Problem Set 4: Forecasting the Unemployment Rate

i. Load the dataset

```
library("fredr")
library("dynlm")
```

```
## Loading required package: zoo
```

```
##
## Attaching package: 'zoo'
```

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

```
library("timetk")
```

ii. Set API key

```
fredr_set_key("5bc428e1039cc5ca47ae6cee02e5fe02")
```

iii. Download and read the dataset

```
Claims <-fredr( series_id = "ICSA",frequency = "m", observation_start=as.Date("1990-01-01"),
                observation_end=as.Date("2023-04-01"))
Unempl <-fredr( series_id = "UNRATE",frequency = "m", observation_start=as.Date("1990-01-01"),
                observation_end=as.Date("2023-04-01"))
Claims_w <-fredr( series_id = "ICSA",frequency = "w", observation_start=as.Date("1990-01-01"),
                  observation_end=as.Date("2023-04-01"))
Claims = tk_zoo_(Claims)
```

```
## Warning: Non-numeric columns being dropped: date, series_id, realtime_start,
## realtime_end
```

```
## Using column `date` for date_var.Using column `realtime_start` for date_var.Using column `realtime_e
```

```
Unempl = tk_zoo_(Unempl)
```

```
## Warning: Non-numeric columns being dropped: date, series_id, realtime_start,
## realtime_end
```

```
## Using column `date` for date_var.Using column `realtime_start` for date_var.Using column `realtime_e
```

```
data = merge(Claims,Unempl)
colnames(data)= c("Claims", "Unempl")
data["2023-03-01"]
```

```
##            Claims Unempl
## 2023-03-01 242000    3.5
```

The unemployment rate in March 2023 is 3.5%.

iv.

The initial claims number is a flow variable that measures the number of individuals who have filed for unemployment benefits in a given week or month. In contrast, the unemployment rate is a stock variable that measures the total number of individuals who are currently unemployed at a particular point in time. Using the initial claims number to predict changes in the unemployment rate is a good strategy because it captures the inflow of newly unemployed workers into the labor market, which can provide insight into the health of the labor market. By analyzing the change in the initial claims number over time, economists can gain a sense of whether the labor market is improving or deteriorating, and whether the unemployment rate is likely to increase or decrease in the future. The initial claims number can also serve as a leading indicator of changes in the unemployment rate, making it a useful tool for forecasting. Overall, focusing on the flow variable of initial claims rather than the stock variable of unemployment numbers can provide a better understanding of the direction of the labor market and help predict changes in the unemployment rate more accurately.

v. ADL models

```r
# ADL(0,1) model
model01 <- dynlm(diff(Unempl) ~ L(Claims,0:1), data=data)
vcov(model01)
```

```
##                    (Intercept) L(Claims, 0:1)1 L(Claims, 0:1)2
## (Intercept)      9.156368e-04   -8.083010e-10   -8.142850e-10
## L(Claims, 0:1)1 -8.083010e-10    8.107973e-15   -5.971890e-15
## L(Claims, 0:1)2 -8.142850e-10   -5.971890e-15    8.112648e-15
```

The given model is an ADL(0,1) model with two lagged values of jobless claims and no lags of unemployment rate. The coefficients for both lagged jobless claims are statistically significant at the 5% level, but the intercept is not. The R-squared value is quite low, suggesting that the model does not explain a large portion of the variation in the unemployment rate.

```r
# ADL(0,2) model
model02 <- dynlm(diff(Unempl) ~ L(Claims,0:2), data=data)
vcov(model02)
```

```
##                    (Intercept) L(Claims, 0:2)1 L(Claims, 0:2)2 L(Claims, 0:2)3
## (Intercept)      6.651663e-04   -6.628870e-10    8.207893e-11   -6.671685e-10
## L(Claims, 0:2)1 -6.628870e-10    5.428271e-15   -4.887595e-15    1.208715e-15
## L(Claims, 0:2)2  8.207893e-11   -4.887595e-15    9.560221e-15   -4.888008e-15
## L(Claims, 0:2)3 -6.671685e-10    1.208715e-15   -4.888008e-15    5.432714e-15
```

The ADL(0,2) model includes only the lagged values of the Claims variable up to order 2. The coefficients for the intercept and the lagged variables are statistically significant.

```r
# ADL(0,3) model
model03 <- dynlm(diff(Unempl) ~ L(Claims,0:3), data=data)
vcov(model03)
```

```
##                    (Intercept) L(Claims, 0:3)1 L(Claims, 0:3)2 L(Claims, 0:3)3
## (Intercept)      7.020030e-04   -5.205210e-10   -1.536878e-10   -1.505698e-10
## L(Claims, 0:3)1 -5.205210e-10    5.665374e-15   -5.411314e-15    2.515478e-15
## L(Claims, 0:3)2 -1.536878e-10   -5.411314e-15    1.049161e-14   -7.195119e-15
## L(Claims, 0:3)3 -1.505698e-10    2.515478e-15   -7.195119e-15    1.049095e-14
## L(Claims, 0:3)4 -5.279006e-10   -1.394098e-15    2.517576e-15   -5.411874e-15
##                 L(Claims, 0:3)4
## (Intercept)       -5.279006e-10
## L(Claims, 0:3)1   -1.394098e-15
## L(Claims, 0:3)2    2.517576e-15
## L(Claims, 0:3)3   -5.411874e-15
## L(Claims, 0:3)4    5.671638e-15
```

The intercept coefficient estimate is 0.000702003, which represents the expected value of the response variable (Unempl) when all predictor variables are equal to zero.

