Problem set 7

Due on Tuesday December 6, 2022 (by 11:59 PM)

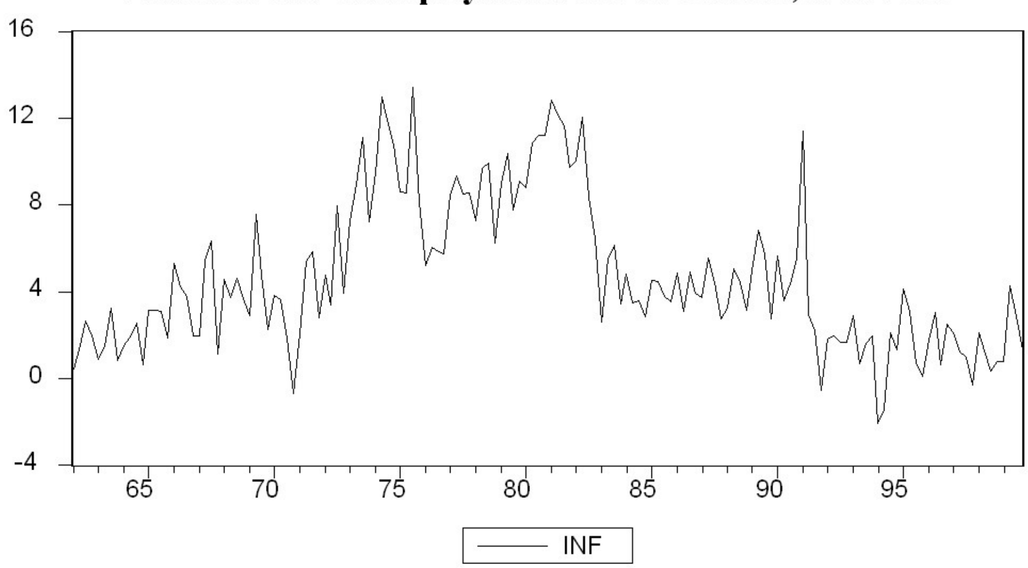
Note: No credit will be given if you report only the final answers without showing formulas and calculations when appropriate. This applies to both theoretical and empirical questions. For the empirical questions, make sure to upload the R script and output on Latte. No credit will be given if the R output is missing.

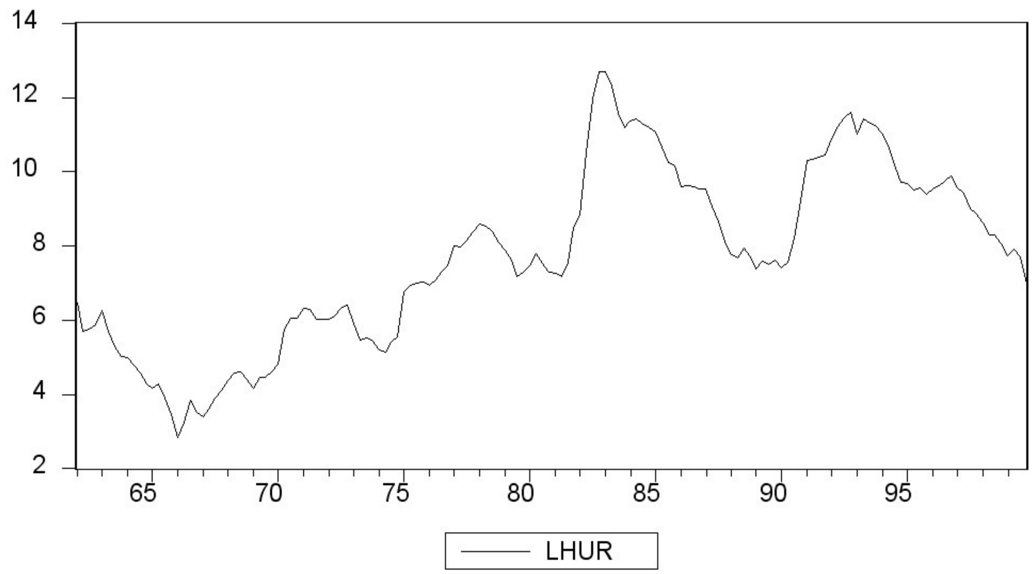
Problem 1

You have collected quarterly data on Canadian unemployment (*UrateC*) and inflation (*InfC*) from 1962 to 1999 with the aim to forecast Canadian inflation.

(a) To get a better feel for the data, you first inspect the plots for the series.

Inflation and Unemployment Rate in Canada, 1962-1999





Inspecting the Canadian inflation rate plot and having calculated the first autocorrelation to be 0.79 for the sample period, do you suspect that the Canadian inflation rate has a stochastic trend? What more formal methods do you have available to test for a unit root?

