ECN702 - Assignment 4

2023-04-09

ECN702 - Econometrics II

Professor Hyunhak Kim

Name: Ha Doan

Student ID: 501026801

Question a.

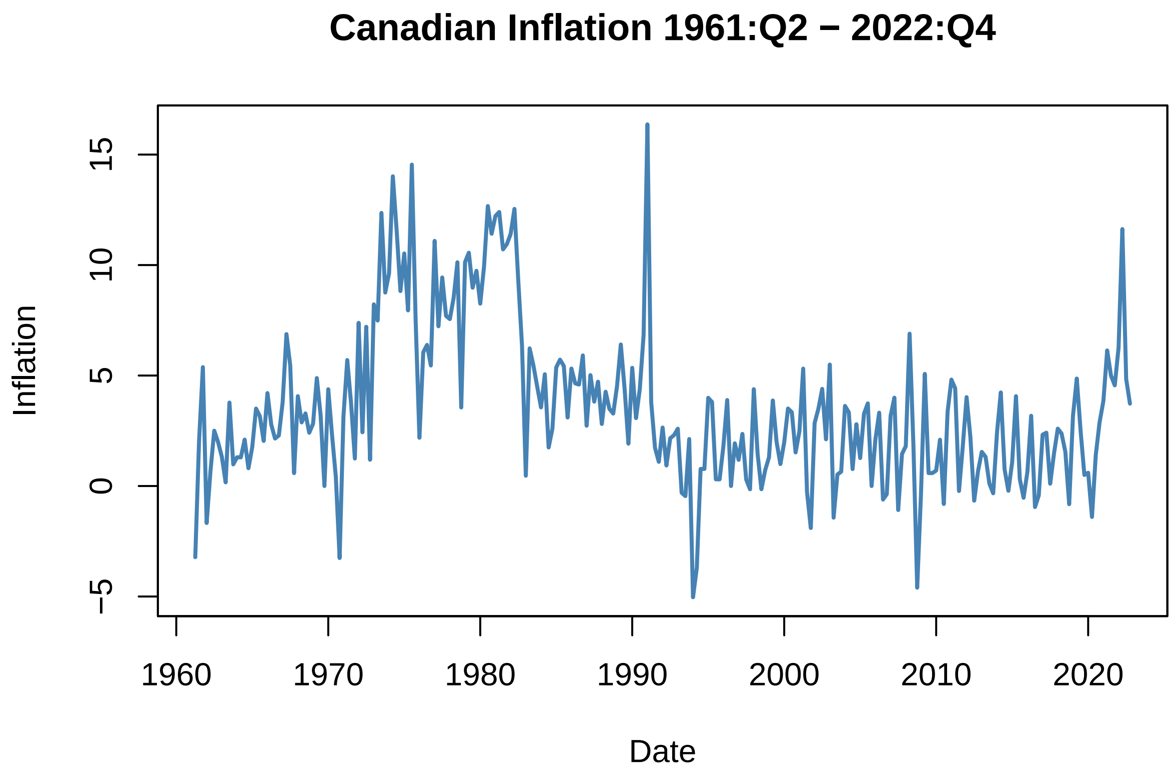
(i) Compute the inflation rate, $Infl$. What are the units of $Infl$?

```
CPI <- ts(df$CPI, start = c(1961,1), frequency = 4)
Infl <- diff(log(CPI),1)*400
```

The unit of $Infl$ is percentage per year. Because it is calculated by lagging the CPI index with its previous period, which is in percentage. If $100 * ln\Delta CPI$ is quarterly percentage change, then $400 * ln\Delta CPI$ is annually percentage year.

(ii) Plot the value of $Infl$ for whole sample period. Based on the plot, do you think that $Infl$ has a stochastic trend? Explain.

```
plot(as.zoo(Infl), col = "Steelblue", lwd = 2,
     ylab = "Inflation", xlab = "Date", main = "Canadian Inflation 1961:Q2 - 2022:Q4")
```



Based on the plot, $Infl$ might have a stochastic trend. It shows an upward trend from 1961-1981, and a downward after 1981. However, to determine whether there is a stochastic trend in $Infl$ or not needs to use the ADF test.

