BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE, PILANI

SECOND SEMESTER 2020-2021

ECON F241 ECONOMETRIC METHODS

PROBLEM SHEET 2 24 March 2021

Multiple Choice Questions

1. If an independent variable in a multiple linear regression model is an exact linear combination of other independent variables, the model suffers from the problem of _____.

A. perfect collinearity
B. homoskedasticity
C. heteroskedasticty
D. omitted variable bias.

2. The assumption that there are no exact linear relationships among the independent variables in a multiple linear regression model fails if _____, where n is the sample size and k is the number of parameters.

A. n>2
B. n=k+1
C. n>k
D. n<k+1

3. Exclusion of a relevant variable from a multiple linear regression model leads to the problem of _____.

A. misspecification of the model
B. multicollinearity
C. perfect collinearity
D. homoskedasticity

4. Suppose the variable x2 has been omitted from the following regression equation,

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$$

~β_1 is the estimator obtained when x_2 is omitted from the equation. The bias in ~β_1 is positive if:

A. $\beta 2 > 0$ and x_1 and x_2 are positively correlated
B. $\beta 2 < 0$ and x_1 and x_2 are positively correlated
C. $\beta 2 > 0$ and x_1 and x_2 are negatively correlated
D. $\beta 2 = 0$ and x_1 and x_2 are negatively correlated

5. Which of the following statements is true?

A. Taking a log of a nonnormal distribution yields a distribution that is closer to normal.
B. The mean of a nonnormal distribution is 0 and the variance is σ2.
C. The CLT assumes that the dependent variable is unaffected by unobserved factors.
D. OLS estimators have the highest variance among unbiased estimators.

6. The Gauss-Markov theorem will not hold if _____.

A. the error term has the same variance given any values of the explanatory variables
B. the error term has an expected value of zero given any values of the independent variables
C. the independent variables have exact linear relationships among them
D. the regression model relies on the method of random sampling for collection of data

7. If $\widehat{\beta_j}$, an unbiased estimator of β_j, is consistent, then the:

A. distribution of $\hat{\beta}_j$ becomes more and more loosely distributed around β_j as the sample size grows.
B. distribution of $\hat{\beta}_j$ becomes more and more tightly distributed around β_j as the sample size grows.
C. distribution of $\hat{\beta}_j$ tends toward a standard normal distribution as the sample size grows.
D. distribution of $\hat{\beta}_j$ remains unaffected as the sample size grows.

True or False

1. R squared can decrease when an independent variable is added to a multiple regression model.

2. An explanatory variable is said to be exogenous if it is correlated with the error term.

3. A larger error variance makes it difficult to estimate the partial effect of any of the independent variables on the dependent variable.

4. Standard errors must always be positive.

5. H1: $\beta_j \neq 0$, where β_j is a regression coefficient associated with an explanatory variable, represents a one-sided alternative hypothesis.

Numericals/Long answer type

1. A researcher investigating the determinants of the demand for public transport in a certain city has the following data for 100 residents for the previous calendar year: expenditure on public transport, E, measured in dollars; number of days worked, W; and number of days not worked, NW. By definition NW is equal to 365 − W. He attempts to fit the following model:

$$E = \beta_1 + \beta_2 W + \beta_3 NW + u$$

Explain why he is unable to fit this equation. (Give both intuitive and technical explanations.) How might he resolve the problem?

2. The researcher in the above question decides to divide the number of days not worked into the number of days not worked because of illness, *I*, and the number of days not worked for other reasons, *O*. The mean value of *I* in the sample is 2.1 and the mean value of *O* is 120.2. He fits the regression (standard errors in parentheses):

$$\underset{}{\hat{E}} = \underset{(8.3)}{-9.6} + \underset{(1.98)}{2.10\text{W}} + \underset{(1.77)}{0.45\text{O}}$$

$$R^2=0.72$$

Perform *t* tests on the regression coefficients and an *F* test on the goodness of fit of the equation. Explain why the *t* tests and *F* test have different outcomes.

