Introductory Econometrics

Tutorial 7

<u>PART A:</u> To be done before you attend the tutorial. The solutions will be made available at the end of the week.

1. *Using difference in logarithms to compute the growth rate (in this exercise you need to use the same formulae repeatedly, so it is best to use Excel rather than a hand calculator):* The seasonally adjusted values of the consumer price index in Australia from the last quarter (conventionally referred to as the December quarter) of 2016 to the second quarter (conventionally referred to as the June quarter) of 2018 are given in the following table

Quarter	CPI
Dec/2016	109.8
Mar/2017	110.5
Jun/2017	110.9
Sep/2017	111.3
Dec/2017	112
Mar/2018	112.7
Jun/2018	113.3

Compute the quarterly inflation rate for each quarter from Mar/2017 to Jun/2018 in two ways: once by calculating the percentage change in the CPI the usual way, i.e. $100 \times \frac{CPI_t - CPI_{t-1}}{CPI_{t-1}}$, and another time by using difference in logarithms, i.e. $100 \times (\ln(CPI_t) - \ln(CPI_{t-1}))$. The inflation rate is often reported with one decimal point as an annualised rate. This rate is sometimes calculated by simply multiplying the quarterly growth in prices by 4.

(a) Would the reported annualised inflation rate be different for the two ways that you computed it?

(b) If you compute the annual growth by working out how much the CPI will grow in one year if it grows by the same quarterly rate each quarter, would the two ways of defining quarterly rate lead to answers that are different when rounded to one decimal point? Recall that if X grows by say 0.5% every quarter for 4 quarters, its value at the end of the year will be $X(1.005)^4$, and hence its annual growth rate will be $100 \times \frac{X(1.005)^4 - X}{X} = 100 \times (1.005^4 - 1) = 2.015\%$.

2. *Logarithmic transformation:* In the following models, x, y and z are variables and α, β and γ are parameters, and u is unobserved population error. In which models can the parameters be estimated (following a suitable transformation if necessary) using ordinary least squares (OLS)?

$$y_i = \alpha + \beta x_i + \gamma z_i + u_i \quad (1)$$

$$y_i = e^{\alpha} x_i^{\beta} z_i^{\gamma} e^{u_i} \quad (2)$$

$$y_i = e^{\alpha + \beta x_i + \gamma z_i + u_i} \quad (3)$$

$$y_i = \alpha + \beta\gamma x_i z_i + u_i \quad (4)$$

$$y_i = \alpha + \beta x_i z_i + u_i \quad (5)$$

$$y_i = \alpha + \beta x_i z_i^{\gamma} + u_i \quad (6)$$

3. *Models with quadratic terms:* Using data from a random sample of 305 women, we have estimated the following model that relates a woman's sleep time in a week (in minutes) to her work time (in minutes) and age.

$$\begin{aligned} \widehat{SLEEP} &= \underset{(333.60)}{4206.17} - \underset{(0.03)}{0.13}\ WRK - \underset{(17.46)}{37.64}\ AGE + \underset{(0.21)}{0.47}\ AGE^2 \\ n &= 305,\ R^2 = 0.092 \end{aligned} \tag{7}$$

   (a) Explain the insights that the regression results provide for the effect of age on sleep, all else equal. In particular, all else equal, at what age women are predicted to sleep the least on average according to this estimated equation?

   (b) Explain how you would test each of the following two hypotheses (no need to perform the test, only need to state the null and the alternative, the test statistics and its distribution under the null, and if needed, the regression that need to be estimated to calculate the test statistic)

      i. Expected value of sleep time conditional on work time and age is a linear function of age.

      ii. Given work time, age is not a significant predictor of sleep time.

Do not forget to bring your answers to PART A and a copy of the tutorial questions to your tutorial.

<u>Part B:</u> This part will be covered in the tutorial. It is still a good idea to attempt these questions before the tutorial.
The purpose of this tutorial is to practice statistical inference based on regression results. In addition to practicing using the t and F tables, you will also learn how to use Eviews to get the critical values for t and F tests. You can use Eviews to find critical values in tutorials and assignments, but you need to know how to use the tables for the exam.

