REED TUTORIALS (Pty) LTD

ECS3706 EXAM PACK

ECONOMETRICS STUDY PACK

MAY/JUNE 2016

Question 1

(a) List any three uses of econometrics (3)

(a) (i) Describing economic reality

(ii) Testing hypothesis about economic theory

(iii) Forecasting future economic activity

(b) Briefly explain the meaning of the following

(i) Regression analysis

(b) (i) regression analysis is used to make quantitative estimates of economic relationship that previously have been completely theoretical in nature.

(ii) The estimated regression equation

(ii) usually population size may be too large or unknown therefore inference is made on the sample drawn from the population using the estimated regression equation which is the statistical technique that attempts to explain movements in one dependent variable given independent variables

(iii) Total, explained and residual sum of squares (7)

(iii) TSS is the squared variation of Y around its mean it is measured by

$\sum_{i=1}^{n}(Y - Y)$. ESS is the variation of Y around its mean which is explained by the equation it is measured by $\sum_{i=1}^{n}(Yi - Y)$. RSS is the variation of Y around its mean which is not explained by the equation, measured by $\sum_{i=1}^{n} e^2$.

TSS= ESS+RSS

c) Carefully distinguish between R^2 and $\overline{R}^2$ Can $\overline{R}^2$ be misused in regression analysis? Explain (5)

(c) R^2 is the co-efficient of determination which measures the goodness of fit of the whole equation. The problem with R^2 is that whenever an independent variable is added it will increase even when it is an irrelevant variable. R^2 measures the percentage of the variation of Y around its mean that is explained by the regression equation, but adjusted for the degrees of freedom. Usually a higher fit does not guarantee a good estimation because the fit can be only measure used.

(a) State four valid criteria for determining whether a given variable belongs in an equation (4)

Question 2

(a)(i) Theory which underlies the model

(ii) The statistical significance of the t-test

(iii) The R^2 which determines the goodness of fit of the whole equation

(iv) Bias

(b) Distinguish between sequential specification searches and data mining (4)

(b) Sequential specification involves the initial specification of an equation and estimates it, then adding other independent variables until a plausible equation is found while data mining involves estimating a variety of alternative specifications before that best equation has been chosen.

(c) Econometricians are often careful to avoid specification bias (7)

What exactly is specification bias?

i What is known to cause specification bias?

(c) (i) Specification bias occurs when an independent variable is omitted which is relevant in the equation.

ii Are unbiased estimates always better than biased estimates? Why or why not?

(ii) Unbiased estimates could mean the equation is correctly specified or an independent variable is omitted those results in a high variance while biased estimates may decrease the variance. The importance of each of the two would depend on the importance of either type 1 or type 2 errors.

iii What's the best way to avoid specification bias?

(iii) To avoid specification the four important specification criteria must be followed (refer to question 2 (a).)

Question 3

(a) What is the stochastic error term and why does it have to be included in a regression equation? (4)

(a) The stochastic error term is added to the regression equation because it helps capture the minor variations in Y which are not captured by the Xs. It is virtually

impossible to avoid some sort of measurement error hence the stochastic error captures that. The underlying theory may differ from the chosen functional form hence the need for an error term.

(b) Briefly explain what OLS means and what it tries to achieve Why should researchers use it? (5)

(b) OLS means ordinary least squares, it is a technique that calculates the Bs so as to minimize the sum of p squared residuals which is calculated as $(Yi - Yi)^2$.

(c) List the steps applied in regression analysis (6)

(c) (i) Review the literature and develop the theoretical model

(ii) Specify the model

(iii) Hypothesize the expected signs of the coefficients

(iv) Collect the data, inspect and clean the data

(v) Estimate and evaluate the equation

(vi) Document the results

Question 4

QUESTION 4 (15 marks)

Suppose you are a junior researcher working for a research institute and your line manager gives you the following information that has been used to establish the impact of income on household consumption

$$\sum_{i=1}^{N} Y_i^2 = 23136,\ \sum_{i=1}^{N} X_i^2 = 76000,\ \sum_{i=1}^{N} X_i Y_i = 48925,\ \sum_{i=1}^{N} Y_i = 5650,\ \sum_{i=1}^{N} X_i = 2800,\ \sum_{i=1}^{N} x_i y_i = 1830$$

$$\sum_{i=1}^{N} y_i^2 = 3390,\ \sum_{i=1}^{N} x_i^2 = 2424,\ \sum_{i=1}^{N} e_i^2 = 3279,\ \bar{X} = 11\,2,\ \bar{Y} = 22\,6$$

