

Course name:	Empirical Methods in Economics 2
Course code:	EC2404
Semester:	Autumn 2018
Type of exam:	MAIN
Examiner:	Peter Skogman Thoursie
Number of credits:	7,5 credits
Date of exam:	Thursday November 1st, 2018
Examination time:	3 hours (15:00-18:00)

Write your identification number on each answer sheet. Only use printed answer sheets for your answers: Multiple-choice answer sheets for the multiple-choice questions and general answer sheets for all other questions. Do not answer more than one question on each 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. **No aids are allowed.**

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.

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!

Question 1 – Multiple choice (20 points, 4 points each)

- 1) If $E[u_i|X_i] = 0$,
- A) there is no omitted variable bias but there could still be a bias from reversed causality
 - B) there is no omitted variable bias but there could still be a bias from measurement error
 - C) omitted variables and reversed causality are not present but measurement error in X_i is possible
 - D) there are no endogeneity problems
- 2) The following OLS assumption is most likely violated by omitted variables bias:
- A) $E[u_i|X_i] = 0$
 - B) $(X_i, Y_i) i = 1, \dots, n$ are i.i.d draws from their joint distribution
 - C) there are no outliers for X_i, u_i
 - D) there is heteroskedasticity
- 3) In a multiple regression, both X_1 and X_2 are determinants of the dependent variable. When omitting X_2 from the regression, then there will be omitted variable bias for $\hat{\beta}_1$
- A) if X_1 and X_2 are correlated
 - B) always
 - C) if X_2 is measured in percentages
 - D) if X_2 is a dummy variable
- 4) The interpretation of the slope coefficient in the estimated model $\ln Y = 10 + 0.7\ln X$ is as follows:
- A) a 10% change in X is associated with a 7% change in Y
 - B) a 10% change in X is associated with a 0.7% change in Y
 - C) a change in X by one unit is associated with a 0.7% change in Y
 - D) a 1% change in X is associated with a change in Y of 0.7 units
- 5) The average marginal effect in the model $\ln Y = \beta_0 + \beta_1 X + \beta_2 X^2 + u$ can be formulated as
- A) $\beta_1 + \beta_2 \bar{X}$
 - B) $\beta_1 \bar{X} + \beta_2 \bar{X}^2$
 - C) $\beta_1 + 2\beta_2 \bar{X}$
 - D) $\beta_1 \bar{X} + 2\beta_2 \bar{X}$

Question 2 – Multiple choice (20 points, 4 points each)

1) Consider a regression model with two variables, in which X_1 is the variable of interest and X_2 is a valid control variable. Conditional mean independence requires

- A) $E(u|X_1, X_2) = E(u|X_2)$
- B) $E(u|X_1, X_2) = E(u|X_1)$
- C) $E(u|X_1) = E(u|X_2)$
- D) $E(u|X_1, X_2) = 0$

2) The equation of interest is $Y_i = \beta_0 + \beta_1 X_i + u_i$ where $V(u_i|X_i) = \sigma^2 X_i^2$. Which of the following transformations would yield a homoskedastic model

- A) $Y_i/X_i = \beta_0/X_i + \beta_1 + u_i/X_i$
- B) $Y_i/X_i^2 = \beta_0/X_i^2 + \beta_1/X_i + u_i/X_i^2$
- C) $Y_i/\sqrt{X_i} = \beta_0/\sqrt{X_i} + \beta_1/\sqrt{X_i} + u_i/\sqrt{X_i}$
- D) $Y_i/\sqrt{X_i} = \beta_0 + \beta_1/\sqrt{X_i} + u_i/\sqrt{X_i}$

3) When formulating the model $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1}(X_{i2} - \bar{X}_2) + u_i$, β_1 is

- A) difficult to interpret since the effect depends on the value of X_2
- B) the marginal effect of X_1 when $X_2 = 0$
- C) the marginal effect of X_1 evaluated at then mean of X_2
- D) the marginal effect of X_1 evaluated at then mean of X_1

4) Estimating the panel data model $Y_{it} = \alpha_i + \beta_1 X_{it} + u_{it}$ when $T = 2$,

- A) on deviations from individual means is preferable since it is most efficient
- B) on first differences is preferable since it is most efficient
- C) is not possible since more than two time periods is required
- D) gives the same results if using deviations from individual mean or first differences

5) When estimating the panel data model $Y_{it} = \alpha_i + \beta_1 X_{it} + u_{it}$

- A) you need to cluster standard errors at the time level
- B) you need to use heteroskedasticity robust standard errors to account for any form of heteroscedasticity
- C) you need to cluster standard errors at the individual level
- D) you need to cluster standard errors and using heteroskedasticity robust standard errors at the same time

Question 3 – Panel and Difference-in-differences (20 points)

Economic models predict that a higher unemployment benefit level will make unemployed to search less intensively for a job. This implies that higher unemployment benefits increase the unemployment rate since unemployed stay unemployed for a longer period. Your task is to investigate if the prediction of the model is supported by an empirical analysis.

