Final Exam - 8:30 a.m. group

Eco 231 - Undergraduate Econometrics

12/18/2011 (Prof. Carolina Caetano)

INSTRUCTIONS

Reading and understanding the instructions is your responsibility. Failure to comply may result in loss of points, and there will be no leniency on that respect.

1. You have received one booklet. This booklet contains the exam instructions, the exam questions, and the numbered pages where you will answer the questions. You must sign this booklet in the space provided.

2. This exam has 2 questions. Question 1 is worth 75% of the grade (all items are worth the same), and question 2 is worth 25% of the grade. You have nearly 3 hours to answer it. You have until 11:25 a.m to answer it.

3. You must answer each question exactly in the space provided for it. You may use the back of the pages if they are empty. If you answer a question out of the order, or otherwise not on the space provided for it, your question will not be graded. If you need more space, you must ask for extra paper from the TA. It is your responsibility at the end of the exam to staple the extra page exactly in the right place in your exam. You may ask for draft paper if you like.

4. You are not allowed the use of notes, cheat sheets, calculators, or electronic devices of any kind. Turn your cell phone off, and put it away. If you did not bring a watch, check the board. The TAs will write down the time in the board every 15 minutes. If your answers are unclear or illegible you may lose points. You may answer in pencil.

5. You may return the exam at any time until 11:20 a.m. If you finis after that, you must remain seated. Do not get up when the TA announces the time is up. Remain seated and follow the TA's instructions.

6. Sign and print your name below. You signature demonstrates that you have read and understood the instructions. An exam without the signature will not be graded.

1. **Name (print):** _____

2. **Class time:** _____

3. **Signature:** _____

Questions 1-18 below are worth the same. Here is a breakdown of what I anticipate are the difficulty levels of the questions:

- Easy: 1, 2, 3, 5, 6,
- Medium-easy: 7, 8, 10, 12, 13, 18
- Medium: 4, 11, 15, 16, 17
- Hard: 9, 14

Question 1: Suppose that someone would like to understand how television watching may affect child development. The proposed model is:

$$tscore = \beta_0 + \beta_1 tvhours + \beta_2 age + \beta_3 motheduc + \beta_4 fatheduc + \beta_5 sibs + u \qquad (1)$$

where $tscore$ is the child score at a standardized test which is age-specific, $tvhours$ is the number of hours the child spends per week watching TV, age is the child's age, $motheduc$ and $fatheduc$ are the years of education of the mother and father of that child respectively, and $sibs$ is the number of siblings of that child.

1. What is the question which the model above tries to answer? Does it answer all the ways in which television watching affects child development? Are there other models you would consider?
2. Why is the relationship proposed by the model a hard one to obtain? In other words, why is the question implicitly proposed in the model hard?
3. Why is $sibs$ included in the model?
4. The tests are age-specific, so a five year old child will be taking the 5 year old test, a 6 year old child will be taking a different test, and so on. Why then is age part of the regression?
5. Argue that the variable $white$, which is equal to 1 if the child is white, and 0 otherwise, should be included in the model. As a consequence of this omission, do you think that $\hat{\beta}_1$ is higher or lower than the true β_1?
6. Suggest 5 other omitted variables. Here you must make a judgement call about how to explain your choices. Some variables that you suggest may be obvious examples of omitted variable. In this case, just naming them is enough. However, you may have an unusual idea about a variable that has been omitted. In this case, you must argue briefly why it should be included, just enough so that the TA understands the reason

for your choice. The focus of this question is not to show that you can argue that a variable has been omitted, but rather to measure whether you can quickly think about potential problems with a model design.

7. Suppose that you have a random sample

$$\{tscore_i, tvhours_i, age_i, motheduc_i, fatheduc_i, sibs_i\}_{i=1}^{n}.$$

How would you answer the question: "Does TV watching affect child development?" scientifically?

8. How would you test whether the assertion: "Mother and father are equally important in the way they affect the child development." is true?

9. How would you test whether the assertion: "The child's socio-economic characteristics do not have any impact in how TV watching affects child development." is true?

