Final Exam - Booklet 1

Eco 231 - Undergraduate Econometrics

Group 1 - 8:30 A.M. - 12/17/2010 (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 two booklets. This booklet contains the exam instructions and the exam questions. The second booklet contains the numbered pages where you will answer the questions. You must sign booklet 2 in the space provided.

2. This exam has 5 questions and it is worth 100 points.

3. You have until 11:25 a.m. to complete this exam. You may return the exam at any time until 11:20 a.m. If you finish this exam after 11:20 a.m., you must remain seated. Do not get up when the TA announces the time is up. Remain seated and follow the TA's instructions.

4. You must answer each question in the Booklet 2, 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 in the second booklet, 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.

5. You are not allowed to use notes, cheat sheets, calculators, or electronic devices of any kind. If your answers are unclear or illegible you may lose points. You may answer in pencil.

6. At the end of the exam you must hand booklets 1 and 2.

7. Sign and print your name in booklet 2. You signature demonstrates that you have read and understood the instructions. An exam without the signature will not be graded.

Question 1 (16 points):

1.a) (4 points) What is the variance of the multivariate OLS estimator $\hat{\beta}_j$ under heteroskedasticity?

1.b) (4 points) Write in detail all the assumptions necessary to derive the variance of the OLS estimator $\hat{\beta}_j$ under heteroskedasticity.

1.c) (4 points) Explain all the terms that affect the variance of the OLS (under heteroskedasticity). Try your best, you will be judged by how well you show that you understand each term.

1.d) (4 points, hard) How does the variance of the OLS under homoskedasticity compare with the variance of the OLS under heteroskedasticity? Attention: the BLUE property of the OLS under homoskedasticity has nothing to do with this question.

Question 2 (24 points): Suppose that you wish to estimate the effect of class attendance on student performance. A basic model is

$$G_f = \beta_0 + \beta_1 C + \beta_2 GPA + \beta_3 GPA^2 + u.$$

where G_f is the final grade in the course, C is the percentage of classes attended, and GPA is the GPA in previous courses.

2.a) (4 points) Why did we include GPA and GPA^2 in the model?

2.b) (4 points) Can you suggest one omitted variable in the equation above? Argue that the variable you suggested is indeed omitted.

2.c) (4 points) Which variable(s) in the model are prone to measurement error, and why? What are the consequences for the estimation under the classical error-in-variables model?

2.d) (4 points) Let D denote the distance from the student living quarters to the lecture hall. Can you use D as an instrumental variable? Defend your answer.

2.e) (4 points) Suppose that only C is endogenous, and that both D and D^2 are instrumental variables in this model. Write the equations for the derivation of the 2SLS estimator.

2.f) (4 points) Which assumptions must the variables in the model satisfy for the 2SLS estimators to be unbiased? Express the assumptions (in detail) in terms of the variables specific to this model. Do not use generic y, x_1, z, etc.

Question 3 (20 points): Suppose that you are studying the marriage market, and you are interested in the determinants of divorce from a socio-economic point-of-view. You model the probability of divorce as

$$P(D = 1) = \beta_0 + \beta_1 Ed_M + \beta_2 Ed_F + \beta_3 In + \beta_4 B_M + \beta_5 B_F + u.$$

where D is a dummy which equals 1 if the couple has divorced, Ed_M is the education of the male in the couple, and Ed_F is the education of the female in the couple at the time of marriage, In is the total family income of the couple at the time of marriage, and B_M and B_F are dummies that are equal to one if the male or the female in the couple is black.

3.a) (2 points) What is the interpretation of β_0?

3.b) (2 points) How do you expect β_0 to change if instead of using a dummy for black male B_M, we used a dummy for white male W_M?

3.c) (3 points) How would you test the hypothesis that race of the female does not impact the probability of divorce at the 95% confidence level?

3.d) (3 points, hard) How would you test the hypothesis that, independent of the specific race of the male and the female, interracial marriages have the same probability of divorce at the 95% confidence level?

3.e) (3 points) Your model may be accused of violating MLR 4. In fact, someone argued that the probability of divorce is also affected by how good a "catch" one of the partners is. Let's analyze the husband. The argument is that the better looking the husband, the more the female will perceive the union as a desirable one, decreasing the probability of divorce. Do you agree that the level of beauty of the male in a couple is an omitted variable in this model?

