Universidad Carlos III de Madrid
ECONOMETRICS I
Academic year 2006/07
FINAL EXAM
February 9, 2008

Tipo de examen: 1

TIME: 2 HOURS 30 MINUTES

1. **(Problem 2.)** In the Linear Probability Model,

   (i) The error term would present heteroskedasticity (conditional on the explanatory variables).

   (ii) The error term (conditional on the explanatory variables) would be normally distributed.

   (iii) The predicted probabilities could be greater than one, but not lower than zero.

   (a) The three statements are true.
   (b) Only (i) is true.
   (c) Only (i) and (iii) are true.
   (d) Only (i) and (ii) are true.

2. **(Problem 2.)** Using the linear probability model estimates,

   (a) The probability to participate in sports increases with age up to the age of 50, approximately.
   (b) The probability to participate in sports always diminishes with age.
   (c) The probability to participate in sports diminishes after the age of 39,5, approximately.
   (d) The probability to participate in sports diminishes with age after the age of 35, approximately.

3. **(Problem 2.)** Using the linear probability model estimates, the probability that a 30 year old male with 10 years of education participates in sports is:

   (a) 0.52.
   (b) 0.94.
   (c) 0.81.
   (d) 0.36.

4. **(Problem 2.)** Using the linear probability model estimates, we can conclude that

   (i) Females have a lower probability to participate in sports, holding all else equal.

   (ii) Age does not affect the decision to participate in sports because the coefficient on *age*2 is not significant at the 5% level.

   (iii) A male with 10 years of education has a higher probability to do sports than a female with 20 years of education, everything else being held equal.

   (a) Only (i) is true.
   (b) (ii) and (iii) are true.
   (c) (i) and (iii) are true.
   (d) (ii) is true.

5. **(Problem 2.)** The estimated linear probability model specification (Output 1),

   (i) Permits measuring the effect of sex on the probability to do sports.

   (ii) Permits that the partial effect of age varies with its value.

   (iii) Permits that the effect of education on the probability to do sports varies by gender.

   (a) Only (i) is true.
   (b) Only (iii) is true.
   (c) Only (i) and (iii) are true.
   (d) Only (i) and (ii) are true.

6. **(Problem 2.)** In a Logit model,

   (i) The partial effect of variable x_j on the probability that $y = 1$ (conditional on the explanatory variables), depends on β_j and x_j, but not on the rest of the explanatory variables or parameters.

   (ii) The partial effect of variable x_j on the probability that $y = 1$ (conditional on the explanatory variables), depends on $\beta_0, \beta_1, \ldots, \beta_k$ and on x_j, but not on the rest of the explanatory variables.

   (iii) The partial effect of variable x_j on the probability that $y = 1$ (conditional on the explanatory variables), depends on β_j and on $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_k$, but not on the rest of the parameters.

   (a) The three statements are false.
   (b) Only (iii) is true.
   (c) Only (ii) is true.
   (d) Only (i) is true.

7. **(Problem 2.)** Using the Logit model estimates, the effect on the probability to participate in sports, of a 10 year increase in education (ceteris paribus) for an individual with the mean characteristics in the sample, is

   (a) 0.74.

   (b) 0.017.

   (c) 0.074.

   (d) 0.17.

8. **(Problem 2.)** Using the Logit model estimates,

   (i) For some females it is possible that the predicted probability to do sports is negative.

   (ii) The difference in the probability to do sports between a male and a female is always 0.29 in favour of the male, independently of his age.

   (iii) The magnitude of the partial effect of years of education on the probability to do sports, diminishes when *yedu* increases if *yedu* is sufficiently high, all else held equal.

   (a) Only (*ii*) and (*iii*) are true.

   (b) Only (*iii*) is true.

   (c) Only (*i*) and (*ii*) are true.

   (d) Only (*i*) is true.

9. **(Problem 2.)** Using the Logit model estimates, to test that the effect of education on the probability to do sports does not depend on gender:

   (a) I would have tested if the coefficients of $female$ and *yedu* are jointly significant with a likelihood ratio test.

   (b) I would have tested if the coefficient on $female$ is significant using a t test.

   (c) It is necessary to estimate a model that is inclusive of interactions of *yedu* with the variable $female$.

   (d) I would have used the log-likelihoods of this model and another which is not inclusive of variables $female$ and *yedu*.

10. **(Problem 2.)** Using the Logit model estimates,

    (i) We can conclude that the explanatory variables aid in predicting the probability to do sports because the percent correctly predicted is fairly high.

    (ii) We can reject the hypothesis that the coefficients on $female$, age, $age2$ and $yedu$ are equal to zero at any usual significance level.

    (iii) The model is not properly specified since the pseudo-R^2 is very low.

    (a) Only (i) and (ii) are true.
    (b) Only (i) is true.
    (c) Only (i) and (iii) are true.
    (d) Only (ii) is true.

11. **(Problem 2.)** Using the Logit model estimates, and assuming that the coefficient on $age2$ is exactly zero,

    (a) The partial effect on the probability to do sports of an additional year of education is equivalent to being 0,88 years younger.
    (b) The partial effect on the probability to do sports of an additional year of education is equivalent to being 0,88 years younger, but only if all explanatory variables are evaluated at the sample means.
    (c) The partial effect on the probability to do sports of an additional year of education is equivalent to being 0,88 years younger, but only for females.
    (d) The partial effect on the probability to do sports of an additional year of education is equivalent to being 0,88 years younger, but only for males.

12. **(Problem 2.)** Using the Logit model estimates, the probability that a 30 year old female 30 with 20 years of education participates in sports is

    (a) 0.29 less than that of a male with equal characteristics.
    (b) 0.27 less than that of a male with equal characteristics.
    (c) 0.29 less than that of a male with the mean characteristics in the sample.
    (d) 0.31 less than that of a male with equal characteristics.

13. **(Problem 3.)** Out of the four suggested specifications:

    (a) Only (2) and (3) can be estimated by OLS.
    (b) None of the specifications can be estimated by OLS because they omit relevant variables.
    (c) The only specification that cannot be estimated by OLS is (4).
    (d) All can be estimated by OLS, but only (2) provides consistent estimators.

14. **(Problem 3.)** Consider the following statements with respect to the previous models:

   (i) It is possible that assumption $E(u|x) = 0$ where x is a vector that contains all the explanatory variables included in each equation, is fulfilled in some of these specifications since they consider single and married working mothers separately.

   (ii) In specification (2), parameter β_3 is interpreted as the difference between the wage rate of married working mothers and single females.

   (iii) There are problems of perfect multicollinearity in Model (2).

   (a) All statements are false.

   (b) (ii) and (iii) are false.

   (c) (i) and (iii) are false.

   (d) (i) and (ii) are false.

15. **(Problem 3.)** In specification (2), the impact of being married on the wage rate is:

   (a) β_2.

   (b) $\beta_2 + \beta_3 m$.

   (c) $\beta_2 + \beta_3 c$.

   (d) $\beta_0 + \beta_2 + \beta_3 m$.

16. **(Problem 3.)** Select the correct answer with respect to specification (1):

   (a) OLS will only provide an unbiased estimator of the constant term.

   (b) All OLS estimators can be biased with a bias equal to $\delta_j \beta_1$ where δ_j are the OLS estimators from the regression of the excluded variable on the included variable, and β_1 is the coefficient of the excluded variable.

   (c) All OLS estimators can be biased with a bias equal to $\delta_j x_1$ where δ_j are the OLS estimators from the regression of the included variables on the omitted variable.

   (d) All OLS estimators can be biased with a bias equal to $\delta_1 x_1$ where δ_1 is the OLS estimator from the regression of the wage rate w on the omitted variable.

17. **(Problem 3.)** We assume that the decision to become a working mother depends on the level of family income, that is why we suggest another equation:

$$m = \gamma_0 + \gamma_1 w + \gamma_2 sm + \gamma_3 c + \varepsilon \tag{6}$$

where sm is the wage rate of the partner with whom she lives. It is assumed that $E(\varepsilon|sm, c) = 0$. Considering specification (2) for w, and that sm and c are exogenous in the aforementioned equation, $E(u|sm, c) = 0$, select the correct response:

(a) Both equations, (2) and (6), satisfy the order condition but fail to satisfy the rank condition.

(b) None of the two equations satisfies the rank nor the order conditions, since in general we expect that $m \times c$ would also be endogenous in equation (2).

(c) Both equations satisfy the rank and order conditions.

(d) Equation (2) satisfies the order condition, but equation (6) does not satisfy the order condition.

18. **(Problem 3.)** Consider the following statements referring to equation (6) for m:

(a) It is equivalent to estimating a Linear Probability Model by OLS, and hence there are no problems with the bias and inconsistency of the estimators.

(b) We can only obtain efficient estimators of the model if instead of OLS we use FGLS (Feasible Generalised Least Squares) in the 2SLS estimation.

(c) First, we need to estimate the reduced form of w by OLS, so that we are able to estimate the structural form by OLS replacing w by the estimated values in the first regression, but this is not possible since we have not got enough instruments.

(d) It is impossible to estimate by OLS given the presence of the evident simultaneity problems. It can only be estimated by maximum likelihood i.e. using a logit or probit model according to the assumptions about the errors distribution.

19. **(Problem 3.)** What would occur if instead of using m as a dependent variable we used $50m$ in (6)?

(i) The only effect on the OLS estimators will be that all are multiplied by 50 as well. The coefficient of determination R^2 will not be affected.

(ii) The sum of the squared residuals will be multiplied by 2.500.

(iii) The OLS estimators will now be biased upwards by 50.

(a) Only (i) is true.

(b) All statements are true.

(c) Only (i) and (ii) are true.

(d) Only (iii) is true.

20. **(Problem 3.)** With respect to the assumption of heteroskedasticity in model (6) for m (and ignoring any **endogeneity** problems), consider the following statements:

   (i) We can estimate the equation by OLS, but for inference purposes we will use the robust t statistics, even though this estimations would not be efficient if heteroskedasticity is present.

   (ii) In this type of models there is no need to test for heteroskedasticity, so that it will no matter whether we use the usual standard errors or the corrected ones.

   (iii) We can estimate the model by Feasible Generalized LS, without the need to run in first place a regression for the squares of the OLS residuals.

   (a) Only (i) and (ii) are true.
   (b) Only (ii) and(iii) are true.
   (c) Only (i) is true.
   (d) Only (i) and (iii) are true.

21. **(Problem 3.)** Assuming that there are no simultaneity problems, but that w was measured with error in equation (6) such that:

$$w = w_{true} + e_1,$$

where e_1 denotes the measurement error and w_{true} is the true level of the wage rate.

   (i) The measurement error will always cause a decrease in the variance of the random disturbance.

   (ii) If $Cov\,(w, e_1) = 0$ holds, the OLS estimators will be unbiased and consistent.

   (iii) Under the CEV (Classical Errors in Variables) assumption, the OLS estimators will be unbiased, but with lower variance than without measurement error due to the attenuation bias.

   (a) Only (i) is true.
   (b) Only (i) and (ii) are true.
   (c) Only (ii) is true.
   (d) Only (ii) and (iii) are true.

22. **(Problem 3.)** We suspect that equation (6) is misspecified (assuming that there is no simultaneity problem). Consider the following statements:

   (i) If assumption $E(\varepsilon|x) = 0$ is satisfied, where x is a vector containing the explanatory variables, then none of the added non-linear functions of the $x's$ should be significant.

   (ii) The RESET test results indicate whether it is necessary to use heteroskedasticity robust standard errors.

   (iii) The RESET test does not inform us on how to proceed if the null hypothesis of correct specification is rejected.

   (a) (i) and (iii) are true.

   (b) All statements are true.

   (c) Only (i) is true.

   (d) (ii) and (iii) are true.

23. **(Problem 3.)** $(Continues)$

   (i) An advantage of the RESET test is that it can compare models with different specifications of the dependent variable.

   (ii) The disadvantage of the Mizon-Richard and Davidson-MacKinnon tests is that they can give inconclusive results regarding which model is preferable.

   (iii) White's test has the capacity to detect more general specification errors compared to the Breusch-Pagan test.

   (a) Only (i) is true.

   (b) All statements are true.

   (c) Only (ii) is true.

   (d) Only (i) and (iii) are true.

24. **(Problem 1.)** Regarding equation (1):

(i) The expected difference in years of education between a child that grew up in a household with a good economic situation and one that grew up in a household with a very poor economic situation is given by $\beta_7 - \beta_0$, if we assume that all other factors are held constant.

(ii) The expected difference in years of education between a child that studied in an urban area inside a household with a good economic situation and another that studied in a rural area inside a household with a very poor economic situation is given by $\beta_5 + \beta_7$, if we assume that all other factors are held constant.

(iii) The expected difference in years of education between a child that grew up in a household with a good economic situation and a working mother and another that grew up in a household with a very poor economic situation, for which we do not know whether the mother was working or not, is given by $\beta_4 + \beta_7$, if we assume that all other factors are held constant.

(a) Only (i) is true.

(b) All statements are false.

(c) Only (ii) and (iii) are true.

(d) Only (ii) is true.

25. **(Problem 1.)** Consider the following statements with respect to Output 1:

(i) We can conclude that the number of siblings has an effect on the total number of years of education of an individual, since $n1childg1$ is significant in Output 1.

(ii) The endogeneity of $n1childg1$ affects the estimated standard errors of the OLS estimators in Output 1, but does not affect the consistency of the estimators.

(iii) All estimators in Output 1 are consistent, except that of $n1childg1$, because it is endogenous.

(a) Only (i) is true.

(b) All statements are false.

(c) Only (ii) and (iii) are true.

(d) Only (ii) is true.

26. **(Problem 1.)** The validity of instruments $boy12$ and $girl12$,

(i) Solely depends on the zero correlation between the instrument and the omitted factor in the equation.

(ii) Can have the implication that the equation is over-identified.

(iii) Guarantees that the OLS estimators are consistent.

(a) Only (i) is true.
(b) (i) and (iii) are true.
(c) (i) and (ii) are true.
(d) Only (ii) is true.

27. **(Problem 1.)** Consider the following statements:

(i) The instruments have to be exogenous variables whose OLS coefficient in the original equation is not significant.

(ii) The instruments can be correlated with the dependent variable.

(iii) The instruments can be correlated with the exogenous variables.

(a) (i) and (iii) are true.
(b) All statements are false.
(c) (ii) and (iii) are true.
(d) Only (ii) is true.

28. **(Problem 1.)** Using instruments $boy12$ and $girl12$, and assuming that they are uncorrelated with u, we can conclude that:

(a) Equation (1) is identified because $boy12$ and $girl12$ are exogenous.
(b) Equation (1) is not identified because $boy12$ is not significant in Output 2.
(c) Equation (1) is not identified because $boy12$ and $girl12$ are not jointly significant in the reduced form of $n1childg1$, since the value of the corresponding test statistic, 55.27, is not significant at the 1% level.
(d) Equation (1) is identified because $boy12$ and $girl12$ are jointly significant in the reduced form of $n1childg1$, since the value of the corresponding test statistic, 13.71, is significant at the 5% level.

29. **(Problem 1.)** Assuming that the instruments $boy12$ and $girl12$ are uncorrelated with u, and using the relevant information from the Outputs,

(i) The 2SLS estimation of Output 4 is more efficient than the 2SLS estimation using only $girl12$ because having more instruments is always better.

(ii) The 2SLS estimation of Output 4, is not consistent because $boy12$ is not significant in Output 2.

(iii) Since we know that $boy12$ is correlated with $n1childg1$, it is always best to use it as an instrument in the 2SLS estimation.

(a) Only (i) is true.
(b) All statements are false.
(c) Only (i) and (iii) are true.
(d) Only (ii) is true.

30. **(Problem 1.)** Consider the following statements:

(i) Hausman's test compares two alternative forms to estimate the parameters of a model on the basis that, at least one of the two forms is not consistent under H_0.

(ii) The estimators of Output 5 are consistent, but less efficient than those of Output 4 if equation(1) is exactly identified.

(iii) The endogeneity test requires that the equation is over-identified.

(a) Only (ii) is true.
(b) All are false.
(c) Only (i) and (iii) are true.
(d) Only (iii) is true.

31. **(Problem 1.)** The results of Output 5:

(i) Estimate the parameters by Two Stages Least Squares.

(ii) Allow us to confirm the endogeneity of $n1childg1$ at the 1% significance level.

(iii) Indicate that $n1childg1$ is uncorrelated with u in equation (1).

(a) Only (i) is true.
(b) (i) and (ii) are true.
(c) All statements are true.
(d) Only (iii) is true.

