Stockholm University

Department of Economics

Course name:	Empirical Methods in Economics 2
Course code:	EC2404
Type of exam:	Re-take
Examiner:	Peter Skogman Thoursie
Number of credits:	7,5
Date of exam:	2018-12-15
Examination time:	09:00-12:00

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

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.

No aids are allowed.

The exam consists of 5 questions. Each question is worth 20 points, 100 points in total. 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 – Multiple choice (20 points, 4 points each)

Please tick (*Kryssa för*) the correct answer, only one answer is correct

1) If the true β_1 is positive, random measurement error in the independent variable X will
A) overestimate β_1
B) underestimate β_1
C) not give an inconsistent estimate of β_1
D) only be a problem in presence of omitted vairables

2) An omitted variable X_2 yields an *inconsistent* estimate of β_1 if
A) if X_1 and X_2 are correlated but X_2 has no effect on Y_i
B) if X_1 and X_2 are correlated and X_2 affects Y_i
C) if X_1 and X_2 are uncorrelated and X_2 has no effect on Y_i
D) if X_1 and X_2 are uncorrelated but X_2 affects Y_i

3) If values of X_i are randomized within two categories represented by the dummy variable W_i, which of the following equations would yield the causal effect of X_i on Y_i
A) $Y_i = \beta_0 + \beta_1 X_i^2 + u_i$
B) $Y_i = \beta_0 + \beta_1 X_i + \beta_2 W_i + u_i$
C) $Y_i = \beta_0 + \beta_1 X_i + \beta_2 X \times W_i + u_i$
D) $Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + u_i$

4) The interpretation of the slope coefficient in the estimated model $ln\, Y = 10 + 0.7 lnX$ is
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 $ln\, Y = \beta_0 + \beta_1 X + \beta_2 X^2 + u$ is
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) $\beta_1 \bar{X} + 2\beta_2 \bar{X}$

Stockholms universitet

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

Please tick (*Kryssa för*) the correct answer, only one answer is correct

1) All of the following are true, with the exception of one condition:
A) a high R^2 or $\bar{R}^2$ does not mean that the regressors are a true cause of the dependent variable
B) a high R^2 or $\bar{R}^2$ does not mean that there is no omitted variable bias
C) a high R^2 or $\bar{R}^2$ always means that an added variable is statistically significant
D) a high R^2 or $\bar{R}^2$ does not necessarily mean that you have the most appropriate set of regressors

2) Including an interaction term between two independent variables, X_1 and X_2, allows for the following except:
A) the interaction term lets the effect on *Y* of a change in X_1 depend on the value of X_2
B) the interaction term coefficient is the effect of a unit increase in X_1 and X_2 above and beyond the sum of the individual effects of a unit increase in the two variables alone
C) the interaction term coefficient is the effect of a unit increase in $\sqrt{X_1 \times X_2}$
D) the interaction term lets the effect on Y of a change in X_2 depend on the value of X_1

3) In the regression model $Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_{2i} + \beta_3 X_{1i} \times (X_{2i} - \bar{X}_2) + u_i$, where X_{1i} and X_{2i} are continuous variables. β_1
A) represents the effect of X_{1i} when $X_{2i} = 1$
B) represents the effect of X_{1i} when $X_{2i} = 0$.
C) represents the mean effect of X_{1i}
D) represents the effect of X_{1i} when X_{2i} is equal to its mean

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) 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) $Y_i/\sqrt{X_i} = \beta_0 + \beta_1/\sqrt{X_i} + u_i/\sqrt{X_i}$

Stockholms universitet

Question 3 (20 points)
Say that we are interested in estimating the returns to college education. We specify the following equation:

$$\ln(Wage_i) = \beta_0 + \beta_1 College_i + u_i$$

where $Wage_i$ is the hourly wage rate, $College_i$ is a dummy variable taking the value 1 if the individual has obtained a college education, 0 otherwise.

You have access to an instrument, Z_i, which indicates whether the individual grow up near a college or not (i.e., a dummy variable). The OLS-estimate of β_1 in the above equation is 0.05.

You estimate the following two equations

$$\ln(Wage_i) = \pi_0 + \pi_1 Z_i + \varepsilon_i$$

$$College_i = \gamma_0 + \gamma_1 Z_i + v_i$$

The OLS-estimate of π_1 is 0.06.

(i) 80 percent of those who grew up close to college went to college. 20 percent of those who did not grew up close to college went to college. Calculate the IV estimate of returns to college. Explain why the IV estimate might differ from the OLS estimate. (12 points)

(ii) Say that effects are heterogeneous. Explain who are the never takers, always takers and compliers. How would you interpret the IV-estimate under heterogeneous effects? (8 points)

Stockholms universitet

Question 4 (20 points)

Suppose you are interested in estimating the effect of an alcohol tax ($AlcTax$) on fatality rates ($FatalityRate$). That is, the generic equation of interest is

$$FatalityRate = \beta_0 + \beta_1 AlcTax + u$$

You have access to municipality data for two years, 2000 and 2015. Fatality rate is measured as the fatalities per 10,000 and the alcohol tax is measured as per cent (from 1-100%).

(i) Suppose you estimate the following equation with OLS

$$FatalityRate_{mt} = \beta_0 + \beta_1 AlcTax_{mt} + u_{mt}$$

where m is municipality and t is year. Explain intuitively why, reversed causality and omitted variables might make $AlcTax_{mt}$ correlated with u_{mt} and in what way they are correlated (i.e., the signs of the correlations) due to these two problems. (7 points)

(ii) Say that would like to solve the problem of time invariant omitted variables. Explain how you would specify the empirical equation in order to solve the endogeneity problem of time invariant omitted variables. (5 points)

(iii) Say that there were only 2 municipalities that changed their alcohol tax, both from 20 to 30 percent. These two municipalities also experienced a reduction in fatality rates by 2 persons per 10,000 municipality citizens between 2000 and 2015. *Ignoring* general time effects, what is the estimate of β_1 using the transformation you suggested in (ii)? Interpret the estimate! (8 points)

Stockholms universitet

Question 5 – credit question (20 points)

Think of the Angrist & Evans (1998) paper analyzing the effect of fertility and labour supply.

- Explain the econometric difficulties in estimating this relationship
- Give a detailed description of their research design and how it solves these difficulties
- Discuss how they argue whether the main assumptions for their strategy to work are fulfilled
- Discuss to what extent they show that their strategy is valid

Write a maximum of 1 ½ pages.