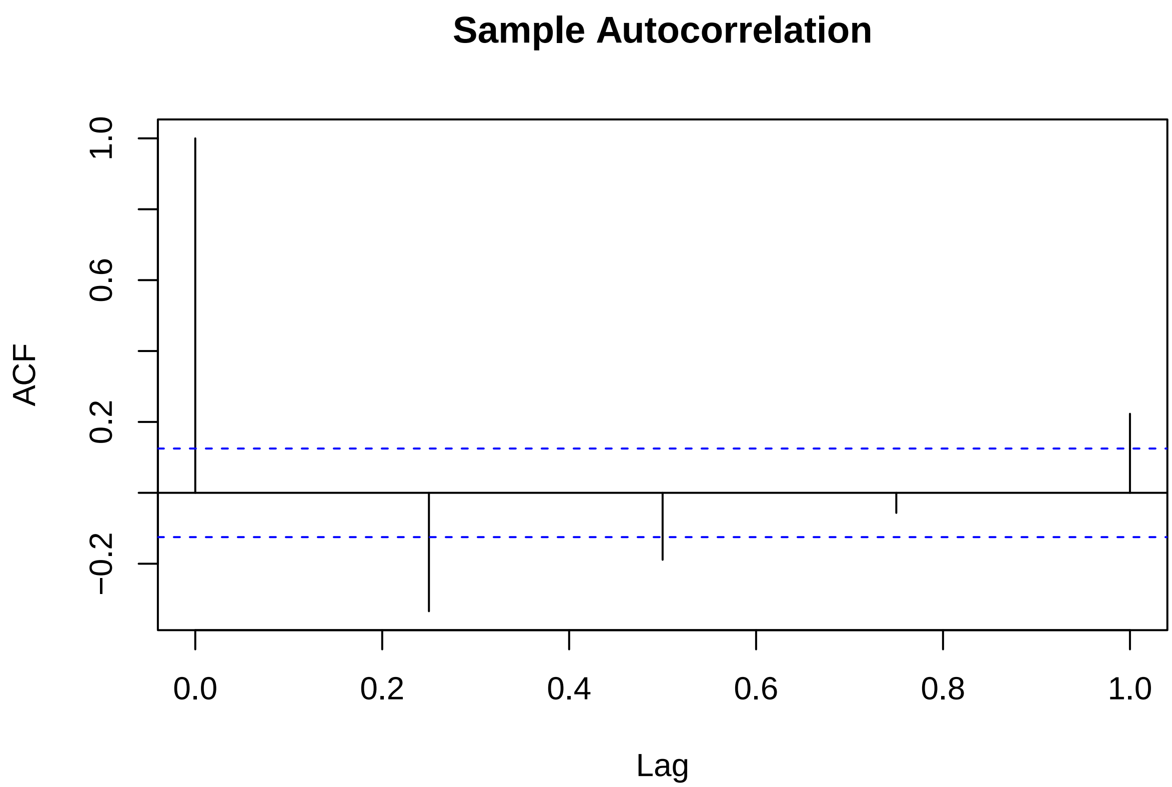
Assuming that $Infl$ has a trend, we compute a lag for $Infl$ which is $\Delta Infl$.

```
delta_Infl = diff(Infl)
```

Question b.