32. **(Problem 1.)** If we had at our disposal the residuals from a 2SLS estimation of equation (1) using as a sole instrument $girl2$, and the results of the OLS regression where those residuals are the dependent variable as a function of all the exogenous variables:

   (i) It would have been possible to test if $boy12$ is uncorrelated with u in equation (1).

   (ii) It would have been possible to test if $girl12$ is uncorrelated with u in equation (1).

   (iii) It wold have been possible to simultaneously test if $boy12$ and $girl12$ are uncorrelated with u in equation (1).

   (a) Only (i) is true.
   (b) All statements are false.
   (c) Only (iii) is true.
   (d) Only (ii) is true.

33. **(Problem 1.)** Using the appropriate outputs, we can conclude that:

   (i) The number of siblings has a negative effect on the individual's education.

   (ii) The economic situation of the family does not affect education because *pobre* is not significant in Output 1 at the 5% level.

   (iii) The number of siblings has a positive significant effect on *yedu* at the 10% level but not at the 5% level.

   (a) Only (i) is true.
   (b) Only (iii) is true.
   (c) All statements are false.
   (d) (i) and (ii) are true.

34. **(Problem 1.)**

   (i) The fact that the standard deviation of the residuals in Output 1 is lower than in Output 4 confirms that OLS is consistent.

   (ii) The R^2 in the 2SLS output has the same interpretation and use as in the OLS case.

   (iii) The asymptotic variance of the 2SLS estimators is always greater or equal to that of the OLS estimators.

   (a) (i) and (ii) are false.
   (b) Only (ii) is false.
   (c) Only (ii) is true.
   (d) (ii) and (iii) are true.

Solution to Exam Type: 1

Universidad Carlos III de Madrid

ECONOMETRICS I

Academic year 2006/07
FINAL EXAM

February 9, 2008

TIME: 2 HOURS 30 MINUTES

1. **(Problem 2.)** In the Linear Probability Model,

   (i) The error term would present heteroskedasticity (conditional on the explanatory variables).

   (ii) The error term (conditional on the explanatory variables) would be normally distributed.

   (iii) The predicted probabilities could be greater than one, but not lower than zero.

   (a) The three statements are true.
   [(b)] Only (i) is true.
   (c) Only (i) and (iii) are true.
   (d) Only (i) and (ii) are true.

2. **(Problem 2.)** Using the linear probability model estimates,

   [(a)] The probability to participate in sports increases with age up to the age of 50, approximately.
   (b) The probability to participate in sports always diminishes with age.
   (c) The probability to participate in sports diminishes after the age of 39,5, approximately.
   (d) The probability to participate in sports diminishes with age after the age of 35, approximately.

3. **(Problem 2.)** Using the linear probability model estimates, the probability that a 30 year old male with 10 years of education participates in sports is:

   [(a)] 0.52.
   (b) 0.94.
   (c) 0.81.
   (d) 0.36.

4. **(Problem 2.)** Using the linear probability model estimates, we can conclude that

   (i) Females have a lower probability to participate in sports, holding all else equal.

   (ii) Age does not affect the decision to participate in sports because the coefficient on $age2$ is not significant at the 5% level.

   (iii) A male with 10 years of education has a higher probability to do sports than a female with 20 years of education, everything else being held equal.

   (a) Only (i) is true.
   (b) (ii) and (iii) are true.
   [(c)] (i) and (iii) are true.
   (d) (ii) is true.

5. **(Problem 2.)** The estimated linear probability model specification (Output 1),

   (i) Permits measuring the effect of sex on the probability to do sports.

   (ii) Permits that the partial effect of age varies with its value.

   (iii) Permits that the effect of education on the probability to do sports varies by gender.

   (a) Only (i) is true.
   (b) Only (iii) is true.
   (c) Only (i) and (iii) are true.
   [(d)] Only (i) and (ii) are true.

6. **(Problem 2.)** In a Logit model,

   (i) The partial effect of variable x_j on the probability that $y = 1$ (conditional on the explanatory variables), depends on β_j and x_j, but not on the rest of the explanatory variables or parameters.

   (ii) The partial effect of variable x_j on the probability that $y = 1$ (conditional on the explanatory variables), depends on $\beta_0, \beta_1, \ldots, \beta_k$ and on x_j, but not on the rest of the explanatory variables.

   (iii) The partial effect of variable x_j on the probability that $y = 1$ (conditional on the explanatory variables), depends on β_j and on $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_k$, but not on the rest of the parameters.

   [(a)] The three statements are false.
   (b) Only (iii) is true.
   (c) Only (ii) is true.
   (d) Only (i) is true.

7. **(Problem 2.)** Using the Logit model estimates, the effect on the probability to participate in sports, of a 10 year increase in education (ceteris paribus) for an individual with the mean characteristics in the sample, is

   (a) 0.74.

   (b) 0.017.

   (c) 0.074.

   [(d)] 0.17.

8. **(Problem 2.)** Using the Logit model estimates,

   (i) For some females it is possible that the predicted probability to do sports is negative.

   (ii) The difference in the probability to do sports between a male and a female is always 0.29 in favour of the male, independently of his age.

   (iii) The magnitude of the partial effect of years of education on the probability to do sports, diminishes when *yedu* increases if *yedu* is sufficiently high, all else held equal.

   (a) Only (*ii*) and (*iii*) are true.

   [(b)] Only (*iii*) is true.

   (c) Only (*i*) and (*ii*) are true.

   (d) Only (*i*) is true.

9. **(Problem 2.)** Using the Logit model estimates, to test that the effect of education on the probability to do sports does not depend on gender:

   (a) I would have tested if the coefficients of $female$ and *yedu* are jointly significant with a likelihood ratio test.

   [(b)] I would have tested if the coefficient on $female$ is significant using a t test.

   (c) It is necessary to estimate a model that is inclusive of interactions of *yedu* with the variable $female$.

   (d) I would have used the log-likelihoods of this model and another which is not inclusive of variables $female$ and *yedu*.

10. **(Problem 2.)** Using the Logit model estimates,

   (i) We can conclude that the explanatory variables aid in predicting the probability to do sports because the percent correctly predicted is fairly high.

   (ii) We can reject the hypothesis that the coefficients on $female$, age, $age2$ and $yedu$ are equal to zero at any usual significance level.

   (iii) The model is not properly specified since the pseudo-R^2 is very low.

   (a) Only (i) and (ii) are true.
   (b) Only (i) is true.
   (c) Only (i) and (iii) are true.
   [(d)] Only (ii) is true.

11. **(Problem 2.)** Using the Logit model estimates, and assuming that the coefficient on $age2$ is exactly zero,

   [(a)] The partial effect on the probability to do sports of an additional year of education is equivalent to being 0,88 years younger.
   (b) The partial effect on the probability to do sports of an additional year of education is equivalent to being 0,88 years younger, but only if all explanatory variables are evaluated at the sample means.
   (c) The partial effect on the probability to do sports of an additional year of education is equivalent to being 0,88 years younger, but only for females.
   (d) The partial effect on the probability to do sports of an additional year of education is equivalent to being 0,88 years younger, but only for males.

12. **(Problem 2.)** Using the Logit model estimates, the probability that a 30 year old female 30 with 20 years of education participates in sports is

   (a) 0.29 less than that of a male with equal characteristics.
   (b) 0.27 less than that of a male with equal characteristics.
   (c) 0.29 less than that of a male with the mean characteristics in the sample.
   [(d)] 0.31 less than that of a male with equal characteristics.

13. **(Problem 3.)** Out of the four suggested specifications:

   (a) Only (2) and (3) can be estimated by OLS.
   (b) None of the specifications can be estimated by OLS because they omit relevant variables.
   [(c)] The only specification that cannot be estimated by OLS is (4).
   (d) All can be estimated by OLS, but only (2) provides consistent estimators.

14. **(Problem 3.)** Consider the following statements with respect to the previous models:

   (i) It is possible that assumption $E(u|x) = 0$ where x is a vector that contains all the explanatory variables included in each equation, is fulfilled in some of these specifications since they consider single and married working mothers separately.

   (ii) In specification (2), parameter β_3 is interpreted as the difference between the wage rate of married working mothers and single females.

   (iii) There are problems of perfect multicollinearity in Model (2).

   [(a)] All statements are false.
   (b) (ii) and (iii) are false.
   (c) (i) and (iii) are false.
   (d) (i) and (ii) are false.

15. **(Problem 3.)** In specification (2), the impact of being married on the wage rate is:

   (a) β_2.
   [(b)] $\beta_2 + \beta_3 m$.
   (c) $\beta_2 + \beta_3 c$.
   (d) $\beta_0 + \beta_2 + \beta_3 m$.

16. **(Problem 3.)** Select the correct answer with respect to specification (1):

   (a) OLS will only provide an unbiased estimator of the constant term.
   [(b)] All OLS estimators can be biased with a bias equal to $\delta_j \beta_1$ where δ_j are the OLS estimators from the regression of the excluded variable on the included variable, and β_1 is the coefficient of the excluded variable.
   (c) All OLS estimators can be biased with a bias equal to $\delta_j x_1$ where δ_j are the OLS estimators from the regression of the included variables on the omitted variable.
   (d) All OLS estimators can be biased with a bias equal to $\delta_1 x_1$ where δ_1 is the OLS estimator from the regression of the wage rate w on the omitted variable.

17. **(Problem 3.)** We assume that the decision to become a working mother depends on the level of family income, that is why we suggest another equation:

$$m = \gamma_0 + \gamma_1 w + \gamma_2 sm + \gamma_3 c + \varepsilon \qquad (6)$$

where sm is the wage rate of the partner with whom she lives. It is assumed that $E(\varepsilon|sm, c) = 0$. Considering specification (2) for w, and that sm and c are exogenous in the aforementioned equation, $E(u|sm, c) = 0$, select the correct response:

(a) Both equations, (2) and (6), satisfy the order condition but fail to satisfy the rank condition.

[(b)] None of the two equations satisfies the rank nor the order conditions, since in general we expect that $m \times c$ would also be endogenous in equation (2).

(c) Both equations satisfy the rank and order conditions.

(d) Equation (2) satisfies the order condition, but equation (6) does not satisfy the order condition.

18. **(Problem 3.)** Consider the following statements referring to equation (6) for m :

(a) It is equivalent to estimating a Linear Probability Model by OLS, and hence there are no problems with the bias and inconsistency of the estimators.

(b) We can only obtain efficient estimators of the model if instead of OLS we use FGLS (Feasible Generalised Least Squares) in the 2SLS estimation.

[(c)] First, we need to estimate the reduced form of w by OLS, so that we are able to estimate the structural form by OLS replacing w by the estimated values in the first regression, but this is not possible since we have not got enough instruments.

(d) It is impossible to estimate by OLS given the presence of the evident simultaneity problems. It can only be estimated by maximum likelihood i.e. using a logit or probit model according to the assumptions about the errors distribution.

19. **(Problem 3.)** What would occur if instead of using m as a dependent variable we used $50m$ in (6)?

(i) The only effect on the OLS estimators will be that all are multiplied by 50 as well. The coefficient of determination R^2 will not be affected.

(ii) The sum of the squared residuals will be multiplied by 2.500.

(iii) The OLS estimators will now be biased upwards by 50.

(a) Only (i) is true.

(b) All statements are true.

[(c)] Only (i) and (ii) are true.

(d) Only (iii) is true.

20. **(Problem 3.)** With respect to the assumption of heteroskedasticity in model (6) for m (and ignoring any **endogeneity** problems), consider the following statements:

   (i) We can estimate the equation by OLS, but for inference purposes we will use the robust t statistics, even though this estimations would not be efficient if heteroskedasticity is present.

   (ii) In this type of models there is no need to test for heteroskedasticity, so that it will no matter whether we use the usual standard errors or the corrected ones.

   (iii) We can estimate the model by Feasible Generalized LS, without the need to run in first place a regression for the squares of the OLS residuals.

   (a) Only (i) and (ii) are true.
   (b) Only (ii) and(iii) are true.
   (c) Only (i) is true.
   [(d)] Only (i) and (iii) are true.

21. **(Problem 3.)** Assuming that there are no simultaneity problems, but that w was measured with error in equation (6) such that:

$$w = w_{true} + e_1,$$

where e_1 denotes the measurement error and w_{true} is the true level of the wage rate.

   (i) The measurement error will always cause a decrease in the variance of the random disturbance.

   (ii) If $Cov\,(w, e_1) = 0$ holds, the OLS estimators will be unbiased and consistent.

   (iii) Under the CEV (Classical Errors in Variables) assumption, the OLS estimators will be unbiased, but with lower variance than without measurement error due to the attenuation bias.

   (a) Only (i) is true.
   (b) Only (i) and (ii) are true.
   [(c)] Only (ii) is true.
   (d) Only (ii) and (iii) are true.

22. **(Problem 3.)** We suspect that equation (6) is misspecified (assuming that there is no simultaneity problem). Consider the following statements:

   (i) If assumption $E(\varepsilon|x) = 0$ is satisfied, where x is a vector containing the explanatory variables, then none of the added non-linear functions of the $x's$ should be significant.

   (ii) The RESET test results indicate whether it is necessary to use heteroskedasticity robust standard errors.

   (iii) The RESET test does not inform us on how to proceed if the null hypothesis of correct specification is rejected.

   [(a)] (*i*) and (*iii*) are true.
   (b) All statements are true.
   (c) Only (*i*) is true.
   (d) (*ii*) and (*iii*) are true.

23. **(Problem 3.)** (*Continues*)

   (i) An advantage of the RESET test is that it can compare models with different specifications of the dependent variable.

   (ii) The disadvantage of the Mizon-Richard and Davidson-MacKinnon tests is that they can give inconclusive results regarding which model is preferable.

   (iii) White's test has the capacity to detect more general specification errors compared to the Breusch-Pagan test.

   (a) Only (*i*) is true.
   (b) All statements are true.
   [(c)] Only (*ii*) is true.
   (d) Only (*i*) and (*iii*) are true.

24. **(Problem 1.)** Regarding equation (1):

(i) The expected difference in years of education between a child that grew up in a household with a good economic situation and one that grew up in a household with a very poor economic situation is given by $\beta_7 - \beta_0$, if we assume that all other factors are held constant.

(ii) The expected difference in years of education between a child that studied in an urban area inside a household with a good economic situation and another that studied in a rural area inside a household with a very poor economic situation is given by $\beta_5 + \beta_7$, if we assume that all other factors are held constant.

(iii) The expected difference in years of education between a child that grew up in a household with a good economic situation and a working mother and another that grew up in a household with a very poor economic situation, for which we do not know whether the mother was working or not, is given by $\beta_4 + \beta_7$, if we assume that all other factors are held constant.

(a) Only (i) is true.

(b) All statements are false.

(c) Only (ii) and (iii) are true.

[(d)] Only (ii) is true.

25. **(Problem 1.)** Consider the following statements with respect to Output 1:

(i) We can conclude that the number of siblings has an effect on the total number of years of education of an individual, since $n1childg1$ is significant in Output 1.

(ii) The endogeneity of $n1childg1$ affects the estimated standard errors of the OLS estimators in Output 1, but does not affect the consistency of the estimators.

(iii) All estimators in Output 1 are consistent, except that of $n1childg1$, because it is endogenous.

(a) Only (i) is true.

[(b)] All statements are false.

(c) Only (ii) and (iii) are true.

(d) Only (ii) is true.

26. **(Problem 1.)** The validity of instruments $boy12$ and $girl12$,

(i) Solely depends on the zero correlation between the instrument and the omitted factor in the equation.

(ii) Can have the implication that the equation is over-identified.

(iii) Guarantees that the OLS estimators are consistent.

(a) Only (i) is true.
(b) (i) and (iii) are true.
(c) (i) and (ii) are true.
[(d)] Only (ii) is true.

27. **(Problem 1.)** Consider the following statements:

(i) The instruments have to be exogenous variables whose OLS coefficient in the original equation is not significant.

(ii) The instruments can be correlated with the dependent variable.

(iii) The instruments can be correlated with the exogenous variables.

(a) (i) and (iii) are true.
(b) All statements are false.
[(c)] (ii) and (iii) are true.
(d) Only (ii) is true.

28. **(Problem 1.)** Using instruments $boy12$ and $girl12$, and assuming that they are uncorrelated with u, we can conclude that:

(a) Equation (1) is identified because $boy12$ and $girl12$ are exogenous.
(b) Equation (1) is not identified because $boy12$ is not significant in Output 2.
(c) Equation (1) is not identified because $boy12$ and $girl12$ are not jointly significant in the reduced form of $n1childg1$, since the value of the corresponding test statistic, 55.27, is not significant at the 1% level.
[(d)] Equation (1) is identified because $boy12$ and $girl12$ are jointly significant in the reduced form of $n1childg1$, since the value of the corresponding test statistic, 13.71, is significant at the 5% level.