3.f) (4 points) Suppose that you would like to satisfy the critic that argued that beauty of the husband should be included. It is hard to have a good measurement of beauty, but you have the height of the husbands in your data-set. If you use the height of the husband as a proxy for beauty, how do you expect $\hat{\beta}_3$ to change?

3.g) (3 points) Is height a good proxy for beauty? How can your proxy be improved? Which other variables could you use to account for beauty, and how would they impact $\hat{\beta}_3$?

Question 4 (20 points): Consider the model

$$y = \beta_0 + \beta_1 x + u$$

where x is endogenous and z is an IV.

4.a) (4 points) Show that

$$\beta_1 = \frac{Cov(z, y)}{Cov(z, x)}.$$

4.b) (4 points) Remember that

$$Cov(A, B) = E\Big[(A - E(A))(B - E(B))\Big]$$

Use the method of moments to derive the formula of the IV estimator $\hat{\beta}_1$.

4.c) (4 points) What is the variance of $\hat{\beta}_1$ from the last question?

4.d) (4 points) Suppose that w is also an IV in this model. However,

$$Cov(w, x) = \frac{1}{2} Cov(z, x).$$

What is the variance of $\hat{\beta}_1$ calculated using w as an IV instead of z?

4.e) (4 points) Rank the following options from best (# 1), to worst (# 3), and defend your ranking.

A - Use z as the IV

B - Use w as the IV

C - Use z and w as IVs

Question 5 (20 points): Public school is a public good. Everyone in the school attendance area can send their children to the public school of that area. How much do parents value public school quality, i.e. how much would the average parent be willing to pay for an extra unit of school quality?[1]

Economists usually determine the value of goods by looking at their price. The price reflects exactly the amount that people are willing to pay for that good, and hence its

[1] There are many ways to measure school quality. For example, we may consider an index which takes into account the ratio of students to teachers, the teachers' qualifications, and the students' scores in standardized Math and English tests.

economic value. For example, the value of one extra unit of school quality at a private school is given by the comparison of the tuition among private schools with quality one unit apart. Public schools are free. You cannot buy your way into a public school, and therefore we cannot gauge its value by looking at how much people are paying for public schools of different quality.

Economists have tried to measure the value of public school by looking at the house prices in the school attendance areas. In order to attend a given public school, the family must reside in that school's attendance area. The idea is that people do pay for public school. They pay for it inside of the house price. House prices reflect the house characteristics (number of bedrooms, square footage, year built etc.), neighborhood characteristics (school district quality, violence, highway proximity, public transportation access, etc.), and neighbors' characteristics (income, education, race, age, number of children per family, etc.) As you can see, the quality of the school district is one of the amenities that characterize a house, and this is such a well known fact that when realtors advertise a house in a good school district such as Brighton or Pittsford, they always stress that fact in their flyers.

Most families would like to live in the best school districts, which increases the demand for houses in those areas. This is partly why houses are so expensive in the best school districts. The problem is that the best school districts are usually safe areas, where houses are pretty, and income and education levels of neighbors are high, among other amenities. All of these positive characteristics increase the demand for houses in those areas as well, and it is not clear how much of the extra price is due to the public school, and how much is due to the other positive characteristics of the neighborhood.

Can you figure out how much people value school quality by looking at the house prices in different school districts? How can you disentangle the part of the price that is due to the school quality from the part of the price that is due to the other house characteristics? How would you come up with this information? Be very specific. Describe your model, which data you need, how you imagine you would search for it, and your procedure step by step. Describe also the problems that you suppose that you will encounter.

Booklet 2 - Final (12/17/10) - Eco 231W

Group 1 - 8:30 A.M.

1. **Name:** _____
2. **Student ID:** _____
3. **Class time:** _____

May your holidays (and well deserved rest) be great!

1.a) (4 points)

1.b) (4 points)

1.c) (4 points)

1.d) (4 points)

2.a) (4 points)

2.b) (4 points)

2.c) (4 points)

2.d) (4 points)

2.e) (4 points)

2.f) (4 points)

3.a) (2 points)

3.b) (2 points)

3.c) (3 points)

3.d) (3 points)

3.e) (3 points)

3.f) (4 points)

3.g) (3 points)

4.a) (4 points)

4.b) (4 points)

4.c) (4 points)

4.d) (4 points)

4.e) (4 points)

5) (20 points)