Time Series Analysis
Of
A. Black Rock Stock Returns
B. Unemployment rate in California

Course:

MA 541 : Time Series Analysis

By

Hrushikesh Thanikonda

Date

December 12, 2025

Table of Contents

1. Introduction

Time series data generated in many financial and economic applications, where observations are collected sequentially over time. Understanding the dynamic behaviour of such data and producing reliable forecasts are essential for decision-making in finance, policy analysis, and economic planning. The objective of this project is to apply time series modelling techniques to real-world data to understand their underlying patterns and to forecast predictions of future.

The topics that are considered for this project are

- **Daily stock prices of BlackRock(BLK)** , a major global asset and invest management company. Stock market data is not interesting because of their relevance to investors and finance analysts but also as the data is noisy and behaves complex and random which might make analysis challenging. Also, predicting future stock prices could assist investors in making optimal investment decisions and developing trading strategies.
- **Monthly unemployment rates in the state of California** are also analysed in this project. Unemployment data are of interest because they provide important insights into labour market conditions and reflect the availability of jobs that affect job seekers. Unlike stock market data, unemployment rates often exhibit clear trends and seasonal patterns due to recurring economic and hiring . Analysing and forecasting unemployment rates can help jobseekers understand the state of labour market and more informed regarding job search, career planning and timing of employment opportunities.

Both time series exhibit different behaviours i.e., BLK stock prices are non-seasonal time series data and unemployment rate in California exhibit seasonal time series behaviour.

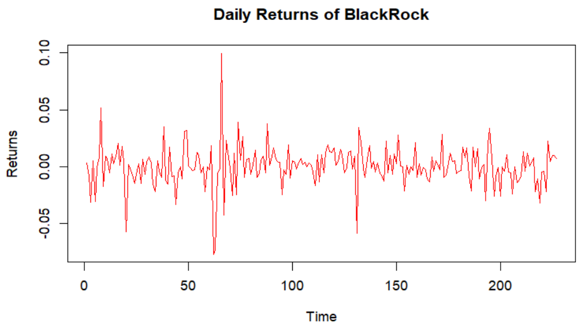
The primary goal of the project is to model each dataset using Box-Jenkins approach and find the suitable time series model while performing diagnostics checks and generate forecast of both dataset and, the project aims to demonstrate how statistical techniques are selected and justified based on data.

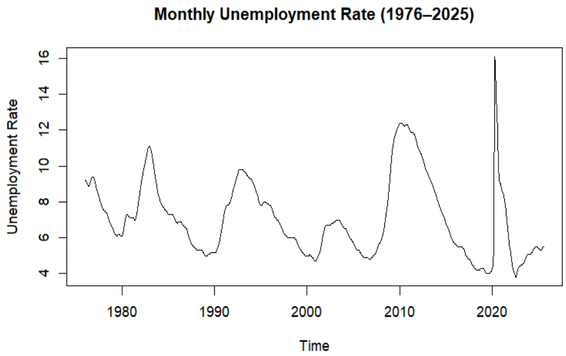
2. Data Description and Exploratory Analysis

A. Non-Seasonal Data: BlackRock Stock returns :

Stock prices of BlackRock (BLK), which is a major global asset and investment management company. The data were collected from Yahoo Finance (https://finance.yahoo.com/quote/BLK/history/). The dataset consists of daily stock price data for 229 trading days, from January 02, 2025, through November 28, 2025. Since the data are recorded on trading days there were no missing values in the dataset. Stock price data usually show trends over time and are non-stationary in nature, which makes them difficult to model directly using time series methods. So, the stock prices were converted into daily returns that measure the percentage change in price from one day to the next. This transformation helps remove trends and makes the series more stable over time and normalise the data.

The plot of the daily returns shows that the values fluctuate around zero and do not follow any clear trend. The series appears noisy and random There is no visible seasonal pattern in the return series, and therefore the BlackRock stock return data are treated as a non-seasonal time series in this project.





B. Seasonal Data:

The seasonal dataset used in this project is the Unemployment Rate in California, collected from the Federal Reserve Bank of St. Louis Economic Data (FRED) database (https://fred.stlouisfed.org/series/CAUR). These series record the monthly unemployment rate in California from January 1976 through August 2025, with no missing observations over this long period and contains 597 data points. Since the data points are monthly, they exhibit seasonal patterns, particularly because employment levels vary systematically throughout the year due to factors such as holiday hiring, school calendars, and cyclical industry activities. This seasonality makes modeling the raw unemployment rate challenging without explicitly accounting for the periodic effects seen at regular intervals. Therefore, in the seasonal analysis we focus on identifying and modeling both the trend and the seasonal components that are present in the time series to accurately characterize its behaviour and check forecast and evaluate it.

3. Box-Jenkins Models

Non-Seasonal Data : BlackRock Stock Returns:

- **Identification:**

  The stationarity of the BlackRock daily returns was analysed. In ACF and PACF plots it was observed that there were no significant spikes at any lag, suggesting weak dependence. To confirm this statistically we used the Augmented Dickey-Fuller (ADF) test was performed. This test yielded a p-value(0.01) of less than 0.05 from which we concluded that the rejection of the null hypothesis. This confirmed that the series is stationary, and therefore, no differencing was required.

- **Estimation:**

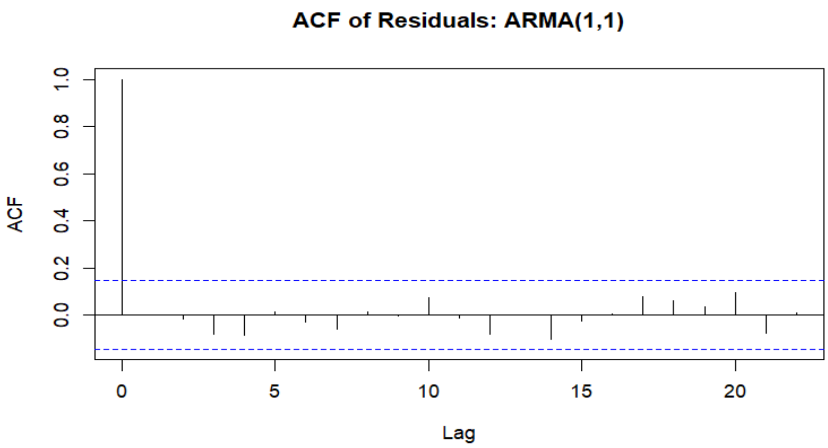
  The data was split into training and testing data in the ratio of 80:20 to evaluate predictive performance. 3 models were fitted to the training data:

  - AR(1)
  - MA(1)
  - ARMA(1,1)

  The Akaike Information Criterion (AIC) values for these models were calculated and compared, and values were found to be similar, but ARMA(1,1) had significantly less AIC value of - 926.79. Therefore, for forecasting phase ARMA(1,1) model was selected to ensure both autoregressive and moving average dynamics were captured.

- **Diagnostic Checking:**

  The ACF of the residuals for these models were plotted and there are no significant patterns or spikes, confirming that the residuals behave like white noise. This indicates that the ARMA(1,1) successfully captured the available information from the data.



- **Forecasting:**

  Forecast for remaining test data was generated using the ARMA(1,1) model and it appeared as a flat line near zero, with confidence intervals. These predictions were compared with the actual returns, and it was noted that most actual values fell within the 95% confidence intervals.

Seasonal Data: Unemployment Rate in California :

- **Identification:**

  From ACF and PACF plots we observed that plots of the monthly unemployment rate exhibited a clear trend and strong seasonality. Differencing (d=1) was applied to remove the trend then followed by seasonal differencing (D=1) at lag 12. The resulting ACF and PACF plots of the differenced series displayed significant spikes, indicating the necessity for both seasonal AR and MA terms.

- **Estimation:**

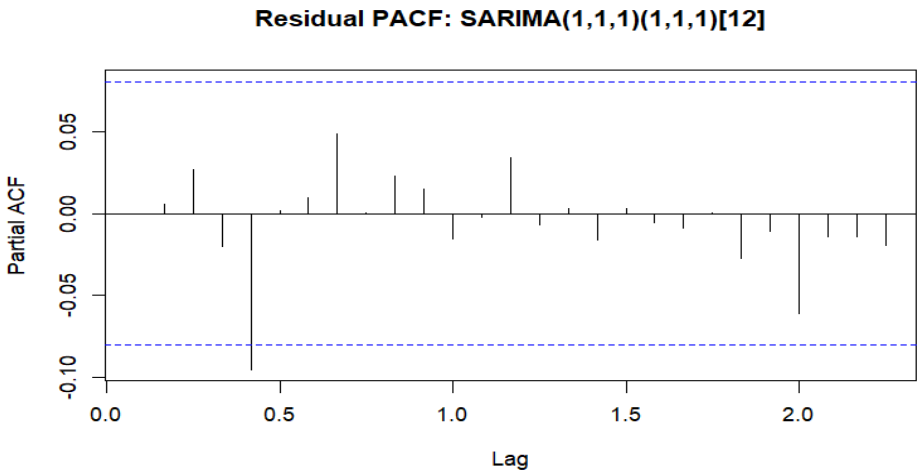
  The dataset was utilized for model fitting instead of having test and train sets to maximize the capture of long-term seasonal trends. Based on the identification step, three Seasonal ARIMA (SARIMA) models were fitted with a fixed seasonal structure of (1,1,1)[12]:

  - SARIMA(1,1,0)(1,1,1)[12]
  - SARIMA(0,1,1)(1,1,1)[12]
  - SARIMA(1,1,1)(1,1,1)[12]

  The AIC was calculated for each model for finding the best model that can capture most data. The SARIMA(1,1,1)(1,1,1)[12] model had the least AIC value and then this model was selected for analysis and forecasting.

