UNIVERSIDAD CARLOS III DE MADRID
ECONOMETRÍA I
Academic year 2008/09
FINAL EXAM

January, 19, 2009

Very important: Take into account that:

1. Each question, unless otherwise stated, requires a complete analysis of all the outputs shown in the corresponding problem.
   For example, to answer those questions referring to "appropriate estimates", or "given the estimates" or "given the problem conditions", the results based on the consistent and more efficient among outputs, must be used.

2. Each output includes all the explanatory variables used in the corresponding estimation.

3. Some results in the output shown may have been omitted.

4. The dependent variable can vary among outputs within the same problem.

5. For the sake of brevity, we will say that a model is well specified if it is linear in the conditioning variables (as they appear in the model) and its error term is mean-independent of such variables.

6. OLS and 2SLS are the corresponding abreviations of ordinary least squares and two stage least squares, respectively.

7. Statistical tables are included at the end of this document.

PROBLEM TEXT

PROBLEM 1: Return to education

We are interested in estimating a wage equation using a sample of male individuals. The dependent variable is the natural logarithm of the monthly wage, *lwage*. The explanatory variables for each individual are his years of education, *educ*, his age, *age*, a binary variable indicating his civil status, *married* (which equals one if the individual is married and zero otherwise), and his ability, *abil*.

$$lwage = \beta_0 + \beta_1 educ + \beta_2 age + \beta_3 married + \beta_4 abil + \varepsilon \tag{1.1}$$

where $E\left(\varepsilon \middle| educ, age, married, abil\right) = 0$. We would expect $\beta_4 > 0$.

Since *abil* is unobservable, the equation that we can estimate omits such variable, and we thus estimate the following wage equation:

$$lwage = \delta_0 + \delta_1 educ + \delta_2 age + \delta_3 married + u \tag{1.2}$$

We have strong reasons to believe that the years of education, *educ*, are correlated with the unobserved ability, *abil*, which is omitted. On the contrary, we can assume that such omitted variable is uncorrelated with the remaining observed explanatory variables, *age* and *married*.

As possible instruments for *educ*, we have two observed variables available, not included in the model, which are uncorrelated with ability:

- *urban* (binary variable that equals one if the individual lives in a city with more than 50000 inhabitants and zero otherwise);
- *feduc* (father's years of education).

Our aim is to obtain consistent estimates of the parameters of (1.1), β_1, β_2, β_3, from the observed information.

Output 1: OLS estimates using the 663 observations 1–663
Dependent variable: *lwage*

Variable	Coefficient	Std. Error	$t-$ statistic	p-value
const	5.0578	0.1831	27.62	0.0000
educ	0.0597	0.0066	9.03	0.0000
age	0.0228	0.0048	4.71	0.0000
married	0.2101	0.0494	4.25	0.0000

Sum of squared residuals	94.4723
Unadjusted R^2	0.1601
Adjusted $\bar{R}^2$	0.1563
$F(3, 659)$	41.88

Output 2: OLS estimates using the 663 observations 1–663
Dependent variable: *educ*

Variable	Coefficient	Std. Error	$t-$ statistic	p-value
const	12.4497	0.9632	12.92	0.0000
age	0.0383	0.0282	1.35	0.1761
married	−0.4256	0.2890	−1.47	0.1413
urban	0.4888	0.4161	1.17	

Sum of squared residuals	3244.91
Unadjusted R^2	0.0156
$F(3, 659)$	3.47

Output 3: OLS estimates using the 663 observations 1–663
Dependent variable: *educ*

Variable	Coefficient	Std. Error	$t-$ statistic	p-value
const	9.1774	0.9140	10.04	0.0000
age	0.0565	0.0257	2.20	0.0281
married	−0.3868	0.2623	−1.47	0.1407
feduc	0.2907	0.0239	12.18	0.0000

