Stockholm University

Department of Economics

Course name:	**Empirical Methods in Economics 2**
Course code:	**EC2404**
Type of exam:	**REGULAR**
Examiner:	**Peter Fredriksson**
Number of credits:	**7,5 credits**
Date of exam:	**Saturday March 23rd 2013**
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 question. 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 1 contains 20 multiple choice question worth 2 points each. Part 2 contains 2 discussion questions worth 30 points each. 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 Friday the 12th of April 2013 at the latest.

Good luck!

Part 1: Multiple Choice Questions (40 points)

Circle the right answer. 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: $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$. To interpret β_1 and β_2 as causal effects, you must assume that
   a. $E(u_i|X_{1i}, X_{2i}) = E(u_i|X_{2i})$.
   b. $E(u_i|X_{1i}, X_{2i}) = 0$.
   c. $E(u_i|X_{1i}) = 0$.
   d. $E(u_i) = 0$.

2) Consider the regression model $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u$. Suppose it is reasonable to assume that $E(u|X_1, X_2) = E(u|X_2)$. This assumption implies that we can causally interpret OLS estimates of
   a. β_0.
   b. β_1.
   c. β_2.
   d. β_1 and β_2.

3) Consider the regression model $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u$, where $\beta_1 > 0$. Suppose that X_2 is unobserved. The OLS estimate of β_1 is biased upwards if
   a. X_1 and X_2 are positively correlated and $\beta_2 > 0$.
   b. X_1 and X_2 are positively correlated and $\beta_2 < 0$.
   c. X_1 and X_2 are positively correlated.
   d. X_1 and X_2 are negatively correlated.

4) Consider testing the hypothesis $\beta_1 = \beta_2$, in a large sample. Your chosen level of significance is 5 %. One of the following statements is <u>not correct</u>:
   a. This hypothesis can be tested using a t-test or an F-test.
   b. In large samples, you reject the hypothesis if the computed F-statistic > 3.84.
   c. This hypothesis can only be tested via an F-test.
   d. In large samples, you reject the hypothesis if the computed t-statistic > 1.96.

5) The slope coefficient in the model $\ln Y_i = \beta_0 + \beta_1 \ln X_i + u_i$ can be interpreted as follows:
   a. a 1 percent change in *X* is associated with a β_1 percent change in *Y*.
   b. a change in *X* by one unit is associated with a ($100\times \beta_1$) percent change in *Y*.
   c. a 1 percent change in *X* is associated with a change in *Y* of ($0.01\times \beta_1$).
   d. a change in *X* by one unit is associated with a β_1 change in *Y*.

6) In the regression model $Y_i = \beta_0 + \beta_1 C_i + \beta_2 F_i + \beta_3 (C_i \times F_i) + u_i$, where *Y* denotes earnings, *C* a dummy variable for having a college degree and *F* a gender dummy variable, β_2
   a. is the gender difference in earnings for someone with a college degree.
   b. is the gender difference in earnings for someone without a college degree.
   c. is the difference in earnings between those with and without a college degree when $F_i = 0$.
   d. cannot be estimated since F_i and $(C_i \times F_i)$ are perfectly collinear when $F_i = 0$.

7) To test whether the regression model in question 6) is identical for males and females you must use the
   a. t-statistic separately for β_1 and β_3.
   b. F-statistic for the joint hypothesis that $\beta_1 = 0$ and $\beta_3 = 0$.
   c. t-statistic separately for β_2 and β_3.
   d. F-statistic for the joint hypothesis that $\beta_2 = 0$ and $\beta_3 = 0$.

8) Examining whether the results from the Tennessee STAR class size experiment extend to primary school students in California or Sweden is an example of looking for
   a. sample selection bias.
   b. internal validity.
   c. simultaneous causality bias.
   d. external validity.