Where $x_i = X_i - \bar{X}$ and $y_i = Y_i - \bar{Y}$ for $i = 1, 2, \quad , N$

(a) How many households were considered for this research analysis? (2)

(a) They are 2800 households in the sample which is given by $\sum_{i=1}^{n} X = 2800$

(b) Use the equation $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ to compute the OLS estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ (3)

(b) Given Y=B0 +B1X1 + Ɛ

Solving for B1=$\frac{\varepsilon xy}{\varepsilon x^2}$= $\frac{1830}{2424}$ = 0.7550

Solving for B0= Y- B1 X1

= 22.6-0.755(11.2)

= 14.144

Y = 14.14+0.755X

(c) What is the meaning of $\hat{\beta}_0$ and the slope coefficient $\hat{\beta}_1$ calculated in part (b)? (4

B0 is the estimated constant which could be called the y-intercept. While B1 is the estimated slope coefficient of the equation.

(d) Compute the coefficient of determination (R^2) and the standard error of the slope coefficient SE($\hat{\beta}_1$) Do you think that the explanatory variable X is statistically significant? *Hint Use SE obtained to calculate the t-statistic* (6

$$R^2 = 1 - \frac{\mathcal{E}e^2}{\mathcal{E}(Yi-Y)^2}$$

$$= 1 - \frac{3279}{(5650-22.6)^2}$$

$$=0.90.$$

The explanatory variable are a good fit because 40% of the variation in Y is explained by the Xs.

Section B

QUESTION B1 (20 marks)

A researcher has tried to establish the relationship between money supply (M3) and its two determinants, interest rates (IR) and economic growth (GDP) and estimated the following regression results

$$\hat{M}3_t = \underset{(0\ 2916)}{21\ 460} - \underset{(-0\ 107)}{0\ 4206}\ IR_t + \underset{(13\ 1632)}{0\ 7336}\ GDP_t$$

$R^2 = 0\ 9470$, $\bar{R}^2 = 0\ 9399$, Durbin-Watson (DW) statistic = 0 8049

Where M3 = money supply
IR = interest rate
GDP = gross domestic product

M and GDP are measured in millions and IR is measured in percentages The values in the parenthesis represent the t-statistics

(a) Fully interpret the above regression results, taking into account the units used Do you think the signs obtained conform to economic theory? (10)

Question B1

(a). When it comes to the interest rate using the rule of 2 we fail to reject the null hypothesis although the test statistic carries the correct negative sign but interest rates when coming to the determination of money supply they are statistically insignificant. This could be because both money supply and interest rates are the proprieties (instruments) of monetary policy with the authorities being able to control the supply of money which is done through the accommodation policy while on the other hand the authorities control interest rate via the repo rate, therefore the direct influence of one over the other may be limited.

Using the rule of 2 on the test statistic of GDP we reject the null hypothesis. Plus the test carries the correct expected sign. This means GDP is statistically significant when it comes to explaining the changes in money supply. Theoretically changes in GDP would refer to changes in money supply via inflation effect plus the changes in economic growth. The overall goodness of fit as measured by the R^2 seems to be good enough to infer that it is a good equation.

(b) What econometric problem(s) does this equation appear to have? Explain (3)

(b) The equation seems to carry an irrelevant variable which is interest rate in the determination of money supply.

