Stockholm University

Department of Economics

Course name:	**Empirical Methods in Economics 2**
Course code:	**EC2404**
Type of exam:	**RETAKE**
Examiner:	**Peter Skogman Thoursie**
Number of credits:	**7,5 credits**
Date of exam:	**Monday 24 April 2017**
Examination time:	**3 hours (09:00-12:00)**

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. Note Question 5 is the credit question

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 "My Studies" account **(**www.mitt.su.se) on 16 May at the latest.

Good luck!

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 variables X_i will

A) overestimate β_1
B) underestimate β_1
C) underestimate the standard errors
D) overestimate the standard errors

2) The earnings equation of interest is $Y_i = \beta_0 + \beta_1 fem_i + \beta_2 marr_i + u_i$, where fem_i is a dummy variable taking the value 1 if individual is a female (0 if male) and $marr_i$ is dummy variable taking the value 1 if individual is married (0 otherwise). This specification implies

A) That the earnings premium of being married β_2 is different between females and males
B) $marr_i$ is a valid control variable for fem_i
C) $E[u_i|marr_i, fem_i] = E[u_i|fem_i]$
D) Females and males have the same earnings premium of being married

3) 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

4) If Z_i is an instrumental variable to X_{1i}. The conditional mean assumption $E[u_i|Z_i, X_{2i}] = E[u_i|X_{2i}]$ says that

A) you can obtain a consistent estimate of the effect of X_{1i}
B) you can obtain a consistent estimate of the effect of X_{2i}
C) Z_i is a valid control variable
D) you can't obtain consistent estimates of neither the effects of X_{1i} nor X_{2i}

5) Controlling for outcomes variables can

A) give important insights on mechanisms
B) could make the estimated treatment effect more precise
C) leads to a collinearity problem
D) is problematic since it destroys the initial exogenous variation in the variable of interest

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

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

1) The interpretation of the slope coefficient in the model $lnY_i = \beta_0 + \beta_1 X_i + u_i$ is as follows:

A) a 1% change in X is associated with a β_1 % change in Y.
B) a change in X by one unit is associated with a β_1 change in Y.
C) a change in X by one unit is associated with a 100 β_1% change in Y.
D) a 1% change in X is associated with a change in Y of 0.01 β_1.

2) 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) $Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + u_i$

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) In the linear probability model, the interpretation of the slope coefficient is

A) the change in odds associated with a unit change in X, holding other regressors constant
B) not all that meaningful since the dependent variable is either 0 or 1
C) the change in probability that $Y = 1$ associated with a unit change in X, holding others regressors constant
D) the response in the dependent variable to a percentage change in the regressor.

5) The major flaw of the linear probability model is that

A) the actuals can only be 0 and 1, but the predicted are almost always different from that.
B) the regression R^2 cannot be used as a measure of fit
C) people do not always make clear-cut decisions
D) the predicted values can lie above 1 and below 0.

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)

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 equation of interest in levels and then how you would transformation of the equation 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)

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

Write maximum 1 ½ A4-page for the answers!

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?