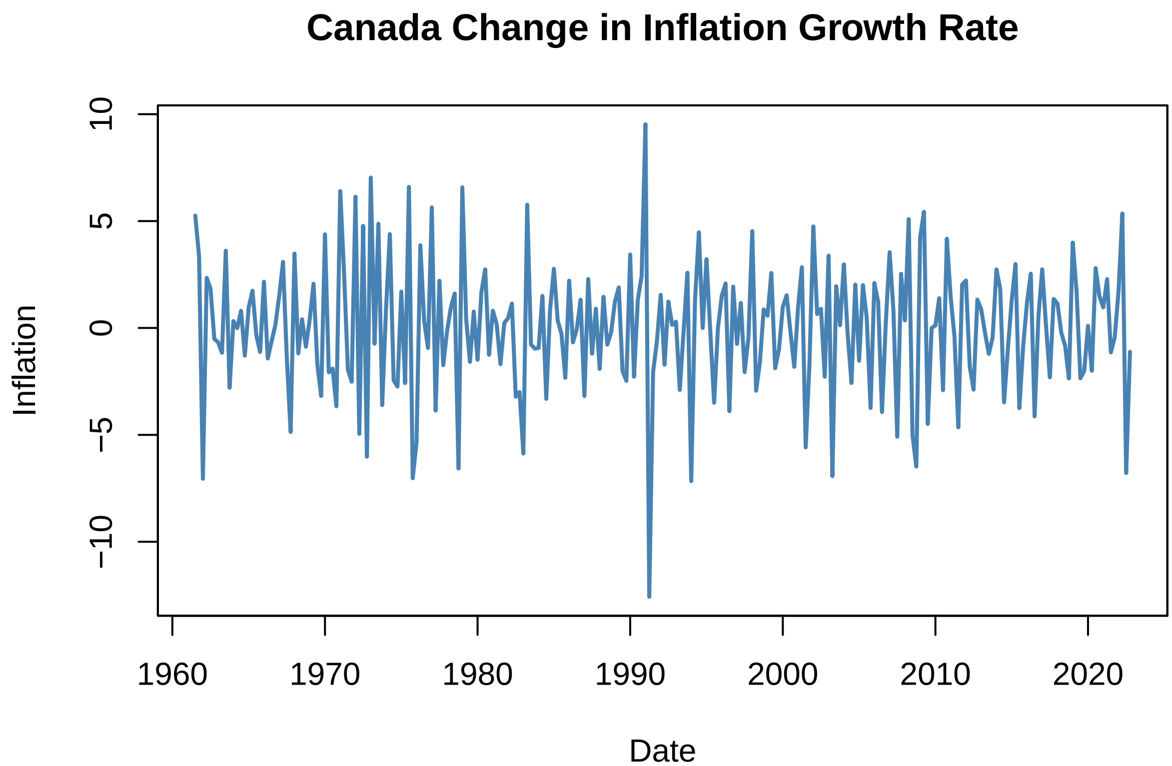
(i) Compute the first four autocorrelation of $\Delta Infl$.

```
acf(na.omit(delta_Infl), lag.max = 4, main = "Sample Autocorrelation")
```



(ii) Plot the value of $\Delta Infl$ for whole sample periods. The plot should look choppy and jagged. Explain why this behavior is consistent with the first autocorrelation that you computed in (i).

```
plot(as.zoo(delta_Infl), col = "steelblue", lwd = 2,
     ylab = "Inflation", xlab = "Date",
     main = "Canada Change in Inflation Growth Rate")
```



Based on (i), the ACF graph shows that the first three lags are significant. The ACF plot shows coefficients of correlation between a time series and it lagged values. Thus, as the first lag has the largest correlation based on the ACF plot, When plotting $\Delta Infl$, we eliminate the trends in $Infl$, so that the plot becomes choppy and jagged (stationary).

Question c.

(i) Run an OLS regression of $\Delta Infl_t$ on $\Delta Infl_{t-1}$. Does knowing the change in inflation over the current quarter help predict the change in inflation over the next quarter? Explain.

```
ar1<- dynlm(delta_Infl ~ L(delta_Infl))
summary(ar1)
```

```
##
## Time series regression with "ts" data:
## Start = 1961(4), End = 2022(4)
##
## Call:
## dynlm(formula = delta_Infl ~ L(delta_Infl))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.4087 -1.5191  0.0969  1.6078 10.3260
##
## Coefficients:
```

```
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.01790    0.18014   0.099    0.921
## L(delta_Infl) -0.33425    0.06006  -5.566 6.89e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.82 on 243 degrees of freedom
## Multiple R-squared:  0.1131, Adjusted R-squared:  0.1094
## F-statistic: 30.98 on 1 and 243 DF,  p-value: 6.889e-08
```

The SER for this model is 2.8195, which implies Leaving aside forecast uncertainty due to estimation of the model coefficients, the RMFSE must be at least 2.82% the estimate of the standard deviation of the errors. The model only explains only little variation in the change of inflation. Thus, this forecast model is not really accurate.

