UNIVERSIDAD CARLOS III DE MADRID

ECONOMETRICS I

Academic year 2007/08

FINAL EXAM

February, 9th, 2008

- TIME: 2 HOURS 30 MINUTES

Directions:

- BEFORE YOU START TO ANSWER THE EXAM:
  - Fill in your personal information in the optical reading form, which will be the only valid answering document. Remember that you must complete all your identifying data (name and surname(s), and **NIU**, which has 9 digits and always begins by 1000) both in letters and in the corresponding optical reading boxes.
  - Fill in, both in letters and in the corresponding optical reading boxes, the course code (10188) and your group (65 or 75). Also check that you have indicated the **type of exam** you are answering.
- AT THE END OF THE EXAM, YOU MUST HAND OUT THE THE OPTICAL READING FORM, TOGETHER WITH THE QUESTIONNAIRE AND THE PROBLEM SET.
- **Check that this document contains 3 exercises and the questionnaire 34 questions sequentially numbered.**
- Check that the number of exam type that appears in the questionnaire matches the number indicated in the optical reading form.
- Read the questions carefully.
  Whenever a question is referred to a Problem included in the enclosed document, the question will include within parentheses at the beginning of the question the corresponding problem number.
  It is advised to read carefully the text of the problem before answering its corresponding questions.
- For each row regarding the number of each question, fill the box which corresponds with your chosen option in the optical reading form .
- **Each question, that must be answered filling the box which corresponds to the chosen option, only has one correct answer (A, B, C or D).**
  Any question in which more than one answer is selected will be considered incorrect and its score will be zero.
- To obtain a grade of 5 over 10 you must correctly answer **20 questions**.

- If you wish, you may use the answer table as a draft, although such table does not have any official validity.
- You can use the back side of the problem text as a draft (no additional sheets will be handed out).
- The relevant **statistical tables** are attached at the end of the Problem text document.

- **Any student who were found talking or sharing any sort of material during the exam will be expelled out immediately and his/her overall score will be zero, independently of any other measure that could be undertaken.**

- **Date of grades publication**: Friday, February, 15th.
- **Date of exam revision**: Tuesday, February, 19th, at 18:00 h, classrooms 14.1.1 y 14.1.4.
- **Rules for exam revision**:
  - Its only purpose will be to check that the number of correct answers is right.
  - To be entitled for revision, the student should bring a *printed copy* of the exam solutions, which will be available in Aula Global from the date of grades publication.
  - Any complaint about hypothetical errors in the exam contents with respect to the official solutions must be done in writing. The complaint sheet must be hand in at the moment of the exam revision, indicating name and surname, NIU, and university e-mail address. After the exam revision, no complaint will be accepted in any case. If, in a five days time, the complaint were not answered and/or the grade were not amended in Aula Global, it must be understood that the complaint has been disregarded, what would end up the complaint procedure to the course professor.

Borrador de RESPUESTAS									
PREGUNTA	**(a)**	**(b)**	**(c)**	**(d)**	**PREGUNTA**	**(a)**	**(b)**	**(c)**	**(d)**
1.					**18.**				
2.					**19.**				
3.					**20.**				
4.					**21.**				
5.					**22.**				
6.					**23.**				
7.					**24.**				
8.					**25.**				
9.					**26.**				
10.					**27.**				
11.					**28.**				
12.					**29.**				
13.					**30.**				
14.					**31.**				
15.					**32.**				
16.					**33.**				
17.					**34.**				

Problem 1. Long term effect of family size.

We wish to study the long term effects of family size. To do so we use a sample of adults aged between 25 and 55 years, that grew up in households with at least one sibling.

For these individuals, we have collected information over the total number of siblings ($n1childg1$) with which they grew up and years of education ($yedu$).

