



Department of Economics

Course name: Empirical Methods in Economics

Course code: EC2404

Type of exam: Main

Examiner: Ferenc Szucs

Number of credits: 7.5

Date of exam: Tuesday April 28th, 2020

Examination time: 3 hours (9:00-12:00)

Aids: Calculator, literature and lecture notes.

Write your personal identity number on each answer sheet.

Start each new question 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. 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 normally be made available on your Ladok account (www.student.ladok.se) within 15 working days from the date of the examination.

Good luck!

Question 1 – Omitted variables bias

We are interested in evaluating the effect of education (Edu) on wages ($\log(w)$). Consider the following equation:

$$\log(w_i) = \alpha_0 + \alpha_1 Edu_i + u_i.$$

We suspect though that wage is also affected by the unobserved ability (a) of individual i . The full relationship is given by the following equation:

$$\log(w_i) = \beta_0 + \beta_1 Edu_i + \beta_2 a_i + v_i$$

Assume that education and ability are also related as follows $a_i = \gamma Edu_i$

1. Derive the relationship between $\hat{\alpha}_1$ and $\hat{\beta}_1, \hat{\beta}_2, \gamma$. Give details of your derivation. (10 points)
2. Explain when omitted variables bias would not be a problem in this case, intuitively and mathematically. (5 points)
3. If there is an omitted variables bias what is its sign and how is it related to the signs of $\hat{\beta}_2$ and γ ? (5 points)

Question 2 – Randomized control trial

A cross-country comparison suggests that BCG vaccination (the vaccination against tuberculosis) may provide some protection against the new coronavirus. Indeed, countries where BCG vaccination has been compulsory for a longer period experience a smaller death rate on average.

In order to establish a causal link the government organize an RCT for healthcare workers. They select 2000 healthcare workers working in the frontline of the epidemic and a random half of them are offered a BCG vaccination. Notice that BCG vaccination was compulsory in Sweden until the 1970's so doctors and nurses who were born before 1970 received vaccination anyway. Some people from the treatment group (who were assigned to get the vaccination) refused to get it.

1. What is the problem with the cross-country evidence? Why can't we interpret it as evidence on the causal relationship between the vaccination and corona induced deaths? (3 points)
2. If you have access to the data of the RCT what regression would you run? Explain all terms, including how each variable is defined and the sample used in the analysis. (6 points)
3. How would you check that the control and the treatment groups were similar prior to the treatment? (4 points)
4. Who are the compliers, always and never takers in this example? (4 points)
5. Do you think that we should control for risk factors like age or diabetes in our equation? Why or why not? (3 points)

Question 3 – Instrumental variables

You are interested in studying how poor households respond to positive income shocks, specifically whether they save or consume the extra money that comes their way. You will explore this question by analyzing the decisions of fishermen of a small town. The price of the fish depends on the quantity of the catch in both the ocean and the nearby lake. You learn that the price fisherman receive for their catch fluctuates on a daily basis based on ocean currents.

For simplicity, we'll say that you construct a "bad day for ocean fishing" current index, which increases when ocean conditions are worse for fishing. The higher the current index, the higher prices (and the more money a fisherman fishing in the lake receives (since they are not affected by the currents)). You collect data on the current index and fish and survey some fisherman fishing in the lake to find out their spending and other home characteristics.

1. Explain how you would use this set-up to examine the impact of positive fluctuations in income on spending (hint: use bad day for fishing index as an instrument). (4 points)
2. Why bad day for fishing index might be a good instrument? Are there any concerns about it? (4 points)
3. How would you formulate the first stage? Spell out the exact equation you would run. (6 points)
4. Outline two different ways you would estimate the treatment effect of extra income on spending. Be specific and describe the equations. (6 points)

Question 4 – Regression Discontinuity Design (20 points)

In many countries a substantial fraction of municipal resources originates from the national budget in the form of grants. 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 as that in 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, Municipality-level vote share for the government party in the last national election.

1. Explain why we can't measure the effect of aligned mayors by simply regressing grants on an indicator for a politically 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 and 50% as a threshold (assume that only two candidates run). (4 points)
3. Explain in detail what equation would you run. (5 points)
4. How can we test that 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)

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.”

For full credit give detailed answers to the following questions:

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)