Introductory Econometrics

Tutorial 9

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

1. In the following equation, *gdp* refers to gross domestic product, and FDI refers to foreign direct investment.

$$\log(gdp) = \underset{(0.13)}{2.65} + \underset{(0.022)}{0.527}\log(bankcredit) + \underset{(0.017)}{0.222}FDI.$$

Which of the following statements is then true?

a. If FDI increases by 1%, *gdp* increases by approximately 22.2%, the amount of bank credit remaining constant.

b. If FDI increases by 1%, *gdp* increases by approximately 26.5%, the amount of bank credit remaining constant.

c. If FDI increases by 1%, *gdp* increases by approximately 24.8%, the amount of bank credit remaining constant.

d. If FDI increases by 1%, *gdp* increases by approximately 52.7%, the amount of bank credit remaining constant.

2. Assume that the following equation has been estimated:

$$\begin{aligned} \widehat{\log(wage)} &= \underset{(0.011)}{2.289} - \underset{(0.015)}{0.357}female + \underset{(0.003)}{0.50}\,tot\,\mathrm{col}\,l + \underset{(0.005)}{0.030}female \times tot\,\mathrm{col}\,l, \\ n &= 6763,\ R^2 = 0.202. \end{aligned}$$

i) Using this equation, find the value of $tot\,\mathrm{col}\,l$ (total number of years in college) such that the predicted values of $\log(wage)$ are the same for men and women.

ii) Based on the equation in part i), can women realistically get enough years of college so that their earnings catch up to those of men? Explain.

3. Using the data in rmchem.wf1, run the following OLS regression of R&D expenditure as a percentage of sales ($rdintens$) on sales (measured in millions of dollars):

$$rdintens = \beta_0 + \beta_1 sales + \beta_2 sales^2.$$

i) Report the estimated equation. At what point does the marginal effect of *sales* on $rdintens$ become negative?

ii) Would you keep the quadratic term in the model? Explain.

iii) Define *salesbil* as sales measured in billions of dollars. Rewrite the estimated equation with *salesbil* and $salesbil^2$ as the independent variables. Be sure to report standard errors and the R^2. [Hint: note that $salesbil = sales/1000$ and $salesbil^2 = sales^2/1000^2$].

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 use dummy variables when using time series data.

1. The data set SeattleElectric2005-6.xlsx contains data on average hourly electricity usage in cold months in Seattle, USA, from 1 Oct 2005 to 31 March 2006. The variables in this data set are:

date	:	date in American format mm-dd-yyyy
dow	:	day of the week (1 for Sunday to 7 for Saturday)
pubhol	:	=1 if that day was a public holiday, =0 otherwise
avetemp	:	average daily temperature
aveload	:	average hourly electricity load on that day

   (a) Use *dow* to create dummy variables for different days of the week. Especially, create dummy variables for Saturday and Sunday and also a dummy variable called *wknd* which is 1 if the day is a Saturday or a Sunday, and 0 for any other day.

   (b) The goal is to predict *aveload* in winter months (Seattle, USA, is in the Northern hemisphere) based on temperature and other available information (our assumption is that the weather bureau can produce temperature forecasts that are pretty accurate). Given this goal, have an appropriate look at the data. In particular, investigate differences in average load in different days of the week and look at the scatter plot of *aveload* against *avetemp* to familiarise yourself with the data. Remember that a good analyst does not confine himself or herself to any particular software. If a pivot chart would be better for visualising differences in average load in different days of the week, then use it! If you have learnt another software (for example R, SPSS or STATA) in another unit that you think will be more helpful, use it!

   (c) Start the sample from October 4, 2005 (remember that dates are in American format: mm/dd/yyyy). Estimate a regression of *aveload* on a constant, *avetemp*, weekend and public holiday dummies. We know that electricity usage, like anything else in life, is likely to be related to what the usage was in previous days. Use the R^2 and $\bar{R}^2$ (and other information criteria) to investigate from the set of additional predictors $\{aveload_{t-1}, aveload_{t-2}, aveload_{t-3}\}$, which combination gives us the best predictive model. To use lags in a regression in eviews, you simply have to write the name of the variable followed by $(-x)$ where x specifies how many periods back you want to go. For example, the specification you need to enter in the equation window to get the regression on all 3 lags of *aveload* is:

```
aveload c avetemp pubhol wknd aveload(-1) aveload(-2) aveload(-3)
```

   (d) Is the OLS estimator unbiased here? Is it consistent?

   (e) Using the best model you found in part (c), investigate if having dummies for Saturday and Sunday separately instead of *wknd* would improve the model. In other words, test the hypothesis $H_0 : \beta_{SAT} = \beta_{SUN}$.

   (f) If there is time left in the tutorial, create dummy variables for different months of the year and investigate if adding monthly dummies could improve the model further. Given that we are accounting for temperature and public holidays, what additional information could month of the year provide for forecasting electricity usage? (Note: you are free to form monthly dummies in Excel and then read the data in Eviews, or do it in Eviews. If you want to do it in eviews `series oct = (@datepart(date,"mm")=10)` extracts the month

part of the date and if it is equal to 10, it returns a 1 and if not, it returns a zero. You can alternatively use `series oct = @recode(@datepart(date,"mm")=10,1,0)`. Remember that the data set only includes October, November, December, January, February and March.)