The coefficient estimates for the lagged Claims variables are negative, which indicates that an increase in the Claims variable in the previous period is associated with a decrease in the Unempl variable in the current period.

```
# ADL(1,1) model
model11 <- dynlm(diff(Unempl) ~ L(diff(Unempl),1) + L(Claims,0:1), data=data)
vcov(model11)
```

```
##                       (Intercept) L(diff(Unempl), 1) L(Claims, 0:1)1
## (Intercept)          1.049957e-03       5.914212e-04   -2.748785e-10
## L(diff(Unempl), 1)  5.914212e-04       1.481809e-03    1.110040e-09
## L(Claims, 0:1)1    -2.748785e-10       1.110040e-09    8.031452e-15
## L(Claims, 0:1)2    -1.780642e-09      -2.649557e-09   -7.287842e-15
##                     L(Claims, 0:1)2
## (Intercept)           -1.780642e-09
## L(diff(Unempl), 1)   -2.649557e-09
## L(Claims, 0:1)1      -7.287842e-15
## L(Claims, 0:1)2       1.194160e-14
```

The ADL(1,1) model has two lagged variables, one for the unemployment rate and one for initial claims. The intercept represents the expected value of the dependent variable when both the unemployment rate and initial claims are zero.

The coefficient of L(diff(Unempl), 1) is positive, indicating that an increase in the unemployment rate in the previous period is associated with an increase in the dependent variable in the current period, holding initial claims constant.

The coefficient of L(Claims, 0:1)1 is negative, indicating that an increase in initial claims in the current period is associated with a decrease in the dependent variable in the current period, holding the unemployment rate constant.

The coefficient of L(Claims, 0:1)2 is also negative, indicating that an increase in initial claims in the previous period is associated with a decrease in the dependent variable in the current period, holding the unemployment rate constant.

```
# ADL(1,2) model
model12 <- dynlm(diff(Unempl) ~ L(diff(Unempl),1) + L(Claims,0:2), data=data)
vcov(model12)
```

```
##                       (Intercept) L(diff(Unempl), 1) L(Claims, 0:2)1
## (Intercept)          7.102976e-04       4.552551e-04   -1.063215e-10
## L(diff(Unempl), 1)  4.552551e-04       2.459648e-03    2.796311e-09
## L(Claims, 0:2)1    -1.063215e-10       2.796311e-09    8.287981e-15
## L(Claims, 0:2)2    -1.450658e-09      -8.254969e-09   -1.398492e-14
## L(Claims, 0:2)3     1.655211e-10       4.286791e-09    6.011150e-15
##                     L(Claims, 0:2)2 L(Claims, 0:2)3
## (Intercept)           -1.450658e-09    1.655211e-10
## L(diff(Unempl), 1)   -8.254969e-09    4.286791e-09
## L(Claims, 0:2)1      -1.398492e-14    6.011150e-15
## L(Claims, 0:2)2       3.670278e-14   -1.898760e-14
## L(Claims, 0:2)3      -1.898760e-14    1.258433e-14
```

The ADL(1,2) model suggests that past values of the first difference of unemployment and the second lag of claims have a significant impact on current unemployment. Specifically, an increase in the first difference of unemployment by 1 unit leads to an increase in current unemployment by 0.455 unit, while an increase in the second lag of claims by 1 unit leads to a decrease in current unemployment by 1.451 units.

```r
# ADL(1,3) model
model13 <- dynlm(diff(Unempl) ~ L(diff(Unempl),1) + L(Claims,0:3), data=data)
vcov(model13)
```

```
##                       (Intercept) L(diff(Unempl), 1) L(Claims, 0:3)1
## (Intercept)          6.079860e-04       3.324673e-04   -1.200116e-10
## L(diff(Unempl), 1)   3.324673e-04       3.249893e-03    2.987054e-09
## L(Claims, 0:3)1     -1.200116e-10       2.987054e-09    7.377617e-15
## L(Claims, 0:3)2     -1.110680e-09      -9.628653e-09   -1.327434e-14
## L(Claims, 0:3)3      1.751419e-10       2.915428e-09    4.736354e-15
## L(Claims, 0:3)4     -1.369212e-10       2.880741e-09    1.507911e-15
##                     L(Claims, 0:3)2 L(Claims, 0:3)3 L(Claims, 0:3)4
## (Intercept)           -1.110680e-09    1.751419e-10   -1.369212e-10
## L(diff(Unempl), 1)    -9.628653e-09    2.915428e-09    2.880741e-09
## L(Claims, 0:3)1       -1.327434e-14    4.736354e-15    1.507911e-15
## L(Claims, 0:3)2        3.710558e-14   -1.452061e-14   -6.476516e-15
## L(Claims, 0:3)3       -1.452061e-14    1.119303e-14   -1.840607e-15
## L(Claims, 0:3)4       -6.476516e-15   -1.840607e-15    7.190786e-15
```

