BUEC 333, Test 3

August 5, 2015, 14:30-17:20

Instructions

- On the front page of **your answer sheet**, write (i) your name; (ii) your student ID; (iii) your tutorial number.
- Once you **finish** this exam, **signal** it to us, and we will come to pick up your exam. Stay seated until somebody comes to collect your exam.
- No bathroom breaks after the first person has finished this test.
- No explanation = no points. A correct answer **with correct explanation** earns 1 point for each subquestion.
- For a "compute" question, an explanation can consist of starting from **an appropriate formula**, and working towards the correct numerical answer.

Mini fomula sheet

$$\hat{\beta}_1 \xrightarrow{p} \beta_1 + \rho_{Xu}\frac{\sigma_u}{\sigma_x},$$
$$\bar{R}^2 = 1 - \frac{n-1}{n-k-1}\frac{SSR}{TSS} = 1 - \frac{s_{\hat{u}}^2}{s_Y^2}$$

where ρ_{Xu} is the correlation between the included regressor X and the error term in the model with only one regressor, u, and σ_u and σ_x are the associated standard deviations of X and u. Furthermore, SSR stands for "sum of squares of residuals", TSS stands for "total sum of squares", s_u^2 is the sample variance of u, s_y^2 is the sample variance of y.

Questions

1. Data were collected from a random sample of 13 home sales from a community in 2003. Let $Price$ denote the selling price (in \$1000). BDR denote the number of bedrooms. $Bath$ denote the number of bathrooms. $Hsize$ denote the size of the house (in square feet). $Lsize$ denote the lot size (in square feet). Age denote the age of the house (in years), and $Poor$ denotes a binary variable that is equal to 1 if the condition of the house is reported as "poor". An estimate regression for the for the **log** of house prices yields:

$$\begin{aligned}
\widehat{\ln(Price)} &= -9.2 + 0.047BDR + 0.23Bath + 0.16\ln(Hsize) + 0.002\ln(Lsize) \\
&\quad + 0.0091Age - 0.11Poor, \\
\bar{R}^2 &= 0.50, \\
TSS &= 100.
\end{aligned}$$

(a) Suppose that a homeowner converts part of an existing family room into a new bathroom. What is the expected increase in the value of the house?

(b) Suppose that a homeowner adds a bathroom, which also increases the size of the house from 2000 square feet to 2200 square feet. What is the expected increase in the value of the house?

(c) Interpret the estimated coefficient on $Poor$. Use the words "house price".

(d) Compute the R^2.

2. A researcher plans to study the causal effect of police on crime using data from a random sample of U.S. counties. He plans to regress the county's crime rate on the (per capita) size of the country's police force.

   (a) Explain why this regression is likely to suffer from omitted variable bias. Which variable would you add to the regression?

   (b) For the variable under (a), explain why it satisfies the two conditions for omitted variable bias to occur.

   (c) Using your answers to (a) and (b) to determine whether the regression will likely over- or underestimate the effect of police on the crime rate.

3. This problem is inspired by the study of the gender gap in top corporate jobs in Bertrand and Hallock (2001). The study compares total compensation among top executives in a large set of U.S. public corporations in the 1990s.

   (a) Let $Female$ be an indicator variable that is equal to 1 for females and to 0 for males. A regression of the logarithm of earnings onto $Female$ yields

$$\log(\widehat{Earnings}) = 6.48 - 0.44Female$$

   where the estimated regression coefficient -0.44 has a standard error of 0.05. Explain what the -0.44 means.

   (b) Does this regression suggest that there is gender discrimination? Explain.

   (c) Two new variables are added to the regression: $\log(MarketValue)$, where $MarketValue$ is a measure of firm size, in millions; and $Return$, the stock return, in percentage points. The resulting estimated regression line is

$$\log(\widehat{Earnings}) = 3.68 - 0.28Female + 0.37\log(MarketValue) + 0.004Return$$

   where the standard errors for the three regressors are 0.04, 0.004, and 0.003, respectively. The coefficient estimate for $Female$ has changed from -0.44 to -0.28. Why has it changed?

   (d) Are large firms more likely to have female top executives than small firms? Explain.

4. Using data on the SAT scores, personal and high school characteristic of 4000 U.S. students, we estimate the following regression:

$$\widehat{SAT} = 1028 + 19.30Hsize - 2.19Hsize^2 - 45.09Female - 169.81VisMin + 62.31 \times Female \times VisMin$$

where SAT is the student's SAT score, $Hsize$ is the number of students in their high school graduating class (in hundreds), $Female$ is a gender dummy (1 if student is female, 0 otherwise) and $VisMin$ is a dummy variable that is 1 if the student is part of a visible minority, and 0 otherwise.

(a) Can you estimate these coefficients using OLS?

(b) Interpret the coefficients on $Hsize$ and $Hsize^2$.

(c) What is the best high school graduating class size for achieving a maximal SAT score, all else equal?

(d) Interpret the coefficient estimate of -45.09 on $Female$.

(e) Interpret the coefficient 62.31 on $Female \times VisMin$.

5. A researcher is interested in the effect of military service on human capital. He collects data from a random sample of 4000 workers aged 40 and runs the OLS regression

$$Y_i = \beta_0 + \beta_1 X_i + u_i,$$

where Y_i is the worker's annual earnings and X_i is a binary variable that is equal to 1 if the person served in the military and is equal to 0 otherwise.

(a) Explain why the OLS estimates are likely to be unreliable.

(b) During the Vietnam War there was a draft, where priority for the draft was determined by a national lottery. (Birthdates were randomly selected and ordered 1 through 365. Those with birthdates ordered first were drafter before those with birthdates ordered second, and so forth.) Explain how the lottery might be used as an instrument to estimate the effect of military service on earnings.

(c) Explain why you think that the instrument in (b) satisfies the conditions an instrumental variable needs to satisfy. First state the conditions, then explain why you think they hold for this application.

6. Evans and Schwab (1995) studied the effect of attending a Catholic high school on college grades (measured by *college*). Let $CathHS$ be a binary variable equal to one if the student attends a Catholic high school. The associated regression model is:

$$college = \beta_0 + \beta_1 CathHS + u.$$

(a) Why might $CathHS$ be correlated with u?

(b) Let $CathRel$ be a binary variable equal to one if the student is Catholic. Discuss the requirements needed for this to be a valid instrumental variable for $CathHS$.

(c) Being Catholic has a significant effect on attending a Catholic high school. Do you think $CathRel$ is a convincing instrument for $CathHS$? Explain.