Sum of squared residuals	2674.57
Unadjusted R^2	0.1886
$F(3, 659)$	51.06

Output 4: 2SLS estimates using the 663 observations 1–663
Dependent variable: *lwage*
Instruments: *urban*

Variable	Coefficient	Std. Error	$t-$ statistic	p-value
const	-0.2134	2.2910	-0.09	0.9258
educ	0.4700	0.1753	2.68	0.0073
age	0.0073	0.0142	0.51	0.6087
married	0.3964	0.1515	2.62	0.0089

Sum of squared residuals	646.232
$F(3, 659)$	29.27

Hausman Test –
Null hypothesis: The OLS estimates are consistent
Asymptotic test statistic: $\chi_1^2 = 40.42$
with p-value = 2.04972e-010

Output 5: 2SLS estimates using the 663 observations 1–663
Dependent variable: *lwage*
Instruments: *feduc*

Variable	Coefficient	Std. Error	$t-$ statistic	p-value
const	4.4176	0.2669	16.55	0.0000
educ	0.1095	0.0161	6.81	0.0000
age	0.0209	0.0051	4.13	0.0000
married	0.2327	0.0519	4.49	0.0000

Sum of squared residuals	102.610
$F(3, 659)$	38.79

Hausman Test –
Null hypothesis: The OLS estimates are consistent
Asymptotic test statistic: $\chi_1^2 = 13.11$
with p-value = 0.000293096

PROBLEM 2: Determinants of the SAT

The variable *sat* is the grade obtained in the scholar aptitude test (SAT), *hsize* is the size of the class-year (measured in hundred students) to which the student belongs, *female* is a binary gender variable (that equals 1 if the student is a women and 0 otherwise), and *black* is a race binary variable (that equals 1 if the student is black and 0 otherwise). The following model was proposed to estimate the effects of several factors on the SAT grades,

$$sat = \beta_0 + \beta_1 hsize + \beta_2 hsize2 + \beta_3 female + \beta_4 black + \beta_5 femaleblack + u \quad (2.1)$$

where $hsize2$ is the square of $hsize$ and the variable $femaleblack$ is the interaction variable $female \times black$.

A more general model was also considered, which includes a further effect that the student is an athlete, through the variable *athlete* (that equals 1 if the student is an athlete and 0 otherwise), as well as the interaction variable $athleteblack = athlete \times black$.

Output 1: OLS estimates using the 4137 observations 1–4137
Dependent variable: *sat*

	Coefficient	Std. Error	$t-$ statistic	p-value
const	1028.1000	6.2902	163.44	0.0000
hsize	19.2971	3.8323	5.03	0.0000
*hsize*2	−2.1948	0.5272	−4.16	0.0000
female	−45.0910	4.2911	−10.51	0.0000
black	−169.8100	12.7131	−13.36	0.0000
femaleblack	62.3064	18.1542	3.43	0.0006

Mean of dependent variable	1030.33
S.D. of dependent variable	139.401
Sum of squared residuals	7.34791e+07
Standard error of residuals ($\hat{\sigma}$)	133.369
Unadjusted R^2	0.0858
Adjusted $\bar{R}^2$	0.0847
$F(5, 4131)$	77.52

Output 2: OLS estimates using the 4137 observations 1–4137
Dependent variable: *usq*1 (squared residuals of *Output 1*)

	Coefficient	Std. Error	$t-$ statistic	p-value
const	19456.50	1195.80	16.27	0.0000
hsize	25.94	728.54	0.04	0.9716
*hsize*2	−43.98	100.22	−0.44	0.6608
female	−3226.80	815.76	−3.96	0.0001
black	7445.69	2416.85	3.08	0.0021
femaleblack	−9217.30	3451.23	−2.67	0.0076

Sum of squared residuals	3.2907e+12
Unadjusted R^2	0.0357
Adjusted $\bar{R}^2$	0.0346
$F(5, 4131)$	30.61

