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):*

Sep-16	109.3					compund	compound
Dec-16	109.8			4*usual	4*dlog	usual	dlog
Mar-17	110.5	0.637523	0.635499	2.6	2.5	2.6	2.6
Jun-17	110.9	0.361991	0.361337	1.4	1.4	1.5	1.5
Sep-17	111.3	0.360685	0.360036	1.4	1.4	1.5	1.4
Dec-17	112	0.628931	0.626961	2.5	2.5	2.5	2.5
Mar-18	112.7	0.625	0.623055	2.5	2.5	2.5	2.5
Jun-18	113.3	0.532387	0.530975	2.1	2.1	2.1	2.1

The moral of the story is that differencing the logarithms is an easy way of calculating growth rates.

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)?

$$y_i = \alpha + \beta x_i + \gamma z_i + u_i \tag{1}$$

$$y_i = e^{\alpha} x_i^{\beta} z_i^{\gamma} e^{u_i} \tag{2}$$

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

$$y_i = \alpha + \beta\gamma x_i z_i + u_i \tag{4}$$

$$y_i = \alpha + \beta x_i z_i + u_i \tag{5}$$

$$y_i = \alpha + \beta x_i z_i^{\gamma} + u_i \tag{6}$$

Model 1 is the usual multuple regression model, so yes it can be estimated using OLS. Model 2 after taking logarithms of both sides becomes

$$\ln y_i = \alpha + \beta \ln x_i + \gamma \ln z_i + u_i$$

which is again linear in parameters and can be estimated using OLS, where the dependent variable is $\ln y_i$ and independent variables are a constant, $\ln x_i$ and $\ln z_i$. Model 3 after taking logarithms of both side becomes:

$$\ln y_i = \alpha + \beta x_i + \gamma z_i + u_i$$

which is again linear in parameters and can be estimated using OLS, where the dependent variable is $\ln y_i$ and independent variables are a constant, x_i and z_i. Model 4 is not linear in parameters (β and γ are multiplied by each other). Model 5 is fine and can be estimated as is using OLS , where the dependent variable is y_i and independent variables are a constant, $x_i z_i$. Model 6 is not linear in parameters and cannot be linearised by any transformation.

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?
   This shows that keeping work time constant, the effect of age on sleep time varies with age. Specifically, keeping work time constant, as people age by one year their predicted sleep changes by $-37.64 + 0.94AGE$. This means that initially women sleep less as they age up to a certain age and then sleep more as their age. The age that predicted age is at its minimum is
   $$\frac{37.64}{0.94} = 40.04$$

   (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.
      This can be done by a simple t-test for the significance of AGE^2.
      $$\begin{aligned} H_0 &: \beta_{AGE^2} = 0 \\ H_1 &: \beta_{AGE^2} \neq 0 \\ t &= \frac{\hat{\beta}_{AGE^2}}{se\left(\hat{\beta}_{AGE^2}\right)} \sim t_{301} \text{ under } H_0 \\ \text{we reject if } t_{calc} &< -t_{crit} \text{ or } t_{calc} > t_{crit} \end{aligned}$$

      ii. Given work time, age is not a significant predictor of sleep time.
      This would be a test of joint significance of AGE and AGE^2. We need to run the restricted regression and obtain the sum of squared residuals or the R^2 of the estimated model
      $$\widehat{SLEEP} = \hat{\beta}_0 + \hat{\beta}_1\ WRK \Longrightarrow R_r^2$$
      $$\begin{aligned} H_0 &: \beta_{AGE} = \beta_{AGE^2} = 0 \\ H_1 &: \text{at least one of the above is not zero} \\ F &= \frac{\left(R_{ur}^2 - R_r^2\right)/2}{\left(1 - R_{ur}^2\right)/301} \sim F_{2,301} \text{ under } H_0 \\ \text{we reject if } F_{calc} &> F_{crit} \end{aligned}$$

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

PART B: You do not need to hand this part in. It 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 $DEMOCA$.