(c) Suppose that you are now told that the simple correlation coefficient between IR and GDP is - 0 89 and the values of the variance inflation factors exceed 5 Does this change your answer to part (b) above? How? (3

(c) Yes it changes the explanations to the answer above. A simple correlation coefficient of -0.89 shows a strong negative relationship between GDP and interest rates. This means multicollinearity exists with a variance inflation factor exceeding S that means the multicollinearity is severe

(d) What remedies for the problems you identify in parts (b) and (c) do you recommend? Explain (4)

We could drop the redundant variable in this case interest rate since it is statistically insignificant, we could increase the sample size to see if multicollinearity fades away

Question B2

QUESTION B2 (20 marks)

You are an economist working in the research department at your institution and in estimation the demand for beef, you obtained the following results The dependent variable is Q Per capita consumption of beef (kilograms per month)

Variable	Coefficient	Standard error	t-statistic
Constant	32 85	4 24	7 74
H Household income	0 0103	0 0053	194
PB price beef	-0 2979	0 1348	-2 21
PC price chicken	-0 0140	0 0642	-0 22
PS price of substitutes	0 1134	0 0820	1 38

Method Ordinary Least Squares, Number of observations in sample 23, R^2=0 92, Durbin-Watson statistic = 0 67, F = 51 8 The variables are as follows

H Per capita disposable household income (Rands)
PB Price of beef (cents/kg)
PC Price of chicken (cents/kg)
PS Price of beef substitutes (weighted average of chicken and pork, cents per kg)

(a) Interpret the meaning of the constant term and β_{PC} Use proper units of measurement if applicable (3)

(a) Bpc is the coefficient which estimates the substitute for keep; it implies that a R1 increase in the price of chicken results in a 1.4kg decrease in the quantity of beef. Although this is theoretical incorrect showing signs of statistically insignificance.

(b) Explain how the t-value of -2 21 in the table has been calculated (1

(b) T-value is calculated as $\frac{Bpb-O}{SE(PB)}$= $\frac{-0.2972-0}{0.1348}$= -2.21

(c) Test each relevant coefficient for statistical significance State its expected value (on theoretical grounds), the null and alternative hypothesis and your conclusion (include the reason) Use a 5% level of significance for testing Please use a table to present your results (13)

variable	Expected sign	Ho and Ha	T-test	Decision
Household income	Positive	Ho! B=0 Ha! B>0	1.94	We reject Ho (t-test>t-critically)
Price of Beef (PB)	Negative	Ho! B=0 Ha! B<0	-2.21	We reject Ho (t-test<t-critical)
Price of chicken (PC)	Positive	Ho B=0 Ha! B>0	-0.22	We do not reject Ho (unexpected sign)
Price of substitutes (PS)	Negative	Ho! B=0 Ha! B>0	1.38	We do not reject Ho (t-test<t critical)

(d) Identify a major shortcoming of this regression and explain what can be done to correct it (3)

(d) This regression equation has been incorrectly specified when it comes to PC therefore PC should be dropped or its coefficient specified. Multicollinearity is present in the independent variables with both PS and PC explaining almost the same outlook.

Question 3

QUESTION B3 (20 marks)

Regression results for the annual demand for money in country XYZ are given as follow

$$\hat{M}_t = 0\ 14 + 1\ 05\ Y_t^* + 0\ 01Y_t + 0\ 75IR_t$$
$$(0\ 16) \quad (0\ 11) \quad (0\ 07)$$
$$\bar{R}^2 = 0\ 986 \quad DW = 0\ 98 \quad N = 70 \quad r_{Y^*,Y} = 0\ 96$$

Where M_t = the log of the money stock
Y_t^* = the log of average income
Y_t = the log of current income
IR_t = the log of the rate of interest
$r_{y^*,y}$ = correlation coefficient between Y^* and Y
Values in parentheses are the standard errors

(a) Interpret the results given above (4)

(a)The above equation is a double log. The values of Y and the Xs have been logged so as to draw a meaning-full inference out of them because naturally the values of Y and X are relatively too large.

(b) Hypothesise signs and test the appropriate null hypotheses at the 5% level of significance (9)

Variable	Expected sign	Ho and Ha	t-value	Decision
Log of average income (Y)	Positive	Ho! B=0 Ha! B>0	1.05/0.16=6.56	Reject Ho (t-test>t-critical)
Log of current income (Y)	Positive	Ho! B=0 Ha! B>0	0.01/0.11=0.09	Do not reject Ho (t-test<t-critical)
Log of the rate of interest (IR)	Negative	Ho! B=0 Ha! B<0	0.75/0.07= 10.7	Do not reject Ho (unexpected sign)

(c) What econometric problems seem likely to be in this equation? (3)

This econometric problem carries multicollinearity and misspecification of equations

(d) In particular, are there are any problems related to the coefficient of Y? If so, are these problems more likely to have been caused by multicollinearity, serial correlation, or heteroskedasticity? (4)

(d) Yes they are problems related to Y which have been caused by multicolinearity between the log of average income and the log of current income. Again serial correlation exists between the log of the rate of interest and the log of average income.

