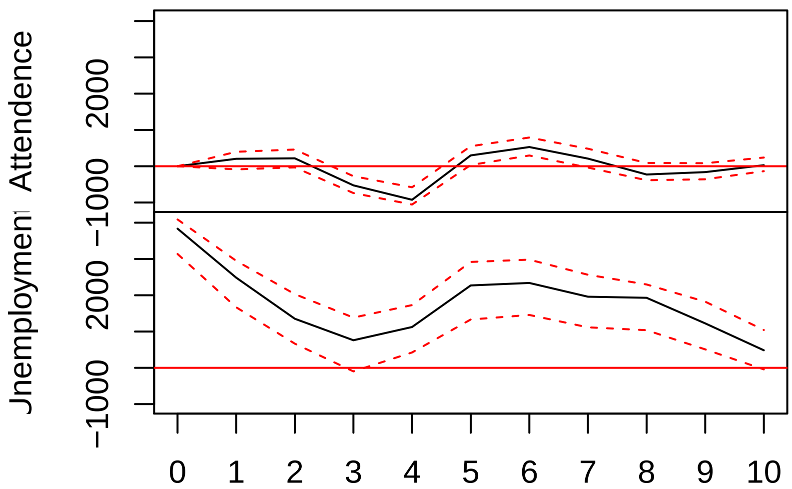
The variable impulse response for attendence shows an initial spike then a quick decay. Cross variable impulse response that shows the effect of attendence's shock on umemployment is initially a negative spike and stays negative through out.

Cross variable impulse response shows the effect of unemployment's shock on attendence is small and next to no effect Own variable impulse response for unemployment shows an initial spike then decay until lag 3 then spike and peak at around lag 6 then slowly decays.

```r
plot(irf(var_model))
```



Orthogonal Impulse Response from Unemployment



95 % Bootstrap CI, 100 runs

k. Granger Test

Granger causality test shows that they both reject the null hypothesis and suggests causality between each other.Therefore the test is inconclusive.

```
grangertest(Attendence ~ Unemployment, order = 6)
```

```
## Granger causality test
##
## Model 1: Attendence ~ Lags(Attendence, 1:6) + Lags(Unemployment, 1:6)
## Model 2: Attendence ~ Lags(Attendence, 1:6)
##   Res.Df Df      F    Pr(>F)
## 1     93
## 2     99 -6 13.373 7.101e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
grangertest(Unemployment ~ Attendence, order = 6)
```

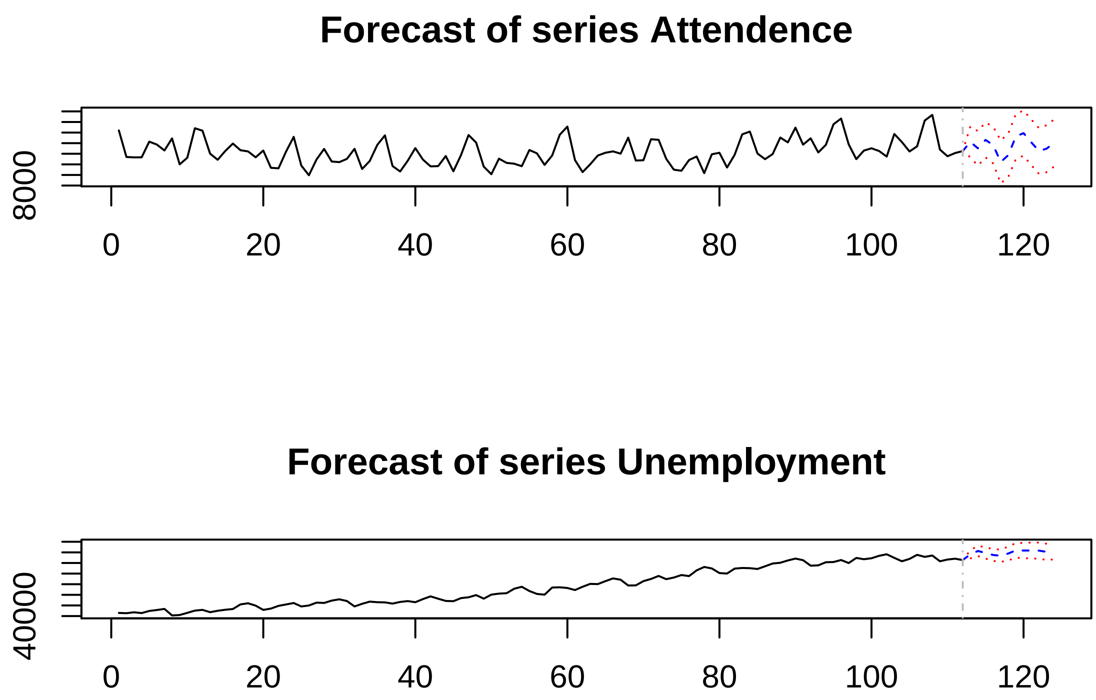
```
## Granger causality test
##
## Model 1: Unemployment ~ Lags(Unemployment, 1:6) + Lags(Attendence, 1:6)
## Model 2: Unemployment ~ Lags(Unemployment, 1:6)
##   Res.Df Df      F    Pr(>F)
## 1     93
## 2     99 -6 7.7426 9.149e-07 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

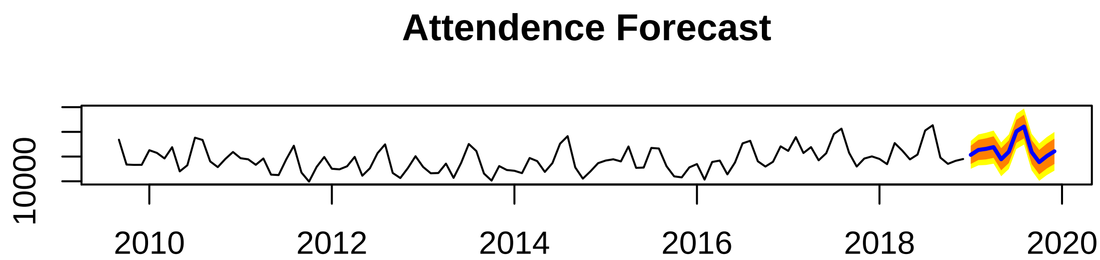
l. VAR forecast

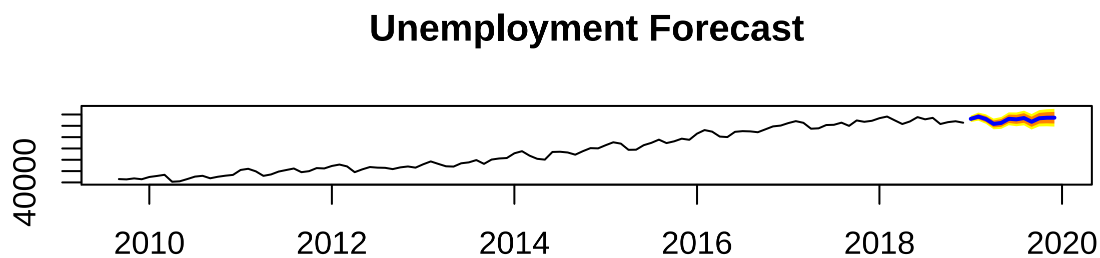
The VAR prediction seems to have similar pattern as the ARMA prediction but smaller fluctuations.

