



Stockholm  
University

Department of Economics

**Course name:** Empirical Methods in Economics 2  
**Course code:** EC2404  
**Type of exam:** Main  
**Examiner:** Daniel Knutsson  
**Number of credits:** 7,5 credits  
**Date of exam:** 12-03-2018  
**Examination time:** 3 hours (09:00-12:00)

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

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. **No aids are allowed.**

-----

**The exam consists of 5 questions. The first two contain multiple choice questions which are to be answered on separate sheets. Each question is worth 20 points, 100 points in total. Note that Question 5 is the credit question.**

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](http://www.student.ladok.se)) within 15 working days from the date of the examination.

-----

**Good luck!**

**1. Multiple choice (20 p, 4 points each).**

Please tick (kryssa för) the correct answer. Only one answer is correct.

Suppose I run an experiment and randomly assign individuals to a treatment ( $T_i$ ) which is equal to 1 for those in the treatment group and 0 for those in the control group. I am interested in an outcome ( $Y_i$ ) and estimate the following regression:

$$Y_i = \beta_0 + \beta_1 T_i + u_i$$

1. Does  $\beta_1$  describe the causal effect of  $T_i$  on the outcome?
  - A. Yes, because the outcome variable is randomly assigned.
  - B. No, because there are no control variables to support the conditional mean independence assumption.
  - C. Yes, because the treatment variable is randomly assigned.
  - D. No, because  $E(u_i|T_i)$  is unknown in the population.
  
2. Consider the regression model above and say that the outcome is binary (either zero or one). Your regression estimate,  $\widehat{\beta}_1 = 0.05$ . How do you interpret this estimate?
  - A. The treatment increased Y by 0.05 treatment units.
  - B. The treatment increased Y by 5 persons.
  - C. The treatment increased Y by 5 percentage points.
  - D. The treatment increased Y by 0.05 %.
  
3. Which of the following statements is correct regarding measurement errors?
  - A. Random measurement error in the outcome variable leads to inconsistent and biased estimates.
  - B. Random measurement error in the treatment variable leads to attenuation bias.
  - C. Random measurement error in the treatment variable only increase standard errors.
  - D. Random measurement error is not a problem in any variable as long as it is strictly random.
  
4. Which of the following statements is correct regarding what the  $R^2$  (or  $\overline{R^2}$ ) can tell us about an estimated regression model?
  - A. If an included variable is statistically significant.
  - B. If there is omitted variables bias.
  - C. If you have chosen the most appropriate set of variables in the model.
  - D. If the regression model is good at explaining the outcome variable.
  
5. Including outcome variables in a regression model as control variables is:
  - A. Problematic as it leads to a multicollinearity problem.
  - B. Valuable as it can give insights into the mechanism of the treatment effect.
  - C. Problematic as it can lead to bias in the estimated coefficient of interest.
  - D. Valuable as it can reduce the standard error of the estimated coefficient of interest.

2. **Multiple choice (20 p, 4 points each).**

Please tick (kryssa för) the correct answer. Only one answer is correct.

1. Suppose  $X_i$  is randomly assigned and I estimate the following regression model:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i^3 + \epsilon_i.$$

What is the expected change in  $Y_i$  from changing  $X_i$  from  $X_i = 1$  to  $X_i = 2$ ?

- A.  $\beta_1 + \beta_2 + \beta_3$
  - B.  $\beta_1 + 2\beta_2 + 4\beta_3$
  - C.  $\beta_1 + 3\beta_2 + 7\beta_3$
  - D.  $2\beta_1 + 4\beta_2 + 8\beta_3$
2. Which of the following statements is **false** when it comes to heteroscedasticity?
- A. Heteroscedasticity tend to make standard errors larger than if the error was homoscedastic.
  - B. Heteroscedasticity should (almost) always be accounted for in regressions using heteroscedasticity robust standard errors.
  - C. Heteroscedasticity means that the error term exhibits serial correlation.
  - D. Heteroscedasticity means that the conditional variance of the error term is not constant.
3. You estimate a bivariate regression model  $Y_i = \beta_0 + \beta_1 X_i + u_i$ . However, you do not include the variable  $W_i$  as a control variable.  $W_i$  is positively correlated with both  $X_i$  and  $Y_i$ . Given this information, the omitted variables bias in your coefficient of interest due to  $W_i$  is:
- A. Negative.
  - B. Positive.
  - C. It does not create bias.
  - D. Insufficient information to assess from the information above.
4. Suppose you estimate the regression model:  $Y_i = \beta_0 + \beta_1 X_i + u_i$ , where  $X_i$  is assigned by randomization. This implies that
- A.  $\beta_1$  can suffer from omitted variables bias if no control variables are included.
  - B.  $\beta_1$  can still be biased by reversed causality (Y affects X).
  - C. An F-test on  $\beta_1 = 0$  must be higher than 10 for a causal interpretation.
  - D.  $\beta_1$  is the causal effect of X on Y.
5. You want to estimate the effect of an intervention providing information on the importance of a vaccine. Measured before the intervention, 14 % of children in the control group and 13 % in the treatment group were vaccinated with this vaccine. After the intervention, 15 % in the control group and 51 % of children in the treatment group were vaccinated. What is the difference-in-differences estimate of the effect of the information program on child vaccinations?
- A. 36 percentage points.
  - B. 37 percentage points.
  - C. 38 percentage points.
  - D. 42 percentage points.