(b) You run the following regression, where the numbers in parenthesis are homoskedasticity-only standard errors:

$$\widehat{\Delta InfC_t} = \underset{(0.28)}{0.49} - \underset{(0.05)}{0.10}\, Inf_{t-1} - \underset{(0.09)}{0.39}\, \Delta InfC_{t-1} - \underset{(0.09)}{0.33}\, \Delta InfC_{t-2} - \underset{(0.09)}{0.21}\, \Delta InfC_{t-3} + \underset{(0.08)}{0.05}\, \Delta InfC_{t-4}$$

Test for the presence of a stochastic trend. Should you have used heteroskedasticity-robust standard errors? Does the fact that you use quarterly data suggest including four lags in the above regression, or how should you determine the number of lags?

(c) To forecast the Canadian inflation rate for 2000:I, you estimate an AR(1), AR(4), and an ADL(4,1) model for the sample period 1962:I to 1999:IV. The results are as follows:

$$\widehat{\Delta InfC_t} = \underset{(0.014)}{0.002} - \underset{(0.10)}{0.31}\, \Delta InfC_{t-1}$$

$$\widehat{\Delta InfC_t} = \underset{(0.158)}{0.021} - \underset{(0.10)}{0.46}\, \Delta InfC_{t-1} - \underset{(0.11)}{0.39}\, \Delta InfC_{t-2} - \underset{(0.08)}{0.25}\, \Delta InfC_{t-3} + \underset{(0.07)}{0.03}\, \Delta InfC_{t-4}$$

$$\widehat{\Delta InfC_t} = \underset{(0.57)}{1.279} - \underset{(0.10)}{0.51}\, \Delta InfC_{t-1} - \underset{(0.11)}{0.44}\, \Delta InfC_{t-2} - \underset{(0.09)}{0.30}\, \Delta InfC_{t-3} - \underset{(0.08)}{0.02}\, \Delta InfC_{t-4}$$

$$- \underset{(0.07)}{0.16}\, UrateC_{t-1}$$

In addition, you have the following information on inflation in Canada during the four quarters of 1999 and the first quarter of 2000:

Inflation and Unemployment in Canada, First Quarter 1999 to First Quarter 2000

Quarter	Unemployment Rate ($UrateC_t$)	Rate of Inflation at an Annual Rate (Inf_t)	First Lag (Inf_{t-1})	Change in Inflation (ΔInf_t)
1999:I	7.7	0.8	0.8	0.0
1999:II	7.9	4.3	0.8	3.5
1999:III	7.7	2.9	4.3	-1.4
1999:IV	7.0	1.3	2.9	-1.5
2000:I	6.8	2.1	1.3	0.8

For each of the three models, calculate the predicted inflation rate for the period 2000:I and the forecast error.

(d) Perform a test on whether or not Canadian unemployment rates Granger-cause the Canadian inflation rate.

Problem 2

You want to determine whether or not the unemployment rate for the United States has a stochastic trend using the Augmented Dickey Fuller Test (ADF). The BIC suggests using 3 lags, while the AIC suggests 4 lags.

(a) Which of the two will you use for your choice of the optimal lag length?

(b) After estimating the appropriate equation, the *t*-statistic on the lag level unemployment rate is (–2.186) (using a constant, but not a trend). What is your decision regarding the stochastic trend of the unemployment rate series in the United States?

(c) Having worked in the previous exercise with the unemployment rate level, you repeat the exercise using the difference in United States unemployment rates. Write down the appropriate equation to conduct the Augmented Dickey-Fuller test here. The *t*-statistic on relevant coefficient turns out to be (-4.791). What is your conclusion now?

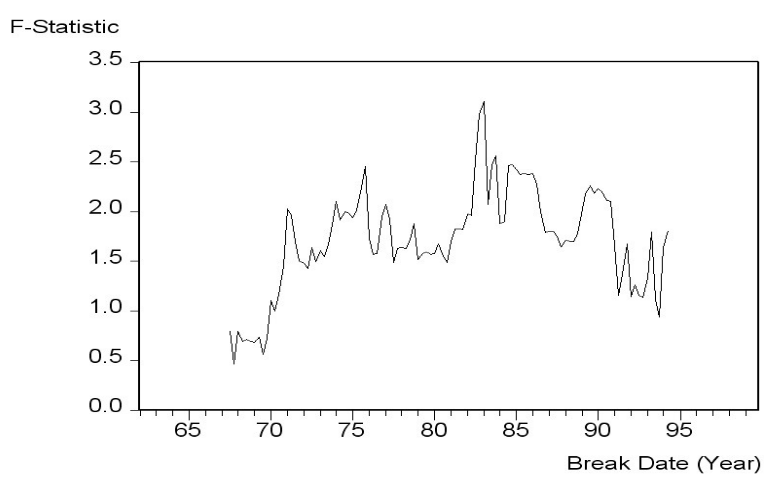
Problem 3

There is some evidence that the Phillips curve has been unstable during the 1962 to 1999 period for the United States, and in particular during the 1990s. You set out to investigate whether or not this instability also occurred in other places. Canada is a particularly interesting case, due to its proximity to the United States and the fact that many features of its economy are similar to that of the U.S.

(a) Reading up on some of the comparative economic performance literature, you find that Canadian unemployment rates (URCAN) were roughly the same as U.S. unemployment rates from the 1920s to the early 1980s. The accompanying figure shows that a gap opened between the unemployment rates of the two countries in 1982, which has persisted to the end of the sample.