Further, we have information regarding the years of mother´s education ($myedu$), father´s education ($pyedu$), a dummy variable to control for whether the mother was working while they were kids ($mwork$, taking the value of 1 if the mother worked; 0 otherwise), a dummy variable for whether they studied in a rural area during their childhood ($purbanm$ taking the value of 1 if they studied in an urban area and 0 if they studied in a rural area), three dummy variables ("$pobre$" (poor), "$buena$" (good) and "$muy\ buena$" (very good)) referring to the economic state of the household where the individual grew up (the omitted category is "indigente" (very poor)).

$$\begin{aligned} yedu &= \beta_0 + \beta_1 n1childg1 + \beta_2 myedu + \beta_3 pyedu + \beta_4 mwork \\ &\quad + \beta_5 purbanm + \beta_6 pobre + \beta_7 buena + \beta_8 mbuena + u. \end{aligned} \tag{1}$$

We have evidence that the number of siblings, $n1childg1$ is an endogenous variable in the model. To solve the endogeneity problem, we have access to two instrumental variable candidates. The first instrument is a dummy variable, $boy12$, taking the value of one if the parents of the individual had two boys in the first two births. The second instrument, $girl12$, is a dummy variable taking the value of one if the parents of the individual had two girls in the first two births.

Output 1: OLS estimates using 4077 observations from 1–4993
Missing or incomplete observations dropped: 916
Dependent variable: yedu

Variable	Coefficient	Std. Error	t-statistic	p-value
const	5.637	0.274	20.538	0.0000
n1childg1	−0.151	0.0235	−6.415	0.000
myedu	0.126	0.0126	10.0129	0.0000
pyedu	0.134	0.012	11.462	0.000
mwork	−0.039	0.078	−0.500	0.617
purbanm	5.192	0.108	48.143	0.000
pobre	0.413	0.246	1.676	0.094
buena	0.943	0.246	3.836	0.0001
mbuena	1.310	0.326	4.013	0.0001

Mean of dependent variable	11.895
S.D. of dependent variable	3.7644
Sum of squared residuals	23836.4
Standard error of residuals ($\hat{\sigma}$)	2.4206
Unadjusted R^2	0.5873
Adjusted $\bar{R}^2$	0.5865
$F(8, 4068)$	723.69
Log-likelihood	−9384.7

Output 2: OLS estimates using 4083 observations from 1–4993
Missing or incomplete observations dropped: 910
Dependent variable: n1childg1

Variable	Coefficient	Std. Error	t-statistic	p-value
const	5.251	0.165	31.873	0.0000
myedu	−0.0396	0.0084	−4.7312	0.0000
pyedu	−0.0194	0.0077	−2.5031	0.0123
mwork	−0.3121	0.0515	−6.0578	0.0000
purbanm	−0.7499	0.0709	−10.576	0.000
pobre	−0.241	0.164	−1.470	0.142
buena	−0.446	0.163	−2.731	0.006
mbuena	−0.580	0.217	−2.672	0.008
boy12	−0.0078	0.0617	−0.126	0.899
girl12	0.296	0.062	4.77	0.00

Mean of dependent variable	3.748
S.D. of dependent variable	1.704
Sum of squared residuals	10566.3
Standard error of residuals ($\hat{\sigma}$)	1.61
Unadjusted R^2	0.109
Adjusted $\bar{R}^2$	0.106
$F(9, 4073)$	55.267

Output 3: OLS estimates using 4083 observations from 1–4993
Missing or incomplete observations dropped: 910
Dependent variable: n1childg1

Variable	Coefficient	Std. Error	t-statistic	p-value
const	5.316	0.1631	32.594	0.000
myedu	−0.0407	0.00839	−4.85	0.000
pyedu	−0.0186	0.0077	−2.392	0.0168
mwork	−0.3103	0.0516	−6.007	0.0000
purbanm	−0.738	0.071	−10.390	0.000
pobre	−0.254	0.164	−1.542	0.123
buena	−0.446	0.164	−2.718	0.0066
mbuena	−0.579	0.218	−2.654	0.008

Mean of dependent variable	3.75
S.D. of dependent variable	1.70
Sum of squared residuals	10637.3
Standard error of residuals ($\hat{\sigma}$)	1.61
Unadjusted R^2	0.103
Adjusted $\bar{R}^2$	0.102
$F(7, 4075)$	66.94

Output 4: TSLS estimates using 4077 observations from 1–4993
Missing or incomplete observations dropped: 916
Dependent variable: yedu
Instruments: boy12 girl12

Variable	Coefficient	Std. Error	t-statistic	p-value
const	1.490	1.800	0.828	0.408
n1childg1	0.630	0.335	1.881	0.060
myedu	0.158	0.0196	8.043	0.000
pyedu	0.148	0.0145	10.195	0.000
mwork	0.202	0.135	1.494	0.135
purbanm	5.768	0.275	20.997	0.000
pobre	0.611	0.291	2.104	0.035
buena	1.291	0.3146	4.104	0.000
mbuena	1.761	0.4156	4.238	0.000