Dependent Variable: VOTEA
Method: Least Squares
Sample: 1 173
Included observations: 173

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	51.13410	2.903327	17.61224	0.0000
LOG(EXPENDA)	6.299279	0.375274	16.78582	0.0000
LOG(EXPENDB)	-6.666045	0.391187	-17.04054	0.0000
DEMOCA	1.208824	1.241612	0.973593	0.3317

R-squared	0.786385	Mean dependent var	50.50289
Adjusted R-squared	0.782593	S.D. dependent var	16.78476
S.E. of regression	7.826209	Akaike info criterion	6.975683
Sum squared resid	10351.17	Schwarz criterion	7.048592

In answering the question, we use the notation that the true population model is $VOTEA = \beta_0 + \beta_1 LOG(EXPENDA) + \beta_2 LOG(EXPENDB) + \beta_3 DEMOCA + u$, and the estimated model is $\widehat{VOTEA} = \hat{\beta}_0 + \hat{\beta}_1 LOG(EXPENDA) + \hat{\beta}_2 LOG(EXPENDB) + \hat{\beta}_3 DEMOCA$ (please train students to define the notation that they want to use before they use it)

(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.
$\hat{\beta}_1 = 6.299$: For a candidate of a specific party (this is keeping DEMOCA constant), a 1% increase in the candidate's campaign expenditure is expected to increase his/her share of votes by 0.063 percentage points (note that the dependent variable is % of total vote), keeping the campaign expenditure of the opponent constant.
$\hat{\beta}_2 = -6.666$: For a candidate of a specific party (this is keeping DEMOCA constant), no change in the candidate's campaign expenditure (this is keeping EXPENDA constant) and a 1% increase in the opponent's campaign expenditure is expected to decrease the candidate's share of votes by 0.067 percentage points.

$\hat{\beta}_3 = 1.209$: Controlling for campaign expenditures, a democratic candidate's share of votes is expected to be 1.2% higher than a republican candidate's share of votes. (mention that it is statistically not different from zero).

(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.

$$
\begin{aligned}
H_0 &: \beta_1 = \beta_2 = \beta_3 = 0 \\
H_1 &: \text{at least one of } \beta_1, \beta_2 \text{ or } \beta_3 \text{ is not zero} \\
F &= \frac{R^2/3}{(1-R^2)/(173-3-1)} \sim F_{3,169} \text{ under } H_0 \\
& \quad \text{please make sure that students understand that the distribution is} \\
& \quad \text{valid only if } H_0 \text{ is true} \\
F_{calc} &= \frac{0.786385/3}{(1-0.786385)/169} = 207.38 \\
\alpha &= 0.01 \Longrightarrow 169 \text{ is not in the table, we use 120 which is the closest one less than 169} \\
& \Longrightarrow cv = 3.95 \text{ from the table, 3.8996 using @qfdist(0.99,3,169)} \\
F_{calc} &> cv \Longrightarrow \text{ we reject the null} \\
\text{Conclusion} &: \text{At least one of the explanatory variables is significant in explaining } VOTA
\end{aligned}
$$

(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.

$$
\begin{aligned}
H_0 &: \beta_3 = 0 \\
H_1 &: \beta_3 \neq 0 \\
t_{\hat{\beta}_3} &= \frac{\hat{\beta}_3}{se_{\hat{\beta}_3}} =\sim t_{169} \text{ under } H_0 \\
& \quad \text{please make sure that students understand that the distribution is} \\
& \quad \text{valid only if } H_0 \text{ is true} \\
t_{calc} &= \frac{1.208824}{1.241612} = 0.973593 \\
\alpha &= 0.05 \Longrightarrow cv = \text{ again choose 1.98 from the table, or 1.9741 using @qtdist(0.975,169)} \\
-cv &< t_{calc} < cv \text{ or } pvalue = 0.33 > \alpha \Longrightarrow \text{we cannot reject the null} \\
\text{Conclusion} &: \text{After accounting for campaign expenses, political party is insignificant in explaining}
\end{aligned}
$$

(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.)

