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:	**Saturday, 9th of December, 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 the 3rd of January 2018 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) 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

2) In a pure experimental setting, controlling for pre-treatment characteristics

A) is always a bad idea since initial randomization will be destroyed
B) could make the estimated treatment effect more precise
C) leads to a collinearity problem
D) would change the interpretation of the estimated treatment effect

3) When there are omitted variables in the regression, which are determinants of the dependent variable, then

A) you cannot measure the effect of the omitted variable, but the estimator of your included variable(s) is (are) unaffected
B) this has no effect on the estimator of your included variable because the other variable is not included
C) this will always bias the OLS estimator of the included variable
D) the OLS estimator is biased if the omitted variable is correlated with the included variable

4) All of the following are examples of joint hypotheses on multiple regression coefficients, with the exception of

A) $H_0: \beta_1 + \beta_2 = 1$
B) $H_0: \frac{\beta_3}{\beta_2} = 1$ and $\beta_4 = 0$
C) $H_0: \beta_2 = 0$ and $\beta_3 = 0$
D) $H_0: \beta_1 = -\beta_2$ and $H_0: \beta_1 + \beta_2 = 1$

5) When testing joint hypothesis, you should

A) use *t*-statistics for each hypothesis and reject the null hypothesis is all of the restrictions fail
B) use the *F*-statistic and reject all the hypothesis if the statistic exceeds the critical value
C) use *t*-statistics for each hypothesis and reject the null hypothesis once the statistic exceeds the critical value for a single hypothesis
D) use the *F*-statistics and reject at least one of the hypothesis if the statistic exceeds the critical value

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 $lnY = \beta_0 + \beta_1 X + u$ 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 100 β_1 % change in Y
C) a 1% change in X is associated with a change in Y of 0.01 β_1
D) a change in X by one unit is associated with a β_1 change in Y

2) In the model $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + u$, the marginal effect $\frac{dY}{dX_1}$ is
A) $\beta_1 + \beta_3 X_2$
B) β_1
C) $\beta_1 + \beta_3$
D) $\beta_1 + \beta_3 X_1$

3) If you estimate the following model $Y_{it} = \alpha_i + \beta_1 X_{it} + u_{it}$ with $T = 2$ using a within group (individual) transformation (i.e., a fixed effect model, FE) or first differences, FD, your results

A) will differ because in the FE model you explicitly include individual dummies
B) will differ because in the FE model you use both time periods but when using FD, you lose the first observation
C) will be the same
D) will differ because FD exploits differences over time and FE compares average before and after

4) If the exclusion restriction holds when using IV, this implies

A) that you can estimate the equation of interest
B) that only the reduced form equation has a causal interpretation
C) is another way of saying that the instrument is as good as randomly assigned
D) that the first stage and the reduced from equation both have causal interpretation but that you cannot estimate the equation of interest

5) The Difference-in-differences method relies on the assumption that the treatment and control groups

A) have equal average outcomes before treatment takes place
B) have parallel outcome trends in absence of treatment
C) are randomly assigned
D) have equal outcomes in all periods prior to treatment

Question 3. Casual effect of class size (20 points)

A social experiment has been conducted in order to evaluate the effect of class size on student's school performance. Class size was randomly distributed within schools. You have access to the following information:

Y_i = *average test score in class* i *(ranging from 0-100)*
$Y_{i,pre}$ = *average test score in class* i*, before experiment took place*
TE_i = *teacher experience in class* i *(in number of years)*
$ShFem_i$ = share female students in class i
PE_i = average parental education (measured in years) in class i
PH_i = average hours parents' help their children with homework in class i (measured after the experiment)

(i) Would estimation of the following equation yield the causal effect of class size in average test score? (5 points)

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

(ii) Explain how you would check whether randomisation was carried out correctly (5 points)

(iii) If you include teacher experience in the equation of interest, would the effect of teacher experience be a causal effect? (4 points)

(iv) Explain intuitively whether or not you think it is a good idea to control for PH_i in order to track out the effect of class size net of the effect of parents helping their children with their homework. Explicitly state the conditional mean independence assumption required in order to be able to estimate the causal effect of class size in this case (6 points)

Question 4. IV (20 points)

Say that you are interested in the estimating the returns to schooling. The equation of interest is:

$$wage_i = \beta_0 + \beta_1 sch_i + u_i$$

where sch_i is years of schooling and $wage_i$ is the hourly wage rate in SEK.

For simplification, say that years schooling is endogenous only because there are ability differences between big cities and the country side. In other words, sch_i is as good as randomized within big cities and within the country side. This means that $BigCity_i$ (1 if individual lives in a big city and 0 otherwise) is a valid control variable.

(i) Explicitly state the conditional mean independence assumption in order for $BigCity_i$ to be a valid control variable (4 points)

Now, you don't really believe that controlling for $BigCity_i$ really solves the endogeneity problem. Rather you try an instrument instead which is whether or not an individual grew up in a big city, $Z = GrUpBigCity_i = 1$ is individual grew up in a big city and 0 otherwise.

You estimate following equation using $GrUpBigCity_i$ as an instrument for years of schooling

$$wage_i = \beta_0 + \beta_1 sch_i + u_i$$

(ii) The estimated coefficient of the instrument in the first stage regression is 0.12. Interpret this coefficient estimate (4 points)

(iii) The estimated coefficient of the instrument in the reduced form outcome equation is 2. Interpret this coefficient estimate (4 points)

(iv) What is the IV estimate of returns to schooling? Interpret the IV-estimate taking into account that effects are heterogeneous. Motivate whether or not you think the instrument satisfies the exclusion restriction (8 points)

Question 5. Credit question. Angrist & Evans (1998) paper (20 points)

The Angrist & Evans (1998) estimates the effect of having more than 2 kids ($morekids_i$=1 if more than 2 kids, 0 otherwise) on e.g., mothers' labour supply ($weeksw_i$= number of weeks worked during a year). As an instrument, the sex composition of the first two children is used ($samesex_i = 1$ if the first two kids have the same sex, 0 otherwise).

(i) Explicitly state the equation of interest, the first stage regression and the reduced form outcome equation. (5 points)

(ii) Interpret the main coefficient (i.e., the slope coefficient) each regression. (5 points)

(iii) How would you interpret the IV estimate using this set-up if effects are heterogeneous? (10 points)