### 3. Instrumental Variables (20 p)

Consider the paper by Card (1993) estimating the returns to education using geographical proximity to a college as an instrumental variable for education. The equation of interest, using the natural logarithm of wages as the outcome, is:

$$\ln(\text{wage}_i) = \beta_0 + \beta_1 \text{Education}_i + u_i$$

- A. Write down the first stage and reduced form outcome equations. Give an interpretation of the main coefficient of interest in each equation. Write down the IV estimate using the two equations ( $\beta_1^{IV}$ ). (4 points)
- B. Which two assumptions needs to hold for the instrument to be valid? Which of these assumptions is it possible to formally test? In which equation can you test this assumption? (4 points)
- C. Card argues that if we include control variables for region (as dummy variables), the instrument is conditionally exogenous. Explicitly state the conditional mean independence assumption necessary for the dummy variables to be valid control variables. (4 points)
- D. Today, the instrumental variable used by Card would have been criticized. Argue briefly why the “near college” variable might not be a valid instrument (even when including the regional dummies). (6 points)

The regression output shows you that the reduced form relationship of interest is 0.12 and the first stage relationship is 1.

- E. Calculate the IV estimate. How do you interpret it? (2 points)

#### 4. Panel data and Difference in Differences: (20 p)

Say you want to assess the importance of increasing the number of police officers on the crime rate. You have access to data at the municipality level and can estimate the equation:

$$Crime_{it} = \beta_0 + \beta_1 Police_{it} + \epsilon_{it} .$$

Both measures (number of crimes and number of police officers) are measured per 1000 inhabitants in each municipality.

- A. Write down the fixed effects regression model based on the regression model above. Based on this equation, describe what kind of problem you can potentially solve using fixed effects compared to the equation above? What kind of problems can a fixed effects model not solve? (5 points)
- B. What must be true (which assumption needs to hold) in order for us to interpret the coefficient from your panel data regression as a causal relationship? Write down the equation or describe the assumption in words. (2 points)
- C. Using yearly panel data for several time periods, the estimated coefficient on *Police* is -0.1 in your fixed effects specification. How do you interpret this estimate? Do you believe that this estimate has a causal interpretation? Motivate your answer. (5 points)
- D. Suppose that around half of the municipalities in your data received central government support to increase police density at the same time period and the other half did not. Using this new information, how could you design a study to answer the question above in a causal way? Write down the equation, describe which parameter you are interested in and describe how you would show your readers that it is plausible that the most important assumption for estimating a causal effect holds. (8 points)

## 5 Credit question (20 p)

**(This question is only for those of you that have not handed in all assignments)**

Consider three papers used in this course that address causal questions using panel data. Each paper claims to estimate a causal effect but use different identifications strategies (methods) based on the question at hand and the data available.

1. **Card and Krueger (1994)**. *Minimum wages and employment: A case study of the fast-food industry in New Jersey and Pennsylvania.*
2. **Burke, Hsiang and Miguel (2015)**. *Global non-linear effect of temperature on economic production.*
3. **Hoynes and Schanzenbach (2009)**. *Consumption responses to in-kind transfers: Evidence from the introduction of the food stamp program.*

- A. Describe briefly the different methods that are used in the papers above. (5 points)
- B. Describe briefly the main underlying assumption in each paper that support a causal interpretation. What are the strengths and weaknesses of each paper? Motivate your answer. (10 points)
- C. In Burke et al. (2015) (paper 2), average temperature is used as the treatment. Do you think that the authors are estimating a causal effect in this paper? Motivate your answer. (5 points)

Try not to write more than 1.5 pages in total.