(ii) Estimate AR(2) model for $\Delta Infl$. Is the AR(2) model better than an AR(1) model? Explain.

```r
ar2 <- dynlm(delta_Infl ~ L(delta_Infl) + L(delta_Infl, 2))
summary(ar2)
```

```
##
## Time series regression with "ts" data:
## Start = 1962(1), End = 2022(4)
##
## Call:
## dynlm(formula = delta_Infl ~ L(delta_Infl) + L(delta_Infl, 2))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.4303 -1.5199  0.0907  1.4993 11.1190
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)        0.02017    0.16839   0.120    0.905
## L(delta_Infl)     -0.46769    0.06002  -7.793 1.96e-13 ***
## L(delta_Infl, 2) -0.35437    0.06028  -5.879 1.37e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.63 on 241 degrees of freedom
## Multiple R-squared:  0.2309, Adjusted R-squared:  0.2245
## F-statistic: 36.17 on 2 and 241 DF,  p-value: 1.826e-14
```

The AR(2) is improved compared to the AR(1). The standard error is smaller, and the R-squared is increased to 23.09%. Both two lags are statistically significant in forecasting the change in inflation rate.

(iii) Estimate AR(p) model for p = 0, ... ,8. What leg length is chosen by BIC? What leg length is chosen by AIC?

```r
order <- 1:8

# generate AICs and BICs
```

```r
AICs <- sapply(order, function(x)
        "AR" = AIC(dynlm(delta_Infl ~ L(delta_Infl, 1:x))))
BICs <- sapply(order, function(x)
        "AR" = BIC(dynlm(delta_Infl ~ L(delta_Infl, 1:x))))

aic_bic = data.frame(rbind(AICs, BICs))
colnames(aic_bic) = 1:8
rownames(aic_bic) = c("AIC", "BIC")
stargazer(t(aic_bic), type = "text")
```

```
##
## =====================
##      AIC       BIC
## ---------------------
## 1 1,207.192 1,217.696
## 2 1,169.251 1,183.240
## 3 1,136.570 1,154.035
## 4 1,134.422 1,155.356
## 5 1,131.357 1,155.750
## 6 1,129.001 1,156.846
## 7 1,123.118 1,154.406
## 8 1,116.817 1,151.540
## ---------------------
```

The BIC chooses AR(3), while the AIC chooses AR(8). For BIC, even though AR(8) has the smallest estimate, the estimate reduces until AR(3) and increases from AR(4). For AIC, the smallest estimate is in AR(8).

(iv) Use the AR(2) model to predict the change in inflation from 2022:Q4 to 2023:Q1 – that is, to predict the value of $\Delta Inflation_{2023:Q1}$.

```r
N <- length(delta_Infl)
delta_Infl[N] # delta_2022Q4
```

```
## [1] -1.116633
```

```r
delta_Infl[N-1] #delta_2022Q3
```

```
## [1] -6.78533
```

```r
forecast <- c("2023:Q1" = coef(ar2) %*% c(delta_Infl[N], delta_Infl[N-1], 1))
```

- $\Delta Inflation_{2023:Q1} = 2.7965$

(V) Use the AR(2) model to predict the change in inflation rate in 2023:Ql – that is, to predict the value of $Inflation_{2023:Q1}$.

```r
n <- length(Infl)
Infl[n] #2022Q4
```

```
## [1] 3.727562
```

```r
q1_23 = forecast + Infl[n]
```

- $Inflation_{2023:Q1} = 6.524$

Question d.

