BUEC 333, Test 3

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

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:

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

   (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? **23%**

   (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? **1x0.23*100% + 10%*0.16 = 23% + 1.6% = 24.6%.**

   (c) Interpret the estimated coefficient on *Poor*. Use the words "house price". **All else equal, if a house goes from not being reported as "poor" to being reported as "poor", then the expected decrease in the house prices is 11%**

   (d) Compute the R^2. **n=13, k=6. R-bar-sq = 1 - (n-1)/(n-k-1) * SSR / TSS so we learn that SSR = 0.25*TSS = 25. So R-squared = 0.75.**

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. **Also, see the answers to the problem set in "Introduction to Linear Regression, q.3".**

   (a) Explain why this regression is likely to suffer from omitted variable bias. Which variable would you add to the regression? **Several correct answers. In what follows, I call the variable that you came up with "vA".**

   (b) For the variable under (a), explain why it satisfies the two conditions for omitted variable bias to occur. **argue that (i) cov(police,vA)$\neq 0$ and that (ii) vA has a non-zero effect on crime after taking into account "police".**

   (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. **use the OV formula. use your answers to (i) and (ii) under (b) to argue for the sign of ρ_{xu}. For example, if you found that your vA has a positive correlation with police, and has a positive effect on crime, then $\rho_{xu} > 0$ and $\hat{\beta}$ will likely overestimate β_1.**

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. **see problem set answer, "Topics in Linear Regression", question 10.**

   (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

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

   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

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

   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:

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

   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? **Yes. Although it looks nonlinear, you can compute the interactions and the squares and run this regression.**

   (b) Interpret the coefficients on $Hsize$ and $Hsize^2$. **Several correct answers are possible. Example of an incorrect answer: if Hsize increase by 1, we expect SAT scores to go up by 19.30 ceteris paribus. Another incorrect answer: same as before, without "ceteris paribus".**

   (c) What is the best high school graduating class size for achieving a maximal SAT score, all else equal? **Find the peak of the quadratic curve 19.30x-2.19x^2, which occurs at x=19.30/4.38. Remember that class size is reported in hundreds.**

   (d) Interpret the coefficient estimate of -45.09 on $Female$. **All else equal, for somebody who is not a visible minority, the expected SAT scores of women are expected to be 45.09 points lower.**

(e) Interpret the coefficient 62.31 on $Female \times VisMin$. **All else equal, for somebody who is a visible minority, the expected SAT scores of women are expected to be 62.31-45.09≈17.22 points higher.**

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. **See problem set answers, "Instrumental variables".**

(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. **Example of an incorrect answer: include the instrument (draft/lottery number) as an additional variable in the regression.**

(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? **Out of many answers, here is one. A catholic school can be thought of as a private school. There are all kinds of reasons that parents that have the required resources send their kids to private school. These reasons, and the presence of resources (high-income) in the family, may be correlated with success in college.**

(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 $CathHS$. **It needs to be correlated with the event of attending a Catholic high school. The only effect it should have on performance in college is through attending a catholic highschool.**

(c) Being Catholic has a significant effect on attending a Catholic high school. Do you think $CathRel$ is a convincing instrument for $CathHS$? Explain. **Several answers are acceptable. If you believe that being Catholic has no effect on performance in college, other than through making it more likely that you attend Catholic High school, then it would be a good instrument. However, it is possible to make arguments against the validity of the instruments as well.**