RETAKE EXAMINATION FOR ECONOMETRICS

JUNE 07, 2006
(10:00 - 13:00)
Instructions

(1) The question paper is in three parts. Each part is worth 60 marks and should be answered completely.

(2) Do not spend longer than 1 hour in total in answering each part.

(3) Please keep the answers to each part separate - i.e have separate pages for the AB, RS and AI parts of the exam, in order to allow us to mark the exam in parallel. The AI part can be saved on the computer by following the specific instructions that you will receive on the day of the exam. PLEASE REMEMBER TO WRITE YOUR ID-NUMBER ON EACH SHEET OF PAPER USED.

(1) **Probability and Statistics - Anindya Banerjee (1 hour)**

Questions (60 marks)

(1) Define carefully:

a. (5 marks) joint probability density function

b. (5 marks) stochastic independence

c. (5 marks) maximum likelihood

(2) a. (2 marks) Write down an expression for the moment generating function (m.g.f.) of a random variable X with density function $f_X(x)$.

b. (5 marks) Suppose the density function (Bernoulli) of X is given by

$$\begin{aligned} X &= 0 \text{ with probability } (1-p) \\ X &= 1 \text{ with probability } p \end{aligned}$$

what is the m.g.f. of X?

c. (20 marks) Now consider a random variable Y (Binomial) which has density function given by

$$B(n,y,p) = P(Y=y) = \frac{n!}{(n-y)y!}p^y(1-p)^{n-y}; y = 0,1,2,...n.$$

Show that the m.g.f. of Y is given by

$$(q+pe^t)^n, \text{where } q = (1-p)$$

d. (5 marks) Next, consider a sequence of n independently and identically distributed random variables $(X_1, X_2, X_3...X_n)$ and let

$$Z = (X_1 + X_2 + ...X_n)$$

What is the m.g.f. of Z?

e. (13 marks) Using your results in b., c. and d. above, show that the binomial density function $B(n,y,p)$ can be derived as the sum of n independent (and identically distributed) Bernoulli densities with parameter p.

Regression - Richard Spady (1 hour)

Questions (60 marks)

(1) The following question is based on U.S. National Longitudinal Survey data contained in NLS80.DAT. It consists of 935 employed American men and their wages, work experience, job tenure, years of completed education, various social and demographic descriptors (including race) and test scores on a conventional IQ test and 'knowledge of the world of work' (kww). For the purposes of this question, assume that the variables on the right-hand side of the regressions to follow are correctly assumed to be exogenous. The left hand side variable is log(wage). The variable 'black' has the value 1 if the subject is African American and 0 otherwise.

a. (15 marks) Explain the regression results in the following table, with particular attention to the estimate of the returns to education and the effect of race on wages. Be sure to discuss the interpretation of every coefficient.

```
================================================
Call:
lm(formula = lwage ~ exper + tenure + educ + iq + kww + black)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.347265   0.125783  42.512  < 2e-16 ***
exper        0.013130   0.003355   3.914 9.75e-05 ***
tenure       0.011080   0.002538   4.365 1.42e-05 ***
educ         0.049011   0.007547   6.494 1.36e-10 ***
iq           0.003455   0.001046   3.303 0.000993 ***
kww          0.005787   0.001907   3.035 0.002472 **
black       -0.133397   0.040550  -3.290 0.001041 **
---

Residual standard error: 0.3772 on 928 degrees of freedom
Multiple R-Squared: 0.203,Adjusted R-squared: 0.1979
F-statistic:  39.4 on 6 and 928 DF,  p-value: < 2.2e-16
================================================
```

To further investigate the effect of race on wages, an additional regression is run that adds the variable 'I(black*educ)' to the previous regression. This variable is obtained by simply multiplying the variables 'black' and 'educ'.

b. (10 marks) Interpret these regression results as they relate to the differences in the effect of education on wages between blacks and others. (In interpreting the coefficient on 'black', it may be useful to remember that 'educ' takes on integer values between 9 and 18 inclusive.) Be sure to illustrate, either graphically or with appropriate computations, the difference between the wages of blacks and non–blacks at different levels of education.