- **Diagnostic Checking:**

  Residual diagnostics were performed on the selected SARIMA(1,1,1)(1,1,1)[12] model. The ACF plots of the residuals showed no significant spikes, and the Ljung-Box test confirmed that the residuals are independent (p-value(0.9826) > 0.05), ensuring that model is statistically adequate.



- **Forecasting:**

  Using SARIMA(1,1,1)(1,1,1)[12] on the full data, a 12-month forecast was generated and predicted values successfully replicated the seasonal wave pattern that was observed in the historical data, and the 95% confidence intervals provided a real-time range for future unemployment rates.

4. Conclusion

BlackRock Stock Analysis :

Analysis suggests that attempting to predict daily price movements of BlackRock stocks based solely on past price history is not a reliable strategy. The market appears to be efficient i.e., new information is instantly reflected in prices, making day-to-day changes random.

So, we cannot predict whether the stock will go up or down tomorrow, our model allows us to estimate the risk i.e., the range of expected fluctuation. Investors should focus on long-term trends and risk management strategies rather than trying to time daily trades based on historical patterns.

California Unemployment Rate analysis

The unemployment rate in California follows a highly predictable annual pattern. We have identified clear seasonal cycle which are driven by agricultural seasons, holiday hiring, and the academic calendar.

This predictability is valuable for government planning. It allows policymakers to distinguish between "normal" seasonal fluctuations and actual economic downturns. For example, a rise in unemployment in a specific month might just be the usual seasonal trend rather than a sign of a recession. Our forecasts provide a reliable baseline for the next year, helping to budget for unemployment benefits and training programs more accurately.

MA 641 Final Project

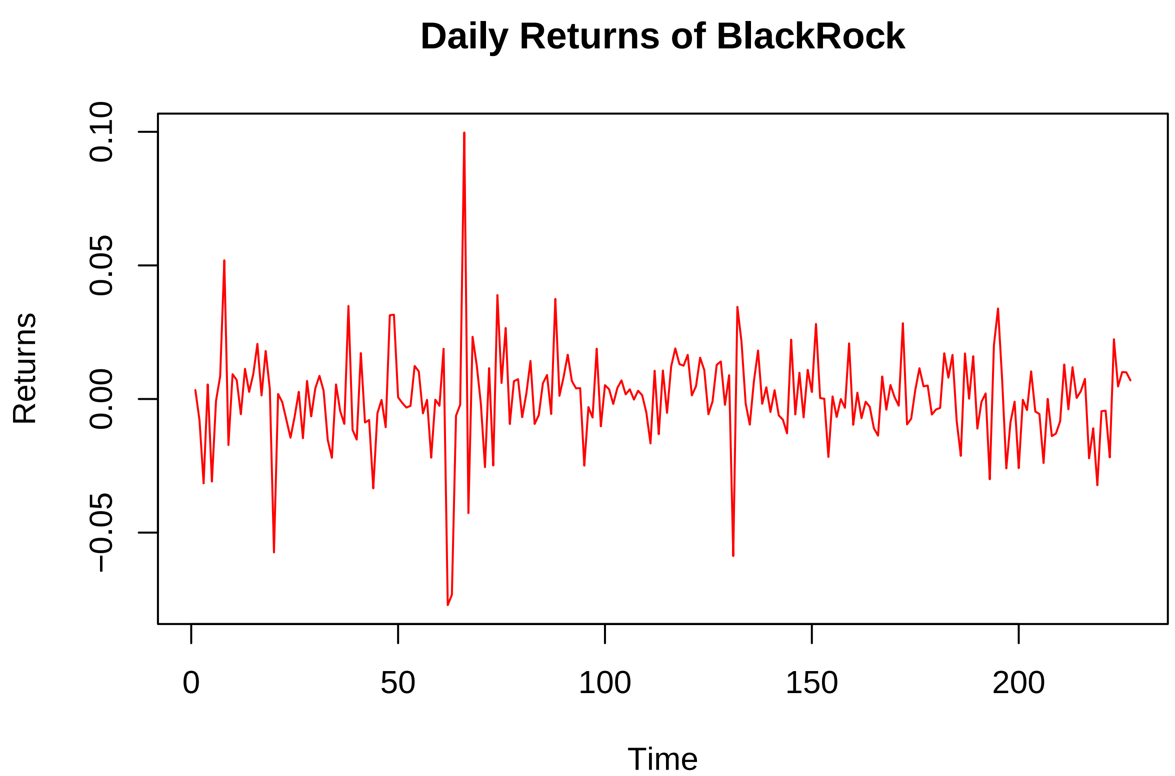
Hrushikesh Thanikonda

2025-12-12

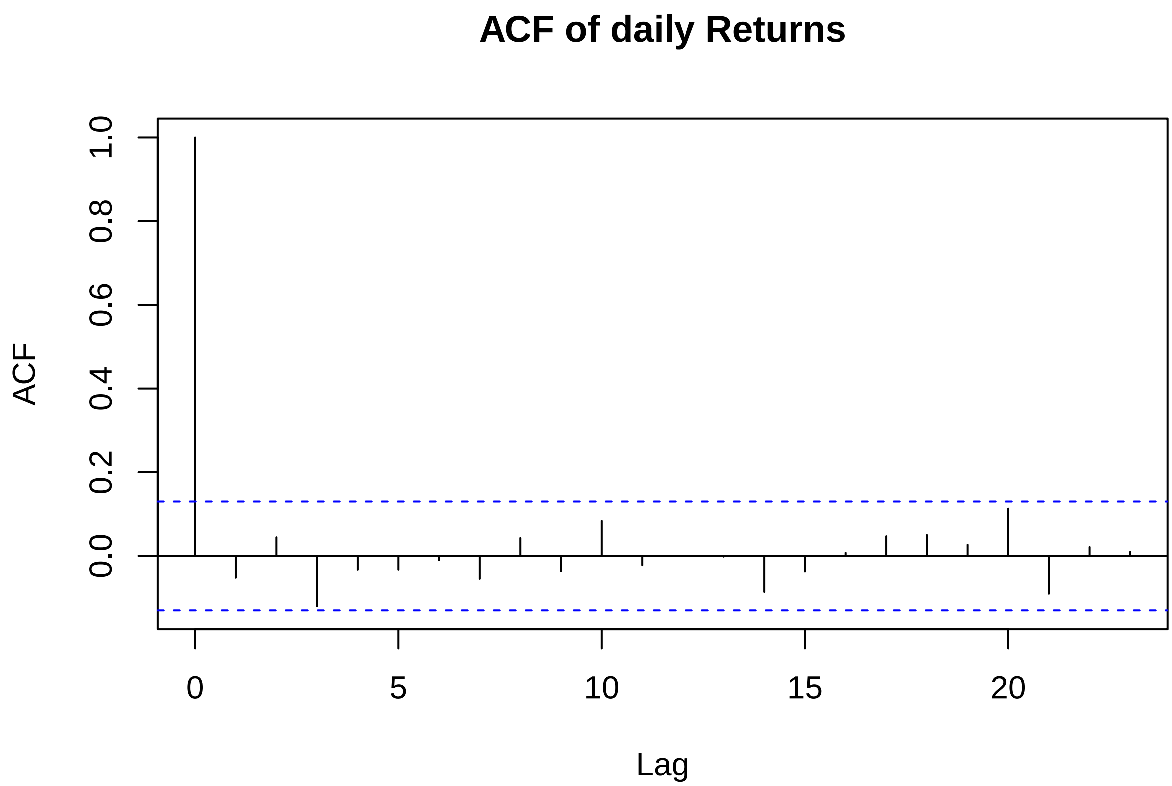
Appendix

Time Series Analysis of Black Rock Daily - Stock Prices (Non- Seasonal)

