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:	**Friday 17 March 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 7 April 2017 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 dependent variables Y_i leads to

A) an endogeneity problem
B) higher standard errors of the estimated regression coefficients
C) inconsistently estimated standard errors
D) heteroskedasticity

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) 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_2affects 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) The conditional mean assumption $E[u_i|X_{1i}, X_{2i}] = E[u_i|X_{1i}]$ 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) you can't obtain consistent estimates of neither X_{2i} nor X_{2i}

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 interpretation of the slope coefficient in the model $lnY_i = \beta_0 + \beta_1 lnX_i + u_i$ 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 $Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_i + \beta_3 D_i \times X_i + u_i$, where X_i is a continuous variable and D_i is a binary variable, β_3

A) indicates the slope of the regression when $D_i = 1$.
B) has a standard error that is not normally distributed even in large samples since *D* is not a normally distributed variable.
C) indicates the difference in the slopes of the two regressions.
D) has no meaning since $D_i \times X_i = 0$ when $D_i = 0$.

4) The Fixed Effects regression model

A) has *n* different intercepts.
B) the slope coefficients are allowed to differ across entities, but the intercept is "fixed" (remains unchanged).
C) has "fixed" (repaired) the effect of heteroskedasticity.
D) in a log-log model may include logs of the binary variables, which control for the fixed effects.

5) The main advantage of using panel data over cross sectional data is that it

A) gives you more observations.
B) allows you to analyze behavior across time but not across entities.
C) allows you to control for some types of omitted variables without actually observing them.
D) allows you to look up critical values in the standard normal distribution.

Question 3 (20 points)
Say that you are interested in replicating the Angrist & Evans (1998) paper on Swedish data and estimate the effect of fertility on female labour supply. You have data with information on whether the mother worked or not during the study year ($worked_i = 1$ if mother worked, 0 otherwise). $SameSex_i$ is a dummy variable taking the value 1 if the first two children have the same sex (i.e., boy-boy or girl-girl), and 0 otherwise. You also have a dummy variable, $More2kids_i$ that takes the value 1 if the mother has more than 2 kids, and 0 otherwise.

The equation of interest is:

$$worked_i = \beta_0 + \beta_1 More2kids_i + u_i$$

By inspection of data you find that mothers with more than 2 kids work to a degree of 60 percent. Mothers with at most 2 children work to a degree of 80 percent.

By further inspection of data, you find that 30 percent of those with the same sex of the first two children have more than 2 kids. The corresponding figure for mothers with a sex mix of the first two children is 20 percent.

You further find that 70 percent of mothers with the same sex of the two first born children worked. 71 percent of the mothers with a sex mix of the first two children worked.

(i) Explicitly state the first stage regression and give the estimated slope coefficient (provide the number) from this first stage regression. Also interpret this slope coefficient (4 points)

(ii) Explicitly state the reduced form outcome equation and give the estimated slope coefficient (provide the number) from this regression. Also interpret this slope coefficient (4 points)

(iii) Estimate β_1 (provide the number) using OLS. Then estimate β_1 (provide the number) using the IV-method. Also interpret the estimated effects in both the OLS as well as the IV cases. Explain intuitively why OLS and IV might differ (7 points).

(iv) Using the IV terminology and discuss who are the never takers, always takers and compliers in the example above (5 points).

Question 4 (20 points)
In Sweden, an important component of the employment protection legislation is the so called "last-in-first-out" principle saying that when firm downsizes the last employed should go first (in practice the rule can be circumvented by negotiations with the employer and the union). In 2001, small firms with less than 11 employees could exempt 2 workers from this principle. According to theory this would decrease the costs for firms. As such, it is hypothesized that the reform increased both hiring and separations for small firms with an ambiguous effect on employment.

You are asked to evaluate the reform and e.g., estimate the reform effect on the number of newly hires on an annual basis.

- Average number of new hires for small firms (g=1) are: $Y_{1,1995} = \cdots = Y_{1,2000} = 1;\ Y_{1,2001} = 3;\ Y_{1,2002} = 4$
- For large firms (g=0) the number of new hires are: $Y_{0,1995} = \cdots = Y_{0,2000} = 11;\ Y_{0,2001} = 12 = Y_{0,2002} = 12$

If you estimate the average reform effect based on the following equation with OLS:

$$Y_{gt} = \beta_0 + \beta_1 T_g + \beta_2 After_t + \gamma After_t \times T_g + u_{gt}$$

where Y_{gt} is the number of new hires for group *g* in year *t*. $T_g = 1$ for treated (i.e., small firms) and 0 for controls (large firms). $After_t = 1$ from year 2001 and onwards, 0 otherwise.

(i) What would be your estimate of γ? Interpret the estimate! (5 points).

Next, estimate the following model with yearly "treatment" effects using OLS:

$$Y_{gt} = \beta_0 + \beta_1 T_g + \lambda_t + \delta_{1996} d1996_t \times T_g + \delta_{1997} d1997_t \times T_g + \delta_{1998} d1998_t \times T_g$$

$$+\delta_{1999} d1999_t \times T_g + \delta_{2000} d2000_t \times T_g + \delta_{2001} d2001_t \times T_g + \delta_{2002} d2002_t \times T_g + u_{gt}$$

where $d1996_t$ is a dummy variable taking the value 1 in year 1996 and zero otherwise, and so on (note 1995 is the reference year).

(ii) What would be your estimates of δ_{2001} and δ_{2002}? (8 points)

(iii) What are the estimates of $\delta_{1996}, \delta_{1997}, \ldots, \delta_{2000}$? Would you claim that the estimates of δ_{2001} and δ_{2002} are causal effects? Motivate! (7 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!