29. **(Problem 1.)** Assuming that the instruments $boy12$ and $girl12$ are uncorrelated with u, and using the relevant information from the Outputs,

(i) The 2SLS estimation of Output 4 is more efficient than the 2SLS estimation using only $girl12$ because having more instruments is always better.

(ii) The 2SLS estimation of Output 4, is not consistent because $boy12$ is not significant in Output 2.

(iii) Since we know that $boy12$ is correlated with $n1childg1$, it is always best to use it as an instrument in the 2SLS estimation.

(a) Only (i) is true.
[(b)] All statements are false.
(c) Only (i) and (iii) are true.
(d) Only (ii) is true.

30. **(Problem 1.)** Consider the following statements:

(i) Hausman's test compares two alternative forms to estimate the parameters of a model on the basis that, at least one of the two forms is not consistent under H_0.

(ii) The estimators of Output 5 are consistent, but less efficient than those of Output 4 if equation(1) is exactly identified.

(iii) The endogeneity test requires that the equation is over-identified.

(a) Only (ii) is true.
[(b)] All are false.
(c) Only (i) and (iii) are true.
(d) Only (iii) is true.

31. **(Problem 1.)** The results of Output 5:

(i) Estimate the parameters by Two Stages Least Squares.

(ii) Allow us to confirm the endogeneity of $n1childg1$ at the 1% significance level.

(iii) Indicate that $n1childg1$ is uncorrelated with u in equation (1).

(a) Only (i) is true.
[(b)] (i) and (ii) are true.
(c) All statements are true.
(d) Only (iii) is true.

32. **(Problem 1.)** If we had at our disposal the residuals from a 2SLS estimation of equation (1) using as a sole instrument $girl2$, and the results of the OLS regression where those residuals are the dependent variable as a function of all the exogenous variables:

(i) It would have been possible to test if $boy12$ is uncorrelated with u in equation (1).

(ii) It would have been possible to test if $girl12$ is uncorrelated with u in equation (1).

(iii) It wold have been possible to simultaneously test if $boy12$ and $girl12$ are uncorrelated with u in equation (1).

[(a)] Only (i) is true.
(b) All statements are false.
(c) Only (iii) is true.
(d) Only (ii) is true.

33. **(Problem 1.)** Using the appropriate outputs, we can conclude that:

(i) The number of siblings has a negative effect on the individual's education.

(ii) The economic situation of the family does not affect education because *pobre* is not significant in Output 1 at the 5% level.

(iii) The number of siblings has a positive significant effect on *yedu* at the 10% level but not at the 5% level.

(a) Only (i) is true.
[(b)] Only (iii) is true.
(c) All statements are false.
(d) (i) and (ii) are true.

34. **(Problem 1.)**

(i) The fact that the standard deviation of the residuals in Output 1 is lower than in Output 4 confirms that OLS is consistent.

(ii) The R^2 in the 2SLS output has the same interpretation and use as in the OLS case.

(iii) The asymptotic variance of the 2SLS estimators is always greater or equal to that of the OLS estimators.

[(a)] (i) and (ii) are false.
(b) Only (ii) is false.
(c) Only (ii) is true.
(d) (ii) and (iii) are true.

Universidad Carlos III de Madrid
<u>ECONOMETRICS I</u>
Academic year 2006/07
FINAL EXAM
February 9, 2008

Tipo de examen: 2

TIME: 2 HOURS 30 MINUTES

1. **(Problem 2.)** In the Linear Probability Model,

   (i) The error term would present heteroskedasticity (conditional on the explanatory variables).

   (ii) The error term (conditional on the explanatory variables) would be normally distributed.

   (iii) The predicted probabilities could be greater than one, but not lower than zero.

   (a) Only (i) and (ii) are true.
   (b) Only (i) and (iii) are true.
   (c) Only (i) is true.
   (d) The three statements are true.

2. **(Problem 2.)** Using the linear probability model estimates,

   (a) The probability to participate in sports diminishes with age after the age of 35, approximately.
   (b) The probability to participate in sports diminishes after the age of 39,5, approximately.
   (c) The probability to participate in sports always diminishes with age.
   (d) The probability to participate in sports increases with age up to the age of 50, approximately.

3. **(Problem 2.)** Using the linear probability model estimates, the probability that a 30 year old male with 10 years of education participates in sports is:

   (a) 0.36.
   (b) 0.81.
   (c) 0.94.
   (d) 0.52.

4. **(Problem 2.)** Using the linear probability model estimates, we can conclude that

   (i) Females have a lower probability to participate in sports, holding all else equal.

   (ii) Age does not affect the decision to participate in sports because the coefficient on $age2$ is not significant at the 5% level.

   (iii) A male with 10 years of education has a higher probability to do sports than a female with 20 years of education, everything else being held equal.

   (a) (ii) is true.
   (b) (i) and (iii) are true.
   (c) (ii) and (iii) are true.
   (d) Only (i) is true.

5. **(Problem 2.)** The estimated linear probability model specification (Output 1),

   (i) Permits measuring the effect of sex on the probability to do sports.

   (ii) Permits that the partial effect of age varies with its value.

   (iii) Permits that the effect of education on the probability to do sports varies by gender.

   (a) Only (i) and (ii) are true.
   (b) Only (i) and (iii) are true.
   (c) Only (iii) is true.
   (d) Only (i) is true.

6. **(Problem 2.)** In a Logit model,

   (i) The partial effect of variable x_j on the probability that $y = 1$ (conditional on the explanatory variables), depends on β_j and x_j, but not on the rest of the explanatory variables or parameters.

   (ii) The partial effect of variable x_j on the probability that $y = 1$ (conditional on the explanatory variables), depends on $\beta_0, \beta_1, \ldots, \beta_k$ and on x_j, but not on the rest of the explanatory variables.

   (iii) The partial effect of variable x_j on the probability that $y = 1$ (conditional on the explanatory variables), depends on β_j and on $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_k$, but not on the rest of the parameters.

   (a) Only (i) is true.
   (b) Only (ii) is true.
   (c) Only (iii) is true.
   (d) The three statements are false.

7. **(Problem 2.)** Using the Logit model estimates, the effect on the probability to participate in sports, of a 10 year increase in education (ceteris paribus) for an individual with the mean characteristics in the sample, is

   (a) 0.17.

   (b) 0.074.

   (c) 0.017.

   (d) 0.74.

8. **(Problem 2.)** Using the Logit model estimates,

   (i) For some females it is possible that the predicted probability to do sports is negative.

   (ii) The difference in the probability to do sports between a male and a female is always 0.29 in favour of the male, independently of his age.

   (iii) The magnitude of the partial effect of years of education on the probability to do sports, diminishes when *yedu* increases if *yedu* is sufficiently high, all else held equal.

   (a) Only (*i*) is true.

   (b) Only (*i*) and (*ii*) are true.

   (c) Only (*iii*) is true.

   (d) Only (*ii*) and (*iii*) are true.

9. **(Problem 2.)** Using the Logit model estimates, to test that the effect of education on the probability to do sports does not depend on gender:

   (a) I would have used the log-likelihoods of this model and another which is not inclusive of variables $female$ and *yedu*.

   (b) It is necessary to estimate a model that is inclusive of interactions of *yedu* with the variable $female$.

   (c) I would have tested if the coefficient on $female$ is significant using a t test.

   (d) I would have tested if the coefficients of $female$ and *yedu* are jointly significant with a likelihood ratio test.

10. **(Problem 2.)** Using the Logit model estimates,

    (i) We can conclude that the explanatory variables aid in predicting the probability to do sports because the percent correctly predicted is fairly high.

    (ii) We can reject the hypothesis that the coefficients on $female$, age, $age2$ and $yedu$ are equal to zero at any usual significance level.

    (iii) The model is not properly specified since the pseudo-R^2 is very low.

    (a) Only (ii) is true.
    (b) Only (i) and (iii) are true.
    (c) Only (i) is true.
    (d) Only (i) and (ii) are true.

11. **(Problem 2.)** Using the Logit model estimates, and assuming that the coefficient on $age2$ is exactly zero,

    (a) The partial effect on the probability to do sports of an additional year of education is equivalent to being 0,88 years younger, but only for males.
    (b) The partial effect on the probability to do sports of an additional year of education is equivalent to being 0,88 years younger, but only for females.
    (c) The partial effect on the probability to do sports of an additional year of education is equivalent to being 0,88 years younger, but only if all explanatory variables are evaluated at the sample means.
    (d) The partial effect on the probability to do sports of an additional year of education is equivalent to being 0,88 years younger.

12. **(Problem 2.)** Using the Logit model estimates, the probability that a 30 year old female 30 with 20 years of education participates in sports is

    (a) 0.31 less than that of a male with equal characteristics.
    (b) 0.29 less than that of a male with the mean characteristics in the sample.
    (c) 0.27 less than that of a male with equal characteristics.
    (d) 0.29 less than that of a male with equal characteristics.

13. **(Problem 1.)** Regarding equation (1):

(i) The expected difference in years of education between a child that grew up in a household with a good economic situation and one that grew up in a household with a very poor economic situation is given by $\beta_7 - \beta_0$, if we assume that all other factors are held constant.

(ii) The expected difference in years of education between a child that studied in an urban area inside a household with a good economic situation and another that studied in a rural area inside a household with a very poor economic situation is given by $\beta_5 + \beta_7$, if we assume that all other factors are held constant.

(iii) The expected difference in years of education between a child that grew up in a household with a good economic situation and a working mother and another that grew up in a household with a very poor economic situation, for which we do not know whether the mother was working or not, is given by $\beta_4 + \beta_7$, if we assume that all other factors are held constant.

(a) Only (ii) is true.

(b) Only (ii) and (iii) are true.

(c) All statements are false.

(d) Only (i) is true.

14. **(Problem 1.)** Consider the following statements with respect to Output 1:

(i) We can conclude that the number of siblings has an effect on the total number of years of education of an individual, since $n1childg1$ is significant in Output 1.

(ii) The endogeneity of $n1childg1$ affects the estimated standard errors of the OLS estimators in Output 1, but does not affect the consistency of the estimators.

(iii) All estimators in Output 1 are consistent, except that of $n1childg1$, because it is endogenous.

(a) Only (ii) is true.

(b) Only (ii) and (iii) are true.

(c) All statements are false.

(d) Only (i) is true.

15. **(Problem 1.)** The validity of instruments $boy12$ and $girl12$,

(i) Solely depends on the zero correlation between the instrument and the omitted factor in the equation.

(ii) Can have the implication that the equation is over-identified.

(iii) Guarantees that the OLS estimators are consistent.

(a) Only (ii) is true.

(b) (i) and (ii) are true.

(c) (i) and (iii) are true.

(d) Only (i) is true.

16. **(Problem 1.)** Consider the following statements:

(i) The instruments have to be exogenous variables whose OLS coefficient in the original equation is not significant.

(ii) The instruments can be correlated with the dependent variable.

(iii) The instruments can be correlated with the exogenous variables.

(a) Only (ii) is true.

(b) (ii) and (iii) are true.

(c) All statements are false.

(d) (i) and (iii) are true.

17. **(Problem 1.)** Using instruments $boy12$ and $girl12$, and assuming that they are uncorrelated with u, we can conclude that:

(a) Equation (1) is identified because $boy12$ and $girl12$ are jointly significant in the reduced form of $n1childg1$, since the value of the corresponding test statistic, 13.71, is significant at the 5% level.

(b) Equation (1) is not identified because $boy12$ and $girl12$ are not jointly significant in the reduced form of $n1childg1$, since the value of the corresponding test statistic, 55.27, is not significant at the 1% level.

(c) Equation (1) is not identified because $boy12$ is not significant in Output 2.

(d) Equation (1) is identified because $boy12$ and $girl12$ are exogenous.

18. **(Problem 1.)** Assuming that the instruments $boy12$ and $girl12$ are uncorrelated with u, and using the relevant information from the Outputs,

(i) The 2SLS estimation of Output 4 is more efficient than the 2SLS estimation using only $girl12$ because having more instruments is always better.

(ii) The 2SLS estimation of Output 4, is not consistent because $boy12$ is not significant in Output 2.

(iii) Since we know that $boy12$ is correlated with $n1childg1$, it is always best to use it as an instrument in the 2SLS estimation.

(a) Only (ii) is true.
(b) Only (i) and (iii) are true.
(c) All statements are false.
(d) Only (i) is true.

19. **(Problem 1.)** Consider the following statements:

(i) Hausman's test compares two alternative forms to estimate the parameters of a model on the basis that, at least one of the two forms is not consistent under H_0.

(ii) The estimators of Output 5 are consistent, but less efficient than those of Output 4 if equation(1) is exactly identified.

(iii) The endogeneity test requires that the equation is over-identified.

(a) Only (iii) is true.
(b) Only (i) and (iii) are true.
(c) All are false.
(d) Only (ii) is true.

20. **(Problem 1.)** The results of Output 5:

(i) Estimate the parameters by Two Stages Least Squares.

(ii) Allow us to confirm the endogeneity of $n1childg1$ at the 1% significance level.

(iii) Indicate that $n1childg1$ is uncorrelated with u in equation (1).

(a) Only (iii) is true.
(b) All statements are true.
(c) (i) and (ii) are true.
(d) Only (i) is true.

21. **(Problem 1.)** If we had at our disposal the residuals from a 2SLS estimation of equation (1) using as a sole instrument $girl2$, and the results of the OLS regression where those residuals are the dependent variable as a function of all the exogenous variables:

   (i) It would have been possible to test if $boy12$ is uncorrelated with u in equation (1).

   (ii) It would have been possible to test if $girl12$ is uncorrelated with u in equation (1).

   (iii) It wold have been possible to simultaneously test if $boy12$ and $girl12$ are uncorrelated with u in equation (1).

   (a) Only (ii) is true.
   (b) Only (iii) is true.
   (c) All statements are false.
   (d) Only (i) is true.

22. **(Problem 1.)** Using the appropriate outputs, we can conclude that:

   (i) The number of siblings has a negative effect on the individual's education.

   (ii) The economic situation of the family does not affect education because *pobre* is not significant in Output 1 at the 5% level.

   (iii) The number of siblings has a positive significant effect on *yedu* at the 10% level but not at the 5% level.

   (a) (i) and (ii) are true.
   (b) All statements are false.
   (c) Only (iii) is true.
   (d) Only (i) is true.

23. **(Problem 1.)**

   (i) The fact that the standard deviation of the residuals in Output 1 is lower than in Output 4 confirms that OLS is consistent.

   (ii) The R^2 in the 2SLS output has the same interpretation and use as in the OLS case.

   (iii) The asymptotic variance of the 2SLS estimators is always greater or equal to that of the OLS estimators.

   (a) (ii) and (iii) are true.
   (b) Only (ii) is true.
   (c) Only (ii) is false.
   (d) (i) and (ii) are false.

24. **(Problem 3.)** Out of the four suggested specifications:

   (a) All can be estimated by OLS, but only (2) provides consistent estimators.

   (b) The only specification that cannot be estimated by OLS is (4).

   (c) None of the specifications can be estimated by OLS because they omit relevant variables.

   (d) Only (2) and (3) can be estimated by OLS.

25. **(Problem 3.)** Consider the following statements with respect to the previous models:

   (i) It is possible that assumption $E(u|x) = 0$ where x is a vector that contains all the explanatory variables included in each equation, is fulfilled in some of these specifications since they consider single and married working mothers separately.

   (ii) In specification (2), parameter β_3 is interpreted as the difference between the wage rate of married working mothers and single females.

   (iii) There are problems of perfect multicollinearity in Model (2).

   (a) (i) and (ii) are false.

   (b) (i) and (iii) are false.

   (c) (ii) and (iii) are false.

   (d) All statements are false.

26. **(Problem 3.)** In specification (2), the impact of being married on the wage rate is:

   (a) $\beta_0 + \beta_2 + \beta_3 m$.

   (b) $\beta_2 + \beta_3 c$.

   (c) $\beta_2 + \beta_3 m$.

   (d) β_2.

27. **(Problem 3.)** Select the correct answer with respect to specification (1):

   (a) All OLS estimators can be biased with a bias equal to $\delta_1 x_1$ where δ_1 is the OLS estimator from the regression of the wage rate w on the omitted variable.

   (b) All OLS estimators can be biased with a bias equal to $\delta_j x_1$ where δ_j are the OLS estimators from the regression of the included variables on the omitted variable.

   (c) All OLS estimators can be biased with a bias equal to $\delta_j \beta_1$ where δ_j are the OLS estimators from the regression of the excluded variable on the included variable, and β_1 is the coefficient of the excluded variable.

   (d) OLS will only provide an unbiased estimator of the constant term.