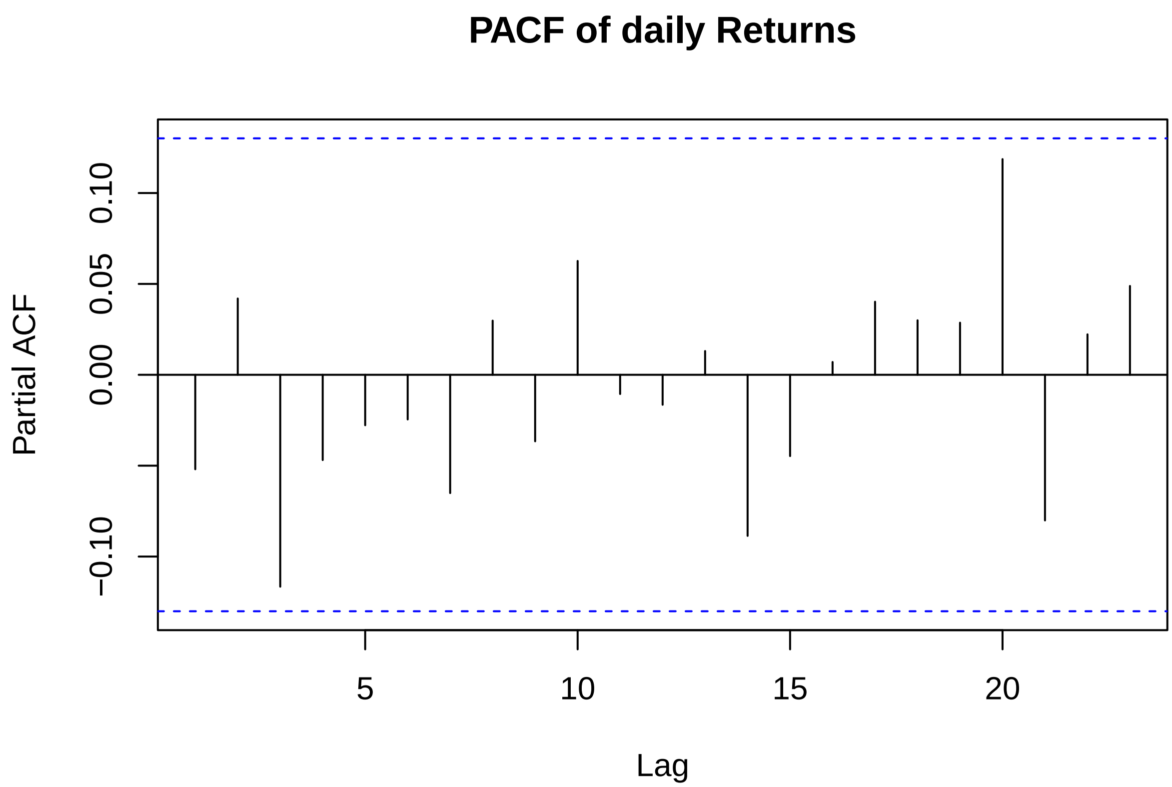
```
# Importing Blackrock data set
blk <- read.csv("C:\\Users\\hrush\\OneDrive\\Desktop\\Time Series Analysis Project\\BlackRockData2025.c
# calculating returns
n <- length(blk$BLK.Adjusted)
blkReturns <-( blk$BLK.Adjusted[2:n]-blk$BLK.Adjusted[1:(n-1)])/blk$BLK.Adjusted[1:(n-1)]
# converting BlackRock returns to ts
blkRetTs <- ts(blkReturns)
plot(blkRetTs,main = "Daily Returns of BlackRock",ylab ="Returns",col = "red")
```



```r
# we notice that mean is around 0 and we'll check for stationarity
acf(blkRetTs,main = "ACF of daily Returns")
```




```r
# we notice that there are no significant spikes in ACF plot => weak dependence
pacf(blkRetTs,main = "PACF of daily Returns")
```




```
# We notice that there are no significant spikes
# Agumented Dickey Fuller Test to confrim stationrity
library(tseries)
```

```
## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo
```

```
adf.test(blkRetTs)
```

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

```
##
##  Augmented Dickey-Fuller Test
##
## data:  blkRetTs
## Dickey-Fuller = -6.614, Lag order = 6, p-value = 0.01
## alternative hypothesis: stationary
```

```
#as p-value <0.05 we reject Ho i.e., Time series are stationary and no differencing required
size <- length(blkRetTs)
trainSize <- floor(0.8 * size)

train <- blkRetTs[1:trainSize]
test  <- blkRetTs[(trainSize + 1):size]

#fitting the modelson training set
ar1<- arima(train, order = c(1, 0, 0))
```

```r
ma1 <- arima(train, order = c(0, 0, 1))
arma11 <- arima(train, order = c(1, 0, 1))
cat("AR1 fit Results:\n")
```

```
## AR1 fit Results:
```

```r
print(ar1)
```

```
##
## Call:
## arima(x = train, order = c(1, 0, 0))
##
## Coefficients:
##           ar1  intercept
##       -0.0744     0.0008
## s.e.   0.0739     0.0013
##
## sigma^2 estimated as 0.0003387:  log likelihood = 466.29,  aic = -926.58
```

```r
cat("MA1 fit Results:\n")
```

```
## MA1 fit Results:
```

```r
print(ma1)
```

```
##
## Call:
## arima(x = train, order = c(0, 0, 1))
##
## Coefficients:
##           ma1  intercept
##       -0.0699     0.0008
## s.e.   0.0725     0.0013
##
## sigma^2 estimated as 0.0003389:  log likelihood = 466.26,  aic = -926.51
```

```r
cat("ARMA(1,1) fit Results:\n")
```

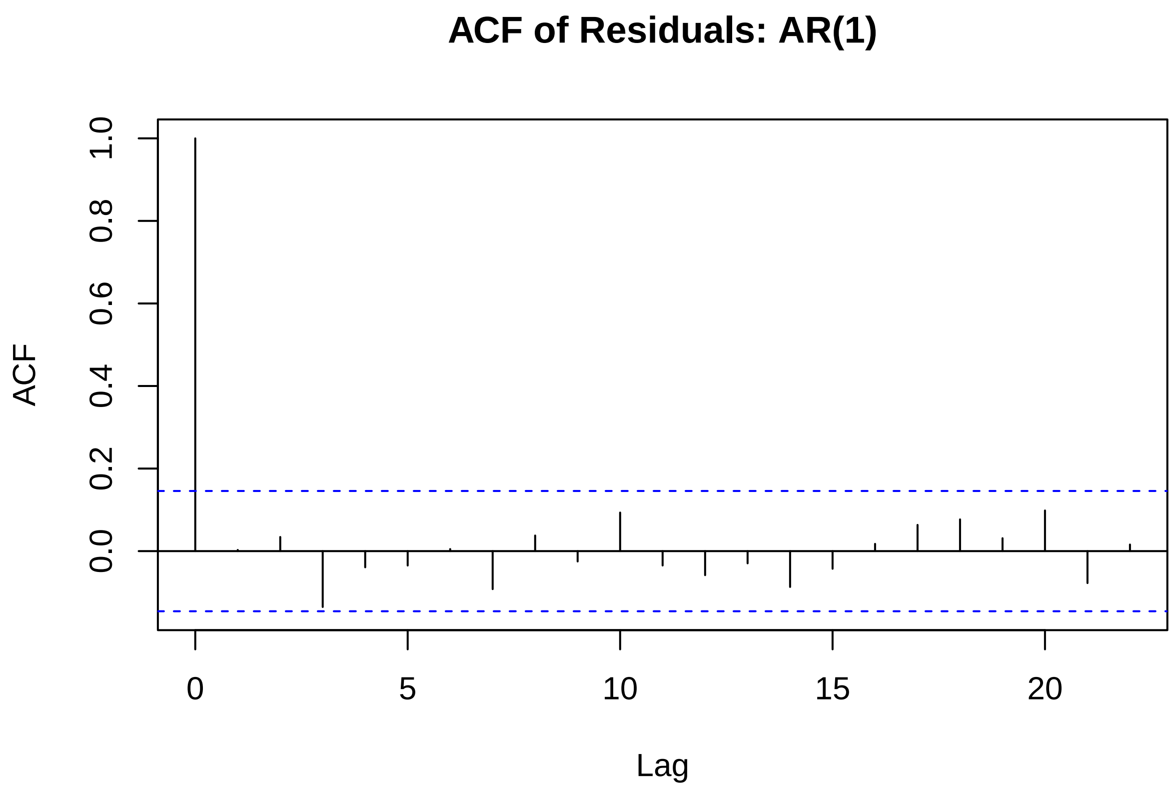
```
## ARMA(1,1) fit Results:
```

```r
print(arma11)
```