1. The file vote1.wf1 contains data on election outcomes and campaign expenditures for 173 two-party competitive races (the two major political parties in the US are Democrats and Republicans, and competitive seats are non-safe seats in which prior to the election no party could be confident that their candidate will win. In Australia, the unsafe seats are called "marginal seats") for the House of Representatives (the "lower house" of the US Congress) in 1988. There are many variables in this data set, but the ones that we are going to use in this exercise are:

$VOTA$	% vote received by Candidate A
$EXPENDA$	Candidate A's campaign expenditure in 1000 dollars
$EXPENDB$	Candidate B's campaign expenditure in 1000 dollars
$DEMOCA$	Dummy variable =1 if Candidate A was a democrat, 0 otherwise

   In each race, Candidate A is the candidate whose last name starts with a letter that is alphabetically above the first letter of the last name of the other candidate. Run a regression of $VOTA$ on a constant $\log(EXPENDA)$, $\log(EXPENDB)$ and $DEMOCA$.

   (a) (*Interpreting the regression results when explanatory variables are logarithms of original variables and also interpreting the coefficient of dummy variables*): Explain what each parameter estimate shows.

   (b) (*Test of the overall significance of a regression*): Test the overall significance of the model at the 1% level of significance (ignore the fact that Eviews produces the F statistic, compute it using the R^2). Explain in words the hypothesis that you are testing.

(c) (*Test of significance of an explanatory variable*): Test the hypothesis that controlling for campaign expenditure, being a democratic candidate is not significant in predicting the % vote received in competitive races at the 5% level of significance. Perform the test by two methods: (i) by comparing the t statistic with the appropriate critical value, and (ii) by using the p-value.

(d) (*Joint test of multiple linear restrictions*): Test the joint hypothesis that controlling for campaign expenditure, being a democratic candidate does not contribute to the % vote received **and** that the effect of every percentage increase in campaign expenditure by Candidate A can be offset exactly by the same percentage increase in the opponent's campaign expenditure. Perform this test at the 5% level of significance. (*Note: the thinking part in these questions is to work out what the restricted regression should be. Exclusion restrictions are easy because we just drop the variables that are hypothesised to not contribute to explaining the dependent variable. Other restrictions, such as $\beta_2 = -\beta_1$ needs forming a linear combination of variables. The advantage of eviews is that it does not require these combinations to be generated as new variables and then entered into a regression. For example, the restricted model for this hypothesis can be estimated by entering "vota c (log(expenda)-log(expendb))" is the equation window.*)

(e) (Testing a single hypothesis about a linear combination of parameters) Drop $DEMOCA$ from the model. In close races each candidate believes that he or she needs to increase their campaign expenditure by more than 1% to offset the effect of a 1% increase in their opponent's expenditure. The null hypothesis is $\beta_1 + \beta_2 = 0$, and although it involves two parameters, it tests only one restriction. The alternative is $\beta_1 + \beta_2 < 0$, so we cannot use the F test because F test provides inference against $\beta_1 + \beta_2 \neq 0$. In such cases that we have only one restriction about a linear combination, we use a reparameterisation trick: Define $\delta = \beta_1 + \beta_2 \Longrightarrow \beta_2 = \delta - \beta_1$. Substitute for β_2 in the population model and rearrange, you will see that δ becomes the coefficient of one of the explanatory variables in the reparameterised model. You can see that testing $\delta = 0$ against $\delta < 0$ can be performed with a simple t test in this reparameterised model. Magic!

Eviews commands to calculate critical values:

- Example: To get the 5 percent one tail critical value of the t-distribution with 30 degrees of freedom, enter `scalar cvt30_5 = @qtdist(0.95,30)`. The critical value is recorded in a scalar variable that we called "cvt30_5". This name is of course your choice and can be anything.
- Example: To get the 5 percent critical value of the F-distribution with 2 and 169 degrees of freedom, enter `scalar cvf2_169_5 = @qfdist(0.95,2,169)`. The critical value is recorded in a scalar variable that we called "cvf2_169_5". This name is of course your choice and can be anything.
- All statistical programs are capable of giving you these critical values, even spreadsheet programs like Excel. Unless you use these commands everyday, you will forget them, but you know that you can always find them using the "help" facility of each program. See if you can find these commands in Eviews help.