The generic regression equation of interest is

$$\ln U = \beta_0 + \beta_1 UB + u$$

where $\ln U$ is the log unemployment rate and UB is the monthly unemployment benefit level. Think of UB as expressed in SEK 1,000 which means that β_1 tells how a 1,000 increase in monthly unemployment benefits change the unemployment rate in percent.

(i) Say that you collect data on unemployment rates and average unemployment benefit level at the municipality level in the year of 2010. Discuss potential endogeneity problems when estimating the above relationship using municipality cross-sectional data. **(5 points)**

(ii) Say hypothetically that municipalities can decide about the unemployment replacement rate in their own municipality. All municipalities changed their unemployment replacement rates at several points in time during the period 2000-2017 (note that unemployment benefits are therefore continuous and can increase and decrease within a municipality). You have access to annual data on municipality unemployment rates and average unemployment benefits levels for this period. Discuss the following:

- What is the variation in data that you can exploit in order to examine how unemployment benefits is related to the unemployment rate
- Explicitly state and explain the equation you would estimate and be careful so that all sub-indexes are correct
- Discuss what type of endogeneity problem you can solve using this estimation strategy
- Discuss what the key identifying assumption is
- Discuss how you would deal with the estimated standard errors in the estimation

(7 points)

(iii) Say now instead that all municipalities had the same unemployment benefit level until 2012 (you still have data for the period 2000-2017). In 2012 half of the municipalities increased the unemployment benefit level but stayed unchanged in the other half of the municipalities. You do not have information on the exact level of average unemployment benefits, only that half of the municipalities raised the benefit level.

- What is the variation in data that you can exploit in this case in order to examine how unemployment benefits is related to the unemployment rate
- Explicitly state and explain the equation you would estimate and be careful so that all sub-indexes are correct
- Discuss what the key identifying assumption for causality is
- Discuss pros and cons regarding how you can handle standard errors when estimating the equation at the municipality-year level versus estimating at the treatment status-year level (i.e. on 18×2 observations)
- Explain how would check that the key identifying assumption is fulfilled

(8 points)

Question 4 – IV (20 points)

You had the possibility to randomize unemployment benefits across municipalities in 2015, so that 200 municipalities received a high benefit level and 200 municipalities received a low benefit level. You don't know the exact benefit level, only whether a municipality was randomized to a low versus high benefit level. However, 40 municipalities who were randomized to have a low benefit level implemented a high benefit level. Think of the instrument (Z) as taking the value 1 if randomized to high benefit municipalities, 0 otherwise. Think of the treatment variable, say HB , as taking the value one if a municipality actually had a high benefit level (0 otherwise).

- (i) Why do we need to apply the IV strategy in this case? (4 points)
- (ii) Write down the first stage equation and explicitly state the estimate (i.e., the number) of the main slope coefficient of this regression. Interpret this coefficient. (4 points)
- (iii) Say that the estimated effect of the instrument in the reduced form outcome equation is 0.16 (note the dependent variable is the log of the unemployment rate, $\ln U$). What is the IV-estimate (i.e., the number)? Interpret this coefficient. (6 points)
- (iv) From a theoretical point of view, unemployment benefits could also affect the quality of the job match. For example, a low benefit level could make unemployed accepting a job "too" quickly, i.e. a bad job match. This could for example be reflected in lower wages among those who belong to municipalities which were randomized to have low unemployment benefits. Say that you have information on wages for the first job unemployed received after the period as unemployed. Assume that all unemployed finally found a job after their unemployment spell.

Someone suggested that you in the *reduced form outcome equation* control for the wage unemployed received after their unemployment spell, with the purpose to control for quality of job when estimating relationship between the unemployment rate and unemployment benefits.

Argue whether or not you think this strategy is a good idea? Base your argument around the conditional mean independence assumption (6 points)

Question 5 – credit question. Acemoglu & Angrist (2001) paper (20 points)

This is the abstract from the Acemoglu & Angrist (2001) paper:

“The Americans with Disabilities Act (ADA) requires employers to accommodate disabled workers and outlaws discrimination against the disabled in hiring, firing, and pay. Although the ADA was meant to increase the employment of the disabled, the net theoretical effects are ambiguous. For men of all working ages and women under 40, Current Population Survey data show a sharp drop in the employment of disabled workers after the ADA went into effect. Although the number of disabled individuals receiving disability transfers increased at the same time, the decline in employment of the disabled does not appear to be explained by increasing transfers alone, leaving the ADA as a likely cause. Consistent with this view, the effects of the ADA appear larger in medium-size firms, possibly because small firms were exempt from the ADA. The effects are also larger in states with more ADA-related discrimination charges.”

Describe how they have econometrically reached to the main conclusion that the ADA seems to have a negative effect on the employment of disabled. Especially, the following issues must be included and explain intuitively as well as using equation notations:

- (i) The main strategy used?
- (ii) What is the key identifying assumption for estimating the causal effect of the ADA?
- (iii) How do they econometrically investigate if this assumption is valid?

Write maximum 1 ½ A4-page for the answers!