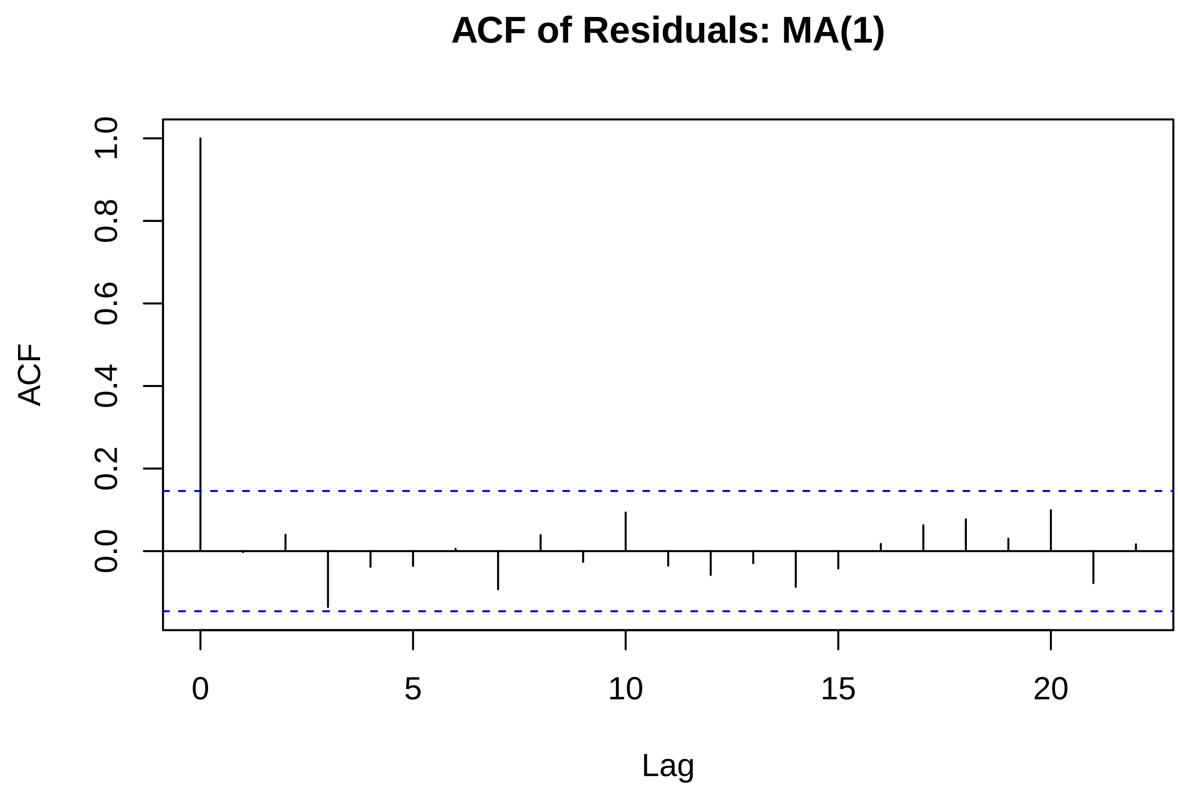
```
##
## Call:
## arima(x = train, order = c(1, 0, 1))
##
## Coefficients:
##           ar1     ma1  intercept
##       -0.9016  0.8438     0.0008
## s.e.   0.1385  0.1719     0.0013
##
## sigma^2 estimated as 0.0003345:  log likelihood = 467.39,  aic = -926.79
```

checking residual ACF

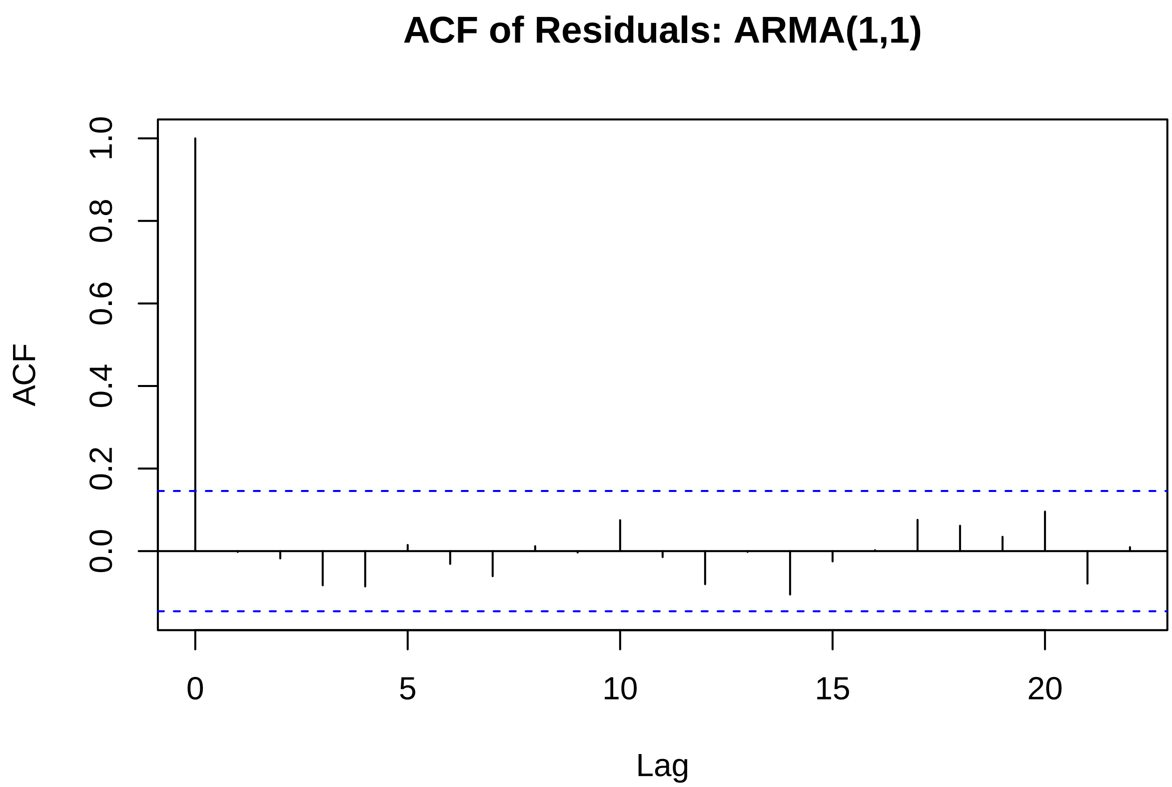
```r
acf(residuals(ar1),main = "ACF of Residuals: AR(1)")
```




```
acf(residuals(ma1),main = "ACF of Residuals: MA(1)")
```




```
acf(residuals(arma11),main = "ACF of Residuals: ARMA(1,1)")
```



We notice that all 3 are similar so we compare AIC's

```r
aicValues <- AIC(ar1, ma1, arma11)
cat("Lowest AIC value is:",min(aicValues$AIC),"for model:",rownames(aicValues)[order(aicValues$AIC)[1]]
```

```
## Lowest AIC value is: -926.7867 for model: arma11
```

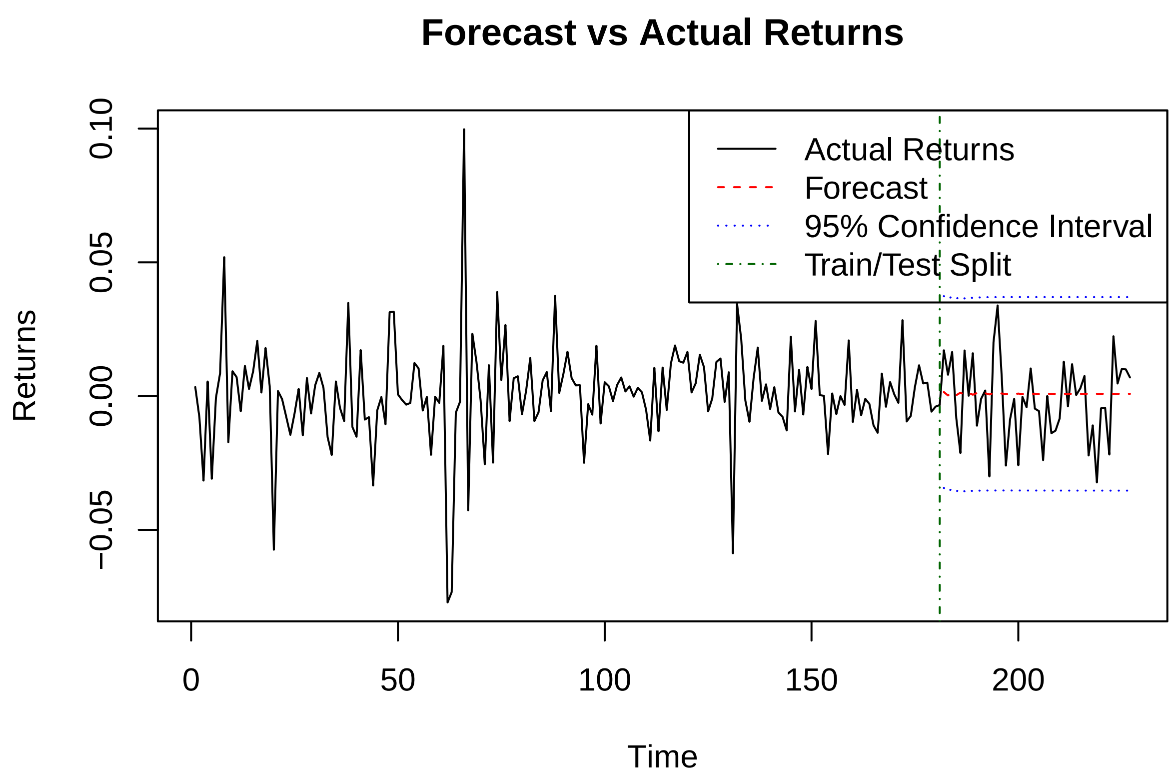
We noitice that phi value is near to 0 so we can forecast

```r
prediction <- predict(arma11, n.ahead = length(test))
fc <- prediction$pred
# 95% confidence intervals for forecasts
upper <- prediction$pred + 1.96 * prediction$se
lower <- prediction$pred - 1.96 * prediction$se
```