MAY/JUNE 2015

Question 1

(a) (i) Zero mean error means the mean is normally distributed and most of its variables are located to zero

(ii) Homoscedastic errors means the variance of the error terms are constant over a wide variety of samples.

Both assumptions ensure that the regression equation BLUE.

(b) May June 2016

(c) May June 2016

Question 2

(a) October November 2015 Question (a) and May June Question 1 (a)

(b) Constant term is the y-intercept which represents. Bo sometimes it may be insignificant economically however in cases it may be important in areas like the cost function due to this it is not tested

(c) The theorem focuses on the classical assumption I and VI with 4 concepts paid attention to:

- Unbiased
- Minimum variance
- Consistent
- Normal distribution

Question 3

(a)

- Hypothesis testing is used to check for statistical difference

- Ho is what the researcher does not expect
- Ha is what the researcher expects
- Critical value is used to draw up the rejection rigor usually given a 45% confidence interval
- Degrees of freedom give us the extent to which our t-distribution can be allowed to vary
- Type I error: we reject a true null hypothesis
- Type II error: we fail to reject a false null
- One sided used when we are sure of the expected sign
- Two sided used when we are not sure of the expected sign

(b) V I F Cooks at the extent to which a given explanatory variable can be explained by all the other explanatory variables in the equation

(c) October November 2015 Question 3 (c)

Question 4

(a)

- ✓ Linear multivariate regression is an equation which contains more than one independent variable Y= Bo+B1X1+B2X2+B3X3+Ɛ
- ✓ Linear univariate regression is an equation which contains only one explanatory Y=Bo+B1X1+Ɛ
- ✓ Steps for regression May June 2016 Question 3 (c)

(b) October November 2015

(c)

- Dummies are used when equation is qualitative
- Assumes value of 1 and 0
- When the dummy is under question we make it 1 otherwise 0

- We introduce m-1 dummies with the other dummy being the slope dummy

Section B

Question B1

(a) May June 2016 Question B3 (b)

(b) October November 2015 Question B2 (a)

(c) The equation might contain multicollinearity because GDP and GDP R(-1) may be one and the same thing but on R^2 of 0.979 implies that 97.9% of changes in Y is explained by the independent variables which implies an overole good fit of the model.

Question B2

(a) October November 2015 Question B3 (e)

(b)

Variables	Expected sign	Ho and Ha	t-value	Decision
Degree of	Freedom	n-k-1-	1.699	
Price of pork	Negative		-0.017415/0.004992=-3.489	Reject Ho
Price of beef	Positive		0.006684/0.00287=2.33	Reject Ho (t-test>t-critical)
Income	Positive		0.271935/0.290844=0.935	Do not reject Ho (t-test <critical)
KG Pork	Positive		3.100112/0.234566=13.2164	Reject Ho (t-test >t-

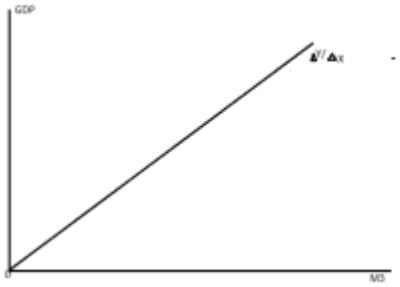
				critical)
Degrees of	Freedom	KV2	n-k-1=2.045	
D1	Unsure		0.007147/0.099133=0.721	Do not reject Ho (t-test<t-critical)
D2	Unsure		0.004917/0.135639=0.036	Do not reject Ho (t-test<t-critical)

Question B3

- Linear: GDP=19000+1.285 (M3)