28. **(Problem 3.)** We assume that the decision to become a working mother depends on the level of family income, that is why we suggest another equation:

$$m = \gamma_0 + \gamma_1 w + \gamma_2 sm + \gamma_3 c + \varepsilon \tag{6}$$

where sm is the wage rate of the partner with whom she lives. It is assumed that $E(\varepsilon|sm, c) = 0$. Considering specification (2) for w, and that sm and c are exogenous in the aforementioned equation, $E(u|sm, c) = 0$, select the correct response:

(a) Equation (2) satisfies the order condition, but equation (6) does not satisfy the order condition.

(b) Both equations satisfy the rank and order conditions.

(c) None of the two equations satisfies the rank nor the order conditions, since in general we expect that $m \times c$ would also be endogenous in equation (2).

(d) Both equations, (2) and (6), satisfy the order condition but fail to satisfy the rank condition.

29. (**Problem 3.**) Consider the following statements referring to equation (6) for m:

(a) It is impossible to estimate by OLS given the presence of the evident simultaneity problems. It can only be estimated by maximum likelihood i.e. using a logit or probit model according to the assumptions about the errors distribution.

(b) First, we need to estimate the reduced form of w by OLS, so that we are able to estimate the structural form by OLS replacing w by the estimated values in the first regression, but this is not possible since we have not got enough instruments.

(c) We can only obtain efficient estimators of the model if instead of OLS we use FGLS (Feasible Generalised Least Squares) in the 2SLS estimation.

(d) It is equivalent to estimating a Linear Probability Model by OLS, and hence there are no problems with the bias and inconsistency of the estimators.

30. **(Problem 3.)** What would occur if instead of using m as a dependent variable we used $50m$ in (6)?

(i) The only effect on the OLS estimators will be that all are multiplied by 50 as well. The coefficient of determination R^2 will not be affected.

(ii) The sum of the squared residuals will be multiplied by 2.500.

(iii) The OLS estimators will now be biased upwards by 50.

(a) Only (iii) is true.

(b) Only (i) and (ii) are true.

(c) All statements are true.

(d) Only (i) is true.

31. **(Problem 3.)** With respect to the assumption of heteroskedasticity in model (6) for m (and ignoring any **endogeneity** problems), consider the following statements:

   (i) We can estimate the equation by OLS, but for inference purposes we will use the robust t statistics, even though this estimations would not be efficient if heteroskedasticity is present.

   (ii) In this type of models there is no need to test for heteroskedasticity, so that it will no matter whether we use the usual standard errors or the corrected ones.

   (iii) We can estimate the model by Feasible Generalized LS, without the need to run in first place a regression for the squares of the OLS residuals.

   (a) Only (i) and (iii) are true.
   (b) Only (i) is true.
   (c) Only (ii) and(iii) are true.
   (d) Only (i) and (ii) are true.

32. **(Problem 3.)** Assuming that there are no simultaneity problems, but that w was measured with error in equation (6) such that:

$$w = w_{true} + e_1,$$

   where e_1 denotes the measurement error and w_{true} is the true level of the wage rate.

   (i) The measurement error will always cause a decrease in the variance of the random disturbance.

   (ii) If $Cov\,(w, e_1) = 0$ holds, the OLS estimators will be unbiased and consistent.

   (iii) Under the CEV (Classical Errors in Variables) assumption, the OLS estimators will be unbiased, but with lower variance than without measurement error due to the attenuation bias.

   (a) Only (ii) and (iii) are true.
   (b) Only (ii) is true.
   (c) Only (i) and (ii) are true.
   (d) Only (i) is true.

33. **(Problem 3.)** We suspect that equation (6) is misspecified (assuming that there is no simultaneity problem). Consider the following statements:

(i) If assumption $E(\varepsilon|x) = 0$ is satisfied, where x is a vector containing the explanatory variables, then none of the added non-linear functions of the $x's$ should be significant.

(ii) The RESET test results indicate whether it is necessary to use heteroskedasticity robust standard errors.

(iii) The RESET test does not inform us on how to proceed if the null hypothesis of correct specification is rejected.

(a) (*ii*) and (*iii*) are true.

(b) Only (*i*) is true.

(c) All statements are true.

(d) (*i*) and (*iii*) are true.

34. **(Problem 3.)** (*Continues*)

(i) An advantage of the RESET test is that it can compare models with different specifications of the dependent variable.

(ii) The disadvantage of the Mizon-Richard and Davidson-MacKinnon tests is that they can give inconclusive results regarding which model is preferable.

(iii) White's test has the capacity to detect more general specification errors compared to the Breusch-Pagan test.

(a) Only (*i*) and (*iii*) are true.

(b) Only (*ii*) is true.

(c) All statements are true.

(d) Only (*i*) is true.

Solution to Exam Type: 2

Universidad Carlos III de Madrid

ECONOMETRICS I

Academic year 2006/07
FINAL EXAM

February 9, 2008

TIME: 2 HOURS 30 MINUTES

1. **(Problem 2.)** In the Linear Probability Model,

   (i) The error term would present heteroskedasticity (conditional on the explanatory variables).

   (ii) The error term (conditional on the explanatory variables) would be normally distributed.

   (iii) The predicted probabilities could be greater than one, but not lower than zero.

   (a) Only (i) and (ii) are true.
   (b) Only (i) and (iii) are true.
   [(c)] Only (i) is true.
   (d) The three statements are true.

2. **(Problem 2.)** Using the linear probability model estimates,

   (a) The probability to participate in sports diminishes with age after the age of 35, approximately.
   (b) The probability to participate in sports diminishes after the age of 39,5, approximately.
   (c) The probability to participate in sports always diminishes with age.
   [(d)] The probability to participate in sports increases with age up to the age of 50, approximately.

3. **(Problem 2.)** Using the linear probability model estimates, the probability that a 30 year old male with 10 years of education participates in sports is:

   (a) 0.36.
   (b) 0.81.
   (c) 0.94.
   [(d)] 0.52.

4. **(Problem 2.)** Using the linear probability model estimates, we can conclude that

   (i) Females have a lower probability to participate in sports, holding all else equal.

   (ii) Age does not affect the decision to participate in sports because the coefficient on *age*2 is not significant at the 5% level.

   (iii) A male with 10 years of education has a higher probability to do sports than a female with 20 years of education, everything else being held equal.

   (a) (ii) is true.
   [(b)] (i) and (iii) are true.
   (c) (ii) and (iii) are true.
   (d) Only (i) is true.

5. **(Problem 2.)** The estimated linear probability model specification (Output 1),

   (i) Permits measuring the effect of sex on the probability to do sports.

   (ii) Permits that the partial effect of age varies with its value.

   (iii) Permits that the effect of education on the probability to do sports varies by gender.

   [(a)] Only (i) and (ii) are true.
   (b) Only (i) and (iii) are true.
   (c) Only (iii) is true.
   (d) Only (i) is true.

6. **(Problem 2.)** In a Logit model,

   (i) The partial effect of variable x_j on the probability that $y = 1$ (conditional on the explanatory variables), depends on β_j and x_j, but not on the rest of the explanatory variables or parameters.

   (ii) The partial effect of variable x_j on the probability that $y = 1$ (conditional on the explanatory variables), depends on $\beta_0, \beta_1, \ldots, \beta_k$ and on x_j, but not on the rest of the explanatory variables.

   (iii) The partial effect of variable x_j on the probability that $y = 1$ (conditional on the explanatory variables), depends on β_j and on $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_k$, but not on the rest of the parameters.

   (a) Only (i) is true.
   (b) Only (ii) is true.
   (c) Only (iii) is true.
   [(d)] The three statements are false.

7. **(Problem 2.)** Using the Logit model estimates, the effect on the probability to participate in sports, of a 10 year increase in education (ceteris paribus) for an individual with the mean characteristics in the sample, is

   [(a)] 0.17.
   (b) 0.074.
   (c) 0.017.
   (d) 0.74.

8. **(Problem 2.)** Using the Logit model estimates,

   (i) For some females it is possible that the predicted probability to do sports is negative.

   (ii) The difference in the probability to do sports between a male and a female is always 0.29 in favour of the male, independently of his age.

   (iii) The magnitude of the partial effect of years of education on the probability to do sports, diminishes when *yedu* increases if *yedu* is sufficiently high, all else held equal.

   (a) Only (*i*) is true.
   (b) Only (*i*) and (*ii*) are true.
   [(c)] Only (*iii*) is true.
   (d) Only (*ii*) and (*iii*) are true.

9. **(Problem 2.)** Using the Logit model estimates, to test that the effect of education on the probability to do sports does not depend on gender:

   (a) I would have used the log-likelihoods of this model and another which is not inclusive of variables $female$ and *yedu*.
   (b) It is necessary to estimate a model that is inclusive of interactions of *yedu* with the variable $female$.
   [(c)] I would have tested if the coefficient on $female$ is significant using a t test.
   (d) I would have tested if the coefficients of $female$ and *yedu* are jointly significant with a likelihood ratio test.

10. **(Problem 2.)** Using the Logit model estimates,

   (i) We can conclude that the explanatory variables aid in predicting the probability to do sports because the percent correctly predicted is fairly high.

   (ii) We can reject the hypothesis that the coefficients on $female$, age, $age2$ and $yedu$ are equal to zero at any usual significance level.

   (iii) The model is not properly specified since the pseudo-R^2 is very low.

   [(a)] Only (ii) is true.
   (b) Only (i) and (iii) are true.
   (c) Only (i) is true.
   (d) Only (i) and (ii) are true.

11. **(Problem 2.)** Using the Logit model estimates, and assuming that the coefficient on $age2$ is exactly zero,

   (a) The partial effect on the probability to do sports of an additional year of education is equivalent to being 0,88 years younger, but only for males.
   (b) The partial effect on the probability to do sports of an additional year of education is equivalent to being 0,88 years younger, but only for females.
   (c) The partial effect on the probability to do sports of an additional year of education is equivalent to being 0,88 years younger, but only if all explanatory variables are evaluated at the sample means.
   [(d)] The partial effect on the probability to do sports of an additional year of education is equivalent to being 0,88 years younger.

12. **(Problem 2.)** Using the Logit model estimates, the probability that a 30 year old female 30 with 20 years of education participates in sports is

   [(a)] 0.31 less than that of a male with equal characteristics.
   (b) 0.29 less than that of a male with the mean characteristics in the sample.
   (c) 0.27 less than that of a male with equal characteristics.
   (d) 0.29 less than that of a male with equal characteristics.

13. **(Problem 1.)** Regarding equation (1):

(i) The expected difference in years of education between a child that grew up in a household with a good economic situation and one that grew up in a household with a very poor economic situation is given by $\beta_7 - \beta_0$, if we assume that all other factors are held constant.

(ii) The expected difference in years of education between a child that studied in an urban area inside a household with a good economic situation and another that studied in a rural area inside a household with a very poor economic situation is given by $\beta_5 + \beta_7$, if we assume that all other factors are held constant.

(iii) The expected difference in years of education between a child that grew up in a household with a good economic situation and a working mother and another that grew up in a household with a very poor economic situation, for which we do not know whether the mother was working or not, is given by $\beta_4 + \beta_7$, if we assume that all other factors are held constant.

[(a)] Only (ii) is true.
(b) Only (ii) and (iii) are true.
(c) All statements are false.
(d) Only (i) is true.

14. **(Problem 1.)** Consider the following statements with respect to Output 1:

(i) We can conclude that the number of siblings has an effect on the total number of years of education of an individual, since $n1childg1$ is significant in Output 1.

(ii) The endogeneity of $n1childg1$ affects the estimated standard errors of the OLS estimators in Output 1, but does not affect the consistency of the estimators.

(iii) All estimators in Output 1 are consistent, except that of $n1childg1$, because it is endogenous.

(a) Only (ii) is true.
(b) Only (ii) and (iii) are true.
[(c)] All statements are false.
(d) Only (i) is true.

15. **(Problem 1.)** The validity of instruments $boy12$ and $girl12$,

   (i) Solely depends on the zero correlation between the instrument and the omitted factor in the equation.

   (ii) Can have the implication that the equation is over-identified.

   (iii) Guarantees that the OLS estimators are consistent.

   [(a)] Only (ii) is true.
   (b) (i) and (ii) are true.
   (c) (i) and (iii) are true.
   (d) Only (i) is true.

16. **(Problem 1.)** Consider the following statements:

   (i) The instruments have to be exogenous variables whose OLS coefficient in the original equation is not significant.

   (ii) The instruments can be correlated with the dependent variable.

   (iii) The instruments can be correlated with the exogenous variables.

   (a) Only (ii) is true.
   [(b)] (ii) and (iii) are true.
   (c) All statements are false.
   (d) (i) and (iii) are true.

17. **(Problem 1.)** Using instruments $boy12$ and $girl12$, and assuming that they are uncorrelated with u, we can conclude that:

   [(a)] Equation (1) is identified because $boy12$ and $girl12$ are jointly significant in the reduced form of $n1childg1$, since the value of the corresponding test statistic, 13.71, is significant at the 5% level.
   (b) Equation (1) is not identified because $boy12$ and $girl12$ are not jointly significant in the reduced form of $n1childg1$, since the value of the corresponding test statistic, 55.27, is not significant at the 1% level.
   (c) Equation (1) is not identified because $boy12$ is not significant in Output 2.
   (d) Equation (1) is identified because $boy12$ and $girl12$ are exogenous.

18. **(Problem 1.)** Assuming that the instruments $boy12$ and $girl12$ are uncorrelated with u, and using the relevant information from the Outputs,

   (i) The 2SLS estimation of Output 4 is more efficient than the 2SLS estimation using only $girl12$ because having more instruments is always better.

   (ii) The 2SLS estimation of Output 4, is not consistent because $boy12$ is not significant in Output 2.

   (iii) Since we know that $boy12$ is correlated with $n1childg1$, it is always best to use it as an instrument in the 2SLS estimation.

   (a) Only (ii) is true.
   (b) Only (i) and (iii) are true.
   [(c)] All statements are false.
   (d) Only (i) is true.

19. **(Problem 1.)** Consider the following statements:

   (i) Hausman's test compares two alternative forms to estimate the parameters of a model on the basis that, at least one of the two forms is not consistent under H_0.

   (ii) The estimators of Output 5 are consistent, but less efficient than those of Output 4 if equation(1) is exactly identified.

   (iii) The endogeneity test requires that the equation is over-identified.

   (a) Only (iii) is true.
   (b) Only (i) and (iii) are true.
   [(c)] All are false.
   (d) Only (ii) is true.

20. **(Problem 1.)** The results of Output 5:

   (i) Estimate the parameters by Two Stages Least Squares.

   (ii) Allow us to confirm the endogeneity of $n1childg1$ at the 1% significance level.

   (iii) Indicate that $n1childg1$ is uncorrelated with u in equation (1).

   (a) Only (iii) is true.
   (b) All statements are true.
   [(c)] (i) and (ii) are true.
   (d) Only (i) is true.

21. **(Problem 1.)** If we had at our disposal the residuals from a 2SLS estimation of equation (1) using as a sole instrument $girl2$, and the results of the OLS regression where those residuals are the dependent variable as a function of all the exogenous variables:

(i) It would have been possible to test if $boy12$ is uncorrelated with u in equation (1).

(ii) It would have been possible to test if $girl12$ is uncorrelated with u in equation (1).

(iii) It wold have been possible to simultaneously test if $boy12$ and $girl12$ are uncorrelated with u in equation (1).

(a) Only (ii) is true.
(b) Only (iii) is true.
(c) All statements are false.
[(d)] Only (i) is true.

22. **(Problem 1.)** Using the appropriate outputs, we can conclude that:

(i) The number of siblings has a negative effect on the individual's education.

(ii) The economic situation of the family does not affect education because *pobre* is not significant in Output 1 at the 5% level.

(iii) The number of siblings has a positive significant effect on $yedu$ at the 10% level but not at the 5% level.

(a) (i) and (ii) are true.
(b) All statements are false.
[(c)] Only (iii) is true.
(d) Only (i) is true.

23. **(Problem 1.)**

(i) The fact that the standard deviation of the residuals in Output 1 is lower than in Output 4 confirms that OLS is consistent.

(ii) The R^2 in the 2SLS output has the same interpretation and use as in the OLS case.

(iii) The asymptotic variance of the 2SLS estimators is always greater or equal to that of the OLS estimators.

(a) (ii) and (iii) are true.
(b) Only (ii) is true.
(c) Only (ii) is false.
[(d)] (i) and (ii) are false.

24. **(Problem 3.)** Out of the four suggested specifications:

(a) All can be estimated by OLS, but only (2) provides consistent estimators.

[(b)] The only specification that cannot be estimated by OLS is (4).

(c) None of the specifications can be estimated by OLS because they omit relevant variables.

(d) Only (2) and (3) can be estimated by OLS.