Comparing forecast results with actual returns

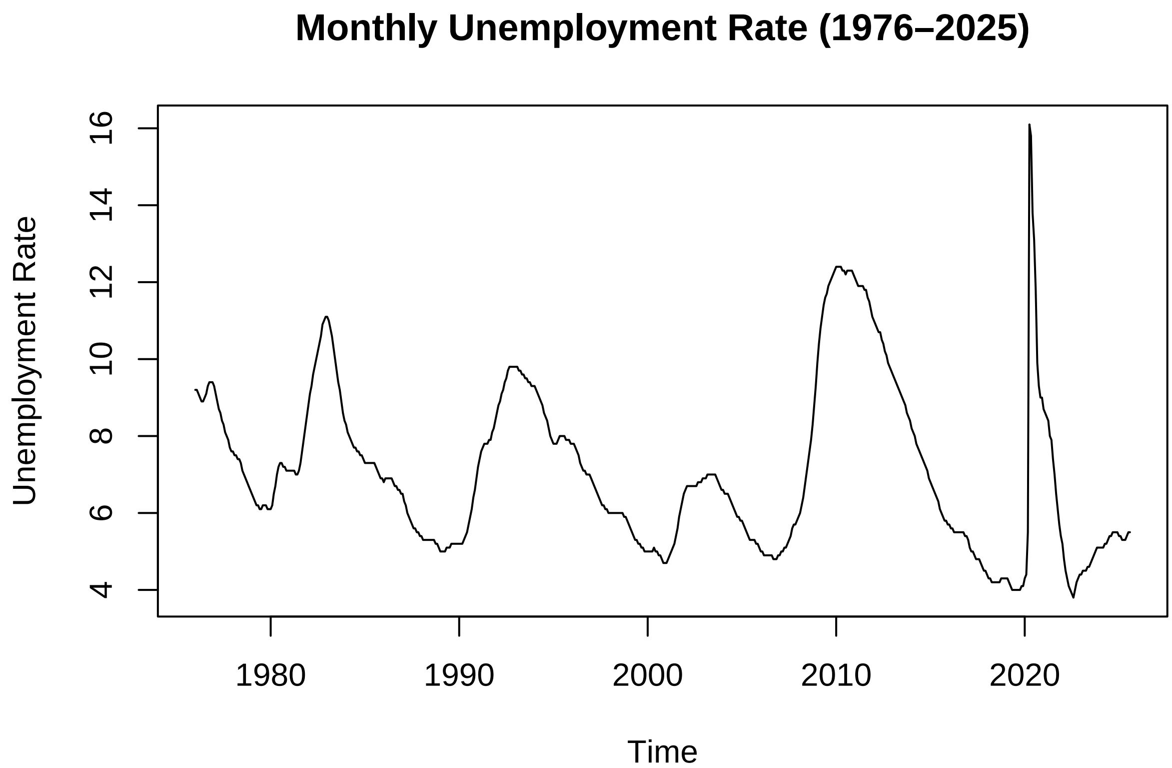
```r
plot(blkRetTs,type = "l",col = "black",main = "Forecast vs Actual Returns",ylab = "Returns",xlab = "Tim

# Forecasted returns
lines((trainSize + 1):(trainSize + length(test)),fc,col = "red",lty = 2)

# Confidence intervals
lines((trainSize + 1):(trainSize + length(test)),upper,col = "blue",
      lty = 3)
```

```r
lines((trainSize + 1):(trainSize + length(test)),
      lower,
      col = "blue",
      lty = 3)

# Train/test split
abline(v = trainSize, col = "darkgreen", lty = 4)

legend("topright",legend = c("Actual Returns","Forecast","95% Confidence Interval","Train/Test Split"),
       col = c("black", "red", "blue", "darkgreen"),
       lty = c(1, 2, 3, 4))
```



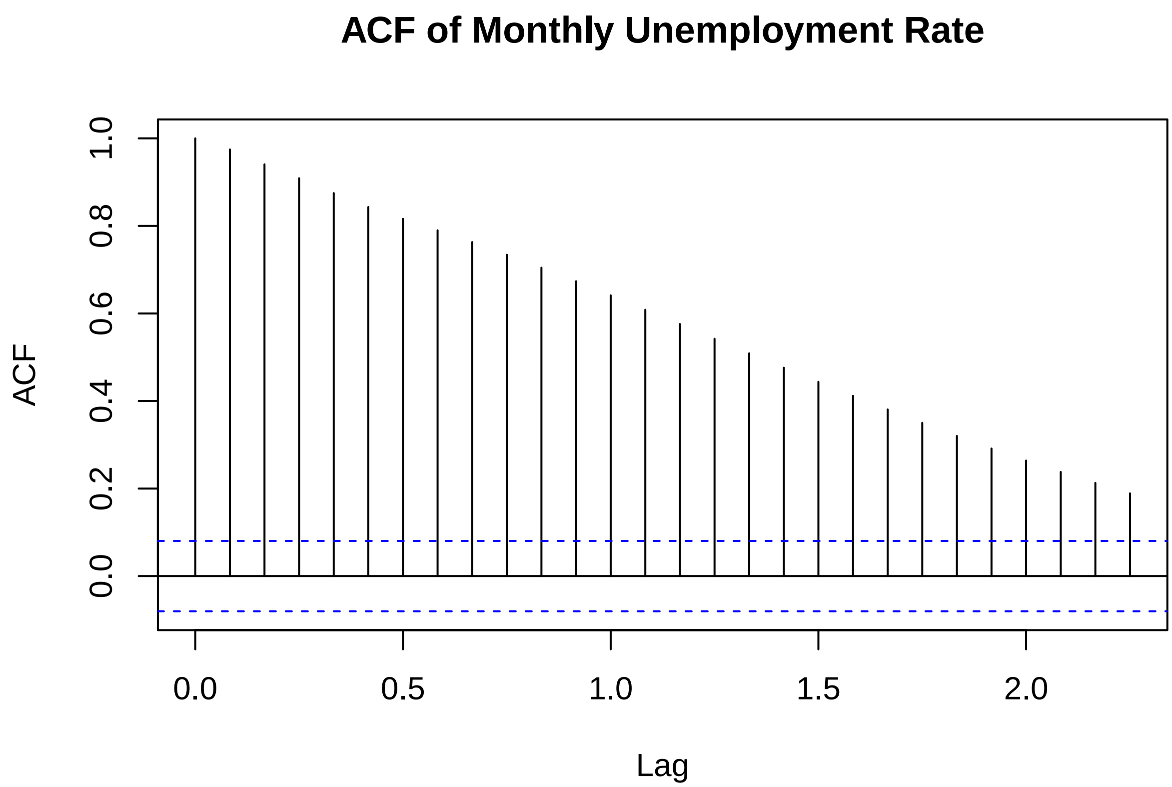
Time Series Analysis of Unemployement in California State (Seasonal)

```r
# Import california unemployment data
caur <- read.csv("C://Users//hrush/OneDrive//Desktop//Time Series Analysis Project//CAUR.csv")

# Convert to monthly time series
cts <- ts(caur$CAUR,start = c(1976, 1),frequency = 12)

# Plot the series
plot(cts,main = "Monthly Unemployment Rate (1976-2025)",ylab = "Unemployment Rate",xlab = "Time")
```



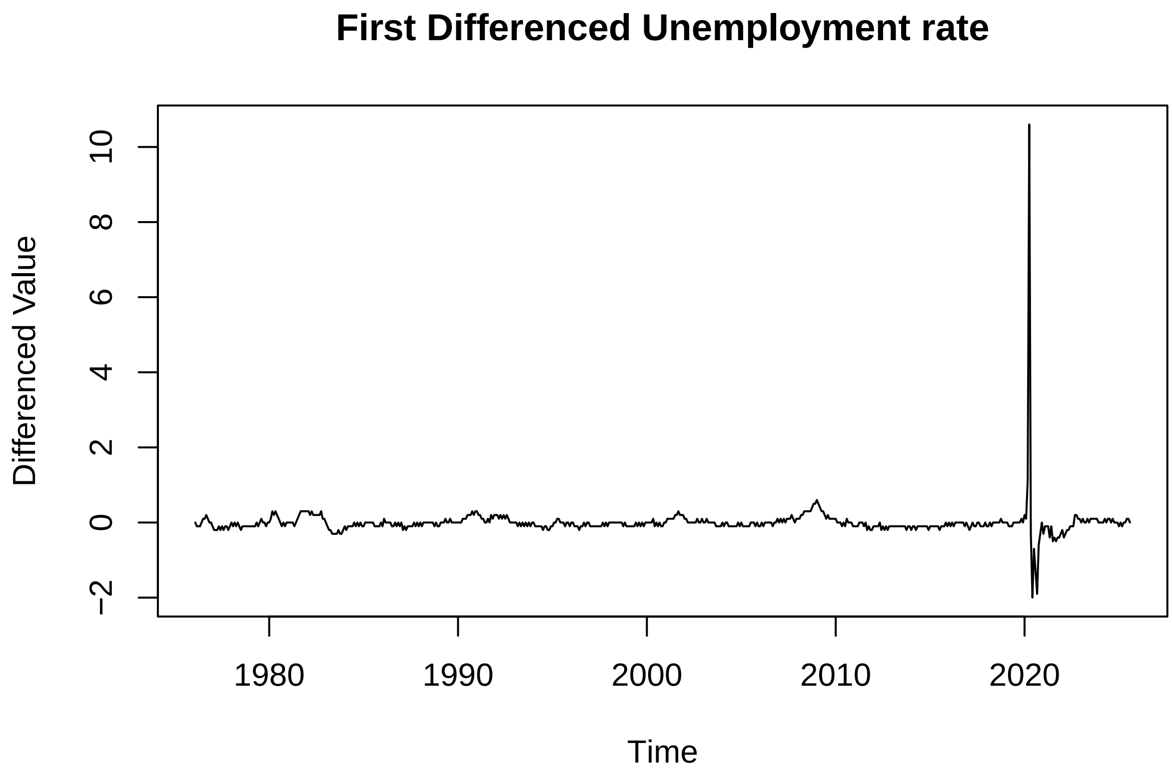
Checking for correlations

```
acf(cts,main = "ACF of Monthly Unemployment Rate")
```