10. The true number of hours a child spends per week watching TV is hard to measure. Suppose that the data comes from a survey that asked parents: "How many hours per week does your child spend watching TV?" Hence, though the true model should use $tvhours^*$, which is the actual average number of hours the child spent watching TV, the actual estimation uses $tvhours$ = answer to the survey question. $tvhours$ is an imperfect measure of the true variable $tvhours^*$. Why? Can you think of better ways to measure $tvhours^*$, either a different way to pose the question, or a different survey design altogether?

11. In the classical error-in-variables model, $x = x^* + e_1$, where $Cov(x^*, e_1) = 0$. Do you think that the model that uses $tvhours$ collected as described in the previous question is a classical error-in-variables model?

12. Suppose that indeed $tvhours = tvhours^* + e_1$, where $Cov(tvhours^*, e_1) = 0$. What are the consequences for the estimation of β_1?

13. Suppose that our data set comes from the Monroe county. The data was collected in the cities of Rochester, Henrietta, and Pittsford, and the data specifies where each child lives. The new model is

$$tscore = \beta_0 + \beta_1 tvhours + \beta_2 age + \beta_3 motheduc + \beta_4 fatheduc + \beta_5 sibs + \beta_6 P + \beta_7 H + u,$$

where $P = 1$ if the child resides in Pittsford, zero otherwise, and $H = 1$ if the child resides in Henrietta, zero otherwise. Interpret $\beta_0 + 5\beta_2 + 12(\beta_3 + \beta_4)$ in this model.

14. The child residence variables P and H can be used as a proxy to several omitted variables. It is an imperfect proxy, but it does provide some information. For example, you showed that *white* should be included in the model. Examine how including P and H accounts for some of the effect of the *white* variable, and what kind of bias still remains.

15. Suppose the survey was conducted in schools. Parents were asked to fill the survey while attending a PTA (Parent-Teacher Association) meeting. What are some of the potential problems originating from this data collection method? Discuss how they may affect the estimate $\hat{\beta}_1$.

16. Suppose that the survey was collected in May of the years of 2010 and 2011. The tests are taken in April, so the test scores collected correspond to the current school-year results. However, in 2010 the Disney channel unexpectedly stopped transmissions for the months of September, October, and November. Argue that the variable $D = 1$, if the observation was collected in the May 2011 survey, and $D = 0$, if the observation was collected in the May 2010 survey is an instrumental variable in the model.

17. Describe the 2SLS estimator in this model (no x_j's and y, use the variables in this model) using D as the instrumental variable for $tvhours$.

18. What are the assumptions required for $\hat{\beta}_1^{2SLS}$ described above to be unbiased?

Question 2: Bullying in high school is perceived as very bad. For the bullied child, the consequences seem clear, for example reduced self-esteem, anxiety, and sometimes even suicide. However, even in a school where bullying is common, not every child is bullied. The majority are very likely not bullied. Suppose that you want to study the effects of bullying for the school at large: how do schools with different levels of bullying compare in terms of education effectiveness? There are many ways in which education effectiveness can be measured. I am proposing that you answer the following question: how does bullying in high schools affects the likelihood of its students going to college? Observe the subtlety of the question: we are not asking whether a child suffering from bullying is more or less likely to go to college. We are asking whether a school where children are bullied more will send proportionally more, less, or the same number of its students to college. Examine why this question is so hard, beginning with the definition of bullying itself. What does "more bullying" mean? It could mean that proportionally more children are bullied, it could mean that bullying is more intense, more violent, etc. Describe how you would go about answering the question, from the determination of how bullying should be measured, through the model and its assumptions, the data set choice and possible problems, the

estimation, hypothesis tests, the analysis of results, and finally the consequences of the conclusions to which your study arrives.

Answers

Q1 (1)

Q1 (2)

Q1 (3)

Q1 (4)

Q1 (5)

Q1 (6)

Q1 (7)

Q1 (8)

Q1 (9)

Q1 (10)

Q1 (11)

Q1 (12)

Q1 (13)

Q1 (14)

Q1 (15)

Q1 (16)

Q1 (17)

Q1 (18)

Q2