Introductory Econometrics: Computer Problem Sheet 3
Instructions and Sketch Answers

Excel is not well designed to work with time series regressions. You can do so, but it involves a lot of copying and pasting to, e.g., create lags of variables. The answers below were prepared by the tutor and offer detailed step by step instructions on how to set up your spreadsheet to do time series regressions for Exercise 1. I have not provided a similar step by step set of instructions for Exercise 2 *

Exercise 1
Use the date set SAFETY.XLS which contains T=60 observations on Y = accident losses and X = hours spent in safety training.

Solution:
Before I even begin answering the distributed lag questions, I find it useful to set up the spreadsheet so that I have all of the potential explanatory variables, X_t, X_{t-1},..,X_{t-qmax} located in columns next to one another. In this question, qmax=4. To do this, note that the variable X_t is located in cells B3:B62 and:
i) Copy cells B3:B61 and paste them into cells C4:C62 (this will create the variable X_{t-1} and line it up properly so that cells in B contain X and cells in C contain X one period ago).
ii) Copy cells B3:B60 and paste them into cells D5:D62 (this will create the variable X_{t-2}).
iii) Copy cells B3:B59 and paste them into cells E6:E62 (this will create the variable X_{t-3}).
iv) Copy cells B3:B58 and paste them into cells F7:F62 (this will create the variable X_{t-4}).

a) Create the explanatory variables you would use in a distributed lag model with lag length equal to 4. How many observations do the explanatory variables have?
Solution:
You can see from the spreadsheet that X_t has 60 observations, X_{t-1} has 59, X_{t-2} has 58, X_{t-3} has 57, X_{t-4} has 56. However, in a regression all variables must have the same number of observations. Hence, in the regression, we only include 56 observations for all variables.

b) Run a regression of Y on X_t, X_{t-1},..,X_{t-4}.
Solution:
If you have set up the spreadsheet as described at the beginning of this exercise, you now simply have to run a regression using the techniques described in previous chapters. Be careful to specify A7:A62 in the box labelled "Input Y Range" and B7:F62 in "Input X Range".

c) Create the explanatory variables you would use in a distributed lag model with lag length equal to 2. How many observations do the explanatory variables have?
Solution:
If you have set up the spreadsheet as described at the beginning of the solution to this exercise then you do not need to do anything more here. You can now run your regression using 58 observations.

d) Estimate the distributed lag model with q=2.
Solution:

Run a regression as described in the solution to part b) except type A5:A62 in the box labelled "Input Y Range" and B5:D62 in "Input X Range".

e) Compare your answers to part d) and part b).
Sketch solution:
The two regressions are a bit different, however you should trust the results from part b) more. The part b) regression contains explanatory variable X_{t-3}. The regression in part d) does not contain this statistically significant variable and, hence, results suffer from the omitted variables bias.

f) Suppose you believe that 6 months is the maximum time that safety training might affect accident losses and accordingly, you set qmax=6. Using the sequential testing strategy described in lectures, select the lag length of the distributed lag model.

Sketch solution:
i) Set up your spreadsheet with 6 lags of the explanatory variable using the method described at the beginning of the solution to this exercise.
ii) Run a regression with q=6, then q=5, then q=4, etc. You should fine that the third lag is the highest order lag that has a coefficient significantly different from zero and choose q=3.

Exercise 2
Excel file LONGGDP.XLS contains annual data on real GDP per capita for four of the largest English speaking countries (USA, UK, Canada and Australia) for the years 1870-1993. Investigate whether there are common movements or trends between GDP in these different countries (I will present results for USA and UK in parts a, b and c).

a) Plot all the data in one time series graph and discuss your results.
Solution:
Not provided.

b) Carry out unit root tests on the time series.
Solution:
The Dickey Fuller test is the one to use (we will not provide details). Use that procedure for USA and UK GDP individually. You should find that both of these series have unit roots. For instance, for the USA if I start with an AR(4) with deterministic trend and then do the sequential testing procedure described on pages 140 and 142 of the book, I end up with an AR(2) with deterministic trend (although, depending on the order in which you omitted explanatory variables, you might have omitted the deterministic trend --- if you have done so that is fine and you will still find a unit root). When I estimate this model I obtain a t-stat on ρ of –1.892 which is NOT more negative than the critical value of –3.45. Hence, we conclude that US GDP has a unit root. For the UK, I find that an AR(4) without deterministic trend is the preferred model and the Dickey fuller test indicates a unit root.

c) For the time series which have unit roots carry out cointegration tests.
Solution:
Y = US and X = UK GDP per capita have unit roots, hence carry out the cointegration test as follows:

i) Run a regression of Y on X and save the residuals. Using Excel's methods for regression which we have discussed before (i.e. Tools/Data Analysis/Regression) you obtain the Regression Menu. Type B3:B126 in the box labelled "Input Y Range" and

C3:C126 in the box labelled “Input X Range” and click on the box labelled “Residuals”. The residuals will be placed after the table with regression output.

ii) For convenience, copy all the residuals and paste them in a free column (I chose column G on the original spreadsheet).

iii) Do a unit root test on these residuals (without including the deterministic trend). If I begin with an AR(4) and then do the sequential test procedure to determine lag length I end up with an AR(2). With this model I obtain a t-stat on ρ of –3.380 which is more negative than the critical value of –2.89. Hence, I conclude that the residuals do not have a unit root and that US and UK GDPs are cointegrated.

d) In this chapter on cointegration, we have considered only testing whether two variables are cointegrated. Using Y = USA and the other countries as explanatory variables, test for cointegration among all the time series.

Solution:

This solution is almost exactly like the one to part c), however we test for cointegration among all four variables. The intuition for cointegration among several variables is the same as that among two. That is, even though each variable individually has a unit root (and is thus trending), some linear combination of them is stationary so they are all trending together. Details follow:

Define X1 = UK GDP, X2 = Canadian GDP and X3 = Australian GDP.

i) Run a regression of Y on X1, X2 and X3 and save the residuals. To be precise, using Excel’s methods for regression which we have discussed before (i.e. Tools/Data Analysis/Regression) you obtain the Regression Menu. Type B3:B126 in the box labelled “Input Y Range” and C3:E126 in the box labelled “Input X Range” and click on the box labelled “Residuals”. The residuals will be placed after the table with regression output.

ii) For convenience, copy all the residuals and paste them in a free column (I chose column G on the original spreadsheet).

Do a unit root test on these residuals (without including the deterministic trend). If I begin with an AR(4) and then do the sequential test procedure to determine lag length I end up with an AR(2). With this model I obtain a t-stat on ρ of –3.474 which is more negative than the critical value of –2.89. Hence, I conclude that the residuals do not have a unit root and that the GDP series for all four countries are cointegrated.