

C29IE Introductory Econometrics
Written assignment (updated 18 Nov 2020 – see Q3)
Due date: Midnight, Friday 4 December
Submit electronic copies via Vision/TurnItIn

The assignment has 3 questions. **Complete all 3 questions.**

Question 1 (theory) [30%]

Argue why

$$\frac{1}{n} \times \frac{\frac{1}{n-2} \sum_{i=1}^n \hat{v}_i^2}{\left[\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \right]^2}$$

is a valid estimator for the variance of $\hat{\beta}_1$, the OLS estimate of β_1 in the linear model $Y_i = \beta_0 + \beta_1 X_i + u_i$.

What is \hat{v}_i in the above formula? Why are the means in the numerator and denominator divided by different scaling factors?

Question 2 (time series) [35%]

To complete this question you need to download the Stata do file from Vision, and then modify and run the do file in Stata. Further instructions are available on Vision.

The Phillips curve describes the relationship between the unemployment rate and inflation, measured either as wage or price inflation. Phillips found a negative relationship between the unemployment rate and inflation. In a nutshell, if the unemployment rate is low, workers have greater bargaining power, which in turn leads to wage increases and higher inflation.

In this exercise we look at the Phillips curve in the US for the period 1983-2008 using quarterly data. On Vision you will find a Stata do file which starts the exercise. However, parts of the do file are missing and you need to fill in the gaps (marked by "..."). The first part of the do file loads the dataset and converts it into the appropriate format and does not need to be altered. We download the data from FRED (Federal Reserve Economic Data) by using the command **freduse**. This is a user-written command and can be installed by saying, within Stata, **ssc install freduse**. The dataset contains the following data for the UK:

Variable	Description
Unemployment	Unemployment rate from 1983Q1 to 2007Q4
Inflation	Quarterly inflation rate from 1983Q1 to 2007Q4
quarter	Year/Quarter indicator (numeric variable)
date	Date (text variable)

Note that inflation for 2008Q1 is missing from the data. This is because you will be forecasting it below. Note also that these are quarterly rates of inflation (so, for example, a quarterly rate of 1% means an annual rate of about 4%).

The form of the Phillips curve you will be estimating is the following ARDL(p,q) model:

$$\Delta \ln f_t = \beta_0 + \beta_1 \Delta \ln f_{t-1} + \dots + \beta_p \Delta \ln f_{t-p} + \delta_1 \Delta Unem_{t-1} + \dots + \delta_q \Delta Unem_{t-q} + u_t$$

where β is a coefficient on a lagged change (first-difference) in inflation, δ is a coefficient on lagged unemployment, and there are up to p and q lags of each, respectively. The Phillips curve relationship suggests that the change in inflation should be negatively related to lagged unemployment, i.e., $\delta < 0$.

Questions:

- (a) Are the unemployment and inflation series stationary? Argue with the use of a graph and use the appropriate test.
- (b) Next we want to estimate an ARDL(p,q) model. To determine the optimal lag length, we use a generalised BIC for multivariate models. For simplicity we assume p=q. The result for the first 12 lags are:

Lags	BIC
1	-2.62
2	-2.65
3	-2.64
4	-3.22
5	-3.13
6	-3.07
7	-2.98
8	-2.99
9	-2.89
10	-3.00
11	-2.90
12	-2.80

Now estimate an ARDL(p,q) model taking into account what you found in part (a) and using the optimal lag length determined by the BIC.

- (c) Does the Phillips curve relationship hold for the US during the period 1983-2007? Comment on the results you obtained in part (b).
- (d) Explain how you would forecast inflation for 2008Q1 (not in the dataset!) and write down the equation for it. In Stata you can use the **predict** command (hint: see Lab 6 and/or the post-lab exercise for Lab 5 to find how to use **predict**). Construct a forecast interval around your forecast for 2008Q1 using the RSS (you can find the RSS in the regression output and is also displayed by the do file code) to obtain an estimate for the mean squared forecast error. How should you interpret your forecast and forecast interval?
- (e) In fact, inflation increased by 0.18 percentage points from 0.57% in 2007Q4 to 0.75% in 2008Q1. What do you conclude about your forecast?