3. From the following data estimate the partial regression coefficients, their standard errors, and the adjusted and unadjusted R^2 values:

$$\bar{Y} = 367.693,\ \overline{X_2} = 402.760,\ \overline{X_3} = 8.0,\ \ n = 15$$

$$\sum(Y_i - \bar{Y})^2 = 66042.249 \quad \sum\left(X_{2_i} - \overline{X_2}\right)^2 = 845855.096 \quad \sum\left(X_{3_i} - \overline{X_3}\right)^2 = 280.00$$

$$\sum(Y_i - \bar{Y})(X_{2_i} - \overline{X_2}) = 74778.346 \quad \sum(Y_i - \bar{Y})(X_{3_i} - \overline{X_3}) = 4250.900 \quad \sum(X_{2_i} - \overline{X_2})(X_{3_i} - \overline{X_3}) = 4796.000$$

4. Show the following:

$$\widehat{\beta_2} = \frac{\sum y_i(x_{2_i} - b_{2_3}x_{3_i})}{\sum(x_{2_i} - b_{2_3}x_{3_i})^2}$$

where b_{23} is the slope coefficient in the regression of $X2$ on $X3$.

5. Give the interpretations of the following regression equations in terms of the coefficient of X:

1. $Y = 0.5 + 200X$
2. $Y = 0.5 + 200\ln(X)$
3. $\ln(Y) = 0.5 + 0.02X$
4. $\ln Y = 0.5 + 2\,ln(X)$

6. Using data on a random sample of 30 houses in a particular neighborhood the following relation between the sales price in 1000$ as dependent variable Y and the interior area of the house in 100 square foot and the lot size in 100 square foot as the independent variables X_1 and X_2 respectively, is estimated with OLS (standard errors in parentheses)

$$\underset{(5.111)}{Y = 20.1} + \underset{(2.221)}{13.1X_1} + \underset{(3.419)}{9.12X_2} \qquad s = 2.22$$

a) What are the interpretations of the coefficient of X_1 and X_2?
b) Can X_2 be omitted from the relation? Why (not)?
c) A homeowner in the neighborhood wants to build an addition that adds 100 square foot to the interior area of the house. The cost of this addition is 7000$. Test whether the net gain (increase in sales price minus cost) is significantly different from 0 at the 5% level, i.e. with level of significance of 5%. Hint: As in a) pay attention to the units of measurement of the variables.
d) Test with level of significance 5% whether the coefficients of interior area and lot size are both equal to 0. The critical value of test statistic is 3.35. Given:

$$\sum_{i=1}^{30}(Y_i - \bar{Y})^2 = 300$$

7. Following Leamer's example, you estimate the demand for tangerines (substitute one for oranges) in China County over the period 1962 to 1991 and obtain the following results (standard errors in parentheses):

$$(1)\quad lnY_t = \underset{(1.38)}{6.18} - \underset{(0.29)}{0.52\,lnPT_t} + \underset{(0.25)}{0.61\,lnX_t} - \underset{(0.18)}{0.23\,lnPO_t} \qquad R^2 = 0.74,\ RSS = 23.19$$

where Y is the consumption of tangerines in thousands of bags, PT is the price of tangerines in Rs. per bag, X is per capita income in Rs.1000, and PO is the price or oranges in Rs. per bag. (Note that oranges are a substitute for tangerines.)
Using the same data, you also get the following results:

$$(2)\quad lnY_t = \underset{(1.23)}{6.62} - \underset{(0.30)}{0.71\,ln(PT_t/PO_t)} + \underset{(0.19)}{0.49\,ln(X_t/PO_t)} \qquad R^2 = 0.68,\ RSS = 28.54$$

a) How do you interpret the three slope coefficients from equation (1)? Comment on which of them are economically reasonable and which of them are statistically significant.

b) What is it about equation (1) that leads you to suspect multicollinearity?
c) What is the logic of estimating equation (2)? What restriction has been imposed?
d) Test the hypothesis that the restriction is statistically valid. Show all calculations. Does it matter whether the significance level is 5% or 1%?

Interesting read

In this article, Emmanuel Derman, a professor of financial engineering at Columbia University and author of *My Life as a Quant: Reflections on Physics and Finance* writes about modelling economic crises.

http://blogs.reuters.com/great-debate/2011/11/03/the-physics-of-an-economic-crisis/