A 1% increase in M3 leads to a 1.3% increase in GDP

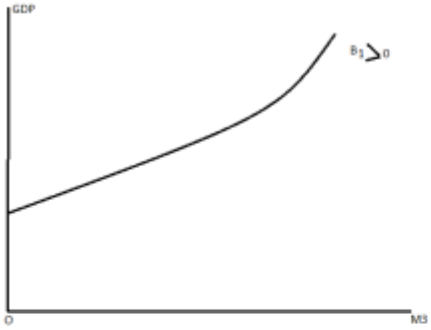
Change in Y divided by Change in X



- Double log: log (GDP)=0.63+0.682(log(M3))

Slope: $\frac{\Delta Y/Y}{\Delta X/X}$

Percentage change in Y is brought about by a percentage change in X

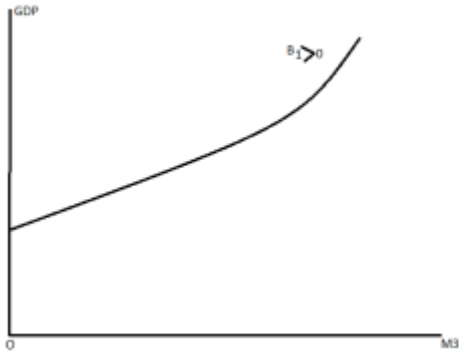


- Semi (log Y)! Log(GDP)=15.68+0.00000275(M3)

A one unit change in M3 leads to a log (0.00000275) change in GDP

Slope: $\frac{\Delta Y/Y}{\Delta X}$

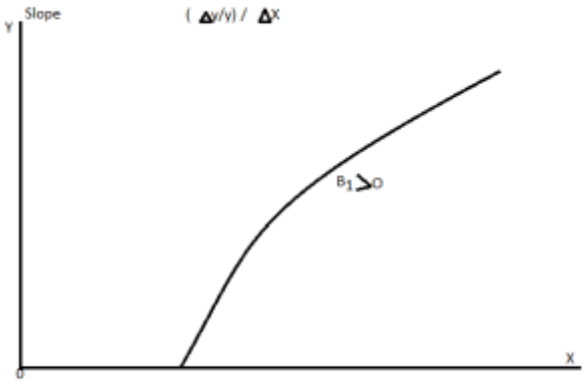
A unit change in M3 leads too percentage change in GDP



- Semi log (logX): GDP=356214+47254 (log(M3))

A percentage increase in M3 leads to a 5.67 unit increase GDP

Slope= $\frac{\Delta Y/Y}{\Delta X}$


Slope
(Δy/y) / Δx
Y
$B_1 > 0$
X
0

OCTOBER/NOVEMBER 2015

Question 1

(a) Econometrics use regression analysis to make qualitative estimates of economic relationship that previously have been completed theoretically in nature. 3 inputs are: economic theory, statistics and economic data

(b) Usually the population size is too large or unknown therefore estimating the true may be difficult. We use the estimate to draw inference about the true linear regression equation. The true is represented by B while the estimated is represented by B. B-B=0 which means the estimated B is an unbiased estimator of the true B.

(c) r measures the strength of the relationship between two variables. It runs from -1 to 1 with -1 showing a strong negative relationship and +1 showing a strong positive relationship. R^2 measures the goodness of fit of the whole equation.

Question 2

(a) Discrete random variable has a countable number of possible values. Each outcome can only take one possible number as the outcome. When throwing a dice you can only get a number between 1 and 6, one cannot get 2.5. The probability of getting any number when throwing a dice randomly P(X) = $\frac{1}{6}$, there is a one in six chances that the dice will drop on 4.

(b)

- ✓ Stochastic error term is the value which captures the variations in Y not explained by the changes in X
- ✓ May June 2016 Question # (a)
- ✓ October November 2015 Question 1 (b)

Question 3

(a) A linear functional relationship between 2 or more independent variables that is so strong that it can significantly affect estimation of the coefficients of the variables. It could be detected using high simple correlation coefficients and high variance inflation factors (V I Fs)

(b) May June 2016 Question 1 (b) (iii) and (c) combined

(c) Irrelevant variables lead to (i) No bias (ii) then increase the variance. Both for the dependent and independent variables

(d) Histograms are used to read statistical outcome of the data. Usually most economic data are normally distributed the histogram would be well shaped. However discrete random variables which have an equal probability usually follow a uniform distribution and the histogram will be equal throughout.