(i) Use the ADF test for the regression like the following with two lags of $\Delta Infl$ (so that p = 3) to test for a stochastic trend in $Infl$.

```r
coeftest(dynlm(diff(Infl) ~ L(Infl) + diff(L(Infl)) + diff(L(Infl), 2)))
```

```
##
## t test of coefficients:
##
##                    Estimate Std. Error t value  Pr(>|t|)
## (Intercept)        0.721189   0.251165  2.8714 0.0044521 **
## L(Infl)           -0.196726   0.053346 -3.6877 0.0002799 ***
## diff(L(Infl))     -0.063699   0.068522 -0.9296 0.3535033
## diff(L(Infl), 2) -0.280343   0.062100 -4.5144 9.956e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

So the ADF test statistic is $t = -3.69$. At a 1% significance level, the critical value is -3.43. Thus, we reject the null hypothesis that there is a unit root. Thus, there is no stochastic trend in this model.

(ii) Is the ADF test based on the above equation preferred to the test based on the below equation for testing for a stochastic trend in $Infl$? Explain.

```r
coeftest(dynlm(diff(Infl) ~ trend(Infl, scale = F) + L(Infl) + diff(L(Infl)) + diff(L(Infl), 2)))
```

```
##
## t test of coefficients:
##
##                           Estimate Std. Error t value  Pr(>|t|)
## (Intercept)              1.2766962  0.4664305  2.7372  0.006663 **
## trend(Infl, scale = F) -0.0035497  0.0025136 -1.4122  0.159191
## L(Infl)                -0.2276039  0.0575513 -3.9548  0.000101 ***
## diff(L(Infl))          -0.0561511  0.0685888 -0.8187  0.413794
## diff(L(Infl), 2)       -0.2684492  0.0625419 -4.2923 2.571e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

With the time trend included in the model, $t = -3.954$ and the critical value is -3.96 at a 1% significance level. There might be an evidence of a stochastic trend in this model.

(iii) In (i), you used two lags of $\Delta Infl$. Should you use more lags? Fewer lags? Explain. In (i), because there is no evidence of a unit root at a 1% significance level, there is no need to investigate more lags in the model. However, fewer lags (p = 1) can be considered as the plot for $\Delta Infl$ show a stationary pattern.

(iv) Based on the test you carried out in (i), does the AR model for $Infl$ contain a unit root? Explain carefully. (Hint: Does the failure to reject a null hypothesis mean that the null hypothesis is true?)

```
TSA <- adf.test(diff(Infl), nlag = 2)
```

```
## Augmented Dickey-Fuller Test
## alternative: stationary
##
## Type 1: no drift no trend
##      lag   ADF p.value
## [1,]   0 -22.3    0.01
## [2,]   1 -18.5    0.01
## Type 2: with drift no trend
##      lag   ADF p.value
## [1,]   0 -22.2    0.01
## [2,]   1 -18.5    0.01
## Type 3: with drift and trend
##      lag   ADF p.value
## [1,]   0 -22.2    0.01
## [2,]   1 -18.4    0.01
## ----
## Note: in fact, p.value = 0.01 means p.value <= 0.01
```

Based on (i), as we reject the null hypothesis that there is a unit root in AR(2) ($H_0 : \delta = 0$). We can say that $Infl$ is stationary. However, in (ii), at a 1% significance level, there is some evidence of a unit root around a time trend.

Question e.

Use the QLR test with 15% trimming to test the stability of the coefficients in the AR(2) model for $\Delta Infl$. Is the AR(2) model stable? Explain.

```
quantile(time(delta_Infl), 0.15)
```

```
##      15%
## 1970.688
```

```
quantile(time(delta_Infl), 0.85)
```

```
##      85%
## 2013.562
```

```
# set up a range of possible break dates
tau <- seq(1970.75, 2013.5, 0.25)

# initialize vector of F-statistics
Fstats <- numeric(length(tau))

# estimation loop over break dates
for(i in 1:length(tau)) {
```

```
  # set up dummy variable
  D <- time(delta_Infl) > tau[i]

  # estimate AR(2) model with interactions
  test <- dynlm(delta_Infl ~ L(delta_Infl) + L(delta_Infl, 2)
               + D*L(delta_Infl) + D*L(delta_Infl, 2),
               start = c(1962, 1), end = c(2020, 4))

  # compute and save the F-statistic
  Fstats[i] <- linearHypothesis(test, c("L(delta_Infl)", "L(delta_Infl, 2)"),
                              vcov. = sandwich)$F[2]

}

# identify QLR statistic
QLR <- max(Fstats)
QLR
```

```
## [1] 18.96841
```