25. **(Problem 3.)** Consider the following statements with respect to the previous models:

(i) It is possible that assumption $E(u|x) = 0$ where x is a vector that contains all the explanatory variables included in each equation, is fulfilled in some of these specifications since they consider single and married working mothers separately.

(ii) In specification (2), parameter β_3 is interpreted as the difference between the wage rate of married working mothers and single females.

(iii) There are problems of perfect multicollinearity in Model (2).

(a) (i) and (ii) are false.

(b) (i) and (iii) are false.

(c) (ii) and (iii) are false.

[(d)] All statements are false.

26. **(Problem 3.)** In specification (2), the impact of being married on the wage rate is:

(a) $\beta_0 + \beta_2 + \beta_3 m$.

(b) $\beta_2 + \beta_3 c$.

[(c)] $\beta_2 + \beta_3 m$.

(d) β_2.

27. **(Problem 3.)** Select the correct answer with respect to specification (1):

(a) All OLS estimators can be biased with a bias equal to $\delta_1 x_1$ where δ_1 is the OLS estimator from the regression of the wage rate w on the omitted variable.

(b) All OLS estimators can be biased with a bias equal to $\delta_j x_1$ where δ_j are the OLS estimators from the regression of the included variables on the omitted variable.

[(c)] All OLS estimators can be biased with a bias equal to $\delta_j \beta_1$ where δ_j are the OLS estimators from the regression of the excluded variable on the included variable, and β_1 is the coefficient of the excluded variable.

(d) OLS will only provide an unbiased estimator of the constant term.

28. **(Problem 3.)** We assume that the decision to become a working mother depends on the level of family income, that is why we suggest another equation:

$$m = \gamma_0 + \gamma_1 w + \gamma_2 sm + \gamma_3 c + \varepsilon \qquad (6)$$

where sm is the wage rate of the partner with whom she lives. It is assumed that $E(\varepsilon|sm, c) = 0$. Considering specification (2) for w, and that sm and c are exogenous in the aforementioned equation, $E(u|sm, c) = 0$, select the correct response:

(a) Equation (2) satisfies the order condition, but equation (6) does not satisfy the order condition.

(b) Both equations satisfy the rank and order conditions.

[(c)] None of the two equations satisfies the rank nor the order conditions, since in general we expect that $m \times c$ would also be endogenous in equation (2).

(d) Both equations, (2) and (6), satisfy the order condition but fail to satisfy the rank condition.

29. **(Problem 3.)** Consider the following statements referring to equation (6) for m :

(a) It is impossible to estimate by OLS given the presence of the evident simultaneity problems. It can only be estimated by maximum likelihood i.e. using a logit or probit model according to the assumptions about the errors distribution.

[(b)] First, we need to estimate the reduced form of w by OLS, so that we are able to estimate the structural form by OLS replacing w by the estimated values in the first regression, but this is not possible since we have not got enough instruments.

(c) We can only obtain efficient estimators of the model if instead of OLS we use FGLS (Feasible Generalised Least Squares) in the 2SLS estimation.

(d) It is equivalent to estimating a Linear Probability Model by OLS, and hence there are no problems with the bias and inconsistency of the estimators.

30. **(Problem 3.)** What would occur if instead of using m as a dependent variable we used $50m$ in (6)?

(i) The only effect on the OLS estimators will be that all are multiplied by 50 as well. The coefficient of determination R^2 will not be affected.

(ii) The sum of the squared residuals will be multiplied by 2.500.

(iii) The OLS estimators will now be biased upwards by 50.

(a) Only (iii) is true.

[(b)] Only (i) and (ii) are true.

(c) All statements are true.

(d) Only (i) is true.

31. **(Problem 3.)** With respect to the assumption of heteroskedasticity in model (6) for m (and ignoring any **endogeneity** problems), consider the following statements:

   (i) We can estimate the equation by OLS, but for inference purposes we will use the robust t statistics, even though this estimations would not be efficient if heteroskedasticity is present.

   (ii) In this type of models there is no need to test for heteroskedasticity, so that it will no matter whether we use the usual standard errors or the corrected ones.

   (iii) We can estimate the model by Feasible Generalized LS, without the need to run in first place a regression for the squares of the OLS residuals.

   [(a)] Only (i) and (iii) are true.
   (b) Only (i) is true.
   (c) Only (ii) and(iii) are true.
   (d) Only (i) and (ii) are true.

32. **(Problem 3.)** Assuming that there are no simultaneity problems, but that w was measured with error in equation (6) such that:

$$w = w_{true} + e_1,$$

   where e_1 denotes the measurement error and w_{true} is the true level of the wage rate.

   (i) The measurement error will always cause a decrease in the variance of the random disturbance.

   (ii) If $Cov\,(w, e_1) = 0$ holds, the OLS estimators will be unbiased and consistent.

   (iii) Under the CEV (Classical Errors in Variables) assumption, the OLS estimators will be unbiased, but with lower variance than without measurement error due to the attenuation bias.

   (a) Only (ii) and (iii) are true.
   [(b)] Only (ii) is true.
   (c) Only (i) and (ii) are true.
   (d) Only (i) is true.

33. **(Problem 3.)** We suspect that equation (6) is misspecified (assuming that there is no simultaneity problem). Consider the following statements:

(i) If assumption $E(\varepsilon|x) = 0$ is satisfied, where x is a vector containing the explanatory variables, then none of the added non-linear functions of the $x's$ should be significant.

(ii) The RESET test results indicate whether it is necessary to use heteroskedasticity robust standard errors.

(iii) The RESET test does not inform us on how to proceed if the null hypothesis of correct specification is rejected.

(a) (*ii*) and (*iii*) are true.

(b) Only (*i*) is true.

(c) All statements are true.

[(d)] (*i*) and (*iii*) are true.

34. **(Problem 3.)** (*Continues*)

(i) An advantage of the RESET test is that it can compare models with different specifications of the dependent variable.

(ii) The disadvantage of the Mizon-Richard and Davidson-MacKinnon tests is that they can give inconclusive results regarding which model is preferable.

(iii) White's test has the capacity to detect more general specification errors compared to the Breusch-Pagan test.

(a) Only (*i*) and (*iii*) are true.

[(b)] Only (*ii*) is true.

(c) All statements are true.

(d) Only (*i*) is true.

Universidad Carlos III de Madrid
ECONOMETRICS I
Academic year 2006/07
FINAL EXAM
February 9, 2008

Tipo de examen: 3

TIME: 2 HOURS 30 MINUTES

1. **(Problem 3.)** Out of the four suggested specifications:
   (a) None of the specifications can be estimated by OLS because they omit relevant variables.
   (b) All can be estimated by OLS, but only (2) provides consistent estimators.
   (c) Only (2) and (3) can be estimated by OLS.
   (d) The only specification that cannot be estimated by OLS is (4).

2. **(Problem 3.)** Consider the following statements with respect to the previous models:

   (i) It is possible that assumption $E(u|x) = 0$ where x is a vector that contains all the explanatory variables included in each equation, is fulfilled in some of these specifications since they consider single and married working mothers separately.

   (ii) In specification (2), parameter β_3 is interpreted as the difference between the wage rate of married working mothers and single females.

   (iii) There are problems of perfect multicollinearity in Model (2).

   (a) (ii) and (iii) are false.
   (b) (i) and (ii) are false.
   (c) All statements are false.
   (d) (i) and (iii) are false.

3. **(Problem 3.)** In specification (2), the impact of being married on the wage rate is:
   (a) $\beta_2 + \beta_3 m$.
   (b) $\beta_0 + \beta_2 + \beta_3 m$.
   (c) β_2.
   (d) $\beta_2 + \beta_3 c$.

4. **(Problem 3.)** Select the correct answer with respect to specification (1):

   (a) All OLS estimators can be biased with a bias equal to $\delta_j \beta_1$ where δ_j are the OLS estimators from the regression of the excluded variable on the included variable, and β_1 is the coefficient of the excluded variable.

   (b) All OLS estimators can be biased with a bias equal to $\delta_1 x_1$ where δ_1 is the OLS estimator from the regression of the wage rate w on the omitted variable.

   (c) OLS will only provide an unbiased estimator of the constant term.

   (d) All OLS estimators can be biased with a bias equal to $\delta_j x_1$ where δ_j are the OLS estimators from the regression of the included variables on the omitted variable.

5. **(Problem 3.)** We assume that the decision to become a working mother depends on the level of family income, that is why we suggest another equation:

$$m = \gamma_0 + \gamma_1 w + \gamma_2 sm + \gamma_3 c + \varepsilon \tag{6}$$

   where sm is the wage rate of the partner with whom she lives. It is assumed that $E(\varepsilon | sm, c) = 0$. Considering specification (2) for w, and that sm and c are exogenous in the aforementioned equation, $E(u|sm, c) = 0$, select the correct response:

   (a) None of the two equations satisfies the rank nor the order conditions, since in general we expect that $m \times c$ would also be endogenous in equation (2).

   (b) Equation (2) satisfies the order condition, but equation (6) does not satisfy the order condition.

   (c) Both equations, (2) and (6), satisfy the order condition but fail to satisfy the rank condition.

   (d) Both equations satisfy the rank and order conditions.

6. **(Problem 3.)** Consider the following statements referring to equation (6) for m:

   (a) We can only obtain efficient estimators of the model if instead of OLS we use FGLS (Feasible Generalised Least Squares) in the 2SLS estimation.

   (b) It is impossible to estimate by OLS given the presence of the evident simultaneity problems. It can only be estimated by maximum likelihood i.e. using a logit or probit model according to the assumptions about the errors distribution.

   (c) It is equivalent to estimating a Linear Probability Model by OLS, and hence there are no problems with the bias and inconsistency of the estimators.

   (d) First, we need to estimate the reduced form of w by OLS, so that we are able to estimate the structural form by OLS replacing w by the estimated values in the first regression, but this is not possible since we have not got enough instruments.

7. **(Problem 3.)** What would occur if instead of using m as a dependent variable we used $50m$ in (6)?

   (i) The only effect on the OLS estimators will be that all are multiplied by 50 as well. The coefficient of determination R^2 will not be affected.

   (ii) The sum of the squared residuals will be multiplied by 2.500.

   (iii) The OLS estimators will now be biased upwards by 50.

   (a) All statements are true.
   (b) Only (iii) is true.
   (c) Only (i) is true.
   (d) Only (i) and (ii) are true.

8. **(Problem 3.)** With respect to the assumption of heteroskedasticity in model (6) for m (and ignoring any **endogeneity** problems), consider the following statements:

   (i) We can estimate the equation by OLS, but for inference purposes we will use the robust t statistics, even though this estimations would not be efficient if heteroskedasticity is present.

   (ii) In this type of models there is no need to test for heteroskedasticity, so that it will no matter whether we use the usual standard errors or the corrected ones.

   (iii) We can estimate the model by Feasible Generalized LS, without the need to run in first place a regression for the squares of the OLS residuals.

   (a) Only (ii) and(iii) are true.
   (b) Only (i) and (iii) are true.
   (c) Only (i) and (ii) are true.
   (d) Only (i) is true.

9. **(Problem 3.)** Assuming that there are no simultaneity problems, but that w was measured with error in equation (6) such that:

$$w = w_{true} + e_1,$$

where e_1 denotes the measurement error and w_{true} is the true level of the wage rate.

   (i) The measurement error will always cause a decrease in the variance of the random disturbance.

   (ii) If $Cov\,(w, e_1) = 0$ holds, the OLS estimators will be unbiased and consistent.

   (iii) Under the CEV (Classical Errors in Variables) assumption, the OLS estimators will be unbiased, but with lower variance than without measurement error due to the attenuation bias.

   (a) Only (i) and (ii) are true.
   (b) Only (ii) and (iii) are true.
   (c) Only (i) is true.
   (d) Only (ii) is true.

10. **(Problem 3.)** We suspect that equation (6) is misspecified (assuming that there is no simultaneity problem). Consider the following statements:

   (i) If assumption $E\,(\varepsilon|x) = 0$ is satisfied, where x is a vector containing the explanatory variables, then none of the added non-linear functions of the $x's$ should be significant.

   (ii) The RESET test results indicate whether it is necessary to use heteroskedasticity robust standard errors.

   (iii) The RESET test does not inform us on how to proceed if the null hypothesis of correct specification is rejected.

   (a) All statements are true.
   (b) (ii) and (iii) are true.
   (c) (i) and (iii) are true.
   (d) Only (i) is true.

11. **(Problem 3.)** (*Continues*)

   (i) An advantage of the RESET test is that it can compare models with different specifications of the dependent variable.

   (ii) The disadvantage of the Mizon-Richard and Davidson-MacKinnon tests is that they can give inconclusive results regarding which model is preferable.

   (iii) White's test has the capacity to detect more general specification errors compared to the Breusch-Pagan test.

   (a) All statements are true.
   (b) Only (*i*) and (*iii*) are true.
   (c) Only (*i*) is true.
   (d) Only (*ii*) is true.

12. **(Problem 2.)** In the Linear Probability Model,

   (i) The error term would present heteroskedasticity (conditional on the explanatory variables).

   (ii) The error term (conditional on the explanatory variables) would be normally distributed.

   (iii) The predicted probabilities could be greater than one, but not lower than zero.

   (a) Only (i) is true.
   (b) Only (i) and (ii) are true.
   (c) The three statements are true.
   (d) Only (i) and (iii) are true.

13. **(Problem 2.)** Using the linear probability model estimates,

   (a) The probability to participate in sports always diminishes with age.
   (b) The probability to participate in sports diminishes with age after the age of 35, approximately.
   (c) The probability to participate in sports increases with age up to the age of 50, approximately.
   (d) The probability to participate in sports diminishes after the age of 39,5, approximately.

14. **(Problem 2.)** Using the linear probability model estimates, the probability that a 30 year old male with 10 years of education participates in sports is:

   (a) 0.94.
   (b) 0.36.
   (c) 0.52.
   (d) 0.81.

15. **(Problem 2.)** Using the linear probability model estimates, we can conclude that

(i) Females have a lower probability to participate in sports, holding all else equal.

(ii) Age does not affect the decision to participate in sports because the coefficient on *age*2 is not significant at the 5% level.

(iii) A male with 10 years of education has a higher probability to do sports than a female with 20 years of education, everything else being held equal.

(a) (ii) and (iii) are true.
(b) (ii) is true.
(c) Only (i) is true.
(d) (i) and (iii) are true.

16. **(Problem 2.)** The estimated linear probability model specification (Output 1),

(i) Permits measuring the effect of sex on the probability to do sports.

(ii) Permits that the partial effect of age varies with its value.

(iii) Permits that the effect of education on the probability to do sports varies by gender.

(a) Only (iii) is true.
(b) Only (i) and (ii) are true.
(c) Only (i) is true.
(d) Only (i) and (iii) are true.

17. **(Problem 2.)** In a Logit model,

(i) The partial effect of variable x_j on the probability that $y = 1$ (conditional on the explanatory variables), depends on β_j and x_j, but not on the rest of the explanatory variables or parameters.

(ii) The partial effect of variable x_j on the probability that $y = 1$ (conditional on the explanatory variables), depends on $\beta_0, \beta_1, \ldots, \beta_k$ and on x_j, but not on the rest of the explanatory variables.

(iii) The partial effect of variable x_j on the probability that $y = 1$ (conditional on the explanatory variables), depends on β_j and on $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_k$, but not on the rest of the parameters.

(a) Only (iii) is true.
(b) Only (i) is true.
(c) The three statements are false.
(d) Only (ii) is true.

18. **(Problem 2.)** Using the Logit model estimates, the effect on the probability to participate in sports, of a 10 year increase in education (ceteris paribus) for an individual with the mean characteristics in the sample, is

(a) 0.017.

(b) 0.17.

(c) 0.74.

(d) 0.074.

19. **(Problem 2.)** Using the Logit model estimates,

(i) For some females it is possible that the predicted probability to do sports is negative.

(ii) The difference in the probability to do sports between a male and a female is always 0.29 in favour of the male, independently of his age.

(iii) The magnitude of the partial effect of years of education on the probability to do sports, diminishes when *yedu* increases if *yedu* is sufficiently high, all else held equal.

(a) Only *(iii)* is true.

(b) Only *(i)* is true.

(c) Only *(ii)* and *(iii)* are true.

(d) Only *(i)* and *(ii)* are true.

20. **(Problem 2.)** Using the Logit model estimates, to test that the effect of education on the probability to do sports does not depend on gender:

(a) I would have tested if the coefficient on $female$ is significant using a t test.

(b) I would have used the log-likelihoods of this model and another which is not inclusive of variables $female$ and *yedu*.

(c) I would have tested if the coefficients of $female$ and *yedu* are jointly significant with a likelihood ratio test.