Mean of dependent variable	11.89
S.D. of dependent variable	3.764
Sum of squared residuals	30304.1
Standard error of residuals ($\hat{\sigma}$)	2.729
$F(8, 4068)$	484.258

Output 5: OLS estimates using 4077 observations from 1–4993
Missing or incomplete observations dropped: 916
Dependent variable: yedu

Variable	Coefficient	Std. Error	t-statistic	p-value
const	1.516	1.586	0.956	0.339
n1childg1	0.624	0.295	2.119	0.034
myedu	0.158	0.017	9.088	0.000
pyedu	0.148	0.0129	11.506	0.000
mwork	0.201	0.120	1.682	0.093
purbanm	5.764	0.242	23.808	0.000
pobre	0.610	0.257	2.370	0.018
buena	1.289	0.278	4.629	0.000
mbuena	1.758	0.368	4.780	0.000
fs_resid	−0.780	0.296	−2.639	0.008

Mean of dependent variable	11.89
S.D. of dependent variable	3.76
Sum of squared residuals	23795.7
Standard error of residuals ($\hat{\sigma}$)	2.42
Unadjusted R^2	0.588
Adjusted $\bar{R}^2$	0.587
$F(9, 4067)$	645.00

fs resid son los residuos de la Salida 2

PROBLEM 2. Determinants of sporting activity

We wish to study the determinants of participation in sports. We have information for adults with ages ranging between 25 and 55 years. The dummy variable *sport*, takes the value of one if the individual has participated in any sport during the past week and zero otherwise. The explanatory variables in our model are the *binary* variable $female$, taking the value of one if the individual is a female, and the *continuous* variables age (age), age squared ($age2$), and years of education ($yedu$).

Output 1: OLS estimates using 4986 observations from 1–4993
Missing or incomplete observations dropped: 7
Dependent variable: sport
Heteroskedasticity-robust standard errors, variant HC1

Variable	Coefficient	Std. Error	t-statistic	p-value
const	0.793	0.141	5.624	0.000
female	−0.276	0.0127	−21.725	0.000
age	−0.020	0.0074	−2.701	0.007
age2	0.0002	1.086e-04	1.841	0.066
yedu	0.0148	0.0017	8.768	0.000

Mean of dependent variable	0.359807
S.D. of dependent variable	0.479992
Sum of squared residuals	1021.94
Standard error of residuals ($\hat{\sigma}$)	0.452955
Unadjusted R^2	0.110198
Adjusted $\bar{R}^2$	0.109483
$F(4, 4981)$	173.525
Log-likelihood	−3123.6

Output 2: Logit estimates using 4986 observations from 1–4993
Missing or incomplete observations dropped: 7
Dependent variable: sport

Variable	Coefficient	Std. Error	t-statistic	Slope*
const	1.277	0.692	1.845	
female	−1.307	0.065	−19.999	−0.294
age	−0.084	0.0365	−2.296	−0.019
age2	0.00066	0.00047	1.421	0.00015
yedu	0.074	0.0085	8.695	0.0166

*Evaluated at the mean

Mean of sport = 0.360
Number of cases 'correctly predicted' = 3421 (68.6 percent)
McFadden's pseudo-R^2 = 0.0891587
$f(\beta' x)$ at mean of independent vars = 0.225
Log-likelihood = −2968.54
Likelihood ratio test: χ^2_4 = 581.157 (p-value 0.000)

PROBLEM 3. Female wage rates

A researcher wishes to analyse whether the female annual wage rate, w, depends on two factors: if the female is a working mother (measured by a binary variable m, taking the value of 1 if she is), and whether she is married or not (c, that takes the value of 1 she is married). The following specifications have been proposed:

$$w = \beta_0 + \beta_1 m + \beta_2 c + \beta_3 (m \times c) + u \quad (2)$$

$$w = \beta_1 (m \times c) + u \quad (3)$$

$$w = \beta_0 + \beta_1 m + \beta_2 c + \beta_3 (m \times c) + \beta_4 (1 - m) + u \quad (4)$$

$$w = \beta_1 m + \beta_2 c + \beta_3 (m \times c) + \beta_4 (1 - c) + u. \quad (5)$$

In addition, we know that experience ($x_1 = exper$) is also an important variable in wage formation aside from m and c.