# Quantitative Methods for Economics
## Tutorial 6
### Katherine Eyal

# TUTORIAL 6

**13 September 2010**

ECO3021S

## Part A: Problems

1. (a) In 1857, the German statistician Ernst Engel formulated his famous law: Households with higher incomes spend a smaller fraction of their income on food. Suppose you have household budget data for a sample of South African households. Specify the econometric model you would use to test Engel's law with this data. What are the appropriate hypotheses on the parameters of the model?

   (b) An extension of Engel's law states that as income rises, increases in food expenditure grow at a smaller rate. Specify the econometric model you would use to test this extension of Engel's law. What are the appropriate hypotheses on the parameters of the model?

2. Suppose you are asked to find the relationship between weekly hours spent studying (*study*) and weekly hours spent working (*work*) for UCT students. Does it make sense to characterise the problem as inferring whether *study* "causes" *work* or *work* "causes" *study*? Explain.

3. (a) Suppose an econometric analysis finds that the relationship between wickets taken and average annual salary for bowlers in the domestic cricket league is

   $$expected\ salary\ = \text{R}230,000 + \text{R}50,000\,(wickets).$$

   How much do expected salaries rise for each wicket taken?

   (b) Suppose, instead, that expected salaries prove to be quadratic in wickets taken:

   $$\text{E}\,(salary \mid wickets) = \text{R}230,000 + [\text{R}65,000\,(wickets)] - \left[\text{R}600\,(wickets)^2\right]$$

   i. What is the derivative of expected salary with respect to wickets taken for a bowler who takes 10 wickets? 100 wickets? Interpret these derivatives.

   ii. Compute the expected salary of a bowler who takes 10 wickets.

   iii. Compute the expected salary of a bowler who takes 11 wickets.

   iv. Compute the expected salary of a bowler who takes 20 wickets.

   (c) Why might a bowler who takes 10 wickets have a different salary from the expected salary you compute in (bii)?

**Part B: Computer Exercises**

1. The data set BWGHT.DTA contains data on births to women in the United States.

   (a) What type of data structure is contained in this data set?

   (b) How many women are in the sample, and how many report smoking during pregnancy? The following commands might help you (they do not have to be executed in this specific order, and you don't have to execute every command to answer the question):

   ```
   describe
   tab cigs
   count if cigs >0
   ```

   (c) What is the average number of cigarettes smoked per day? (Use the command: `sum cigs`) Is the average a good measure of the "typical" woman in this case? Explain.

   (d) Among women who smoked during pregnancy, what is the average number of cigarettes smoked per day? (Use the command: `sum cigs if cigs> 0`) How does this compare with your answer from part (c), and why?

   (e) Find the average of *fatheduc* in the sample. Why are only 1,192 observations used to compute this average?

   (f) Report the average family income and its standard deviation in dollars.

2. The data set in CEOSAL2.DTA contains information on chief executive officers for U.S. corporations. The variable *salary* is annual compensation, in thousands of dollars, and *ceoten* is prior number of years as company CEO.

   (a) Find the average salary and the average tenure in the sample.

   (b) How many CEOs are in their first year as CEO (i.e., *ceoten* = 0)? What is the longest tenure as a CEO?

   (c) Estimate the simple regression model

   $$\log(salary) = \beta_0 + \beta_1 ceoten + u$$

   by using the command: `reg lsalary ceoten`. What is the (approximate) predicted percentage increase in salary given one more year as a CEO? (In the tutorials, as in the textbook, log always refers to the natural logarithm.)

3. Trygve Haavelmo used consumption and income data from 1922 to 1941 to estimate the marginal propensity to consume. This data is in HAAVELMO.DTA. The measure of consumption in this data set *real_cons* and the measure of income is *real_GDP*.

Haavelmo reported a marginal propensity to consume of 0.73 when he estimated a consumption function of the form

$$consumption = \alpha + \beta(income) + u$$

where $\beta$ is the marginal propensity to consume. Milton Friedman later argued that the consumption function passes through the origin, as in

$$consumption = \beta(income) + u$$

(a) Find the averages of consumption and income in this sample.

(b) Confirm Haavelmo's estimate of the marginal propensity to consume. What does the intercept in the consumption function represent?

(c) Estimate the marginal propensity to consume using the model specified by Milton Friedman. How does this compare to Haavelmo's estimate?

(d) As you know, in Keynesian macroeconomic theory an independent change in government spending, $\Delta G$ increases income by $\Delta G/(1 - \beta)$. This is called the "multiplier effect", and "the multiplier" is $1/(1 - \beta)$. What are the estimated multipliers for the two estimates of the marginal propensity to consume?

# TUTORIAL 6 SOLUTIONS

**13 September 2010**

ECO3021S

## Part A: Problems

1. (a) In 1857, the German statistician Ernst Engel formulated his famous law: Households with higher incomes spend a smaller fraction of their income on food. Suppose you have household budget data for a sample of South African households. Specify the econometric model you would use to test Engel's law with this data. What are the appropriate hypotheses on the parameters of the model?