(d) It is necessary to estimate a model that is inclusive of interactions of *yedu* with the variable $female$.

21. **(Problem 2.)** Using the Logit model estimates,

   (i) We can conclude that the explanatory variables aid in predicting the probability to do sports because the percent correctly predicted is fairly high.

   (ii) We can reject the hypothesis that the coefficients on $female$, age, $age2$ and $yedu$ are equal to zero at any usual significance level.

   (iii) The model is not properly specified since the pseudo-R^2 is very low.

   (a) Only (i) is true.
   (b) Only (ii) is true.
   (c) Only (i) and (ii) are true.
   (d) Only (i) and (iii) are true.

22. **(Problem 2.)** Using the Logit model estimates, and assuming that the coefficient on $age2$ is exactly zero,

   (a) The partial effect on the probability to do sports of an additional year of education is equivalent to being 0,88 years younger, but only if all explanatory variables are evaluated at the sample means.
   (b) The partial effect on the probability to do sports of an additional year of education is equivalent to being 0,88 years younger, but only for males.
   (c) The partial effect on the probability to do sports of an additional year of education is equivalent to being 0,88 years younger.
   (d) The partial effect on the probability to do sports of an additional year of education is equivalent to being 0,88 years younger, but only for females.

23. **(Problem 2.)** Using the Logit model estimates, the probability that a 30 year old female 30 with 20 years of education participates in sports is

   (a) 0.27 less than that of a male with equal characteristics.
   (b) 0.31 less than that of a male with equal characteristics.
   (c) 0.29 less than that of a male with equal characteristics.
   (d) 0.29 less than that of a male with the mean characteristics in the sample.

24. **(Problem 1.)** Regarding equation (1):

(i) The expected difference in years of education between a child that grew up in a household with a good economic situation and one that grew up in a household with a very poor economic situation is given by $\beta_7 - \beta_0$, if we assume that all other factors are held constant.

(ii) The expected difference in years of education between a child that studied in an urban area inside a household with a good economic situation and another that studied in a rural area inside a household with a very poor economic situation is given by $\beta_5 + \beta_7$, if we assume that all other factors are held constant.

(iii) The expected difference in years of education between a child that grew up in a household with a good economic situation and a working mother and another that grew up in a household with a very poor economic situation, for which we do not know whether the mother was working or not, is given by $\beta_4 + \beta_7$, if we assume that all other factors are held constant.

(a) All statements are false.

(b) Only (ii) is true.

(c) Only (i) is true.

(d) Only (ii) and (iii) are true.

25. **(Problem 1.)** Consider the following statements with respect to Output 1:

(i) We can conclude that the number of siblings has an effect on the total number of years of education of an individual, since $n1childg1$ is significant in Output 1.

(ii) The endogeneity of $n1childg1$ affects the estimated standard errors of the OLS estimators in Output 1, but does not affect the consistency of the estimators.

(iii) All estimators in Output 1 are consistent, except that of $n1childg1$, because it is endogenous.

(a) All statements are false.

(b) Only (ii) is true.

(c) Only (i) is true.

(d) Only (ii) and (iii) are true.

26. **(Problem 1.)** The validity of instruments $boy12$ and $girl12$,

(i) Solely depends on the zero correlation between the instrument and the omitted factor in the equation.

(ii) Can have the implication that the equation is over-identified.

(iii) Guarantees that the OLS estimators are consistent.

(a) (i) and (iii) are true.
(b) Only (ii) is true.
(c) Only (i) is true.
(d) (i) and (ii) are true.

27. **(Problem 1.)** Consider the following statements:

(i) The instruments have to be exogenous variables whose OLS coefficient in the original equation is not significant.

(ii) The instruments can be correlated with the dependent variable.

(iii) The instruments can be correlated with the exogenous variables.

(a) All statements are false.
(b) Only (ii) is true.
(c) (i) and (iii) are true.
(d) (ii) and (iii) are true.

28. **(Problem 1.)** Using instruments $boy12$ and $girl12$, and assuming that they are uncorrelated with u, we can conclude that:

(a) Equation (1) is not identified because $boy12$ is not significant in Output 2.
(b) Equation (1) is identified because $boy12$ and $girl12$ are jointly significant in the reduced form of $n1childg1$, since the value of the corresponding test statistic, 13.71, is significant at the 5% level.
(c) Equation (1) is identified because $boy12$ and $girl12$ are exogenous.
(d) Equation (1) is not identified because $boy12$ and $girl12$ are not jointly significant in the reduced form of $n1childg1$, since the value of the corresponding test statistic, 55.27, is not significant at the 1% level.

29. **(Problem 1.)** Assuming that the instruments $boy12$ and $girl12$ are uncorrelated with u, and using the relevant information from the Outputs,

(i) The 2SLS estimation of Output 4 is more efficient than the 2SLS estimation using only $girl12$ because having more instruments is always better.

(ii) The 2SLS estimation of Output 4, is not consistent because $boy12$ is not significant in Output 2.

(iii) Since we know that $boy12$ is correlated with $n1childg1$, it is always best to use it as an instrument in the 2SLS estimation.

(a) All statements are false.
(b) Only (ii) is true.
(c) Only (i) is true.
(d) Only (i) and (iii) are true.

30. **(Problem 1.)** Consider the following statements:

(i) Hausman's test compares two alternative forms to estimate the parameters of a model on the basis that, at least one of the two forms is not consistent under H_0.

(ii) The estimators of Output 5 are consistent, but less efficient than those of Output 4 if equation(1) is exactly identified.

(iii) The endogeneity test requires that the equation is over-identified.

(a) All are false.
(b) Only (iii) is true.
(c) Only (ii) is true.
(d) Only (i) and (iii) are true.

31. **(Problem 1.)** The results of Output 5:

(i) Estimate the parameters by Two Stages Least Squares.

(ii) Allow us to confirm the endogeneity of $n1childg1$ at the 1% significance level.

(iii) Indicate that $n1childg1$ is uncorrelated with u in equation (1).

(a) (i) and (ii) are true.
(b) Only (iii) is true.
(c) Only (i) is true.
(d) All statements are true.

32. **(Problem 1.)** If we had at our disposal the residuals from a 2SLS estimation of equation (1) using as a sole instrument $girl2$, and the results of the OLS regression where those residuals are the dependent variable as a function of all the exogenous variables:

(i) It would have been possible to test if $boy12$ is uncorrelated with u in equation (1).

(ii) It would have been possible to test if $girl12$ is uncorrelated with u in equation (1).

(iii) It wold have been possible to simultaneously test if $boy12$ and $girl12$ are uncorrelated with u in equation (1).

(a) All statements are false.
(b) Only (ii) is true.
(c) Only (i) is true.
(d) Only (iii) is true.

33. **(Problem 1.)** Using the appropriate outputs, we can conclude that:

(i) The number of siblings has a negative effect on the individual's education.

(ii) The economic situation of the family does not affect education because *pobre* is not significant in Output 1 at the 5% level.

(iii) The number of siblings has a positive significant effect on *yedu* at the 10% level but not at the 5% level.

(a) Only (iii) is true.
(b) (i) and (ii) are true.
(c) Only (i) is true.
(d) All statements are false.

34. **(Problem 1.)**

(i) The fact that the standard deviation of the residuals in Output 1 is lower than in Output 4 confirms that OLS is consistent.

(ii) The R^2 in the 2SLS output has the same interpretation and use as in the OLS case.

(iii) The asymptotic variance of the 2SLS estimators is always greater or equal to that of the OLS estimators.

(a) Only (ii) is false.
(b) (ii) and (iii) are true.
(c) (i) and (ii) are false.
(d) Only (ii) is true.

Solution to Exam Type: 3

Universidad Carlos III de Madrid

ECONOMETRICS I

Academic year 2006/07
FINAL EXAM

February 9, 2008

TIME: 2 HOURS 30 MINUTES

1. **(Problem 3.)** Out of the four suggested specifications:

   (a) None of the specifications can be estimated by OLS because they omit relevant variables.

   (b) All can be estimated by OLS, but only (2) provides consistent estimators.

   (c) Only (2) and (3) can be estimated by OLS.

   [(d)] The only specification that cannot be estimated by OLS is (4).

2. **(Problem 3.)** Consider the following statements with respect to the previous models:

   (i) It is possible that assumption $E(u|x) = 0$ where x is a vector that contains all the explanatory variables included in each equation, is fulfilled in some of these specifications since they consider single and married working mothers separately.

   (ii) In specification (2), parameter β_3 is interpreted as the difference between the wage rate of married working mothers and single females.

   (iii) There are problems of perfect multicollinearity in Model (2).

   (a) (ii) and (iii) are false.

   (b) (i) and (ii) are false.

   [(c)] All statements are false.

   (d) (i) and (iii) are false.

3. **(Problem 3.)** In specification (2), the impact of being married on the wage rate is:

   [(a)] $\beta_2 + \beta_3 m$.

   (b) $\beta_0 + \beta_2 + \beta_3 m$.

   (c) β_2.

   (d) $\beta_2 + \beta_3 c$.

4. **(Problem 3.)** Select the correct answer with respect to specification (1):

   - [(a)] All OLS estimators can be biased with a bias equal to $\delta_j \beta_1$ where δ_j are the OLS estimators from the regression of the excluded variable on the included variable, and β_1 is the coefficient of the excluded variable.
   - (b) All OLS estimators can be biased with a bias equal to $\delta_1 x_1$ where δ_1 is the OLS estimator from the regression of the wage rate w on the omitted variable.
   - (c) OLS will only provide an unbiased estimator of the constant term.
   - (d) All OLS estimators can be biased with a bias equal to $\delta_j x_1$ where δ_j are the OLS estimators from the regression of the included variables on the omitted variable.

5. **(Problem 3.)** We assume that the decision to become a working mother depends on the level of family income, that is why we suggest another equation:

$$m = \gamma_0 + \gamma_1 w + \gamma_2 sm + \gamma_3 c + \varepsilon \qquad (6)$$

   where sm is the wage rate of the partner with whom she lives. It is assumed that $E(\varepsilon | sm, c) = 0$. Considering specification (2) for w, and that sm and c are exogenous in the aforementioned equation, $E(u|sm, c) = 0$, select the correct response:

   - [(a)] None of the two equations satisfies the rank nor the order conditions, since in general we expect that $m \times c$ would also be endogenous in equation (2).
   - (b) Equation (2) satisfies the order condition, but equation (6) does not satisfy the order condition.
   - (c) Both equations, (2) and (6), satisfy the order condition but fail to satisfy the rank condition.
   - (d) Both equations satisfy the rank and order conditions.

6. **(Problem 3.)** Consider the following statements referring to equation (6) for m :

   - (a) We can only obtain efficient estimators of the model if instead of OLS we use FGLS (Feasible Generalised Least Squares) in the 2SLS estimation.
   - (b) It is impossible to estimate by OLS given the presence of the evident simultaneity problems. It can only be estimated by maximum likelihood i.e. using a logit or probit model according to the assumptions about the errors distribution.
   - (c) It is equivalent to estimating a Linear Probability Model by OLS, and hence there are no problems with the bias and inconsistency of the estimators.
   - [(d)] First, we need to estimate the reduced form of w by OLS, so that we are able to estimate the structural form by OLS replacing w by the estimated values in the first regression, but this is not possible since we have not got enough instruments.

7. **(Problem 3.)** What would occur if instead of using m as a dependent variable we used $50m$ in (6)?

   (i) The only effect on the OLS estimators will be that all are multiplied by 50 as well. The coefficient of determination R^2 will not be affected.

   (ii) The sum of the squared residuals will be multiplied by 2.500.

   (iii) The OLS estimators will now be biased upwards by 50.

   (a) All statements are true.
   (b) Only (iii) is true.
   (c) Only (i) is true.
   [(d)] Only (i) and (ii) are true.

8. **(Problem 3.)** With respect to the assumption of heteroskedasticity in model (6) for m (and ignoring any **endogeneity** problems), consider the following statements:

   (i) We can estimate the equation by OLS, but for inference purposes we will use the robust t statistics, even though this estimations would not be efficient if heteroskedasticity is present.

   (ii) In this type of models there is no need to test for heteroskedasticity, so that it will no matter whether we use the usual standard errors or the corrected ones.

   (iii) We can estimate the model by Feasible Generalized LS, without the need to run in first place a regression for the squares of the OLS residuals.

   (a) Only (ii) and(iii) are true.
   [(b)] Only (i) and (iii) are true.
   (c) Only (i) and (ii) are true.
   (d) Only (i) is true.

9. **(Problem 3.)** Assuming that there are no simultaneity problems, but that w was measured with error in equation (6) such that:

$$w = w_{true} + e_1,$$

where e_1 denotes the measurement error and w_{true} is the true level of the wage rate.

(i) The measurement error will always cause a decrease in the variance of the random disturbance.

(ii) If $Cov\,(w, e_1) = 0$ holds, the OLS estimators will be unbiased and consistent.

(iii) Under the CEV (Classical Errors in Variables) assumption, the OLS estimators will be unbiased, but with lower variance than without measurement error due to the attenuation bias.

(a) Only (i) and (ii) are true.

(b) Only (ii) and (iii) are true.

(c) Only (i) is true.

[(d)] Only (ii) is true.

10. **(Problem 3.)** We suspect that equation (6) is misspecified (assuming that there is no simultaneity problem). Consider the following statements:

(i) If assumption $E\,(\varepsilon|x) = 0$ is satisfied, where x is a vector containing the explanatory variables, then none of the added non-linear functions of the $x's$ should be significant.

(ii) The RESET test results indicate whether it is necessary to use heteroskedasticity robust standard errors.

(iii) The RESET test does not inform us on how to proceed if the null hypothesis of correct specification is rejected.

(a) All statements are true.

(b) (ii) and (iii) are true.

[(c)] (i) and (iii) are true.

(d) Only (i) is true.

11. **(Problem 3.)** (*Continues*)

   (i) An advantage of the RESET test is that it can compare models with different specifications of the dependent variable.

   (ii) The disadvantage of the Mizon-Richard and Davidson-MacKinnon tests is that they can give inconclusive results regarding which model is preferable.

   (iii) White's test has the capacity to detect more general specification errors compared to the Breusch-Pagan test.

   (a) All statements are true.
   (b) Only (*i*) and (*iii*) are true.
   (c) Only (*i*) is true.
   [(d)] Only (*ii*) is true.

12. **(Problem 2.)** In the Linear Probability Model,

   (i) The error term would present heteroskedasticity (conditional on the explanatory variables).

   (ii) The error term (conditional on the explanatory variables) would be normally distributed.

   (iii) The predicted probabilities could be greater than one, but not lower than zero.

   [(a)] Only (i) is true.
   (b) Only (i) and (ii) are true.
   (c) The three statements are true.
   (d) Only (i) and (iii) are true.

13. **(Problem 2.)** Using the linear probability model estimates,

   (a) The probability to participate in sports always diminishes with age.
   (b) The probability to participate in sports diminishes with age after the age of 35, approximately.
   [(c)] The probability to participate in sports increases with age up to the age of 50, approximately.
   (d) The probability to participate in sports diminishes after the age of 39,5, approximately.

14. **(Problem 2.)** Using the linear probability model estimates, the probability that a 30 year old male with 10 years of education participates in sports is:

   (a) 0.94.
   (b) 0.36.
   [(c)] 0.52.
   (d) 0.81.

15. **(Problem 2.)** Using the linear probability model estimates, we can conclude that

(i) Females have a lower probability to participate in sports, holding all else equal.

(ii) Age does not affect the decision to participate in sports because the coefficient on *age*2 is not significant at the 5% level.

(iii) A male with 10 years of education has a higher probability to do sports than a female with 20 years of education, everything else being held equal.

(a) (ii) and (iii) are true.
(b) (ii) is true.
(c) Only (i) is true.
[(d)] (i) and (iii) are true.

16. **(Problem 2.)** The estimated linear probability model specification (Output 1),

(i) Permits measuring the effect of sex on the probability to do sports.

(ii) Permits that the partial effect of age varies with its value.

(iii) Permits that the effect of education on the probability to do sports varies by gender.

(a) Only (iii) is true.
[(b)] Only (i) and (ii) are true.
(c) Only (i) is true.
(d) Only (i) and (iii) are true.

17. **(Problem 2.)** In a Logit model,

(i) The partial effect of variable x_j on the probability that $y = 1$ (conditional on the explanatory variables), depends on β_j and x_j, but not on the rest of the explanatory variables or parameters.

(ii) The partial effect of variable x_j on the probability that $y = 1$ (conditional on the explanatory variables), depends on $\beta_0, \beta_1, \ldots, \beta_k$ and on x_j, but not on the rest of the explanatory variables.