The ADL(1,3) model suggests that changes in unemployment and up to 3 lagged values of claims are statistically significant predictors of the current value of claims. Specifically, an increase in the previous period's change in unemployment is associated with an increase in the current period's claims, while an increase in lagged claims is associated with a decrease in current claims.

vi.

To estimate the regression models on the first subset of data and compute the forecasted values on the second subset, we need to first create the two subsets of data as follows:

```r
# Subset data into two periods
data1 <- window(data, start = "1990-01-01", end = "2005-12-01")
data2 <- window(data, start = "2006-01-01")
```

```r
# ADL(0,1)
model1 <- dynlm(diff(Unempl) ~ L(Claims, 0:1), data = data1)
forecast1 <- model1$coef[1] + model1$coef[2] * lag(data2$Claims, 0)
```

```r
# ADL(0,2)
model2 <- dynlm(diff(Unempl) ~ L(Claims, 0:2), data = data1)
forecast2 <- model2$coef[1] + model2$coef[2] * lag(data2$Claims, 0) +
            model2$coef[3] * lag(data2$Claims, 1)
```

```r
# ADL(0,3)
model3 <- dynlm(diff(Unempl) ~ L(Claims, 0:3), data = data1)
forecast3 <- model3$coef[1] + model3$coef[2] * lag(data2$Claims, 0) +
            model3$coef[3] * lag(data2$Claims, 1) +
            model3$coef[4] * lag(data2$Claims, 2)
```

```r
# ADL(1,1)
model4 <- dynlm(diff(Unempl) ~ L(diff(Unempl), 1) + L(Claims, 0:1), data = data1)
forecast4 <- model4$coef[1] + model4$coef[2]*lag(data2[, "Unempl"], 1) +
  model4$coef[3]*lag(data2$Claims, 0) + model4$coef[4]*lag(data2$Claims, 1)
```

```r
# ADL(1,2)
model5 <- dynlm(diff(Unempl) ~ L(diff(Unempl), 1) + L(Claims, 0:2), data = data1)
forecast5 <- model5$coef[1] + model5$coef[2]*lag(data2[, "Unempl"], 1) +
```

```r
  model5$coef[3]*lag(data2$Claims, 0) + model5$coef[4]*lag(data2$Claims, 1) +
  model5$coef[5]*lag(data2$Claims, 2)
```

```r
# ADL(1,3)
model6 <- dynlm(diff(Unempl) ~ L(diff(Unempl), 1) + L(Claims, 0:3), data = data1)
forecast6 <- model6$coef[1] + model6$coef[2]*lag(data2[, "Unempl"], 1) +
  model6$coef[3]*lag(data2$Claims, 0) + model6$coef[4]*lag(data2$Claims, 1) +
  model6$coef[5]*lag(data2$Claims, 2) + model6$coef[6]*lag(data2$Claims, 3)
```

vii.

```r
# compute the RMSFE for each model
RMSFE1 <- sqrt(mean((data2$Unempl - forecast1)^2))
RMSFE2 <- sqrt(mean((data2$Unempl - forecast2)^2))
RMSFE3 <- sqrt(mean((data2$Unempl - forecast3)^2))
RMSFE4 <- sqrt(mean((data2$Unempl - forecast4)^2))
RMSFE5 <- sqrt(mean((data2$Unempl - forecast5)^2))
RMSFE6 <- sqrt(mean((data2$Unempl - forecast6)^2))

# display the RMSFE for each model
cat("RMSFE for ADL(0,1) model: ", RMSFE1, "\n")
```

```
## RMSFE for ADL(0,1) model:  NA
```

```r
cat("RMSFE for ADL(0,2) model: ", RMSFE2, "\n")
```

```
## RMSFE for ADL(0,2) model:  NA
```

```r
cat("RMSFE for ADL(0,3) model: ", RMSFE3, "\n")
```

```
## RMSFE for ADL(0,3) model:  NA
```

```r
cat("RMSFE for ADL(1,1) model: ", RMSFE4, "\n")
```

```
## RMSFE for ADL(1,1) model:  NA
```

```r
cat("RMSFE for ADL(1,2) model: ", RMSFE5, "\n")
```

```
## RMSFE for ADL(1,2) model:  NA
```

```r
cat("RMSFE for ADL(1,3) model: ", RMSFE6, "\n")
```

```
## RMSFE for ADL(1,3) model:  NA
```