   (b) An extension of Engel's law states that as income rises, increases in food expenditure grow at a smaller rate. Specify the econometric model you would use to test this extension of Engel's law. What are the appropriate hypotheses on the parameters of the model?

   **SOLUTION:**

   (a) One way to examine whether the fraction of income spent on food falls with income is to estimate the relationship

   $$food\ expenditure = \beta_0 + \beta_1\ income + u$$

   In this formulation, the average food consumption for a given income level is $\beta_0 + \beta_1\ income$, so the fraction of income spent on food, ($food\ expenditure/income$), is

   $$(\beta_0/income) + \beta_1$$

   which falls as income rises if $\beta_0$ is positive, and rises as income rises if $\beta_0$ is negative.

   Another way to examine Engel's law is to use a double logarithmic relationship:

   $$\log(food\ expenditure) = \alpha_0 + \alpha_1 \log(income) + v$$

   In this formulation, $\alpha_1$ is the elasticity of food expenditure with respect to income. If $\alpha_1 < 1$, then average food expenditure grows more slowly than does income, because food demand is income inelastic, and the fraction of income spent on food expenditure falls as income rises, as Engel's law posits.

1

(b) A quadratic specification in which both income and the square of income affect food expenditure allows a test of the hypothesis that as income rises, increases in food expenditure grow at a smaller rate.

$$food\ expenditure = \beta_0 + \beta_1\ income + \beta_2\ income^2 + \varepsilon$$

In this formulation, the rate of change of food expenditure with respect to income is given by the first derivative, $\dfrac{d\,(food\ expenditure)}{d\,(income)} = \beta_1 + \beta_2\ income$. The extension of Engel's law states that this first derivative is a decreasing function of income, i.e. the second derivative $\dfrac{d^2\,(food\ expenditure)}{d\,(income)^2} = \beta_2$ is negative. Thus, if the estimated $\beta_2$ is negative, the data support this extension of Engel's law.

2. Suppose you are asked to find the relationship between weekly hours spent studying (*study*) and weekly hours spent working (*work*) for UCT students. Does it make sense to characterise the problem as inferring whether *study* "causes" *work* or *work* "causes" *study*? Explain.

**SOLUTION:**

It does not make sense to pose the question in terms of causality. Economists would assume that students choose a mix of studying and working (and other activities, such as attending class, leisure, and sleeping) based on rational behavior, such as maximizing utility subject to the constraint that there are only 168 hours in a week. We can then use statistical methods to measure the association between studying and working, including regression analysis that we cover starting in Chapter 2. But we would not be claiming that one variable "causes" the other. They are both choice variables of the student.

3. (a) Suppose an econometric analysis finds that the relationship between wickets taken and average annual salary for bowlers in the domestic cricket league is

$$expected\ salary\ = R230,000 + R50,000\,(wickets).$$

How much do expected salaries rise for each wicket taken?

(b) Suppose, instead, that expected salaries prove to be quadratic in wickets taken:

$$\mathrm{E}\,(salary \mid wickets) = R230,000 + [R65,000\,(wickets)] - \left[R600\,(wickets)^2\right]$$

      i. What is the derivative of expected salary with respect to wickets taken for a bowler who takes 10 wickets? 100 wickets? Interpret these derivatives.

      ii. Compute the expected salary of a bowler who takes 10 wickets.

     iii. Compute the expected salary of a bowler who takes 11 wickets.

     iv. Compute the expected salary of a bowler who takes 20 wickets.

(c) Why might a bowler who takes 10 wickets have a different salary from the expected salary you compute in (bii)?

**SOLUTION:**

(a) The expected increase in the salaries for each wicket taken is $\dfrac{d\ expected\ salary}{d\ wickets} =$ R50, 000.

(b)   i. $\dfrac{d\ \mathrm{E}\left(salary\mid wickets\right)}{d\ wickets} = 65,000 - 1,200\left(wickets\right)$.

Thus, when $wickets = 10$ : $\dfrac{d\ \mathrm{E}\left(salary\mid 10\right)}{d\ wickets} = 65,000 - 1,200(10) = 53,000$.

When $wickets = 100$ : $\dfrac{d\ \mathrm{E}\left(salary\mid 100\right)}{d\ wickets} = 65,000 - 1,200(100) = 55,000$.

Interpretation: rate of change of the expected salary with respect to wickets taken. If a bowler who has already taken 10 wickets takes an additional wicket, then his expected salary increases by R53 000. If a bowler who has already taken 100 wickets takes an additional wicket, then his expected salary increases by R55 000.

  ii. $\mathrm{E}(salary\mid wickets = 10) = 230,000 + 65,000(10) - 600(100) = 820,000$.

 iii. $\mathrm{E}(salary\mid wickets = 11) = 230,000 + 65,000(11) - 600(121) = 872,400$. (The difference in expected salary is $872,400 - 820,000 = 52,400$, which is close to our estimate of $53,000$ from part (bi).

  iv. $\mathrm{E}(salary\mid wickets = 20) = 230,000 + 65,000(20) - 600(400) = 1,290,000$.

(c) There are many other factors that could affect a bowler's salary that are not considered in the econometric analysis. Some of these factors are observable, age, number of years with current team, etc, and some of these factors are unobservable, such as popularity and ability. These factors (as well as any measurement error) would be captured in the error term.

