Stockholm University

Department of Economics

Course name:	**Empirical Methods in Economics 2**
Course code:	**EC2404**
Semester:	**Spring 2015**
Type of exam:	**MAIN**
Examiner:	**Peter Fredriksson**
Number of credits:	**7,5 credits (hp)**
Date of exam:	**Friday 20th of March, 2015**
Examination time:	**3 hours (09:00-12:00)**

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

Use one cover sheet per exam part. 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 2 parts. Part I contains 20 multiple choice questions, worth 2 points each. Record your answers to the Part I multiple-choice questions on the separate answer sheet provided.

Part II contains 2 discussion questions, worth 30 points each. Answer these questions on separate sheets of graph paper.

The total score is 100 points. 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 14th of April at the latest.

Good luck!

Part 1: Multiple Choice Questions (40 points)

Indicate your response on the separate answer sheet provided. Only one answer per question. No credit is given for multiple answers or additional explanations. Two points per question for correct answers.

1) Consider the regression model $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$. Suppose it is reasonable to assume that $E(u_i|X_{1i}, X_{2i}) = E(u_i|X_{2i})$. Then we can causally interpret OLS estimates of
   a. β_0.
   b. β_1.
   c. β_2.
   d. β_1 and β_2.

2) Consider the regression model $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$, where $\beta_1 > 0$. Suppose that X_{2i} is unobserved. The OLS estimate of β_1 is biased upwards if
   a. X_{1i} and X_{2i} are negatively correlated and $\beta_2 < 0$.
   b. X_{1i} and X_{2i} are positively correlated and $\beta_2 < 0$.
   c. X_{1i} and X_{2i} are positively correlated.
   d. X_{1i} and X_{2i} are negatively correlated.

3) Consider the regression model $Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + u_i$. The effect of changing X_i (marginally) on Y_i is given by
   a. $\beta_1 + \beta_2$
   b. β_1
   c. β_2
   d. $\beta_1 + 2\beta_2 X$

4) To test whether or not the population regression function is linear or whether it is better described by the regression in 3), you estimate the regression in 3) and
   a. check whether the regression R^2 is higher than that of the linear regression.
   b. test whether $\beta_1 = 0$ and $\beta_2 = 0$ using an F-test.
   c. test whether $\beta_2 = 0$ using a *t*-test.
   d. check whether the Total Sum of Squares (TSS) is higher than that of the linear regression.

5) The following are all sensible specifications of a non-linear model <u>with the exception</u> of
   a. $Y_i = \beta_0 + \beta_1 X_i + \beta_2 \ln Y_i + u_i$.
   b. $\ln Y_i = \beta_0 + \beta_1 \ln X_i + u_i$.
   c. $\ln Y_i = \beta_0 + \beta_1 X_i + u_i$.
   d. $Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + u_i$.

6) Consider the regression model $Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_i + \beta_3 (X_i \times D_i) + u_i$, where X_i is a continuous variable and D_i is a binary gender dummy variable. To test that the two regressions are identical across the genders, you must use the
   a. *t*-statistic separately for the hypotheses $\beta_2 = 0$, $\beta_3 = 0$.
   b. *F*-statistic for the joint hypothesis that $\beta_0 = 0$, $\beta_1 = 0$.
   c. *t*-statistic for the hypothesis $\beta_3 = 0$.
   d. *F*-statistic for the joint hypothesis that $\beta_2 = 0$, $\beta_3 = 0$.

7) The following tools from multiple regression analysis carry over in a meaningful manner to the linear probability model, with the exception of the
   a. *F*-statistic.
   b. significance test using the *t*-statistic.
   c. 95% confidence interval using 1.96 times the standard error.
   d. regression R^2 .

8) Consider the probit model $\Pr(Y_i = 1 | X_i) = \Phi(\beta_0 + \beta_1 X_i)$, where X_i is an immigrant dummy variable. The marginal effect (ME) of being immigrant (as opposed to native-born) on $\Pr(Y_i = 1)$ is given by
   a. $ME = \Phi(\hat{\beta}_0 + \hat{\beta}_1) - \Phi(\hat{\beta}_0)$.
   b. $ME = \hat{\beta}_1$.
   c. $ME = \Phi(\hat{\beta}_0 + \hat{\beta}_1 \bar{X}) - \Phi(\hat{\beta}_0)$ (where $\bar{X}$ denotes the mean of X_i).
   d. $ME = \Phi(\hat{\beta}_0 + \hat{\beta}_1)$.

9) Errors-in-variables bias
   a. vanishes when sample size is large.
   b. arises from error in the measurement of the independent variable.
   c. can be mitigated with panel data.
   d. arises from error in the measurement of the dependent variable.

10) Sample selection bias occurs when
   a. data are missing at random.
   b. data are missing based on the values of the control variables.
   c. the choice between two samples is made by the researcher.
   d. data are missing based on the unobserved error term in the regression.

11) An empirical analysis is externally valid if
   a. the statistical inference about the causal effect is valid for the population being studied.
   b. the standard errors are heteroskedasticity-robust.
   c. the statistical inference about the causal effect can be generalized from the studied population to other populations.
   d. the results are obtained from a randomized controlled experiment

12) You are interested in the effect of spending on police on crime rates. To address this question you have access to municipality level data for 3 years. The major concern for your study is
   a. simultaneous causality bias.
   b. misreporting of crime rates.
   c. that you cannot rely on large sample inference since there are only 290 municipalities.
   d. that the errors may be correlated over time.