9) Errors-in-variables bias
   a. is only a problem in small samples.
   b. becomes larger as the variance in the explanatory variable increases relative to the error variance.
   c. arises from error in the measurement of the independent variable.
   d. is particularly severe when the source is an error in the measurement of the dependent variable.

10) You are interested in the effects of participating in a training program. You have data on wages after program completion for those who participated in the program and a potential comparison group. A major concern for this study is:
   a. sample selection bias.
   b. misspecification of the functional form.
   c. bias caused by a so-called Hawthorne effect.
   d. that wages are measured with error.

11) 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.

12) One of the following statements is not true. In the linear probability model,
   a. the errors are heteroskedastic.
   b. $\Delta \Pr(Y = 1) / \Delta X_1 \neq \beta_1$.
   c. predicted probabilities can be greater than unity.
   d. predicted probabilities can be less than zero.

13) The following problems could be analyzed using probit and logit estimation with the exception of whether or not
   a. a college student decides to study abroad for one semester.
   b. an applicant will default on a loan.
   c. a college student will attend a certain college after being accepted.
   d. being a female has an effect on earnings.

14) The main advantage of using panel data over cross sectional data is that it
   a. avoids problems associated with mis-specified functional forms.
   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. implies that inference based on the standard normal distribution is more appropriate.

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 can be estimated.
   b. only β_2 can be estimated.
   c. Both β_1 and β_2 can be estimated.
   d. Neither β_1 nor β_2 can be estimated.

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. $\hat{\beta}_1 = \widehat{cov}(Z, Y) / \widehat{cov}(Z, X)$.
   b. $\hat{\beta}_1 = \widehat{cov}(X, Y) / \widehat{var}(X)$.
   c. $\hat{\beta}_1 = \widehat{cov}(Z, X) / \widehat{cov}(Z, Y)$.
   d. $\hat{\beta}_1 = \widehat{cov}(Z, Y) / \widehat{var}(Z)$.

17) In one of the following cases, TSLS estimation is not possible:
   a. the number of instruments equals the number of endogenous regressors.
   b. the model is over-identified.
   c. the model is under-identified.
   d. the model is exactly identified.

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 individual causal effects for all 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 average causal effect for 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

Earnings functions, whereby the log of earnings is regressed on years of education, and individual characteristics, have been used to estimate the returns to education. A major concern has been the fact that ability should enter as a determinant of earnings, but that it is close to impossible to measure and therefore ability represents an omitted variable. The OLS estimator of the returns to education could thus be biased.

a) What is the likely sign of the bias of the OLS estimator?

To overcome this bias, various authors have used instrumental variable estimation techniques. For each of the potential instruments listed below discuss instrument validity.

b) The individual's postal zip code.

c) The individual's IQ or tests core on a work related exam.

d) Years of education for the individual's mother or father.

e) Number of siblings the individual has.

Discussion Question 2

A recent paper investigated whether direct democracy (as compared to representative democracy) yields an increase in public welfare spending using Swedish data.[1] To estimate the effect of direct democracy they used an institutional rule. The Local Government Act from 1918 stipulated that all local governments with a population of more than 1,500 people were required by law to have representative democracy, while those below this limit were given a choice between direct democracy and representative democracy. The rule came into effect starting with the election in 1919.

In order to evaluate the impact of direct democracy on local government spending, the authors had access to data for 2,500 local governments during 1917-1938. During this period there were elections in 1919, 1922, 1926, 1930, and 1934. The data set includes information on, e.g., public welfare spending, population size, direct/representative democracy, and area size (km^2), for the entire time period.

a) Explain how you would estimate the effect of direct democracy on public welfare spending during 1919-1921. Indicate how you would specify the key regression(s) and be clear on how you define key variable(s) of interest.

b) Discuss the validity of the research design. How would you provide evidence on the validity of the design?

[1] Hinnerich, B.T. and P. Pettersson-Lidbom (2012), "Democracy, Redistribution, and Political Participation: Evidence from Sweden 1919-1938", manuscript, Department of Economics, Stockholm University.