Output 3: OLS estimates using the 4137 observations 1–4137
Dependent variable: *usq*1 (squared residuals of *Output 1*)

	Coefficient	Std. Error	$t-$ statistic	p-value
const	2024.31	607.05		
female	−3892.49	907.94		
black	26726.56	2689.55		
femaleblack	−11921.12	3838.87		

Unadjusted R^2	0.0355
$F(3, 4131)$	50.78

Output 4: OLS estimates using the 4137 observations 1–4137
Dependent variable: *sat*

	Coefficient	Std. Error	$t-$ statistic	p-value
const	1033.6200	6.2360	165.75	0.0000
hsize	18.0214	3.7787	4.77	0.0000
*hsize*2	−1.8931	0.5202	−3.64	0.0003
female	−48.4500	4.2450	−11.41	0.0000
black	−130.2500	14.1122	−9.23	0.0000
femaleblack	36.7119	18.3839	2.00	0.0459
athlete	−97.5820	11.0091	−8.86	0.0000
athleteblack	−59.1710	25.2181	−2.35	0.0190

Universidad Carlos III de Madrid
<u>ECONOMETRICS I</u>
Academic year 2008/09
FINAL EXAM
January, 19, 2009

Exam type: 1

TIME: 2 HOURS 30 MINUTES

Directions:

- BEFORE YOU START TO ANSWER THE EXAM:
  - Fill in your personal data 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 digit 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).
- **Check that this document contains 55 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 problem text and the questions carefully.
  Whenever applicable, the question will include within parentheses the corresponding problem number at the beginning of the question.
- For each row regarding the number of each question, fill the box which corresponds with your chosen option in the optical reading form (A, B, C or D).
- **Each question only has one correct answer.**
  Any question in which more than one answer is selected will be considered incorrect and its score will be zero.
- All the questions correctly answered has the same score. Any incorrect answer will score as zero. To obtain a grade of 5 over 10 you must correctly answer **32** 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).
- **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.**
- **Dates of grades publication:** Friday, January, 23.
- **Date of exam revision:** January, 26 at 15 h (the place will be announced in Aula Global).

- **Rules for exam revision:**
  - Its only purpose will be that each student:
    - ∗ check the number of correct answers;
    - ∗ handout in writing, if (s)he wants, the possible claims about the problem text and the questions, that will be attended by writing in the next 10 days since the revision date.
  - 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 grade publication.

Draft of ANSWERS														
	(a)	(b)		(a)	(b)		(a)	(b)		(a)	(b)		(a)	(b)
1.			**12.**			**23.**			**34.**			**45.**		
2.			**13.**			**24.**			**35.**			**46.**		
3.			**14.**			**25.**			**36.**			**47.**		
4.			**15.**			**26.**			**37.**			**48.**		
5.			**16.**			**27.**			**38.**			**49.**		
6.			**17.**			**28.**			**39.**			**50.**		
7.			**18.**			**29.**			**40.**			**51.**		
8.			**19.**			**30.**			**41.**			**52.**		
9.			**20.**			**31.**			**42.**			**53.**		
10.			**21.**			**32.**			**43.**			**54.**		
11.			**22.**			**33.**			**44.**			**55.**		

1. (**Problem 1**) The consequences of heteroskedasticity on 2SLS inference are similar to those on OLS inference, and heteroskedasticity can thus be tested and treated analogously for both estimators.

   (a) False.

   (b) True.

2. (**Problem 1**) If the error term is correlated with any explanatory variable, the OLS estimator is biased and inconsistent, but the inconsistency bias in negligible as the simple size increases.

   (a) False.

   (b) True.

3. (**Problem 1**) Using the appropriate estimates, given an age and a civil status, an additional year of education implies, on average, an estimated wage increase of 10.95%.

   (a) False.

   (b) True.

4. (**Problem 1**) Using the appropriate estimates, given an age and a civil status, an additional year of education implies, on average, an estimated wage increase of 4.7%.