13) Consider the panel data model: $Y_{it} = \alpha_i + \beta_1 X_{it} + u_{it}$. Information on Y_{it} and X_{it} is available for 48 US states over 2 years. You can estimate β_1 in three ways: (i) Define a dummy variable for each US state and estimate the entire model using OLS; (ii) transform the model by "demeaning" the data and estimate the transformed model using OLS; (iii) transform the model by "first-differencing" the data and estimate the transformed model using OLS. Which of the following statements is correct?
   a. (i) and (ii) yield identical estimates of β_1.
   b. (i), (ii), and (iii) yield identical estimates of β_1.
   c. (i) and (iii) yield identical estimates of β_1.
   d. (ii) and (iii) yield identical estimates of β_1.

14) Consider a standard panel data setting. Heteroscedasticity robust standard errors are invalid in large samples if
   a. the errors are homoskedastic
   b. the error variance differs across units
   c. the dependent variable is binary
   d. the errors are serially correlated within unit over time

15) You consider adding individual gender (the effect is measured by β_1) and national unemployment (the effect is measured by β_2) to a panel data model with individual and time fixed effects. Then
   a. only β_1 is identified (i.e. can be estimated).
   b. only β_2 is identified.
   c. Both β_1 and β_2 are identified.
   d. Neither β_1 nor β_2 are identified.

16) When there is a single instrument and a single (endogenous) regressor, the TSLS estimator for the slope can be calculated as follows ($\widehat{cov}(\cdot)$ ($\widehat{var}(\cdot)$) denotes estimated covariance (variance))
   a. $\hat{\beta}_1 = \widehat{cov}(X,Y)/\widehat{var}(X)$.
   b. $\hat{\beta}_1 = \widehat{cov}(Z,X)/\widehat{cov}(Z,Y)$.
   c. $\hat{\beta}_1 = \widehat{cov}(Z,Y)/\widehat{cov}(Z,X)$.
   d. $\hat{\beta}_1 = \widehat{cov}(Z,Y)/\widehat{var}(Z)$.

17) Consider the simple regression model $Y_i = \beta_0 + \beta_1 X_i + u_i$. Suppose X_i and u_i are correlated, and that you have one instrument that you can use to estimate β_1. In this setting
   a. the TSLS is unbiased (provided the instrument is valid).
   b. the TSLS estimator is consistent (provided the instrument is valid).
   c. You can use the *J*-statistic to test the validity of your model.
   d. OLS and TSLS produce the same estimate in large samples.

18) A differences-in-differences (DiD) approach compares a treated group to a comparison group. The key assumption for the validity of the DiD approach is that:
   a. there are no “time effects”.
   b. the time effects do not vary across groups.
   c. the two groups should be similar prior to the intervention.
   d. no other characteristics should affect *Y*.

19) In the fuzzy Regression Discontinuity design:
   a. you get the "treatment" if the assignment variable is above/below a known threshold.
   b. the threshold is unknown.
   c. being above/below a known threshold influences the probability of getting the treatment.
   d. you must control for pre-determined characteristics.

20) In the ideal randomized experiment
   a. you can estimate the average causal effect for individuals participating in the experiment.
   b. you must control for variables that are correlated with the dependent variable.
   c. self-selection bias is a serious issue.
   d. you can estimate the individual causal effects for all individuals participating in the experiment.

Part 2: Discussion Questions (60 points)

Answer the following questions on separate sheets of paper. Answer clearly and concisely. Only legible answers will be considered. If you think that a question is vaguely formulated, specify the conditions used for answering it. Each question is worth 30 points.

Discussion Question 1

The current Swedish government has lowered payroll taxes for youths. A recent paper has analyzed the employment effects of such payroll tax reductions using Swedish data.[1] The authors analyzed a payroll tax reduction implemented in 2007. This reform reduced payroll taxes by 11 percentage points for all individuals who were aged 19-25. The authors had access to annual individual-level data for the time period 2001-2010. The data contain information on employment and standard individual characteristics, such as education, gender, birth year, and immigrant status. The data set included all individuals living in Sweden during these years.

a) Explain how you would estimate the effect of the payroll tax reduction on employment in this setting. Indicate how you would specify the key regression(s) and be clear on how you define key variables of interest.

b) State the key identifying assumption(s). What are the threats to identification? And how would you provide evidence on this (these) assumption(s) in this setting?

[1] Egebark, J. and N. Kaunitz (2013). Do payroll tax cuts raise youth employment? IFAU Working Paper 2013:27.

Discussion Question 2

A recent study by Bennmarker et al. (2013) examined the impact of contracting out job placement services[2] to private providers. To examine this issue, they randomized the *offer* to go to a private provider. Those who got the offer of a private provider could accept or decline the offer. Those who declined the offer got their services from the public provider (*Arbetsförmedlingen*), which was the default option.

The Table below presents a sub-set of the results from Bennmarker et al. (2013). It shows how wage earnings are related to private placement service. The two columns present results from different empirical strategies.

The first strategy (see column (1)) uses the population that got the offer to go to a private provider. The variation in the binary variable "Private provider" comes from individuals accepting or declining the offer. The second strategy (see column (2)) compares those who got the offer to go to a private provider with those that did not get this offer.

Table: Wage earnings and private placement service
(Dependent variable: Wage earnings (SEK); standard errors in parentheses)

	(1)	(2)
Private provider	-3257	--
	(1408)	
Randomized to private provider	--	1050
		(937)
Number of observations	2410	4804

Notes: The regressions also include a constant plus a number of pre-determined control variables (gender, age, education, unemployment history etc.).

a) Which of these two empirical strategies do you prefer? Substantiate your argument.

b) Explain how you would provide evidence in favor of your preferred empirical strategy.

c) What does your preferred empirical strategy suggest about the effect of getting job placement service from a private provider?

[2] In Swedish "job placement services" means "arbetsförmedlingstjänster".