Question 4

(a) Pure serial correlation occurs when classical assumption IV which assumes uncorrelated observations of the error term is violated in a correctly specified equation. Impure social correlation is caused by a specification error such as an omitted variable or an incorrect function form.

(b) Given the following hypothesis

Ho! The dependent is innocent

Ha! The dependent is guilty

Type I error: sending an innocent dependent to jail

Type II error: freeing guilty dependent

(c) B1= $\frac{\mathcal{E}x1y1}{\mathcal{E}x^2}$ = $\frac{1980}{3786}$ = 0.523

Bo= Y-B1X1 = 5.9-0.523(11.6) = -0.1668

The y-intercept us -0.1668 while the slope is 0.523.

Section B

Question B1

(a) A 1% increase in total household income will lead to a 0.036 kilowatt decrease in electricity consumption by households. A 1% increase in the average price of electricity leads to a decrease in electricity consumption by 10%.

(b)

Variable	Expected sign	Ho and Ha	T-value	Decision
Degree of	Freedom	Txn-k-1	1.667	
Inc	Positive	Ho! B=0 Ha! B>0	-0.36/0.012=-3	Do not reject Ho (unexpected sign)
Hh	Positive	Ho! B=0 Ha! B>0	2.25/0.25=9	Reject Ho t-test>t critical
Pop	Positive	Ho! B=0 Ha! B>0	-0.04/0.03=-1.3	Do not reject Ho (unexpected sign)
Price	Negative	Ho! B=0 Ha! B<0	-10/2.5=-4	Reject Ho t-test<t critical
Season	Unsure	Ho! B=0 Ha! B≠0	0.24/0.08=3	Reject Ho t-test>t critical

(c) Reject Ho if t-test > t- critical. Do not reject Ho if t-test ≤ t- critical. T test = 245. T critical = 2.35. Reject Ho because t-test>t-critical.

(d) Electricity has been log alongside income on total number of residence. Advantage is that correct inference can be drawn from the logged numbers since the absolute original values may be too large to draw inference.

Question B2

(a) Positive serial correlation may be present because total population of area (pop) and total number of household resident within area are SD are related.

Positive serial correlation:

Ho! p≤0

Ha! p>0

DW-d= 2.08

dl= 1.49

du= 1.77

Do not reject Ho because d>du

(b) When C is omitted

Qt=Bo+BaAt+BbBt+Ɛ** where the Ɛ** contains the omitted variable which could then be written as Ɛ**= Ɛi-BcCt

(c)

- ✓ Form 1 is a semi log
- ✓ The alternative function for Form 1 is Y= ab^x
- ✓ Y increases at a geometric rate with respect to X given by
- ✓ Y could be the years of experience while X is the salary

- ✓ Form 2 is a double lag
- ✓ Alternative function for Form 2 is
- ✓ A 1% change in X will cause a b% change in Y
- ✓ When both variables change in percentages where Y is electricity consumption and X is income of households

Question B3

(a) They is slope dummy variable and the coefficient dummy variables

(b) The first quarter dummy variable serves as the benchmark dummy variable which is the y-intercept

(c) Yes it makes sense because during the last quarter of the year it is the coldest season in the USA therefore little or no sales are made

(d) Frig= 456.2 + 2.7 (500) + 2.42.5 (0) + 325.1 (1) – 86.1(0)

Sales for the 3rd quarter where 2132 fridges

(e) This is a double lag with elasticity of

The production function shows the percentage change in output brought about by production change in capital input, technology and technology input.

The advantage is that absolute output maybe large to draw inference therefore it is important to log it

(f) Y=a+bX used where Y and X change in real values

Log (Y)= log (a) + b log (X) where percentage change in Y is caused by percentage change in X

(g) R^2 measures well variables of linear form when transformed variables may have a small R^2 which means poor fit however this cold ne am incorrect indication.

FOR FURTHER QUERIES CONTACT

PHILLIP MANGAVA
FOUNDER & CEO
PHONE O787912965
EMAIL info@reedtutorials.co.za
WEB www.reedtutorials.co.za
TWITTER @reedtutorials1