Imperial College Business School

MSc Business Analytics

STATISTICS AND ECONOMETRICS

Mock Exam

Instructions

Answer all **FOUR** questions.

You are supplied with a formula sheet.

College approved calculators can be used.

Question 1 (20 Marks)

Answer the following questions. Be concise and to the point.

(a) What does R^2 measure? Is a regression useless if its R^2 is low? Explain.

(b) Suppose the true return to education model is $wage = \beta_0 + \beta_1 exper + \beta_2 exper^2 + u$. What is the consequence, if we estimate the model without the quadratic term $exper^2$?

(c) Explain in words the difference between a population regression function and an OLS regression line?

(d) What is unbiasedness and why is it a desirable property of the OLS estimators?

Question 2 (30 Marks)

Suppose you have tested a model of rent rates and student population in a college town

$$\begin{aligned}\log\widehat{(rent)} &= \underset{(.844)}{1.39} + \underset{(.039)}{.066}\log(pop) + \underset{(.081)}{.507}\log(avginc) + \underset{(.0017)}{.0056}\, pctstu + u \\ n &= 264,\quad R^2 = .458\end{aligned}$$

where $rent$ is the average monthly rent paid on rental units in a college town, pop denotes the total city population, $avginc$ denotes the average city income, and $pctstu$ denotes the student population as a percentage of the total population.

(a) Suppose you want to test $H_0\colon \beta_{\log(avginc)} = 0.5$ against $H_1\colon \beta_{\log(avginc)} \neq 0.5$. Construct the 95% confidence interval for the parameter $\beta_{\log(avginc)}$. (The 5% critical value for a two-tailed test is 1.96)

(b) Test the hypothesis in part (a) using the calculated confidence interval for $\beta_{\log(avginc)}$.

(c) Define $\theta = \beta_{\log(pop)} - 2\beta_{pctstu}$. If you want to test $H_0\colon \theta = 0$ against $H_1\colon \theta \neq 0$. Rewrite the regression model appropriately, so that you can directly obtain $\hat{\theta}$ and the standard error $se(\hat{\theta})$.

(d) Test the joint significance of $\beta_{\log(pop)}, \beta_{pctstu}$ and $\beta_{\log(avginc)}$ at the 5% level. That is, test $H_0\colon \beta_{\log(pop)} = 0, \beta_{pctstu} = 0$, and $\beta_{\log(avginc)} = 0$. (The 5% $F_{3,260}$ critical value is c = 2.60.)

Question 3 (30 Marks)

Consider the housing price model:

$$\log(price) = \beta_0 + \beta_1 \log(nox) + \beta_2 rooms + \beta_3 rooms^2 + u$$

where $price$ is the median housing price in a community, nox denotes the amount of nitrogen dioxide in the air in the community, in parts per million, and rooms is the average number of rooms in houses in the community.

(a) What is the effect of $rooms$ on $price$ in this model?

(b) Suppose the estimated equation is:

$$\widehat{\log(price)} = \underset{(.57)}{13.39} - \underset{(.115)}{.902} \log(nox) - \underset{(.165)}{.545}\, rooms + \underset{(.013)}{.062}\, rooms^2$$

$n = 506,\ \ R^2 = .603.$

Based on your answer in part (a), what is the predicted difference in median housing prices for a community with $rooms = 5$ and a community with $rooms = 6$?

(c) You want to test the joint hypothesis $H_0: \beta_1 = 0, \beta_2 = 0, \beta_3 = 0$ at the 5% level. The p-value associated with the *F* statistic for that test is 0.0234? What would you conclude?

(d) Is $\widehat{\beta_1}$ economically significant? Explain.

Question 4 (20 Marks)

Consider the following fixed-effects model for a panel dataset

$$y_{it} = \rho y_{i,t-1} + \beta x_{it} + \delta_i + u_{it}\ ,$$

where y_{it} is a continuous variable, x_{it} is an explanatory variable, δ_i is the fixed effect, u_{it} is the idiosyncratic error. $y_{i,t-1}$ is the lagged dependent variable, which is also used as an explanatory variable in the model.

(a) (10 Marks) Explain how to estimate the model using the first-differenced approach as discussed in the class.

(b) (10 Marks) Discuss why the first-differenced estimator always yields biased estimates for this model.