c. (5 marks) Comment on the proposition: 'African–Americans invest less in education because discrimination is a more powerful factor in determining the wages of the educated.' (For this question you may relax the assumption of exogeneity if you wish.)

```
=====================================================
Call:
lm(formula = lwage ~ exper + tenure + educ + iq + kww + black +
I(black * educ))

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)      5.318740   0.127149  41.831  < 2e-16 ***
exper            0.012783   0.003361   3.804 0.000152 ***
tenure           0.011110   0.002537   4.379 1.33e-05 ***
educ             0.051143   0.007676   6.662 4.61e-11 ***
iq               0.003445   0.001045   3.296 0.001017 **
kww              0.005903   0.001907   3.095 0.002026 **
black            0.255798   0.264286   0.968 0.333355
I(black * educ) -0.031010   0.020809  -1.490 0.136498
---

Residual standard error: 0.3769 on 927 degrees of freedom
Multiple R-Squared: 0.2049,Adjusted R-squared: 0.1989
F-statistic: 34.14 on 7 and 927 DF,  p-value: < 2.2e-16
=====================================================
```

2. 'Endogeneity':
   (a) (5 marks) What is it?
   (b) (5 marks) Why does it occur?
   (c) (5 marks) Why is it important?
   (d) (15 marks) What econometric methods can be used to address it?

Microeconometrics - Andrea Ichino (1 hour)

Instructions
Please, rename the folder D:/DATA/EXAM/.../ReName_Me/, substituting ReName_Me with the code you have received from Jessica. In this folder you will also find the dataset to be used in the exam. Pay attention not to make mistakes in writing the code to rename the folder.

Save your programs and a commented log of your computer output in the same folder.
Be concise in your answers.

Introduction

The dataset exametric06.dta, that you can find in D:/DATA/EXAM/.../ReName_Me/, contains data on all the workers on payroll in firm F for 3 years (2003-2005). The variables included are:

- idnumber : Worker identification number
- year : Calendar year
- female : female
- married : married
- schyears : years of schooling
- gwagey : gross real yearly income in Euros
- nweeky : number of weeks worked in the year
- age : age in years
- yten : tenure in years
- cat : main hierarchical level
- level : detailed hierarchical level
- nabs : number of absence episodes in a year
- dabs : days of absence in a year

The variable cat identifies the main hierarchical level of workers as described in the columns of Table 1 (1 = top managers; 6 = low blue collars). The variable level identifies a finer characterization of the hierarchy described in the rows of Table 1 (1 = lowest level; 13 = highest level).

Table 1: Distribution of hierarchical categories and levels in 2004

	Variable *Cat*						
Variable *Level*	manag1	manag2	manag3	wcollar	bcollar1	bcollar2	Total
1	0	0	0	0	0	126	126
2	0	0	0	0	398	16	414
3	0	0	0	541	72	1	614
4	0	0	0	3,382	2	0	3,384
5	0	0	0	2,405	0	0	2,405
6	0	0	0	3,407	0	0	3,407
7	0	0	0	2,499	1	0	2,500
8	0	0	1,754	218	0	0	1,972
9	0	229	55	20	0	0	304
10	0	1,639	69	89	0	0	1,797
11	0	978	0	0	0	0	978
12	0	270	0	0	0	0	270
13	121	0	0	0	0	0	121
Total	121	3,116	1,878	12,561	473	143	18,292

Question (60 marks)

(1) (15 marks) Generate a binary indicator taking value 1 if a subject is a manager. Note that managers are defined as subjects for whom cat ≤ 3.

(2) (15 marks) Generate a binary indicator taking value 1 if a subject is promoted to manager in year 2005. Note that a promotion occurs if someone is a manager in 2005 but not before.

(3) (15 marks) Using data for year 2005 and the econometric model you prefer, provide an estimate of the effect of being female on the probability of becoming a manager. Justify your choice of estimation method.

(4) (15 marks) Do whatever is needed to discuss whether controlling for observables like age and absenteeism makes a difference for your estimates.