   (a) False.

   (b) True.

5. (**Problem 1**) Using the appropriate estimates, given an age and a level of education, a married man earns on average 23.3% more than an unmarried man.

   (a) False.

   (b) True.

6. (**Problem 1**) Using the appropriate estimates, we can assert that age does not significantly affects wages.

   (a) False.

   (b) True.

7. (**Problem 1**) Using the appropriate estimates, given the outputs shown, for a given age and civil status, an additional year of education implies, on average, an estimated wage increase of 5.97%.

   (a) False.

   (b) True.

8. (**Problem 1**) Given the estimates, since age and civil status are uncorrelated with ability, we can conclude that the correlation between education and ability is negative.

   (a) False.

   (b) True.

9. (**Problem 2**) Assume that model (2.1) is well specified. Looking at Output 1, the class-year size after which its effect is negative is about 440 students.

   (a) False.

   (b) True.

10. (**Problem 2**) Assume that model (2.1) is well specified. Looking at Output 1, the class-year size after which its effect is negative is about 44 students.

    (a) False.

    (b) True.

11. (**Problem 2**) Assume that model (2.1) is well specified. Looking at Output 1, for a given class-year size, black female students achieve, on average, a SAT grade 45.09 lower than non black male students.

    (a) False.

    (b) True.

12. (**Problem 2**) Assume that model (2.1) is well specified. Looking at Output 1, for a given class-year size, black female students achieve, on average, a SAT grade 17.22 points lower than non black male students.

    (a) False.

    (b) True.

13. (**Problem 2**) Assume that model (2.1) is well specified. Looking at Output 1, for a given class-year size, the average SAT grade of a black male student is about 170 points lower than for a non black male student.

    (a) False.

    (b) True.

14. (**Problem 2**) Assume that model (2.1) is well specified. Looking at Output 1, for a given class-year size, the average SAT grade of a black male student is about 107 point greater than for a non black male student.

    (a) False.

    (b) True.

15. (**Problem 2**) Assume that model (2.1) is well specified. Looking at Output 1, if we take the SAT exams of a black male and a non black male, both randomly chosen, the grade difference will be about 170 points lower for the first one.

    (a) False.

    (b) True.

16. (**Problem 2**) Assume that model (2.1) is well specified. Looking at Output 1, if we take the SAT exams of a black male and a non black male, both randomly chosen, the grade difference will be about 107 points greater for the first one.

    (a) False.

    (b) True.

17. (**Problem 2**) Assume that model (2.1) is well specified. Looking at Output 1, the null hypothesis that there is no difference in the grades achieved by black and non black males could be tested through the F statistic or the asymptotic statistic $W^0 = 2 \times F$ of joint significance of the variables $black$ and $femaleblack$.

    (a) False.

    (b) True.

18. (**Problem 2**) Assume that model (2.1) is well specified. Looking at Output 1, for a given class-year size, the estimated average difference in SAT grades between black and non black female students is about 107.5 lower for the first ones.

   (a) False.
   (b) True.

19. (**Problem 2**) Assume that model (2.1) is well specified. Looking at Output 1, for a given class-year size, the estimated average difference in SAT grades between black and non black female students is about 152.6 lower for the first ones.

   (a) False.
   (b) True.

20. (**Problem 2**) Assume that model (2.1) is well specified. If we want to test the hypothesis that there are no difference in the average SAT grades between black and non black female students, the null hypothesis is $H_0 : \beta_4 + \beta_5 = 0$.

   (a) False.
   (b) True.

21. (**Problem 2**) Assume that model (2.1) is well specified. If we want to test the hypothesis that there are no difference in the average SAT grades between black and non black female students, the null hypothesis is $H_0 : \beta_4 = \beta_5 = 0$.

   (a) False.
   (b) True.