## Part B: Computer Exercises

1. The data set BWGHT.DTA contains data on births to women in the United States.

(a) What type of data structure is contained in this data set?

(b) How many women are in the sample, and how many report smoking during pregnancy? The following commands might help you (they do not have to be executed in this specific order, and you don't have to execute every command to answer the question):

```
describe
tab cigs
count if cigs >0
```

(c) What is the average number of cigarettes smoked per day? (Use the command: `sum cigs`) Is the average a good measure of the "typical" woman in this case? Explain.

(d) Among women who smoked during pregnancy, what is the average number of cigarettes smoked per day? (Use the command: `sum cigs if cigs>0`) How does this compare with your answer from part (c), and why?

(e) Find the average of *fatheduc* in the sample. Why are only 1,192 observations used to compute this average?

(f) Report the average family income and its standard deviation in dollars.


**SOLUTION:**


(a) Cross-sectional

(b) There are 1,388 observations in the sample. Tabulating the variable *cigs* shows that 212 women have *cigs* > 0.

(c) The average of *cigs* is about 2.09, but this includes the 1,176 women who did not smoke. Reporting just the average masks the fact that almost 85 percent of the women did not smoke. It makes more sense to say that the "typical" woman does not smoke during pregnancy; indeed, the median number of cigarettes smoked is zero.

(d) The average of *cigs* over the women with *cigs* > 0 is about 13.7. Of course this is much higher than the average over the entire sample because we are excluding 1,176 zeros.

(e) The average of *fatheduc* is about 13.2. There are 196 observations with a missing value for *fatheduc*, and those observations are necessarily excluded in computing the average.

(f) The average and standard deviation of *faminc* are about 29.027 and 18.739, respectively, but *faminc* is measured in thousands of dollars. So, in dollars, the average and standard deviation are $29,027 and $18,739.

2. The data set in CEOSAL2.DTA contains information on chief executive officers for U.S. corporations. The variable *salary* is annual compensation, in thousands of dollars, and *ceoten* is prior number of years as company CEO.

   (a) Find the average salary and the average tenure in the sample.
   (b) How many CEOs are in their first year as CEO (i.e., *ceoten* = 0)? What is the longest tenure as a CEO?
   (c) Estimate the simple regression model

   $$\log(salary) = \beta_0 + \beta_1 ceoten + u$$

   by using the command: `reg lsalary ceoten`. What is the (approximate) predicted percentage increase in salary given one more year as a CEO? (In the tutorials, as in the textbook, log always refers to the natural logarithm.)

   **SOLUTION:**

   (a) Average salary is about 865.864, which means $865,864 because salary is in thousands of dollars. Average *ceoten* is about 7.95.
   (b) There are five CEOs with *ceoten* = 0. The longest tenure is 37 years.
   (c) The estimated equation is

   $$
   \begin{aligned}
   \widehat{\log(salary)} &= 6.51 + .0097 \; ceoten \\
   n &= 177, \quad R^2 = .013
   \end{aligned}
   $$

   We obtain the approximate percentage change in salary given $\Delta ceoten = 1$ by multiplying the coefficient on *ceoten* by 100, $100(.0097) = .97\%$. Therefore, one more year as CEO is predicted to increase salary by almost 1%.

3. Trygve Haavelmo used consumption and income data from 1922 to 1941 to estimate the marginal propensity to consume. This data is in HAAVELMO.DTA. The measure of consumption in this data set *real_cons* and the measure of income is *real_GDP*. Haavelmo reported a marginal propensity to consume of 0.73 when he estimated a consumption function of the form

   $$consumption = \alpha + \beta(income) + u$$

   where $\beta$ is the marginal propensity to consume. Milton Friedman later argued that the consumption function passes through the origin, as in

   $$consumption = \beta(income) + u$$

(a) Find the averages of consumption and income in this sample.

(b) Confirm Haavelmo's estimate of the marginal propensity to consume. What does the intercept in the consumption function represent?

(c) Estimate the marginal propensity to consume using the model specified by Milton Friedman. How does this compare to Haavelmo's estimate?

(d) As you know, in Keynesian macroeconomic theory an independent change in government spending, $\Delta G$ increases income by $\Delta G / (1 - \beta)$. This is called the "multiplier effect", and "the multiplier" is $1 / (1 - \beta)$. What are the estimated multipliers for the two estimates of the marginal propensity to consume?

**SOLUTION:**

(a) Stata command: `sum real_cons real_GDP`. The average of consumption and income in the sample are 437.6 and 482.95 respectively.

(b) Stata command: `regress real_cons real_GDP` and the MPC is 0.73, same as Haavelmo. The estimated constant or intercept is 84 which is the level of consumption when income is zero.

(c) Stata command: `regress real_cons real_GDP, noconstant`. The estimated slope is now 0.903, higher than the estimated using the Haavelmo model.

(d) Multiplier effect in Haavelmo: $1/(1 - 0.73) = 3.7$ and in Friedman: $1/(1 - 0.903) = 10.3$ (much higher).