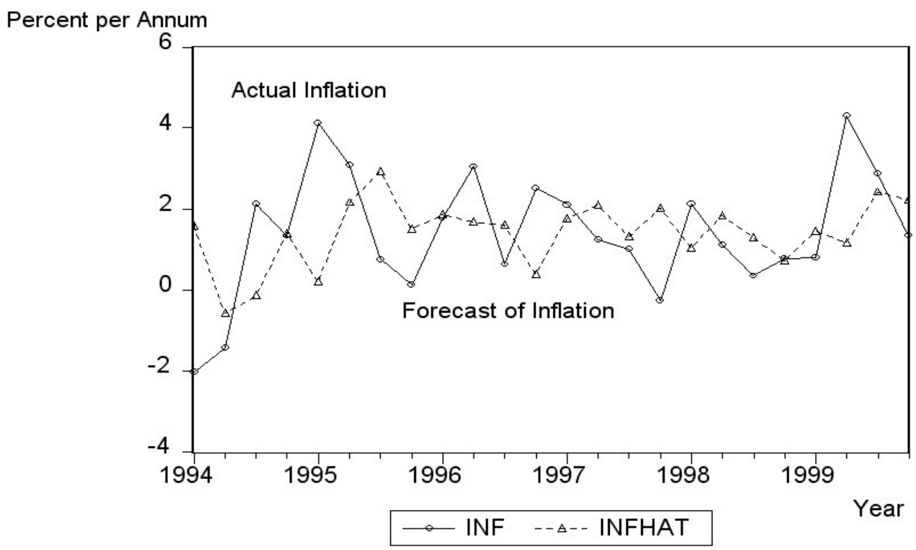
Inspection of the graph and data suggest that the break occurred during the second quarter of 1982. To investigate whether the Canadian Phillips curve shows a break at that point, you estimate an ADL(4,4) model for the sample period 1962:I-1999:IV and perform a Chow test. Specifically, you postulate that the constant and coefficients of the unemployment rates changed at that point. The F-statistic is 1.96. Find the critical value from the F-table and test the null hypothesis that a break occurred at that time. What do you conclude?

(b) You consider alternative ways to test for a break in the relationship. The accompanying figure shows the F-statistics testing for a break in the ADL(4,4) equation at different dates.



The QLR-statistic with 15% trimming is 3.11. Comment on the figure and test for the hypothesis of a break in the ADL(4,4) regression.

(c) To test for the stability of the Canadian Phillips curve in the 1990s, you decide to perform a pseudo out-of-sample forecasting. For the 24 quarters from 1994:I-1999:IV you use the ADL(4,4) model to calculate the forecasted change in the inflation rate, the resulting forecasted inflation rate, and the forecast error. The average forecast error for the 24 inflation rates is 0.003 and the sample standard deviation of the forecast errors is 0.82. Calculate the *t*-statistic and test the hypothesis that the mean out-of-sample forecast error is zero. Comment on the result and the accompanying figure of the actual (INF) and forecasted inflation rate (INFHAT).



Problem 4 (empirical)

In this exercise, you will conduct a Monte Carlo experiment to study the phenomenon of spurious regressions discussed in class. In a Monte Carlo study, artificial data are generated using a computer, and then those artificial data are used to calculate the statistics being studied. This makes it possible to compute the distribution of statistics for known models when mathematical expressions for those statistics are complicated or unknown. In this exercise, you will generate data so that the two time series Y_t and X_t are independently distributed random walks. The specific steps are as follows:

I. Use your computer to generate a sequence of $T = 100$ iid standard normal random variables. Call these variables $e_1, e_2, \ldots, e_{100}$. Set $Y_1 = e_1$ and $Y_t = Y_{t-1} + e_t$ for $t = 2, \ldots, 100$.
II. Use your computer to generate a new sequence of $T = 100$ iid standard normal random variables. Call these variables $a, a_2, \ldots, a_{100}$. Set $X_1 = a_1$ and $X_t = X_{t-1} + a_t$ for $t = 2, \ldots, 100$.
III. Regress $\mathrm{Y_t}$ onto a constant and X_t. Compute the OLS estimator, the regression R-squared and the homoskedastic only t statistics testing the null that the coefficient on X_t in this regression is 0

Now, use this algorithm to answer the following questions

a. Run the algorithm (I) through (III) once. Use the t-statistics from (III) to test the null that the coefficient on X_t is zero using the usual 5% critical value of 1.96. What is the R-squared of this regression?
b. Repeat step (a) above 1,000 times, saving each time the value of the R-squared and t-statistics. Now construct a histogram of the R-squared and t-statistics. What are the 5%, 50%, and 95% percentiles of the distributions of the R-squared and t-statistics? In what fraction of the 1,000 simulated data sets does the t-statistics exceed 1.96 in absolute value?
c. Now repeat (b) for different number of observations, like T=50 and T=200. As the sample size increases, does the fraction of times thar you reject the null hypothesis approach 5% as it should because you have generated *Y* and *X* as two independently distributed series? Does this fraction seem to approach some other limit as *T* gets large? If so, what is that limit?