Steps to get the restricted model

$$\begin{aligned} VOTEA &= \beta_0 + \beta_1 LOG(EXPENDA) - \beta_1 LOG(EXPENDB) + u \\ VOTEA &= \beta_0 + \beta_1 (LOG(EXPENDA) - LOG(EXPENDB)) + u \end{aligned}$$

Dependent Variable: VOTEA
Method: Least Squares
Sample: 1 173
Included observations: 173

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	49.97148	0.594598	84.04253	0.0000
LOG(EXPENDA)-LOG(EXPENDB)	6.545465	0.262391	24.94545	0.0000

R-squared	0.784438	Mean dependent var	50.50289
Adjusted R-squared	0.783178	S.D. dependent var	16.78476
S.E. of regression	7.815690	Akaike info criterion	6.961637
Sum squared resid	10445.54	Schwarz criterion	6.998091

H_0 : $\beta_2 = -\beta_1$ and $\beta_3 = 0$

H_1 : at least one of the above restrictions is false

$F = \dfrac{(SSR_r - SSR_{ur})/2}{SSR_{ur}/(173-3-1)} \sim F_{2,169}$ under H_0

please make sure that students understand that the distribution is valid only if H_0 is true

$F_{calc} = \dfrac{(10445.54 - 10351.17)/2}{10351.17/169} = 0.77$

$\alpha = 0.05 \Longrightarrow cv =$ choose conservative 3.07 from the table, or 3.049 using @qfdist(0.95,2,169)

$F_{calc} < cv \Longrightarrow$ we cannot reject the null

Conclusion : There is not enough evidence in the data to reject the assumption that the effect of every 1% increase in campaign expenditure can be exactly offset by a similar % increase in the opponents expenditure, and that political affiliation is insignificant in marginal seats

(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 re-arrange, 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!

Steps to get the reparameterised model

$$\begin{aligned} VOTEA &= \beta_0 + \beta_1 LOG(EXPENDA) + (\delta - \beta_1) LOG(EXPENDB) + u \\ VOTEA &= \beta_0 + \beta_1 (LOG(EXPENDA) - LOG(EXPENDB)) + \delta LOG(EXPENDB) + u \end{aligned}$$

Dependent Variable: VOTEA
Method: Least Squares
Sample: 1 173
Included observations: 173

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	52.03893	2.750137	18.92231	0.0000
LOG(EXPENDA)-LOG(EXPENDB)	6.341950	0.372649	17.01858	0.0000
LOG(EXPENDB)	-0.414801	0.538688	-0.770019	0.4424

R-squared	0.785187	Mean dependent var	50.50289
Adjusted R-squared	0.782660	S.D. dependent var	16.78476
S.E. of regression	7.825009	Akaike info criterion	6.969716
Sum squared resid	10409.23	Schwarz criterion	7.024397
Log likelihood	-599.8804	Hannan-Quinn criter.	6.991900
F-statistic	310.6936	Durbin-Watson stat	1.662940
Prob(F-statistic)	0.000000		

H_0 : $\delta = 0$ which is the same as $\beta_1 + \beta_2 = 0$

H_1 : $\delta < 0$ which is the same as $\beta_1 + \beta_2 < 0$

$$t_{\hat{\delta}} = \frac{\hat{\delta}}{se_{\hat{\delta}}} \sim t_{170} \text{ under } H_0$$

please make sure that students note that $k = 2$ here

$$t_{calc} = \frac{-0.414801}{0.538688} = -0.77$$

α = 0.05 one tailed $\Longrightarrow$ use the conservative -1.658 from the table, or -1.654 using @qtdist(0.05,170)

$t_{calc} > -1.654 \Longrightarrow$ we cannot reject the null

Conclusion : There is no evidence in the data to suggest that a candidate needs a larger than 1% increase in his or her campaign expenditure to offset the effect of a 1% increase in the opponent's campaign expenditure on the candidate's share of votes.