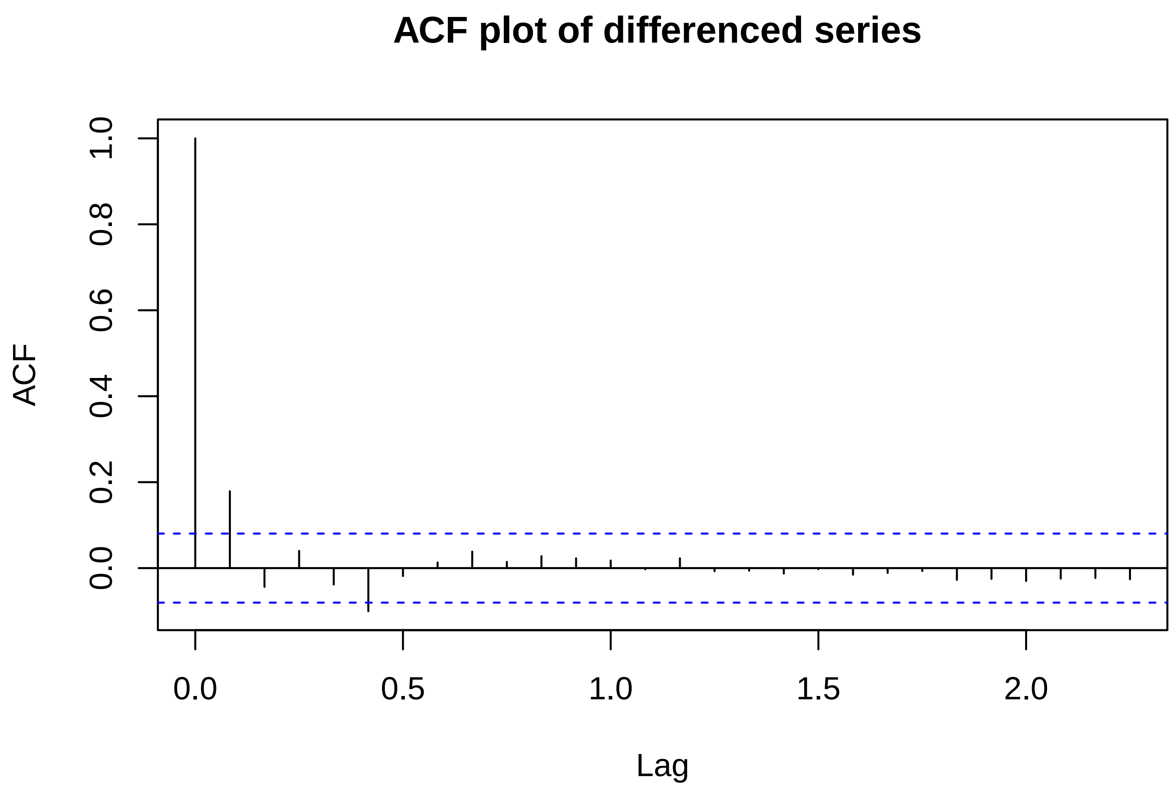
We notice ACF decays slowly => series are not stationary

```
cdiff <- diff(cts,differences = 1)
plot(cdiff,main ="First Differenced Unemployment rate",ylab = "Differenced Value",xlab="Time")
```



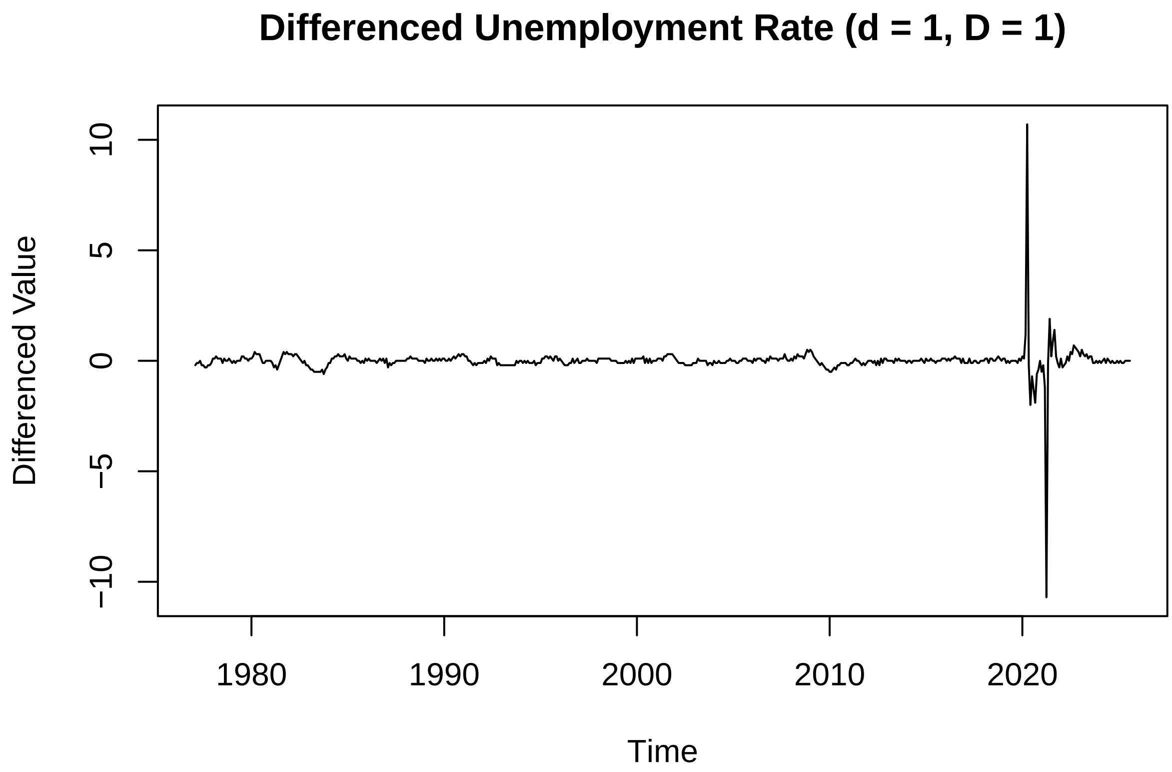
Plot seems to be stationary after first differencing

```
acf(cdiff,main = "ACF plot of differenced series ")
```

