Stockholm University

Department of Economics

Course name:	Empirical Methods in Economics
Course code:	EC2404
Type of exam:	Retake
Examiner:	Ferenc Szucs
Number of credits:	7.5
Date of exam:	Tuesday April 30th, 2019
Examination time:	3 hours (9:00-12:00)
Aids:	No aids are allowed.

Write your identification number on each answer sheet (the number stated in the upper right hand corner on your exam cover).

Start each new questions on a new answer sheet.

Explain notions/concepts and symbols. If you think that a question is vaguely formulated, specify the conditions used for solving it. Only legible exams will be marked.

The exam consists of 5 questions. The first two contain multiple choice questions, worth 4 points each. Questions 3-5 are worth 20 points each. If you submitted 4 accepted home assignments you do not need to solve question 5 (you will get full credit anyway).

The maximum total point is 100. For the grade E 45 points are required, for D 50 points, C 60 points, B 75 points and A 90 points.

Your results will be made available on your Ladok account (www.student.ladok.se) within 15 working days from the date of the examination.

Good luck!

Stockholms universitet

Question 1 (20 points)

1) Consider the following equation: $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$, where we are mostly interested in β_1. Leaving X_2 out of the equation is a problem only if
   A) X_{2i} is correlated with the error term.
   B) $\beta_1 \neq 0$.
   C) $\beta_2 \neq 0$ and X_{2i} is correlated with X_{1i}.
   D) X_{2i} is correlated with X_{1i} and $\beta_2 = 0$.
2) Consider the following equation: $Y_i = \beta_0 + \beta_1 X_i + u_i$, an omitted variable in u_i would imply that
   A) $E[Y_i|X_i] = E[Y_i]$
   B) $E[X_i|u_i] \neq 0$
   C) $E[u_i] \neq 0$
   D) $E[u_i|X_i] \neq 0$
3) Heteroskedasticity is
   A) always a problem when our sample is not i.i.d.
   B) caused by the correlation of different control variables.
   C) making the OLS standard errors incorrect.
   D) related to the measurement error of the main explanatory variable.
4) Assume we randomly allocate students into a treatment group and a control group, where the treatment is taught by a new curriculum while the control with the traditional. We are interested in the effect of the treatment on test scores, which one of the following is a bad control?
   A) Education of the parents.
   B) Time spent solving homework.
   C) Distance of the student's home from the school.
   D) Qualifications of the teacher.
5) Consider the following equation: $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{2i}^2 + \beta_4 X_{1i} X_{2i} + u_i$, what is the marginal effect of X_{1i} on Y_i?
   A) $\frac{\partial Y}{\partial X_1} = \beta_1 + \beta_3$.
   B) $\frac{\partial Y}{\partial X_1} = \beta_1 + 2\beta_4 X_{2i}$.
   C) $\frac{\partial Y}{\partial X_1} = \beta_1 + \beta_3 + \beta_4 X_{1i} X_{2i}$.
   D) $\frac{\partial Y}{\partial X_1} = \beta_1 + 2\beta_3 X_{1i} + \beta_4 X_{2i}$.

Stockholms universitet

Question 2 (20 points)

1) If Y also affects X, this is called
   A) simultaneity.
   B) extreme multicollinearity.
   C) heteroskedasticity.
   D) omitted variables problem.
2) A proxy variable is used
   A) to deal with extreme multicollinearity.
   B) to improve the precision of the estimated coefficient of the explanatory variable of interest.
   C) to decrease or remove the omitted variables bias from the estimated coefficient of the variable of interest.
   D) as an instrument for an endogenous variable
3) Regression discontinuity requires that,
   A) There is a discontinuity in the without-treatment potential outcome at the threshold.
   B) There is no discontinuity in the probability of being treated at the threshold.
   C) There is a discontinuity in pre-determined control variables at the threshold.
   D) There is no discontinuity in the density of the running variable at the threshold.
4) Which of the following statements are true?
   A) A measurement error in Y_i is similar to the omitted variables problem in making our point estimates inconsistent.
   B) A measurement error in X_i is similar to the omitted variables problem in making our point estimates inconsistent.
   C) A measurement error in Y_i is similar to heteroskedasticity in leaving our point estimates consistent but making our t-tests invalid.
   D) A measurement error in X_i is not a big deal it just makes our point estimates less precise.
5) The interpretation of the slope coefficient in the following equation $\log(Y_i) = 3 + 0.15\, X_i$,
   A) A 1 unit change in X_i is associated with 15% change in Y_i.
   B) A 1% change in X_i is associated with 0.15% change in Y_i.
   C) A 1% change in X_i is associated with 15% change in Y_i.
   D) A 1% change in X_i is associated with 0.15 units change in Y_i.

Stockholms universitet

Question 3 – Regression Discontinuity Design (20 points)

In many countries a substantial fraction of municipal resources originates from the national budget in the form of funds. Assume we are interested whether national governments favor aligned local governments (aligned means the same party rules both national and local level). We would like to use an RD design with mayoral elections to see whether governments give more money to places electing mayors from the same party which gives the government. We have access to the last municipal election results of 2000 municipalities, the per capita amount of grants, and a list of other observables like population, average income, average schooling, vote share of the government party in the last national election at the municipality level.

1. Explain why we can't measure the effect of aligned mayors by simply regressing fund on an indicator for aligned mayor? (3 points)
2. Explain the intuition of a RDD with using the vote share of the government party's mayoral candidate as a running variable. (4 points)
3. Explain in detail what equation would you run. (5 points)
4. How can we test there is no manipulation at the threshold? (3 points)
5. What is the identifying assumption? (3 points)
6. How would you test this assumption? (2 points)

Stockholms universitet

Question 4 – Instrumental variables (20 points)

We are interested in the effect of having more children on women's labor market participations. Assume you have access to the registry data of all households in Sweden in 2010. You observe many demographic variables of the members of the households and their labor market participation. Since you have access to the birth dates of children in the household you can identify twins. You want to use twins as an instrument for the number of children a woman has.

1. Why do you think you need an IV to measure the effect of childbearing on labor market participation? (3 points)
2. Why having twins might be a good instrument? (3 points)
3. How would you formulate the first stage? Spell out the exact equation you would run. (4 points)
4. Outline two different ways you would estimate the treatment effect of going into college on income (hint: using first stage and reduced form or the fitted values). Be specific and describe the equations. (6 points)
5. How would you provide supporting evidence for the instrument is good? (4 points)

Stockholms universitet

Question 5 – Angrist and Evans (1998) paper

This is the abstract from the paper titled "Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size":

"Research on the labor-supply consequences of childbearing is complicated by the endogeneity of fertility. This study uses parental preferences for a mixed sibling-sex composition to construct instrumental variables (IV) estimates of the effect of childbearing on labor supply. IV estimates for women are significant but smaller than ordinary least-squares estimates, The IV are also smaller for more educated women and show no impact of family size on husbands' labor supply. A comparison of estimates using sibling-sex composition and twins instruments implies that the impact of a third child disappears when the child reaches age."

1. What is the intuition behind the sibling-sex composition IV (why does it affect family size, why doesn't affect other things)? (6 points)
2. Why does it measure a local average treatment effect? Is it a different local effect than the one measured by the twins IV, if yes how? (8 points)
3. How is their IV used in a Wald estimation? Give me the exact formula they use. (6 points)