22. (**Problem 2**) Assume that model (2.1) is well specified. If the variable $hsize2$ was omitted, the OLS estimators of the remaining coefficients will be, in general, biased and inconsistent if $\beta_2 \neq 0$, since $hsize$ and $hsize2$ will be correlated.

   (a) False.
   (b) True.

23. (**Problem 2**) Assume that model (2.1) is well specified. The White test statistic for such model is nR^2, where R^2 is the determination coefficient of the regression of Output 1 squared residuals over all the explanatory variables, their squares, and the cross products among all of them.

   (a) False.
   (b) True.

24. (**Problem 2**) Assume that model (2.1) is well specified. Output 2 allows to compute the White test of heteroskedasticity for such model.

   (a) False.
   (b) True.

25. (**Problem 2**) Assume that model (2.1) is well specified. If we find that it is heteroskedastic, both the OLS estimators of the coefficients and their corresponding estimated variances based on the conventional formula are inconsistent.

   (a) False.
   (b) True.

26. (**Problem 2**) Assume that model (2.1) is well specified. If we knew that $E(u^2|hsize, female, black)$ is a non constant function of $hsize$, we could not say anything about the existence or not of conditional homoskedasticity, since conditional homoskedasticity requires that the variance of u, conditional on $hsize$, $hsize2$, $female$, $black$, $femaleblack$, be constant.

   (a) False.

   (b) True.

27. (**Problem 2**) Assume that model (2.1) is well specified. Given Output 2 results, we reject the null hypothesis of the heteroskedasticity test, so that we can conclude that there is heteroskedasticity in such model.

   (a) False.

   (b) True.

28. (**Problem 2**) Assume that model (2.1) is well specified. In the presence of heteroskedasticity, the OLS estimated coefficients in Output 1 would be inconsistent, so that we should use 2SLS to estimate such model.

   (a) False.

   (b) True.

29. (**Problem 2**) Assume that model (2.1) is well specified. In the presence of heteroskedasticity, the standard errors of the coefficients in Output 1 would be inconsistent, and we should compute robust standard errors to make valid statistical inference in such model.

   (a) False.

   (b) True.

30. (**Problem 2**) Assume that model (2.1) is well specified. Given the results, the conditional variance of u differs by gender and ethnic origin, but not by class-year size.

   (a) False.

   (b) True.

31. (**Problem 1**) Given the problem conditions, the OLS estimator of the education coefficient in equation (1.2) will be, in general, a biased and inconsistent estimator of β_1.

   (a) False.

   (b) True.

32. (**Problem 1**) Given the problem conditions, the OLS estimators of the coefficients of *age* and *married* in equation (1.2) will be, in general, biased and inconsistent estimators of β_2 and β_3, respectively.

   (a) False.

   (b) True.

33. (**Problem 1**) The OLS estimator of the education coefficient in equation (1.2) will be, in general, a biased estimator of β_1, so that we should calculate heteroskedasticity robust standard errors.

   (a) False.

   (b) True.

34. (**Problem 1**) The fact that the variable *educ* is or is not endogenous in equation (1.2) depends on whether $\beta_4 \neq 0$ and $Cov(educ, abil) \neq 0$.

   (a) False.

   (b) True.

35. (**Problem 1**) If $\beta_4 \neq 0$, the OLS estimation of equation (1.2) will produce consistent estimators of β_1, β_2 and β_3.

   (a) False.

   (b) True.

36. (**Problem 1**) If $Cov(educ, abil) \neq 0$, the OLS estimation of equation (1.2) will produce consistent estimators of β_1, β_2 and β_3.

   (a) False.

   (b) True.

37. (**Problem 1**) Since $Cov(u, urban) = 0, Cov(u, feduc) = 0$, and given Outputs 2 and 3, both $urban$ and $feduc$ are valid instruments.

   (a) False.

   (b) True.

38. (**Problem 1**) Since $Cov(u, urban) = 0, Cov(u, feduc) = 0$, and given Outputs 2 and 3, only $urban$ is a valid instrument.