(iii) The partial effect of variable x_j on the probability that $y = 1$ (conditional on the explanatory variables), depends on β_j and on $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_k$, but not on the rest of the parameters.

(a) Only (iii) is true.
(b) Only (i) is true.
[(c)] The three statements are false.
(d) Only (ii) is true.

18. **(Problem 2.)** Using the Logit model estimates, the effect on the probability to participate in sports, of a 10 year increase in education (ceteris paribus) for an individual with the mean characteristics in the sample, is

(a) 0.017.
[(b)] 0.17.
(c) 0.74.
(d) 0.074.

19. **(Problem 2.)** Using the Logit model estimates,

(i) For some females it is possible that the predicted probability to do sports is negative.

(ii) The difference in the probability to do sports between a male and a female is always 0.29 in favour of the male, independently of his age.

(iii) The magnitude of the partial effect of years of education on the probability to do sports, diminishes when *yedu* increases if *yedu* is sufficiently high, all else held equal.

[(a)] Only *(iii)* is true.
(b) Only *(i)* is true.
(c) Only *(ii)* and *(iii)* are true.
(d) Only *(i)* and *(ii)* are true.

20. **(Problem 2.)** Using the Logit model estimates, to test that the effect of education on the probability to do sports does not depend on gender:

[(a)] I would have tested if the coefficient on $female$ is significant using a t test.

(b) I would have used the log-likelihoods of this model and another which is not inclusive of variables $female$ and *yedu*.

(c) I would have tested if the coefficients of $female$ and *yedu* are jointly significant with a likelihood ratio test.

(d) It is necessary to estimate a model that is inclusive of interactions of *yedu* with the variable $female$.

21. **(Problem 2.)** Using the Logit model estimates,

 (i) We can conclude that the explanatory variables aid in predicting the probability to do sports because the percent correctly predicted is fairly high.

 (ii) We can reject the hypothesis that the coefficients on $female$, age, $age2$ and $yedu$ are equal to zero at any usual significance level.

 (iii) The model is not properly specified since the pseudo-R^2 is very low.

 (a) Only (i) is true.
 [(b)] Only (ii) is true.
 (c) Only (i) and (ii) are true.
 (d) Only (i) and (iii) are true.

22. **(Problem 2.)** Using the Logit model estimates, and assuming that the coefficient on $age2$ is exactly zero,

 (a) The partial effect on the probability to do sports of an additional year of education is equivalent to being 0,88 years younger, but only if all explanatory variables are evaluated at the sample means.
 (b) The partial effect on the probability to do sports of an additional year of education is equivalent to being 0,88 years younger, but only for males.
 [(c)] The partial effect on the probability to do sports of an additional year of education is equivalent to being 0,88 years younger.
 (d) The partial effect on the probability to do sports of an additional year of education is equivalent to being 0,88 years younger, but only for females.

23. **(Problem 2.)** Using the Logit model estimates, the probability that a 30 year old female 30 with 20 years of education participates in sports is

 (a) 0.27 less than that of a male with equal characteristics.
 [(b)] 0.31 less than that of a male with equal characteristics.
 (c) 0.29 less than that of a male with equal characteristics.
 (d) 0.29 less than that of a male with the mean characteristics in the sample.

24. **(Problem 1.)** Regarding equation (1):

(i) The expected difference in years of education between a child that grew up in a household with a good economic situation and one that grew up in a household with a very poor economic situation is given by $\beta_7 - \beta_0$, if we assume that all other factors are held constant.

(ii) The expected difference in years of education between a child that studied in an urban area inside a household with a good economic situation and another that studied in a rural area inside a household with a very poor economic situation is given by $\beta_5 + \beta_7$, if we assume that all other factors are held constant.

(iii) The expected difference in years of education between a child that grew up in a household with a good economic situation and a working mother and another that grew up in a household with a very poor economic situation, for which we do not know whether the mother was working or not, is given by $\beta_4 + \beta_7$, if we assume that all other factors are held constant.

(a) All statements are false.

[(b)] Only (ii) is true.

(c) Only (i) is true.

(d) Only (ii) and (iii) are true.

25. **(Problem 1.)** Consider the following statements with respect to Output 1:

(i) We can conclude that the number of siblings has an effect on the total number of years of education of an individual, since $n1childg1$ is significant in Output 1.

(ii) The endogeneity of $n1childg1$ affects the estimated standard errors of the OLS estimators in Output 1, but does not affect the consistency of the estimators.

(iii) All estimators in Output 1 are consistent, except that of $n1childg1$, because it is endogenous.

[(a)] All statements are false.

(b) Only (ii) is true.

(c) Only (i) is true.

(d) Only (ii) and (iii) are true.

26. **(Problem 1.)** The validity of instruments $boy12$ and $girl12$,

(i) Solely depends on the zero correlation between the instrument and the omitted factor in the equation.

(ii) Can have the implication that the equation is over-identified.

(iii) Guarantees that the OLS estimators are consistent.

(a) (i) and (iii) are true.
[(b)] Only (ii) is true.
(c) Only (i) is true.
(d) (i) and (ii) are true.

27. **(Problem 1.)** Consider the following statements:

(i) The instruments have to be exogenous variables whose OLS coefficient in the original equation is not significant.

(ii) The instruments can be correlated with the dependent variable.

(iii) The instruments can be correlated with the exogenous variables.

(a) All statements are false.
(b) Only (ii) is true.
(c) (i) and (iii) are true.
[(d)] (ii) and (iii) are true.

28. **(Problem 1.)** Using instruments $boy12$ and $girl12$, and assuming that they are uncorrelated with u, we can conclude that:

(a) Equation (1) is not identified because $boy12$ is not significant in Output 2.

[(b)] Equation (1) is identified because $boy12$ and $girl12$ are jointly significant in the reduced form of $n1childg1$, since the value of the corresponding test statistic, 13.71, is significant at the 5% level.

(c) Equation (1) is identified because $boy12$ and $girl12$ are exogenous.

(d) Equation (1) is not identified because $boy12$ and $girl12$ are not jointly significant in the reduced form of $n1childg1$, since the value of the corresponding test statistic, 55.27, is not significant at the 1% level.

29. **(Problem 1.)** Assuming that the instruments $boy12$ and $girl12$ are uncorrelated with u, and using the relevant information from the Outputs,

(i) The 2SLS estimation of Output 4 is more efficient than the 2SLS estimation using only $girl12$ because having more instruments is always better.

(ii) The 2SLS estimation of Output 4, is not consistent because $boy12$ is not significant in Output 2.

(iii) Since we know that $boy12$ is correlated with $n1childg1$, it is always best to use it as an instrument in the 2SLS estimation.

[(a)] All statements are false.
(b) Only (ii) is true.
(c) Only (i) is true.
(d) Only (i) and (iii) are true.

30. **(Problem 1.)** Consider the following statements:

(i) Hausman's test compares two alternative forms to estimate the parameters of a model on the basis that, at least one of the two forms is not consistent under H_0.

(ii) The estimators of Output 5 are consistent, but less efficient than those of Output 4 if equation(1) is exactly identified.

(iii) The endogeneity test requires that the equation is over-identified.

[(a)] All are false.
(b) Only (iii) is true.
(c) Only (ii) is true.
(d) Only (i) and (iii) are true.

31. **(Problem 1.)** The results of Output 5:

(i) Estimate the parameters by Two Stages Least Squares.

(ii) Allow us to confirm the endogeneity of $n1childg1$ at the 1% significance level.

(iii) Indicate that $n1childg1$ is uncorrelated with u in equation (1).

[(a)] (i) and (ii) are true.
(b) Only (iii) is true.
(c) Only (i) is true.
(d) All statements are true.

32. **(Problem 1.)** If we had at our disposal the residuals from a 2SLS estimation of equation (1) using as a sole instrument $girl2$, and the results of the OLS regression where those residuals are the dependent variable as a function of all the exogenous variables:

   (i) It would have been possible to test if $boy12$ is uncorrelated with u in equation (1).

   (ii) It would have been possible to test if $girl12$ is uncorrelated with u in equation (1).

   (iii) It wold have been possible to simultaneously test if $boy12$ and $girl12$ are uncorrelated with u in equation (1).

   (a) All statements are false.
   (b) Only (ii) is true.
   [(c)] Only (i) is true.
   (d) Only (iii) is true.

33. **(Problem 1.)** Using the appropriate outputs, we can conclude that:

   (i) The number of siblings has a negative effect on the individual's education.

   (ii) The economic situation of the family does not affect education because *pobre* is not significant in Output 1 at the 5% level.

   (iii) The number of siblings has a positive significant effect on *yedu* at the 10% level but not at the 5% level.

   [(a)] Only (iii) is true.
   (b) (i) and (ii) are true.
   (c) Only (i) is true.
   (d) All statements are false.

34. **(Problem 1.)**

   (i) The fact that the standard deviation of the residuals in Output 1 is lower than in Output 4 confirms that OLS is consistent.

   (ii) The R^2 in the 2SLS output has the same interpretation and use as in the OLS case.

   (iii) The asymptotic variance of the 2SLS estimators is always greater or equal to that of the OLS estimators.

   (a) Only (ii) is false.
   (b) (ii) and (iii) are true.
   [(c)] (i) and (ii) are false.
   (d) Only (ii) is true.

Universidad Carlos III de Madrid
ECONOMETRICS I
Academic year 2006/07
FINAL EXAM
February 9, 2008

Tipo de examen: 4

TIME: 2 HOURS 30 MINUTES

1. **(Problem 1.)** Regarding equation (1):

   (i) The expected difference in years of education between a child that grew up in a household with a good economic situation and one that grew up in a household with a very poor economic situation is given by $\beta_7 - \beta_0$, if we assume that all other factors are held constant.

   (ii) The expected difference in years of education between a child that studied in an urban area inside a household with a good economic situation and another that studied in a rural area inside a household with a very poor economic situation is given by $\beta_5 + \beta_7$, if we assume that all other factors are held constant.

   (iii) The expected difference in years of education between a child that grew up in a household with a good economic situation and a working mother and another that grew up in a household with a very poor economic situation, for which we do not know whether the mother was working or not, is given by $\beta_4 + \beta_7$, if we assume that all other factors are held constant.

   (a) Only (ii) and (iii) are true.

   (b) Only (i) is true.

   (c) Only (ii) is true.

   (d) All statements are false.

2. **(Problem 1.)** Consider the following statements with respect to Output 1:

   (i) We can conclude that the number of siblings has an effect on the total number of years of education of an individual, since $n1childg1$ is significant in Output 1.

   (ii) The endogeneity of $n1childg1$ affects the estimated standard errors of the OLS estimators in Output 1, but does not affect the consistency of the estimators.

   (iii) All estimators in Output 1 are consistent, except that of $n1childg1$, because it is endogenous.

   (a) Only (ii) and (iii) are true.
   (b) Only (i) is true.
   (c) Only (ii) is true.
   (d) All statements are false.

3. **(Problem 1.)** The validity of instruments $boy12$ and $girl12$,

   (i) Solely depends on the zero correlation between the instrument and the omitted factor in the equation.

   (ii) Can have the implication that the equation is over-identified.

   (iii) Guarantees that the OLS estimators are consistent.

   (a) (i) and (ii) are true.
   (b) Only (i) is true.
   (c) Only (ii) is true.
   (d) (i) and (iii) are true.

4. **(Problem 1.)** Consider the following statements:

   (i) The instruments have to be exogenous variables whose OLS coefficient in the original equation is not significant.

   (ii) The instruments can be correlated with the dependent variable.

   (iii) The instruments can be correlated with the exogenous variables.

   (a) (ii) and (iii) are true.
   (b) (i) and (iii) are true.
   (c) Only (ii) is true.
   (d) All statements are false.

5. **(Problem 1.)** Using instruments $boy12$ and $girl12$, and assuming that they are uncorrelated with u, we can conclude that:

   (a) Equation (1) is not identified because $boy12$ and $girl12$ are not jointly significant in the reduced form of $n1childg1$, since the value of the corresponding test statistic, 55.27, is not significant at the 1% level.

   (b) Equation (1) is identified because $boy12$ and $girl12$ are exogenous.

   (c) Equation (1) is identified because $boy12$ and $girl12$ are jointly significant in the reduced form of $n1childg1$, since the value of the corresponding test statistic, 13.71, is significant at the 5% level.

   (d) Equation (1) is not identified because $boy12$ is not significant in Output 2.

6. **(Problem 1.)** Assuming that the instruments $boy12$ and $girl12$ are uncorrelated with u, and using the relevant information from the Outputs,

   (i) The 2SLS estimation of Output 4 is more efficient than the 2SLS estimation using only $girl12$ because having more instruments is always better.

   (ii) The 2SLS estimation of Output 4, is not consistent because $boy12$ is not significant in Output 2.

   (iii) Since we know that $boy12$ is correlated with $n1childg1$, it is always best to use it as an instrument in the 2SLS estimation.

   (a) Only (i) and (iii) are true.

   (b) Only (i) is true.

   (c) Only (ii) is true.

   (d) All statements are false.

7. **(Problem 1.)** Consider the following statements:

   (i) Hausman's test compares two alternative forms to estimate the parameters of a model on the basis that, at least one of the two forms is not consistent under H_0.

   (ii) The estimators of Output 5 are consistent, but less efficient than those of Output 4 if equation(1) is exactly identified.

   (iii) The endogeneity test requires that the equation is over-identified.

   (a) Only (i) and (iii) are true.

   (b) Only (ii) is true.

   (c) Only (iii) is true.

   (d) All are false.

8. **(Problem 1.)** The results of Output 5:

   (i) Estimate the parameters by Two Stages Least Squares.

   (ii) Allow us to confirm the endogeneity of $n1childg1$ at the 1% significance level.

   (iii) Indicate that $n1childg1$ is uncorrelated with u in equation (1).

   (a) All statements are true.
   (b) Only (i) is true.
   (c) Only (iii) is true.
   (d) (i) and (ii) are true.

9. **(Problem 1.)** If we had at our disposal the residuals from a 2SLS estimation of equation (1) using as a sole instrument $girl2$, and the results of the OLS regression where those residuals are the dependent variable as a function of all the exogenous variables:

   (i) It would have been possible to test if $boy12$ is uncorrelated with u in equation (1).

   (ii) It would have been possible to test if $girl12$ is uncorrelated with u in equation (1).

   (iii) It wold have been possible to simultaneously test if $boy12$ and $girl12$ are uncorrelated with u in equation (1).

   (a) Only (iii) is true.
   (b) Only (i) is true.
   (c) Only (ii) is true.
   (d) All statements are false.

10. **(Problem 1.)** Using the appropriate outputs, we can conclude that:

   (i) The number of siblings has a negative effect on the individual's education.

   (ii) The economic situation of the family does not affect education because *pobre* is not significant in Output 1 at the 5% level.

   (iii) The number of siblings has a positive significant effect on *yedu* at the 10% level but not at the 5% level.

   (a) All statements are false.
   (b) Only (i) is true.
   (c) (i) and (ii) are true.
   (d) Only (iii) is true.

11. **(Problem 1.)**

   (i) The fact that the standard deviation of the residuals in Output 1 is lower than in Output 4 confirms that OLS is consistent.

   (ii) The R^2 in the 2SLS output has the same interpretation and use as in the OLS case.

   (iii) The asymptotic variance of the 2SLS estimators is always greater or equal to that of the OLS estimators.

   (a) Only (ii) is true.
   (b) (i) and (ii) are false.
   (c) (ii) and (iii) are true.
   (d) Only (ii) is false.

12. **(Problem 3.)** Out of the four suggested specifications:

   (a) The only specification that cannot be estimated by OLS is (4).
   (b) Only (2) and (3) can be estimated by OLS.
   (c) All can be estimated by OLS, but only (2) provides consistent estimators.
   (d) None of the specifications can be estimated by OLS because they omit relevant variables.

13. **(Problem 3.)** Consider the following statements with respect to the previous models:

   (i) It is possible that assumption $E(u|x) = 0$ where x is a vector that contains all the explanatory variables included in each equation, is fulfilled in some of these specifications since they consider single and married working mothers separately.

   (ii) In specification (2), parameter β_3 is interpreted as the difference between the wage rate of married working mothers and single females.

   (iii) There are problems of perfect multicollinearity in Model (2).

   (a) (i) and (iii) are false.
   (b) All statements are false.
   (c) (i) and (ii) are false.
   (d) (ii) and (iii) are false.

14. **(Problem 3.)** In specification (2), the impact of being married on the wage rate is:

   (a) $\beta_2 + \beta_3 c$.
   (b) β_2.
   (c) $\beta_0 + \beta_2 + \beta_3 m$.
   (d) $\beta_2 + \beta_3 m$.

