STOCKHOLM UNIVERSITY
Department of Economics

Course name:	**Empirical Methods in Economics 2**
Course code:	**EC2402**
Examiner:	**Peter Skogman Thoursie**
Number of credits:	**7,5 credits (hp)**
Date of exam:	**Saturday, 12 December, 2015**
Examination time:	**3 hours (09-12)**

Write your identification number on each paper and cover sheet (the number stated in the upper right hand corner on your exam cover).

Do not write answers to more than one question in the same cover 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 main questions where each main question includes several sub questions. The points for each sub question are stated after each such question.

Main questions 1 and 2 contain a total of 10 multiple choice sub questions, record your answers to these questions on the separate answer sheet provided. Use separate sheets of graph paper to write your answers to the remaining main questions 3, 4 and 5.

Please give short and precise answers!

The exam can yield 100 points in total. For the grade E 40 points are required, for D 50 points, C 60 points, B 75 points and A 90 points.

Results will be posted on the notice board, House A, floor 3, on the 7th of January 2015 at the latest

---

Good luck!

Question 1: Multiple choice (20 points, 4 points each)

Indicate your answer on the separate answer sheet provided. Only give one answer per question.

1) The notation for panel data is (X_{it}, Y_{it}), i = 1, ..., n and t = 1, ..., T because
A) we take into account that the entities included in the panel change over time and are replaced by others.
B) the X's represent the observed effects and the Y the omitted fixed effects.
C) there are n entities and T time periods.
D) n has to be larger than T for the OLS estimator to exist.

2) The difference between an unbalanced and a balanced panel is that
A) you cannot have both fixed time effects and fixed entity effects regressions.
B) an unbalanced panel contains missing observations for at least one time period or one entity.
C) the impact of different regressors are roughly the same for balanced but not for unbalanced panels.
D) in the former you may not include drivers who have been drinking in the fatality rate/beer tax study.

3) Consider the special panel case where T = 2. If some of the omitted variables, which you hope to capture in the changes analysis, in fact change over time, then the estimator on the included change regressor
A) will be unbiased only when allowing for heteroskedastic-robust standard errors.
B) may still be unbiased.
C) will only be unbiased in large samples.
D) will always be unbiased.

4) 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.

5) In the Fixed Effects regression model, you should exclude one of the binary variables for the entities when an intercept is present in the equation
A) because one of the entities is always excluded.
B) because there are already too many coefficients to estimate.
C) to allow for some changes between entities to take place.
D) to avoid perfect multicollinearity.

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

Indicate your answer on the separate answer sheet provided. Only give one answer per question.

1) Estimation of the IV regression model
A) requires exact identification.
B) allows only one endogenous regressor, which is typically correlated with the error term.
C) requires exact identification or overidentification.
D) is only possible if the number of instruments is the same as the number of regressors.

2) The conditions for valid instruments ***do not*** include the following:
A) each instrument must be uncorrelated with the error term.
B) each one of the instrumental variables must be normally distributed.
C) at least one of the instruments must enter the population regression of X on the Zs and the Ws.
D) perfect multicollinearity between the predicted endogenous variables and the exogenous variables must be ruled out.

3) The IV regression assumptions include all of the following with the exception of
A) the error terms must be normally distributed.
B) $E(ui \mid W_{1i},\ldots, W_{ri}) = 0$.
C) Large outliers are unlikely: the X's, W's, Z's, and Y's all have nonzero, finite fourth moments.
D) $(X_{1i},\ldots, X_{ki}, W_{1i},\ldots,W_{ri}, Z_{1i}, \ldots Z_{mi}, Y_i)$ are i.i.d. draws from their joint distribution.

4) The rule-of-thumb for checking for weak instruments is as follows: for the case of a single endogenous regressor,
A) a first stage F must be statistically significant to indicate a strong instrument.
B) a first stage $F > 1.96$ indicates that the instruments are weak.
C) the t-statistic on each of the instruments must exceed at least 1.64.
D) a first stage $F < 10$ indicates that the instruments are weak.

