Stockholm University

Department of Economics

Course name:	**Empirical Methods in Economics 2**
Course code:	**EC2404**
Type of exam:	**MAIN**
Examiner:	**Peter Skogman Thoursie**
Number of credits:	**7,5 credits**
Date of exam:	**Sunday 29 October 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 20 November 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) Random measurement error in the regressor

A) makes the estimate of the regression coefficient inconsistent
B) makes the estimate less precise
C) implies that standard errors are flawed
D) leads to an upward bias in the regression coefficient

2) The equation of interest is $Y_i = \beta_0 + \beta_1 X_{1i} + u_i$, and X_{2i} is the only omitted variable. Estimating β_1 using the bivariate regression would yield an *inconsistent* estimate

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) The earnings equation of interest is $Y_i = \beta_0 + \beta_1 sch_i + \beta_2 fem_i + \beta_3 fem_i \times sch_i + u_i$, where fem_i is a dummy variable taking the value 1 if individual is a female (0 if male) and sch_i is years of schooling. This specification implies that

A) β_2 represents the average gender wage gap
B) $\beta_2 + \beta_3$ represents the average gender wage gap
C) β_2 represents the gender wage gap when years of schooling is zero
D) $\beta_1 + \beta_3$ represents the average gender wage gap

4) The conditional mean independence assumption $E[u_i|X_{1i}, X_{2i}] = E[u_i|X_{2i}]$ implies that

A) you can obtain a consistent estimate of X_{1i}
B) you can obtain a consistent estimate of X_{2i}
C) you can obtain consistent estimates of both X_{2i} and X_{2i}
D) X_{2i} is a valid instrument for X_{1i}

5) 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

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

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

1) The results of the estimated wage-schooling equation can be summarized by $\widehat{lnWage} = 1.5 + 0.2lnX_i$. This means that:

A) the returns to schooling is 20 percent
B) a 10% change in X is associated with a 20% increase in wages.
C) a change in X by one unit is associated with a 20% of 100 SEK i.e., 20 SEK
D) a 10% change in X is associated with a change in Y of 2%.

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) 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 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 methods relies on the assumptions that the treatment and control groups

A) have equal average outcomes before treatment takes place.
B) have parallel trends in absence of treatment
C) would have had parallel trend in outcomes after treatment takes place if no group actually received treatment
D) have equal outcomes in all periods prior to treatment

Question 3 (20 points)

You have been asked to referee a paper that tries to estimate the effect of class size ($ClassSize = \#students\ in\ class$) on child performance at school. Authors have a representative sample of classes, with only one class per school. The equation of interest is:

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

where y_i is the average test results in class i (where the test score can take values 0-100).

Authors claim that they can use distance (Z_i) to schools, measured in kilometres, as an instrument to class size (Z_i is the average distance for pupils in a class). Their argument is that longer distances to school reflect locations/areas with a smaller population density and therefore smaller class sizes. They further argue that parental background can be correlated with locations and that the instrument is "as good as randomized" condition on family background. They use mothers' years of schooling ($msch_i$) as a measure of family background.

(i) State the conditional mean independence assumption for instrument exogeneity based on their argument. Also interpret this assumption. (**4 points**)

Autors estimate the main coefficient from the first stage regression to -2 and the main coefficient from the reduced form outcome equation to 4.

(ii) Explicitly state the first stage and the reduced form outcome equations. Interpret the estimated main coefficients from these two equations. (**6 points**)

(iii) Calculate the IV-estimate based on the above information. Give an economic interpretation if this estimate. (**2 points**)

Authors further argue that parents help their children with their homework to a larger extent the closer in distance they live to school, in order to compensate for their children being in a large class. They estimate the reduced from outcome equation controlling for how much parents help their children with their homework, represented by PH. They argue that this an interesting type of analysis since they can then detect whether there is a pure class size effect, cleaned from the variation that stems from parents helping their children depending on the class size.

(iv) State the conditional mean independence assumption for PH to be a valid control variable in the reduced from outcome regression. Argute intuitively whether you think this assumption is valid. (**6 points**)

(v) Do you think their instrumental variable strategy is valid in general? Motivate! (**2 points**)

Question 4 (20 points)

Say that you interested in estimating the effects of a beer tax on highway traffic fatalities, that is the equation of interest is:

$$y_{it} = \alpha + \delta X_{it} + \varepsilon_{it}$$

where y_{it} = highway fatality rate (number of annual traffic deaths per 10,000 people in the population) in municipality i and year t, and X_{it} is the beer tax rate. You have access to municipality panel data for the two years 2002 and 2008.

(i) Explain what type of endogeneity problems that could arise if you want to estimate the causal effect of a beer tax on fatality rates based on the equation given above. (**4 points**)

(ii) Explain why a fixed effect (FE) estimation and a first difference (FD) estimation could solve one of these endogeneity problems. Explicitly state the FE and FD equations you would estimate. (**6 points**)

(iii) Say that the beer tax was the same in all municipalities until it was changed and increased in half of the municipalities in 2004. Explain short how you would design a study in order to estimate the causal effect of beer tax on fatality rates. Show the equations you will estimate. Note that you need to convince readers that the key identifying assumption holds – this should be captured by the equation. (**10 points**)

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.