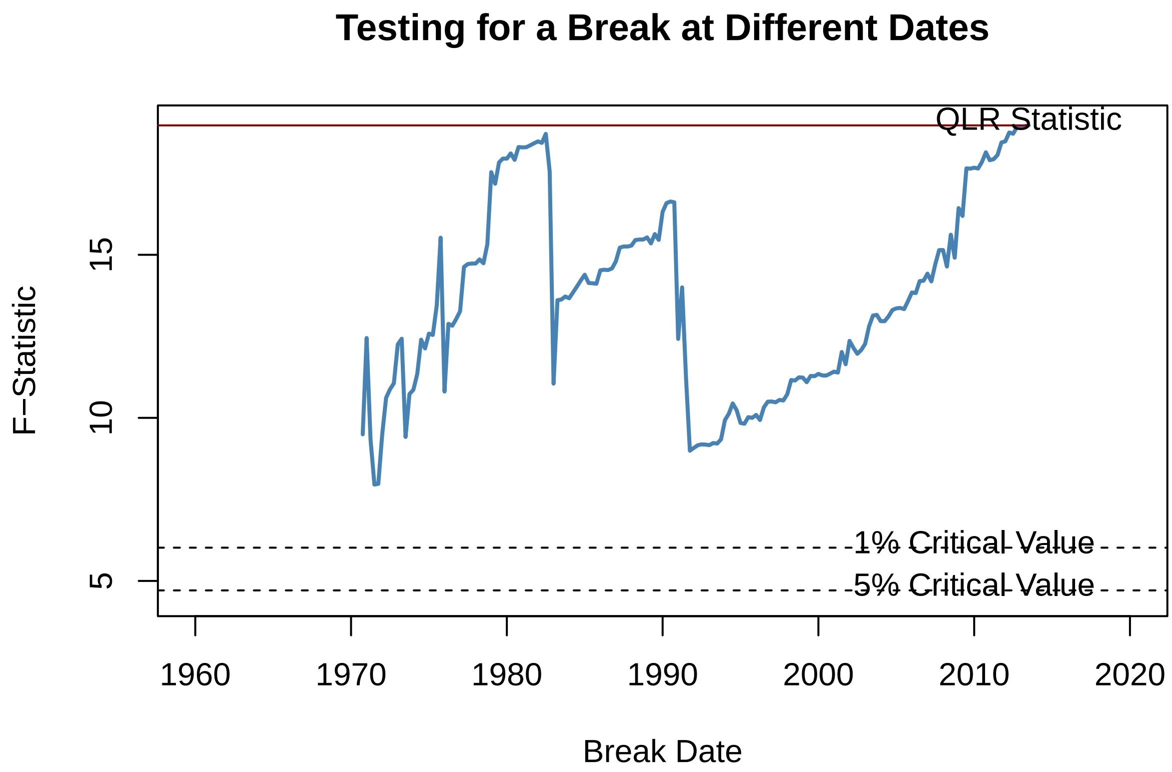
```
# identify the time period where the QLR-statistic is observed
as.yearqtr(tau[which.max(Fstats)])
```

```
## [1] "2013 Q3"
```

```
# series of F-statistics
Fstatsseries <- ts(Fstats, start = tau[1], end = tau[length(tau)], frequency = 4)

# plot the F-statistics
plot(Fstatsseries, xlim = c(1960, 2020), ylim = c(4.5,19), lwd = 2,
     col = "steelblue", ylab = "F-Statistic", xlab = "Break Date",
     main = "Testing for a Break at Different Dates")

# dashed horizontal lines for critical values and QLR statistic
abline(h = 4.71, lty = 2)
abline(h = 6.02, lty = 2)
segments(0, QLR, 2013.5, QLR, col = "darkred")
text(2010, 6.2, "1% Critical Value")
text(2010, 4.9, "5% Critical Value")
text(2013.5, QLR + 0.2, "QLR Statistic")
```