Question 3 (causal inference) [35%]

This question is based on a 1995 study by Kiel and McCain in the *Journal of Environmental Economics and Management* on the impact of the construction of a municipal rubbish incinerator on house prices in North Andover, Massachusetts. In 1979, the local government started investigating the possibility of building an incinerator. Construction of the incinerator was completed in 1981. Kiel and McCain expect that houses near the incinerator could be negatively affected.

Kiel and McCain assemble data on the sales of houses in North Andover in 1978 and 1981. In addition to data on the sales prices of houses, they also collect data on the characteristics of the houses sold.

The variables in this dataset include:

lprice	The log of the house sales price (in US \$)
year	Year of the sale (1978 or 1981)
y81	Dummy variable, =1 if the house was sold in 1981, =0 if 1978.
rooms	The number of rooms in the house.
baths	The number of bathrooms in the house.
age	The age of the house in years.
larea	The log of the house floor area (square feet).
lland	The log of the house property area (square feet).
nearinc	Dummy variable, =1 if the house is near the new incinerator.

To do the assignment you should create and save your own do file. You do not need to submit the do file with your answer.

Start the do file with the following command in order to load the data. Update 18 Nov 2020: there are two ways to do this. You only need to use ONE method.

The first method is to load the data using the full URL. The second method is to load the data using the **bcuse** command. The **bcuse** command loads datasets stored at the Boston College data repository. The command is not part of official Stata but is very easy to install. Just type the following into Stata **once**:

```
ssc install bcuse
```

Installing this command illustrates how easy it is to add additional functionality to Stata. It will remain as part of your Stata installation the next time you start up Stata and want to load the **kielmc** dataset.

Method 1: load the data using the full URL.

```
use http://fmwww.bc.edu/ec-p/data/wooldridge/kielmc, clear
```

Method 2: use the **bcuse** command.

```
bcuse kielmc, clear
```

Reminder – your do file should include only ONE of the above methods.

- (a) First consider the impact of building the incinerator on the price of nearby houses sold in 1981, after the incinerator was built. A naïve researcher estimates a simple OLS regression of the log of the price of houses sold in 1981 on the dummy variable for location near in the incinerator. The estimation is based on the Stata command below. You can run this regression using the following command:

```
regress lprice nearinc if year==1981, rob
```

Report your results. Discuss the estimated impact on sales price of being near the incinerator. Explain why this OLS regression suffers from a severe omitted variables bias problem.

- (b) Now add as controls the variables for house characteristics described above: number of rooms, number of baths, log house floor and log house property area, and house age. Include a quadratic specification for age (i.e., both age and the square of age).

Report your results. Discuss the estimated impact of being near an incinerator on the sales price in this specification vs. the specification in part (a) with no control variables. (Hint: the Stata command to estimate this equation is just a modified version of the one for (a) above; see Lab 4 for how to add a quadratic in age.)

- (c) Next, consider using the approach of “differences-in-differences” to estimate the impact on house prices of building the incinerator. Use the following model based on data on house sales prices in 1978 (before the incinerator was planned) combined with the data used above on house sales prices in 1981:

$$\log(\text{price}) = \beta_1 + \beta_2 y81 + \beta_3 \text{near_incinerator} + \beta_4 (y81 \times \text{near_incinerator}) + u$$

where **y81** is the dummy variable for a 1981 house sale, and (**y81 × nearinc**) is the interaction of the year dummy and the near-to-the-incinerator dummy. You can run this regression using the following command:

```
regress lprice i.y81##i.nearinc, robust
```

Report your results and comment on them.

- (d) Finally, use the difference-in-differences approach as in (c) above, but augmented by the same set of control variables used in (b). Report your results and comment on them.