![Stockholm University logo]

Stockholm University

Department of Economics

Course name:	Empirical Methods in Economics 2
Course code:	EC2404
Type of exam:	Main
Examiner:	Ferenc Szucs
Number of credits:	7.5
Date of exam:	Monday March 18th, 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: $Y_i = \beta_0 + \beta_1 X_i + u_i$, if there are no omitted variables which assumption holds?
   A) $E[Y_i|X_i] = E[Y_i]$
   B) $E[X_i|u_i] = 0$
   C) $E[u_i|X_i] = 0$
   D) $E[u_i] = 0$

2) Reverse causality means,
   A) that the outcome variable also affects the explanatory variable.
   B) that we left out an important variable from the equation.
   C) that the coefficient of the variable of interest is biased by a bad control.
   D) that OLS would measure the causal effect of the outcome on the explanatory variable of interest.

3) Which of the following statements are true?
   A) It is better to have measurement error in X_i than in Y_i because the latter effects the coefficients of all explanatory variables.
   B) It is better to have measurement error in Y_i than in X_i because the former does not make coefficients inconsistent while the latter does.
   C) Measurement error, both in X_i and in Y_i, is a serious problem which results into inconsistent estimates of the coefficients.
   D) It is not a big problem to have a measurement error in X_i or in Y_i, we just need to have a large enough sample to get precise estimates.

4) 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) $\beta_2 \neq 0$ and X_{2i} is correlated with X_{1i}.
   B) X_{2i} is correlated with the error term.
   C) Both $\beta_1 = 0$ and $\beta_2 = 0$.
   D) X_{2i} is correlated with X_{1i} and $\beta_2 = 0$.

5) A proxy variable is useful for
   A) generating exogenous variation in the explanatory variable of interest.
   B) getting a sense of the effect of the omitted variable it stands for.
   C) improving the precision of the estimated coefficient of the explanatory variable of interest.
   D) decreasing or removing the omitted variables bias from the estimated coefficient of the variable of interest.

Stockholms
universitet

Question 2 (20 points)
1) The interpretation of the slope coefficient in the following equation $\log(Y_i) = 3 + 0.15 \log(X_i)$,
   A) A 1% change in X_i is associated with 0.15% change in Y_i.
   B) A 1 unit 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.
2) Heteroskedasticity is
   A) the problem of having linearly dependent explanatory variables in our regression equation.
   B) making the OLS standard errors incorrect.
   C) a form of the endogeneity problem.
   D) the same as having measurement error in the explanatory variables.
3) A bad control in an RCT means,
   A) a control variable which is uncorrelated with the outcome.
   B) a control variable which is affected by the treatment.
   C) a control variable which is uncorrelated with the treatment.
   D) a control variable which affects the outcome.
4) Consider the following equation: $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{1i}^2 + \beta_4 X_{1i} X_{2i} + u_i$, what is the marginal effect of X_{1i} on Y_i?
   A) $\frac{\partial Y}{\partial X_1} = \beta_1 + \beta_3$.
   B) $\frac{\partial Y}{\partial X_1} = \beta_1 + 2\beta_3$.
   C) $\frac{\partial Y}{\partial X_1} = \beta_1 + \beta_3 + \beta_4 X_{1i} X_{2i}$.
   D) $\frac{\partial Y}{\partial X_1} = \beta_1 + 2\beta_3 X_{1i} + \beta_4 X_{2i}$.
5) We cannot use a regression discontinuity design if,
   A) There is no discontinuity in the density of the running variable at the threshold.
   B) There is no discontinuity in the without-treatment potential outcome at the threshold.
   C) There is no discontinuity in the probability of being treated at the threshold.
   D) There is no discontinuity in pre-determined control variables at the threshold.

Question 3 – Panel data methods (20 points)

Many countries obligate sellers to show after tax prices on price tags. The reasoning behind such legislation is that people don't fully take into account taxes if before tax prices are shown. We would like to test whether consumers are really prone to this mistake and making after tax prices more salient helps them to overcome their bias. Assume there are two countries next to each other (Country 1 and Country 2). Before 2010, both countries allow retailers to show before tax prices on price tags but from 2011 Country 1 obligates them to show after tax prices (Country 2 sticks to its old policy). You have access to the monthly sales data of 100 retail stores from both countries for 2005-2015.

1. Explain why we can't measure the effect of after tax prices by regressing sales on whether price tags show before or after tax prices. (2 points)
2. Explain what variation in the data can be used to overcome the problems you outlined above. (3 points)
3. Explain in detail what equation would you run. (5 points)
4. What is your main identifying assumption behind that equation? (3 points)
5. How would you test this assumption? (3 points)
6. Assume you get access to the sales data of 100 stores in Country 3 which introduced the after tax prices law in 2013. How would you modify your empirical strategy to exploit this data too? (4 points)

Question 4 – Instrumental variables (20 points)

We are interested in measuring the effect of finishing college on latter income. Assume you have access to the registry data of a full cohort finishing high school in 1990. You observe whether they went to college and also their income at the age of 40 and a bunch of family background variables (such as the schooling and occupation of parents). You also know their addresses at 1990 when they finished high school so you can create a variable showing their distance from the closest College. This variable is a promising instrument for going into college.

1. Why do you think you need an IV to measure the effect of schooling on income? (2 points)
2. Why distance from the closest college might be a good instrument? What are the concerns about it? (4 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. Be specific and describe the equations. (6 points)
5. How would you provide supporting evidence for the instrument is good? (4 points)

Question 5 – Brollo et al. (2013) paper

This is the abstract from the paper titled "The Political Resource Curse":
"This paper studies the effect of additional government revenues on political corruption and on the quality of politicians, both with theory and data. The theory is based on a political agency model with career concerns and endogenous entry of candidates. The data refer to Brazil, where federal transfers to municipal governments change exogenously at given population thresholds, allowing us to implement a regression discontinuity design. The empirical evidence shows that larger transfers increase observed corruption and reduce the average education of candidates for mayor. These and other more specific empirical results are in line with the predictions of the theory."

1. Why is it hard to measure the effect of revenues on corruption? (5 points)
2. Explain the intuition behind their empirical strategy. (7 points)
3. What is the main assumption behind this strategy? How can it be tested? (8 points)