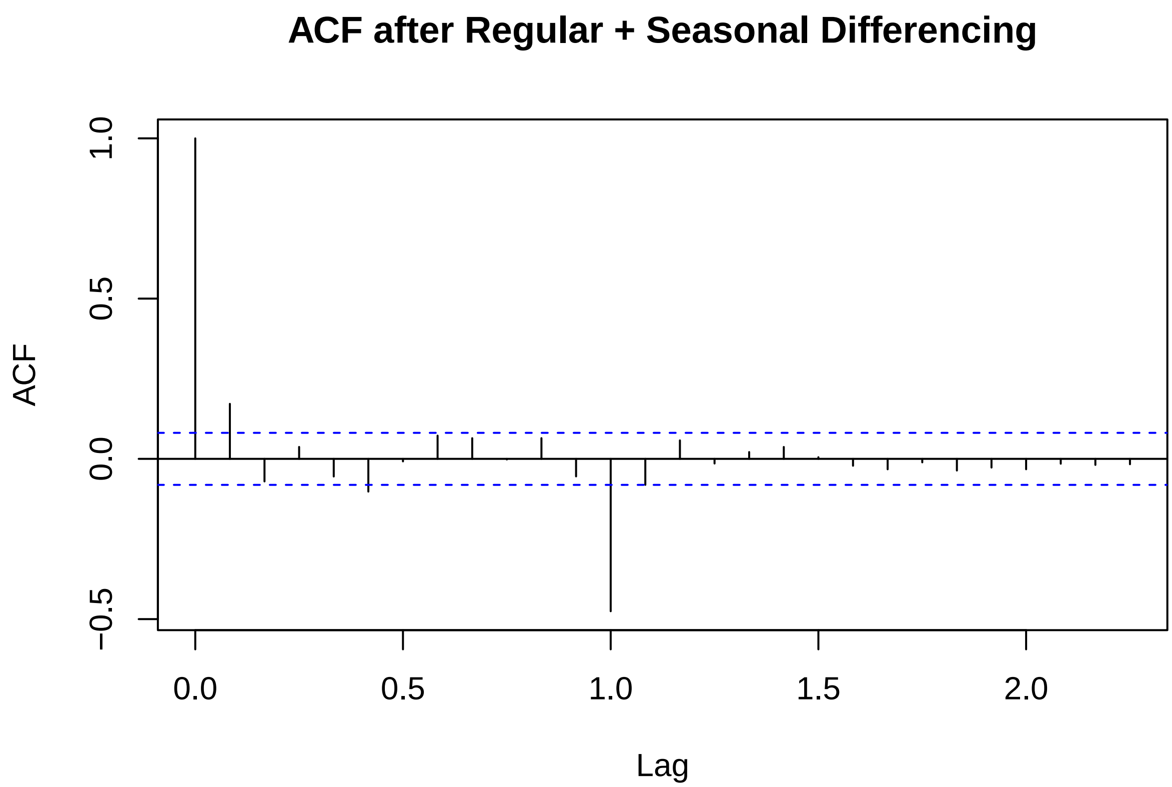


From acf we can still see structure at lag 12(i.e., 1) so we mignt need seasonal differencing

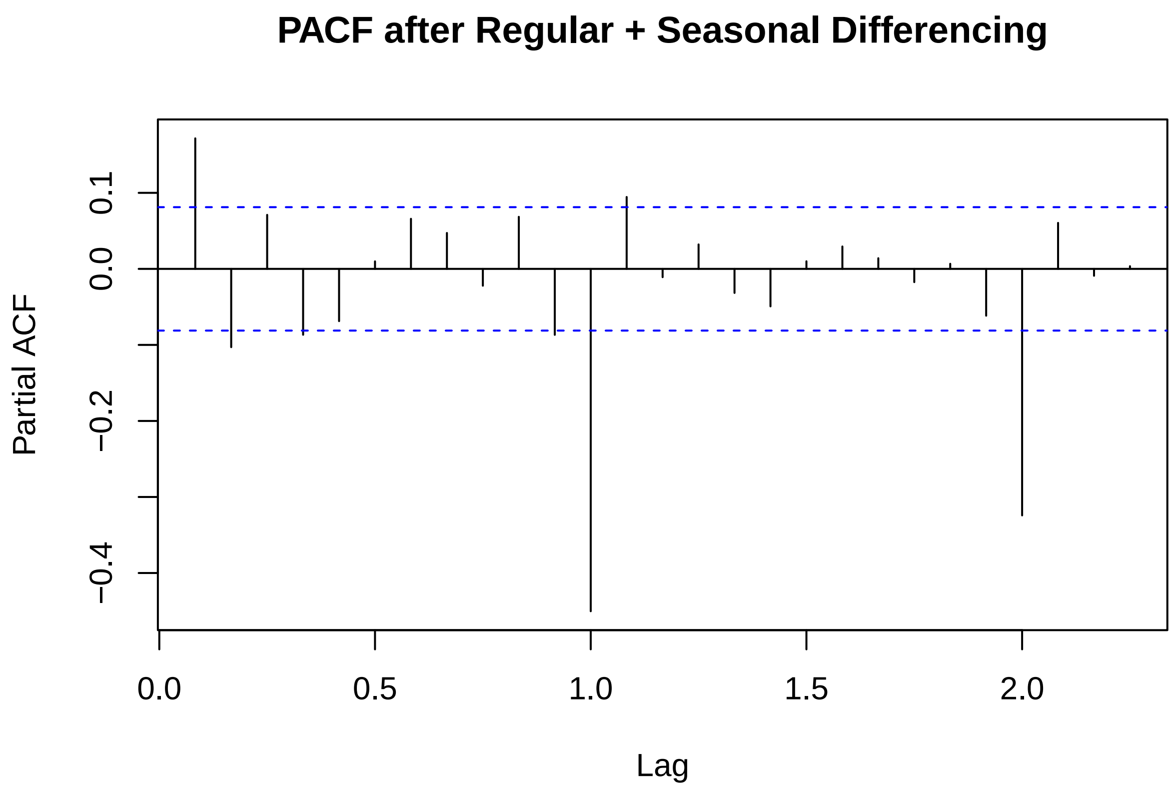
```
cDiff <- diff(cdiff,lag = 12)
plot(cDiff,main = "Differenced Unemployment Rate (d = 1, D = 1)",ylab = "Differenced Value",xlab = "Tim
```




```
acf(cDiff,main = "ACF after Regular + Seasonal Differencing")
```




```
pacf(cDiff,main = "PACF after Regular + Seasonal Differencing")
```




```r
# ARIMA(1,1,0)(1,1,1)[12]
sarimaAr <- arima(cts,
                  order = c(1, 1, 0),
                  seasonal = list(order = c(1, 1, 1), period = 12))

# ARIMA(0,1,1)(1,1,1)[12]
sarimaMa <- arima(cts,
                  order = c(0, 1, 1),
                  seasonal = list(order = c(1, 1, 1), period = 12))

# ARIMA(1,1,1)(1,1,1)[12]
sarimaArma <- arima(cts,
                    order = c(1, 1, 1),
                    seasonal = list(order = c(1, 1, 1), period = 12))
cat("SARIMA(1,1,0)(1,1,1)[12] Results:\n")
```

```
## SARIMA(1,1,0)(1,1,1)[12] Results:
```

```r
print(sarimaAr)
```

```
##
## Call:
## arima(x = cts, order = c(1, 1, 0), seasonal = list(order = c(1, 1, 1), period = 12))
##
## Coefficients:
##          ar1    sar1     sma1
##       0.1820  0.0019  -0.9599
```

```
## s.e.  0.0407  0.0437   0.0241
##
## sigma^2 estimated as 0.2255:  log likelihood = -408.22,  aic = 824.44
```

```r
cat("\nSARIMA(0,1,1)(1,1,1)[12] Results:\n")
```

```
##
## SARIMA(0,1,1)(1,1,1)[12] Results:
```

```r
print(sarimaMa)
```

```
##
## Call:
## arima(x = cts, order = c(0, 1, 1), seasonal = list(order = c(1, 1, 1), period = 12))
##
## Coefficients:
##          ma1    sar1     sma1
##       0.2139  0.0040  -0.9599
## s.e.  0.0437  0.0437   0.0241
##
## sigma^2 estimated as 0.2243:  log likelihood = -406.6,  aic = 821.19
```

```r
cat("\nSARIMA(1,1,1)(1,1,1)[12] Results:\n")
```

```
##
## SARIMA(1,1,1)(1,1,1)[12] Results:
```

```r
print(sarimaArma)
```

```
##
## Call:
## arima(x = cts, order = c(1, 1, 1), seasonal = list(order = c(1, 1, 1), period = 12))
##
## Coefficients:
##           ar1     ma1    sar1     sma1
##       -0.2785  0.4816  0.0087  -0.9608
## s.e.   0.1480  0.1338  0.0438   0.0243
##
## sigma^2 estimated as 0.2232:  log likelihood = -405.28,  aic = 820.56
```

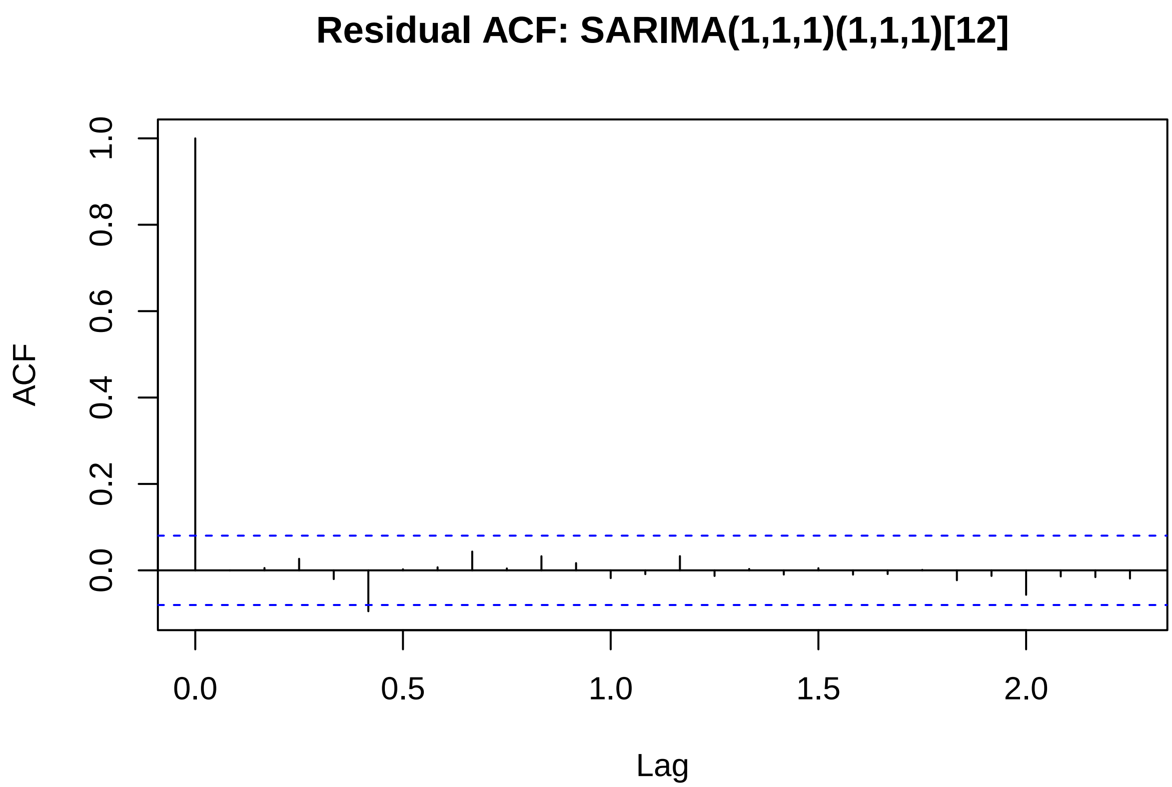
Check AIC values to choose the model

```r
aicValuesS <- AIC(sarimaAr,sarimaMa,sarimaArma)
cat("Lowest AIC value is:",min(aicValuesS$AIC),"for model:",rownames(aicValuesS)[order(aicValuesS$AIC)[
```