```r
varpred = predict(object = var_model, n.ahead = 12, level = .95)
plot(varpred)
```




```r
par(mfrow=c(2,1))
plot(forecast(Attendence_full_model, h = 12),
     shadecol = 'oldstyle', main ="Attendence Forecast")
plot(forecast(Unemployment_full_model, h= 12),
     shadecol = 'oldstyle', main = "Unemployment Forecast")
```






```
print("ARIMA Attendence Model Diagnostics:")
```

```
## [1] "ARIMA Attendence Model Diagnostics:"
```

```
accuracy(Attendence_full_model)
```

```
##                    ME     RMSE      MAE       MPE     MAPE      MASE
## Training set 168.2067 1322.192 1001.043 0.8683353 7.121259 0.6743661
##                      ACF1
## Training set -0.003525128
```

```
print("ARIMA Unemployment Model Diagnostics:")
```

```
## [1] "ARIMA Unemployment Model Diagnostics:"
```

```
accuracy(Unemployment_full_model)
```

```
##                    ME     RMSE      MAE       MPE     MAPE      MASE
## Training set 32.16916 2656.197 2028.982 0.1990982 2.171159 0.1639051
##                     ACF1
## Training set 0.005383186
```

```
print("VAR Attendence Model Diagnostics:")
```

```
## [1] "VAR Attendence Model Diagnostics:"
```

```
accuracy(var_model$varresult$Attendence)
```

```
##                        ME     RMSE      MAE        MPE     MAPE      MASE
## Training set 3.431637e-14 1390.033 1100.437 -0.9248801 7.784846 0.5711279
```

```r
print("VAR Unemployment Model Diagnostics:")
```

```
## [1] "VAR Unemployment Model Diagnostics:"
```

```r
accuracy(var_model$varresult$Unemployment)
```

```
##                         ME     RMSE      MAE        MPE     MAPE
## Training set -6.866792e-13 3689.248 2893.808 -0.2894182 3.227812
##                    MASE
## Training set 0.09117132
```

```r
AIC(var_model$varresult$Attendence,Attendence_full_model)
```

```
## Warning in AIC.default(var_model$varresult$Attendence,
## Attendence_full_model): models are not all fitted to the same number of
## observations
```

```
##                               df      AIC
## var_model$varresult$Attendence 14 1863.076
## Attendence_full_model           5 1733.809
```

```r
AIC(var_model$varresult$Unemployment,Unemployment_full_model)
```

```
## Warning in AIC.default(var_model$varresult$Unemployment,
## Unemployment_full_model): models are not all fitted to the same number of
## observations
```

```
##                                 df      AIC
## var_model$varresult$Unemployment 14 2070.009
## Unemployment_full_model           4 1867.372
```

```r
BIC(var_model$varresult$Attendence,Attendence_full_model)
```

```
## Warning in BIC.default(var_model$varresult$Attendence,
## Attendence_full_model): models are not all fitted to the same number of
## observations
```

```
##                               df      BIC
## var_model$varresult$Attendence 14 1900.365
## Attendence_full_model           5 1746.785
```

```r
BIC(var_model$varresult$Unemployment,Unemployment_full_model)
```

```
## Warning in BIC.default(var_model$varresult$Unemployment,
## Unemployment_full_model): models are not all fitted to the same number of
## observations
```

```
##                                 df      BIC
## var_model$varresult$Unemployment 14 2107.297
## Unemployment_full_model           4 1877.753
```

When we look at the diagnostics for both models, AIC, BIC, and most mean error claims are better for the ARIMA models. Therefore, we should use the ARIMA models instead of the VAR because the time series of unemployment or attendence does not do enough of a job to predict attendence and vice versa. The other problem might be unemployment is essentially uncorrelated to the daily attendence and therefore might not be the best dataset to be used in a VAR model. In the future, we can maybe use retail sales data or other local museum/recreational service's number of visitors as the second dataset.