5) In the case of the simple regression model $Y_i = \beta_0 + \beta_1 X_i + u_i$, $i = 1, \ldots,$ n, when X and u are correlated, then
A) the OLS estimator is biased in small samples only.
B) OLS and TSLS produce the same estimate.
C) X is exogenous.
D) the OLS estimator is inconsistent.

Question 3: Omitted variables (20 points)

We are interested in estimating the returns to schooling. We specify the following equation:

$$Wage_i = \beta_0 + \beta_1 S_i + u_i \qquad (1)$$

where $Wage_i$ is the hourly wage rate, S_i is years of schooling.

(i) When you estimate the equation (1) with OLS you get an estimate of β_1 equal to 0.12. Interpret this estimated coefficient. (**4 points**)

(ii) Assume that $u_i = \beta_2 Ability_i + \varepsilon_i$ and that $cov(S_i, \varepsilon_i) = 0$. Also consider the following help regression

$$Ability_i = \gamma_0 + \gamma_1 S_i + \varepsilon_i$$

Say that we know that $\gamma_1 = 1$ and that $\beta_2 = 0.07$. Based on this information what is the size of the omitted variable bias? (**6 points**)

(iii) You have a measure of an iq test (IQ_i) conducted very early in individual's life (it is therefore not a function of years of schooling) and you estimate the following model

$$Wage_i = \beta_0 + \beta_1 S_i + \beta_2 IQ_i + u_i$$

and your OLS-estimate of β_1 turns out to be equal to 0.10. You then run the following regression

$$S_i = \alpha_0 + \alpha_1 IQ_i + \varepsilon_i$$

by OLS and calculates the residual: $\hat{\varepsilon}_i = S_i - (\hat{\alpha}_0 + \hat{\alpha}_1 IQ_i)$. You then run the following regression by OLS

$$Wage_i = \lambda_0 + \lambda_1 \hat{\varepsilon}_i + v_i$$

What is your OLS-estimate of λ_1? Explain intuitively why. (**10 points**)

Question 4. Interaction model (20 points)

We are interest in gender wage differentials and you estimate the following equation:

$$Wage_i = \beta_0 + \beta_1 S_i + \beta_2 Fem_i + \beta_3 Fem_i \times sch_i + u_i$$

where $Wage_i$ is the hourly wage rate (in SEK), Sch_i is years of schooling and Fem_i is a dummy variable taking the value 1 if the individual is a female and 0 otherwise.

(i) Express the marginal effect of being a female versus a male and interpret this marginal effect. (**10 points**)

(ii) Suggest how you would specify the above equation in order to directly estimate a general wage gap. Express the marginal effect also in this case. (**10 points**)

Question 5: IV (20 points)

We are interested in estimating the effect of a labour market training on labour earnings. We specify the following equation:

$$\ln(Earnings_i) = \beta_0 + \beta_1 D_i + u_i$$

where $Earnings_i$ is annual labour earnings, D_i is a dummy variable taking the value 1 if the individual has participated in the program, 0 otherwise.

The evaluation of the labour market training program was conducted together with the Employment Office and you had the opportunity to randomise individuals to the training program. The randomisation, i.e., the instrument, is represented by Z_i, which takes the value 1 if the individual was randomised to training, otherwise.

The OLS-estimate of β_1 in the above equation is 0.15.

You estimate the following two equations

$$\ln(Earnings_i) = \pi_0 + \pi_1 Z_i + \varepsilon_i$$

$$D_i = \gamma_0 + \gamma_1 Z_i + v_i$$

The OLS-estimate of π_1 is 0.08 and the OLS-estimate of γ_1 is 0.8.

(i) Calculate the IV estimate of the effect of the training program. (**6 points**)

(ii) Explain why the OLS and the IV estimate might differ given that the instrument is valid (**8 points**)

(iii) Explain whether you think whether the instrument is exogenous i.e., uncorrelated with u_i (**6 points**)