Stockholm University

Department of Economics

Course name:	**Empirical Methods in Economics 2**
Course code:	**EC2404**
Semester:	**Autumn 2016**
Type of exam:	**Main**
Examiner:	**Peter Skogman Thourise**
Number of credits:	**7,5 credits**
Date of exam:	**Sunday October 30 2016**
Examination time:	**3 hours (16:00-19:00)**

Write your identification number on each answer sheet. Use the printed answer sheets for all your answers. 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. 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 "My Studies" account **(**www.mitt.su.se) on November 18 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) 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.

2) In a two regressor regression model, if you exclude one of the relevant variables then
A) it is no longer reasonable to assume that the errors are homoskedastic.
B) OLS is no longer unbiased, but still consistent.
C) you are no longer controlling for the influence of the other variable.
D) the OLS estimator no longer exists.

3) 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.

4) If you reject a joint null hypothesis using the *F*-test in a multiple hypothesis setting, then
A) a series of *t*-tests may or may not give you the same conclusion.
B) the regression is always significant.
C) all of the hypotheses are always simultaneously rejected.
D) the *F*-statistic must be negative.

5) In nonlinear models, the expected change in the dependent variable for a change in one of the explanatory variables is given by
A) $\triangle Y = f(X_1 + X_1, X_2, \ldots X_k)$.
B) $\triangle Y = f(X_1 + \triangle X_1, X_2 + \triangle X_2, \ldots, X_k + \triangle X_k) - f(X_1, X_2, \ldots X_k)$.
C) $\triangle Y = f(X_1 + \triangle X_1, X_2, \ldots, X_k) - f(X_1, X_2, \ldots X_k)$.
D) $\triangle Y = f(X_1 + X_1, X_2, \ldots, X_k) - f(X_1, X_2, \ldots X_k)$.

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 $\ln(Y_i) = \beta_0 + \beta_1 X_i + u_i$ 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 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) A polynomial regression model is specified as:
A) $Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \cdots + \beta_r X_i^r + u_i$.

B) $Y_i = \beta_0 + \beta_1 X_i + \beta_1^2 X_i + \cdots + \beta_1^r X_i + u_i$.

C) $Y_i = \beta_0 + \beta_1 X_i + \beta_2 Y_i^2 + \cdots + \beta_r Y_i^r + u_i$.

D) $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_2 + \beta_3 (X_{1i} \times X_{2i}) + u_i$.

3) In the model $Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 (X_1 \times X_2) + u_i$, the expected effect $\frac{\Delta Y}{\Delta X_1}$ is

A) $\beta_1 + \beta_3 X_2$.
B) β_1.
C) $\beta_1 + \beta_3$.
D) $\beta_1 + \beta_3 X_1$.

4) In the log-log model, the slope coefficient indicates
A) the effect that a unit change in X has on Y.
B) the elasticity of Y with respect to X.
C) $\Delta Y / \Delta X$.
D) $\frac{\Delta Y}{\Delta X} \times \frac{Y}{X}$.

5) Consider the population regression of log earnings [Y_i, where $Y_i = \ln(Earnings_i)$] against two binary variables: whether a worker is married (D_{1i}, where D_{1i}=1 if the i^{th} person is married) and the worker's gender (D_{2i}, where D_{2i}=1 if the i^{th} person is female), and the product of the two binary variables
$Y_i = \beta_0 + \beta_1 D_{1i} + \beta_2 D_{2i} + \beta_3 (D_{1i} \times D_{2i}) + u_i$. The interaction term
A) allows the population effect on log earnings of being married to depend on gender
B) does not make sense since it could be zero for married males
C) indicates the effect of being married on log earnings
D) cannot be estimated without the presence of a continuous variable

Question 3 – Difference-in-differences (20 points)

Say that you evaluate the effect of a labor market training program that took place in the beginning 2004. You have access to yearly average outcomes for treated ($g = 1$) and untreated ($g = 0$) during the period 2000-2005. The average outcome, Y_{gt}, for the two groups are the following:

- $Y_{0t} = 100$ in all years for the control group ($g = 0$)
- $Y_{1,2000} = Y_{1,2001} = Y_{1,2002} = Y_{1,2003} = 100$ i.e., the outcome is 100 for the treatment group in all years up to 2003. $Y_{1,2004} = 200$ and $Y_{1,2005} = 400$.

Let $T_g = 1$ for the treated group and 0 for the control group and $After_t = 1$ during years 2004 and 2005 (i.e., the after period) and zero otherwise.

You estimate the following equation with OLS:

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

(i) What would be your estimate of γ? (2 points)

(ii) Interpret the estimated coefficient (3 points)

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

$$Y_{gt} = \beta_0 + \beta_1 T_g + \lambda_{2000} d2000_t + \lambda_{2001} d2001_t + \lambda_{2002} d2002_t + \lambda_{2004} d2004_t$$

$$+\lambda_{2005} d2005_t + \delta_{2000} d2000_t \times T_g + +\delta_{2001} d2001_t \times T_g + \delta_{2002} d2002_t \times T_g$$

$$+\delta_{2004} d2004_t \times T_g + +\delta_{2005} d2005_t \times T_g + u_{gt}$$

where $d2000_t$ is a dummy variable taking the value 1 in year 2000 and zero otherwise, and so on.

(iii) What would be your estimates of δ_{2004} and δ_{2005}? (5 points)

(iv) Interpret these two estimated coefficients (5 points)

(v) Would you claim that the estimates of δ_{2004} and δ_{2005} are causal effects? Motivate! (5 points)

Question 4 – IV (20 points)

Say that you are interested in the estimating the returns to schooling and 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 if individual lives in country side) 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).

(ii) Interpret this conditional mean independence assumption (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

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

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

(iv) The estimated coefficient of the instrument in the reduced form outcome equation is 2.4. Interpret this coefficient estimate (4 points)

(v) What is the IV estimate of returns to schooling? (4 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)