We reject the null hypothesis that all coefficients (the coefficients on both lags of term spread and the intercept) are stable since the computed QLR-statistic exceeds this threshold. Thus evidence from the QLR test suggests that there is a break in the AR(2) model of $\Delta Infl$ in the third quarter of 2013.

Question f.

(i) Using the AR(2) model for $\Delta Infl$ with a sample period that begins in 1980:Q1, compute pseudo out-of-sample forecast for the change in inflation beginning in 2008:Q1 and going through 2022:Q4.

```
# end of sample dates
EndOfSample <- seq(2007.75, 2022.5, 0.25)

# initialize matrix  forecasts
forecasts <- matrix(nrow = 1,
                    ncol = length(EndOfSample))

# estimation loop over end of sample dates
for(i in 1:length(EndOfSample)) {

  # estimate AR(2) model
  m_ar2 <- dynlm(delta_Infl ~ L(delta_Infl) + L(delta_Infl, 2),
                start = c(1980, 1),
                end = EndOfSample[i])

  SER <- summary(m_ar2)$sigma
```

```r
  # sample data for one-period ahead forecast
  s <- window(delta_Infl, EndOfSample[i] - 0.25, EndOfSample[i])

  # compute forecast
  forecasts[i] <- coef(m_ar2) %*% c(1, s[1], s[2])
}

# compute pseudo-out-of-sample forecast errors
poosfe <- c(window(delta_Infl, c(2008,1), c(2022, 4))) - forecasts
```

(ii) Are the pseudo out-of-sample forecasts biased? That is, do the forecast errors have a nonzero mean?

```r
t.test(poosfe)
```

```
##
##  One Sample t-test
##
## data:  poosfe
## t = 0.69337, df = 59, p-value = 0.4908
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -0.4032387  0.8308790
## sample estimates:
## mean of x
## 0.2138202
```

We cannot reject the null hypothesis that the forecast error is zero. The `t.test` suggests that the AR(2) model coefficients are stable and unbiased.

(iii) How large is the RMSFE of the pseudo out-of-sample forecasts? Is this consistent with the AR(2) model for $\Delta Infl$ estimated over the sample period?

```r
sd(poosfe)
```

```
## [1] 2.38867
```

```r
summary(ar2)$sigma
```

```
## [1] 2.629675
```

The root mean squared forecast error of the pseudo-out-of-sample forecasts is somewhat smaller. Based on the plot, the pseudo forecasts track the actual change in inflation rate quite well. Thus, it can explain the consistency in between the RMSFE of the POOP forecast and the AR(2) model's SER.

```r
# series of pseudo-out-of-sample forecasts
PSOSSFc <- ts(c(forecasts), start = 2008, end = 2022.75, frequency = 4)

# plot the GDP growth time series
plot(window(delta_Infl, c(2008, 1), c(2022, 4)),
```

```r
     col = "steelblue", lwd = 2, ylab = "Percent",
     main = "Pseudo-Out-Of-Sample Forecasts")

# add the series of pseudo-out-of-sample forecasts
lines(PSOSSFc, lwd = 2, lty = 2)

# shade area between curves (the pseudo forecast error)
polygon(c(time(PSOSSFc), rev(time(PSOSSFc))),
        c(window(delta_Infl, c(2008, 1), c(2022, 4)), rev(PSOSSFc)),
        col = alpha("blue", alpha = 0.3), border = NA)

# add a legend
legend("bottomleft", lty = c(1, 2, 1), lwd = c(2, 2, 10),
       col = c("steelblue", "black", alpha("blue", alpha = 0.3)),
       legend = c("Actual rate", "Forecasted rate", "Pseudo forecast Error"))
```