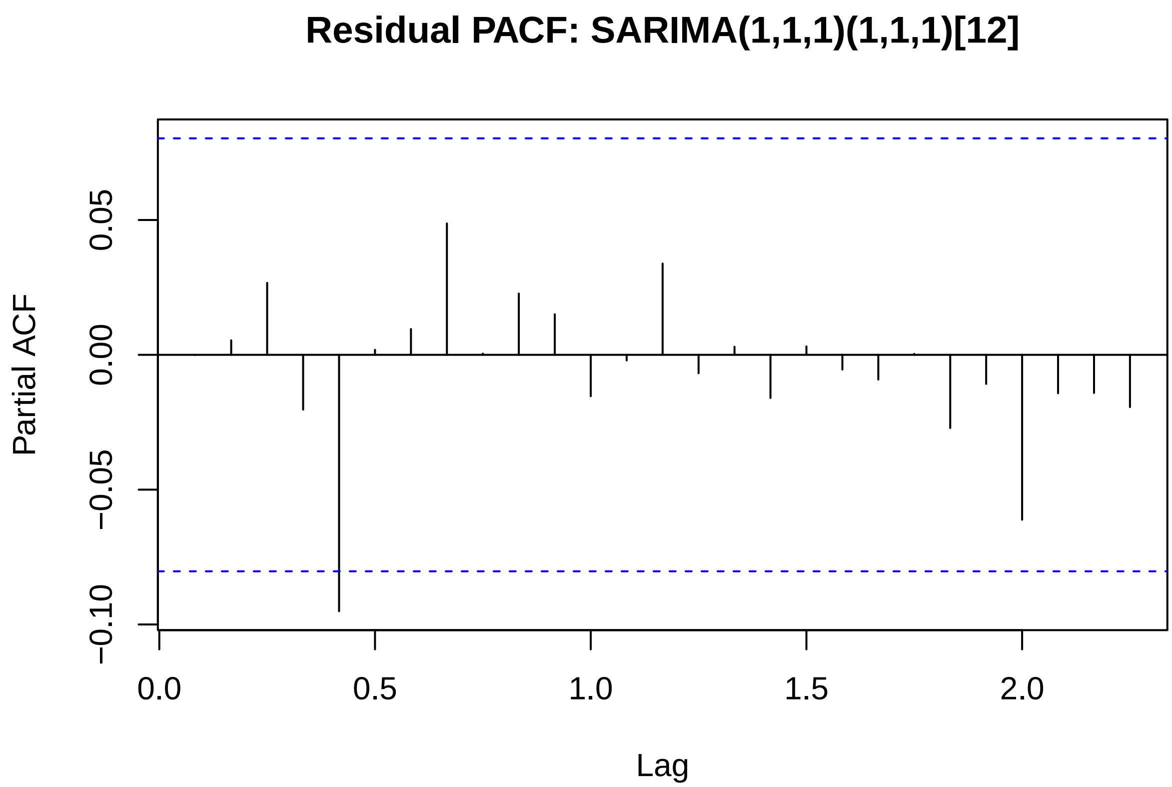
```
## Lowest AIC value is: 820.5599 for model: sarimaArma
```

Residual Diagnostics

```r
acf(residuals(sarimaArma),
    main = "Residual ACF: SARIMA(1,1,1)(1,1,1)[12]")
```



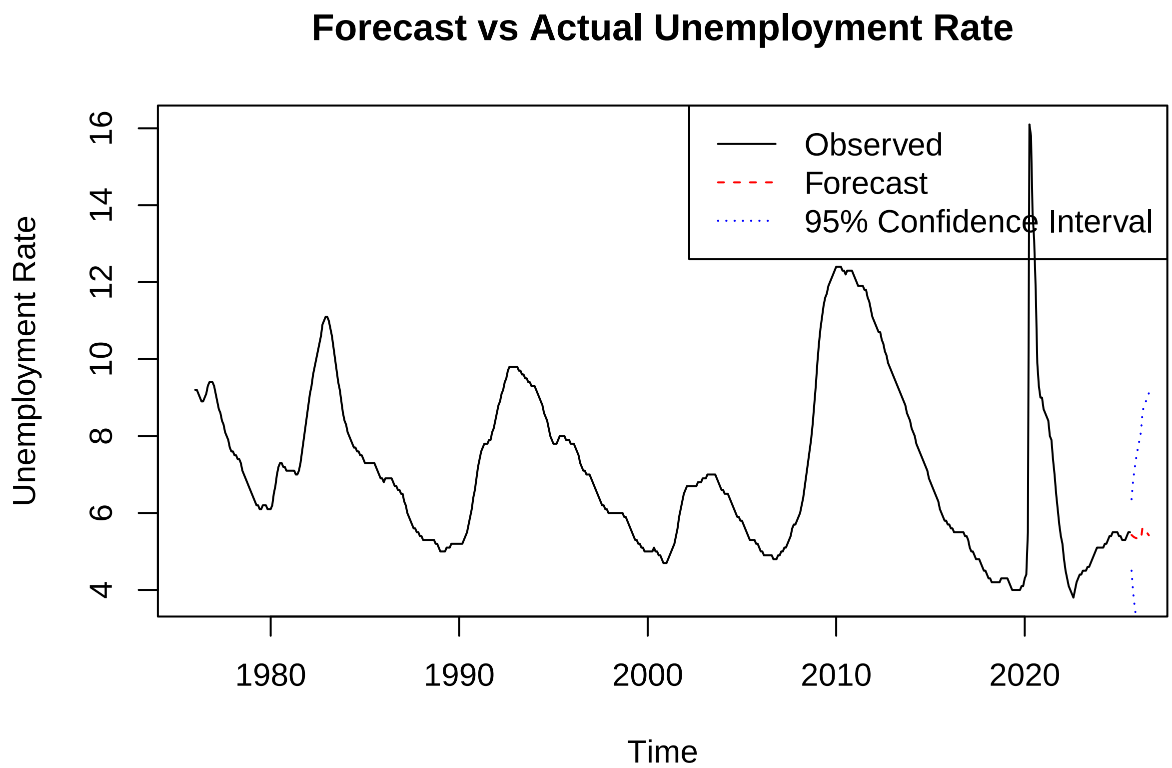

```
pacf(residuals(sarimaArma),
     main = "Residual PACF: SARIMA(1,1,1)(1,1,1)[12]")
```




```r
Box.test(residuals(sarimaArma),lag = 24,type = "Ljung-Box")
```

```
##
##  Box-Ljung test
##
## data:  residuals(sarimaArma)
## X-squared = 11.75, df = 24, p-value = 0.9826
```

```r
sPrediction <- predict(sarimaArma, n.ahead = 12)

sFc <- sPrediction$pred

sUpper <- sPrediction$pred + 1.96 * sPrediction$se
sLower <- sPrediction$pred - 1.96 * sPrediction$se

plot(cts,type = "l",col = "black",main = "Forecast vs Actual Unemployment Rate",ylab = "Unemployment Ra
# Forecasted values
lines(sFc,col = "red",lty = 2)
# Confidence intervals
lines(sUpper,col = "blue",lty = 3)
lines(sLower,col = "blue",lty = 3)

legend("topright",
       legend = c("Observed","Forecast","95% Confidence Interval"),
       col = c("black", "red", "blue"),
       lty = c(1, 2, 3))
```


Forecast vs Actual Unemployment Rate
Observed
Forecast
95% Confidence Interval
Unemployment Rate
4
6
8
10
12
14
16
1980
1990
2000
2010
2020
Time