   (a) False.

   (b) True.

39. (**Problem 1**) Since $Cov(u, urban) = 0, Cov(u, feduc) = 0$, and given Outputs 2 and 3, only $feduc$ is a valid instrument.

   (a) False.

   (b) True.

40. (**Problem 1**) The coefficients of the 2SLS estimation in Output 5 could have been obtained equivalently estimating equation (1.2) by OLS, but substituting *educ* by its prediction based on the estimates of Output 3.

   (a) False.

   (b) True.

41. (**Problem 1**) The coefficients of the 2SLS estimation in Output 5 could have been obtained equivalently estimating equation (1.2) by 2SLS, using as instrument the prediction of *educ* based on its prediction based on the estimates of Output 3.

   (a) False.

   (b) True.

42. (**Problem 1**) Since $Cov(u, urban) = 0, Cov(u, feduc) = 0$, and given the results of the Hausman test (Output 4 and 5) and outputs 2 y 3, we can conclude that the OLS estimators of Output 1 are consistent.

   (a) False.

   (b) True.

43. (**Problem 1**) Since $Cov(u, urban) = 0, Cov(u, feduc) = 0$, and given the results of the Hausman test (Output 4 and 5) and outputs 2 y 3, we can conclude that the 2SLS estimators of Output 5 are consistent.

   (a) False.

   (b) True.

44. (**Problem 1**) Since $Cov(u, urban) = 0, Cov(u, feduc) = 0$, and given the results of the Hausman test (Output 4 and 5) and outputs 2 y 3, we can conclude that the 2SLS estimators of Output 4 are consistent.

   (a) False.

   (b) True.

45. (**Problem 1**) If we estimated equation (1.2) by OLS using the residuals of Output 3 as additional explanatory variable, the t statistic associated to such residuals would yield an exogeneity test of *educ*.

   (a) False.

   (b) True.

46. (**Problem 1**) If we estimated equation (1.2) by OLS using the residuals of Output 3 as additional explanatory variable, the estimated coefficients of *educ*, *age* and *married* would be equal to the corresponding coefficients of the same variables in Output 5.

   (a) False.

   (b) True.

47. (**Problem 1**) If we had a (third) additional valid instrument for *educ*, we could test whether $feduc$ is uncorrelated with u.

   (a) False.

   (b) True.

48. (**Problem 1**) If we had a (third) additional valid instrument for *educ*, we could obtain an alternative 2SLS estimator, more efficient than the obtained in Outputs 4 y 5.

   (a) False.

   (b) True.

49. (**Problem 1**) Assume that we have an appropriate proxy variable for *abil*: the intelligence quotient (IQ). We could not use IQ as valid instrument for *educ* in equation (1.2), since *abil* is included in the error term of that equation.

   (a) False.

   (b) True.

50. (**Problem 2**) The model estimated in Output 4 allows to test whether there exist differences in SAT grades between athlete male students and athlete female students.

   (a) False.

   (b) True.

51. (**Problem 2**) Given Output 4 results, we can conclude that the estimated difference in SAT grades between black and non black athlete students is not statistically significant.

   (a) False.

   (b) True.

52. (**Problem 2**) Given Output 4 results, for a given class-year size, a non-black athlete female student achieves, on average, about 4 points less in the SAT than a black non-athlete female student.

   (a) False.

   (b) True.

53. (**Problem 2**) Given Output 4 results, we can assert that, on average, the worst SAT grades correspond to black athlete male students.

   (a) False.

   (b) True.

54. (**Problem 2**) Given Output 4 results, we can assert that, on average, the best SAT grades correspond to non-black non-athlete male students.

   (a) False.

   (b) True.

55. (**Problem 2**) Given Output 4 results, we can assert that, for a given gender, class-year size, and ethnic origin, athlete students achieve, on average, SAT grades that are systematically worse than those for non athlete students.

   (a) False.

   (b) True.