15. **(Problem 3.)** Select the correct answer with respect to specification (1):

   (a) All OLS estimators can be biased with a bias equal to $\delta_j x_1$ where δ_j are the OLS estimators from the regression of the included variables on the omitted variable.

   (b) OLS will only provide an unbiased estimator of the constant term.

   (c) All OLS estimators can be biased with a bias equal to $\delta_1 x_1$ where δ_1 is the OLS estimator from the regression of the wage rate w on the omitted variable.

   (d) All OLS estimators can be biased with a bias equal to $\delta_j \beta_1$ where δ_j are the OLS estimators from the regression of the excluded variable on the included variable, and β_1 is the coefficient of the excluded variable.

16. **(Problem 3.)** We assume that the decision to become a working mother depends on the level of family income, that is why we suggest another equation:

$$m = \gamma_0 + \gamma_1 w + \gamma_2 sm + \gamma_3 c + \varepsilon \qquad (6)$$

   where sm is the wage rate of the partner with whom she lives. It is assumed that $E(\varepsilon|sm, c) = 0$. Considering specification (2) for w, and that sm and c are exogenous in the aforementioned equation, $E(u|sm, c) = 0$, select the correct response:

   (a) Both equations satisfy the rank and order conditions.

   (b) Both equations, (2) and (6), satisfy the order condition but fail to satisfy the rank condition.

   (c) Equation (2) satisfies the order condition, but equation (6) does not satisfy the order condition.

   (d) None of the two equations satisfies the rank nor the order conditions, since in general we expect that $m \times c$ would also be endogenous in equation (2).

17. (**Problem 3.)** Consider the following statements referring to equation (6) for m :

   (a) First, we need to estimate the reduced form of w by OLS, so that we are able to estimate the structural form by OLS replacing w by the estimated values in the first regression, but this is not possible since we have not got enough instruments.

   (b) It is equivalent to estimating a Linear Probability Model by OLS, and hence there are no problems with the bias and inconsistency of the estimators.

   (c) It is impossible to estimate by OLS given the presence of the evident simultaneity problems. It can only be estimated by maximum likelihood i.e. using a logit or probit model according to the assumptions about the errors distribution.

   (d) We can only obtain efficient estimators of the model if instead of OLS we use FGLS (Feasible Generalised Least Squares) in the 2SLS estimation.

18. **(Problem 3.)** What would occur if instead of using m as a dependent variable we used $50m$ in (6)?

    (i) The only effect on the OLS estimators will be that all are multiplied by 50 as well. The coefficient of determination R^2 will not be affected.

    (ii) The sum of the squared residuals will be multiplied by 2.500.

    (iii) The OLS estimators will now be biased upwards by 50.

    (a) Only (i) and (ii) are true.
    (b) Only (i) is true.
    (c) Only (iii) is true.
    (d) All statements are true.

19. **(Problem 3.)** With respect to the assumption of heteroskedasticity in model (6) for m (and ignoring any **endogeneity** problems), consider the following statements:

    (i) We can estimate the equation by OLS, but for inference purposes we will use the robust t statistics, even though this estimations would not be efficient if heteroskedasticity is present.

    (ii) In this type of models there is no need to test for heteroskedasticity, so that it will no matter whether we use the usual standard errors or the corrected ones.

    (iii) We can estimate the model by Feasible Generalized LS, without the need to run in first place a regression for the squares of the OLS residuals.

    (a) Only (i) is true.
    (b) Only (i) and (ii) are true.
    (c) Only (i) and (iii) are true.
    (d) Only (ii) and(iii) are true.

20. **(Problem 3.)** Assuming that there are no simultaneity problems, but that w was measured with error in equation (6) such that:

$$w = w_{true} + e_1,$$

where e_1 denotes the measurement error and w_{true} is the true level of the wage rate.

   (i) The measurement error will always cause a decrease in the variance of the random disturbance.

   (ii) If $Cov\,(w, e_1) = 0$ holds, the OLS estimators will be unbiased and consistent.

   (iii) Under the CEV (Classical Errors in Variables) assumption, the OLS estimators will be unbiased, but with lower variance than without measurement error due to the attenuation bias.

   (a) Only (ii) is true.
   (b) Only (i) is true.
   (c) Only (ii) and (iii) are true.
   (d) Only (i) and (ii) are true.

21. **(Problem 3.)** We suspect that equation (6) is misspecified (assuming that there is no simultaneity problem). Consider the following statements:

   (i) If assumption $E\,(\varepsilon|x) = 0$ is satisfied, where x is a vector containing the explanatory variables, then none of the added non-linear functions of the $x's$ should be significant.

   (ii) The RESET test results indicate whether it is necessary to use heteroskedasticity robust standard errors.

   (iii) The RESET test does not inform us on how to proceed if the null hypothesis of correct specification is rejected.

   (a) Only (i) is true.
   (b) (i) and (iii) are true.
   (c) (ii) and (iii) are true.
   (d) All statements are true.

22. **(Problem 3.)** (*Continues*)

(i) An advantage of the RESET test is that it can compare models with different specifications of the dependent variable.

(ii) The disadvantage of the Mizon-Richard and Davidson-MacKinnon tests is that they can give inconclusive results regarding which model is preferable.

(iii) White's test has the capacity to detect more general specification errors compared to the Breusch-Pagan test.

(a) Only (*ii*) is true.

(b) Only (*i*) is true.

(c) Only (*i*) and (*iii*) are true.

(d) All statements are true.

23. **(Problem 2.)** In the Linear Probability Model,

(i) The error term would present heteroskedasticity (conditional on the explanatory variables).

(ii) The error term (conditional on the explanatory variables) would be normally distributed.

(iii) The predicted probabilities could be greater than one, but not lower than zero.

(a) Only (i) and (iii) are true.

(b) The three statements are true.

(c) Only (i) and (ii) are true.

(d) Only (i) is true.

24. **(Problem 2.)** Using the linear probability model estimates,

(a) The probability to participate in sports diminishes after the age of 39,5, approximately.

(b) The probability to participate in sports increases with age up to the age of 50, approximately.

(c) The probability to participate in sports diminishes with age after the age of 35, approximately.

(d) The probability to participate in sports always diminishes with age.

25. **(Problem 2.)** Using the linear probability model estimates, the probability that a 30 year old male with 10 years of education participates in sports is:

(a) 0.81.

(b) 0.52.

(c) 0.36.

(d) 0.94.

26. **(Problem 2.)** Using the linear probability model estimates, we can conclude that

(i) Females have a lower probability to participate in sports, holding all else equal.

(ii) Age does not affect the decision to participate in sports because the coefficient on $age2$ is not significant at the 5% level.

(iii) A male with 10 years of education has a higher probability to do sports than a female with 20 years of education, everything else being held equal.

(a) (i) and (iii) are true.
(b) Only (i) is true.
(c) (ii) is true.
(d) (ii) and (iii) are true.

27. **(Problem 2.)** The estimated linear probability model specification (Output 1),

(i) Permits measuring the effect of sex on the probability to do sports.

(ii) Permits that the partial effect of age varies with its value.

(iii) Permits that the effect of education on the probability to do sports varies by gender.

(a) Only (i) and (iii) are true.
(b) Only (i) is true.
(c) Only (i) and (ii) are true.
(d) Only (iii) is true.

28. **(Problem 2.)** In a Logit model,

(i) The partial effect of variable x_j on the probability that $y = 1$ (conditional on the explanatory variables), depends on β_j and x_j, but not on the rest of the explanatory variables or parameters.

(ii) The partial effect of variable x_j on the probability that $y = 1$ (conditional on the explanatory variables), depends on $\beta_0, \beta_1, \ldots, \beta_k$ and on x_j, but not on the rest of the explanatory variables.

(iii) The partial effect of variable x_j on the probability that $y = 1$ (conditional on the explanatory variables), depends on β_j and on $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_k$, but not on the rest of the parameters.

(a) Only (ii) is true.
(b) The three statements are false.
(c) Only (i) is true.
(d) Only (iii) is true.

29. **(Problem 2.)** Using the Logit model estimates, the effect on the probability to participate in sports, of a 10 year increase in education (ceteris paribus) for an individual with the mean characteristics in the sample, is

(a) 0.074.

(b) 0.74.

(c) 0.17.

(d) 0.017.

30. **(Problem 2.)** Using the Logit model estimates,

(i) For some females it is possible that the predicted probability to do sports is negative.

(ii) The difference in the probability to do sports between a male and a female is always 0.29 in favour of the male, independently of his age.

(iii) The magnitude of the partial effect of years of education on the probability to do sports, diminishes when *yedu* increases if *yedu* is sufficiently high, all else held equal.

(a) Only (*i*) and (*ii*) are true.

(b) Only (*ii*) and (*iii*) are true.

(c) Only (*i*) is true.

(d) Only (*iii*) is true.

31. **(Problem 2.)** Using the Logit model estimates, to test that the effect of education on the probability to do sports does not depend on gender:

(a) It is necessary to estimate a model that is inclusive of interactions of *yedu* with the variable $female$.

(b) I would have tested if the coefficients of $female$ and *yedu* are jointly significant with a likelihood ratio test.

(c) I would have used the log-likelihoods of this model and another which is not inclusive of variables $female$ and *yedu*.

(d) I would have tested if the coefficient on $female$ is significant using a t test.

32. **(Problem 2.)** Using the Logit model estimates,

(i) We can conclude that the explanatory variables aid in predicting the probability to do sports because the percent correctly predicted is fairly high.

(ii) We can reject the hypothesis that the coefficients on $female$, age, $age2$ and $yedu$ are equal to zero at any usual significance level.

(iii) The model is not properly specified since the pseudo-R^2 is very low.

(a) Only (i) and (iii) are true.

(b) Only (i) and (ii) are true.

(c) Only (ii) is true.

(d) Only (i) is true.

33. **(Problem 2.)** Using the Logit model estimates, and assuming that the coefficient on $age2$ is exactly zero,

(a) The partial effect on the probability to do sports of an additional year of education is equivalent to being 0,88 years younger, but only for females.

(b) The partial effect on the probability to do sports of an additional year of education is equivalent to being 0,88 years younger.

(c) The partial effect on the probability to do sports of an additional year of education is equivalent to being 0,88 years younger, but only for males.

(d) The partial effect on the probability to do sports of an additional year of education is equivalent to being 0,88 years younger, but only if all explanatory variables are evaluated at the sample means.

34. **(Problem 2.)** Using the Logit model estimates, the probability that a 30 year old female 30 with 20 years of education participates in sports is

(a) 0.29 less than that of a male with the mean characteristics in the sample.

(b) 0.29 less than that of a male with equal characteristics.

(c) 0.31 less than that of a male with equal characteristics.

(d) 0.27 less than that of a male with equal characteristics.

Solution to Exam Type: 4

Universidad Carlos III de Madrid

ECONOMETRICS I

Academic year 2006/07
FINAL EXAM

February 9, 2008

TIME: 2 HOURS 30 MINUTES

1. **(Problem 1.)** Regarding equation (1):

   (i) The expected difference in years of education between a child that grew up in a household with a good economic situation and one that grew up in a household with a very poor economic situation is given by $\beta_7 - \beta_0$, if we assume that all other factors are held constant.

   (ii) The expected difference in years of education between a child that studied in an urban area inside a household with a good economic situation and another that studied in a rural area inside a household with a very poor economic situation is given by $\beta_5 + \beta_7$, if we assume that all other factors are held constant.

   (iii) The expected difference in years of education between a child that grew up in a household with a good economic situation and a working mother and another that grew up in a household with a very poor economic situation, for which we do not know whether the mother was working or not, is given by $\beta_4 + \beta_7$, if we assume that all other factors are held constant.

   (a) Only (ii) and (iii) are true.
   (b) Only (i) is true.
   [(c)] Only (ii) is true.
   (d) All statements are false.

2. **(Problem 1.)** Consider the following statements with respect to Output 1:

   (i) We can conclude that the number of siblings has an effect on the total number of years of education of an individual, since $n1childg1$ is significant in Output 1.

   (ii) The endogeneity of $n1childg1$ affects the estimated standard errors of the OLS estimators in Output 1, but does not affect the consistency of the estimators.

   (iii) All estimators in Output 1 are consistent, except that of $n1childg1$, because it is endogenous.

   (a) Only (ii) and (iii) are true.
   (b) Only (i) is true.
   (c) Only (ii) is true.
   [(d)] All statements are false.

3. **(Problem 1.)** The validity of instruments $boy12$ and $girl12$,

   (i) Solely depends on the zero correlation between the instrument and the omitted factor in the equation.

   (ii) Can have the implication that the equation is over-identified.

   (iii) Guarantees that the OLS estimators are consistent.

   (a) (i) and (ii) are true.
   (b) Only (i) is true.
   [(c)] Only (ii) is true.
   (d) (i) and (iii) are true.

4. **(Problem 1.)** Consider the following statements:

   (i) The instruments have to be exogenous variables whose OLS coefficient in the original equation is not significant.

   (ii) The instruments can be correlated with the dependent variable.

   (iii) The instruments can be correlated with the exogenous variables.

   [(a)] (ii) and (iii) are true.
   (b) (i) and (iii) are true.
   (c) Only (ii) is true.
   (d) All statements are false.

5. **(Problem 1.)** Using instruments $boy12$ and $girl12$, and assuming that they are uncorrelated with u, we can conclude that:

   (a) Equation (1) is not identified because $boy12$ and $girl12$ are not jointly significant in the reduced form of $n1childg1$, since the value of the corresponding test statistic, 55.27, is not significant at the 1% level.

   (b) Equation (1) is identified because $boy12$ and $girl12$ are exogenous.

   [(c)] Equation (1) is identified because $boy12$ and $girl12$ are jointly significant in the reduced form of $n1childg1$, since the value of the corresponding test statistic, 13.71, is significant at the 5% level.

   (d) Equation (1) is not identified because $boy12$ is not significant in Output 2.

6. **(Problem 1.)** Assuming that the instruments $boy12$ and $girl12$ are uncorrelated with u, and using the relevant information from the Outputs,

   (i) The 2SLS estimation of Output 4 is more efficient than the 2SLS estimation using only $girl12$ because having more instruments is always better.

   (ii) The 2SLS estimation of Output 4, is not consistent because $boy12$ is not significant in Output 2.

   (iii) Since we know that $boy12$ is correlated with $n1childg1$, it is always best to use it as an instrument in the 2SLS estimation.

   (a) Only (i) and (iii) are true.

   (b) Only (i) is true.

   (c) Only (ii) is true.

   [(d)] All statements are false.

7. **(Problem 1.)** Consider the following statements:

   (i) Hausman's test compares two alternative forms to estimate the parameters of a model on the basis that, at least one of the two forms is not consistent under H_0.

   (ii) The estimators of Output 5 are consistent, but less efficient than those of Output 4 if equation(1) is exactly identified.

   (iii) The endogeneity test requires that the equation is over-identified.

   (a) Only (i) and (iii) are true.

   (b) Only (ii) is true.

   (c) Only (iii) is true.

   [(d)] All are false.

8. **(Problem 1.)** The results of Output 5:

   (i) Estimate the parameters by Two Stages Least Squares.

   (ii) Allow us to confirm the endogeneity of $n1childg1$ at the 1% significance level.

   (iii) Indicate that $n1childg1$ is uncorrelated with u in equation (1).

   (a) All statements are true.
   (b) Only (i) is true.
   (c) Only (iii) is true.
   [(d)] (i) and (ii) are true.

9. **(Problem 1.)** If we had at our disposal the residuals from a 2SLS estimation of equation (1) using as a sole instrument $girl2$, and the results of the OLS regression where those residuals are the dependent variable as a function of all the exogenous variables:

   (i) It would have been possible to test if $boy12$ is uncorrelated with u in equation (1).

   (ii) It would have been possible to test if $girl12$ is uncorrelated with u in equation (1).

   (iii) It wold have been possible to simultaneously test if $boy12$ and $girl12$ are uncorrelated with u in equation (1).

   (a) Only (iii) is true.
   [(b)] Only (i) is true.
   (c) Only (ii) is true.
   (d) All statements are false.

10. **(Problem 1.)** Using the appropriate outputs, we can conclude that:

    (i) The number of siblings has a negative effect on the individual's education.

    (ii) The economic situation of the family does not affect education because *pobre* is not significant in Output 1 at the 5% level.

    (iii) The number of siblings has a positive significant effect on *yedu* at the 10% level but not at the 5% level.

    (a) All statements are false.
    (b) Only (i) is true.
    (c) (i) and (ii) are true.
    [(d)] Only (iii) is true.

11. **(Problem 1.)**

   (i) The fact that the standard deviation of the residuals in Output 1 is lower than in Output 4 confirms that OLS is consistent.

   (ii) The R^2 in the 2SLS output has the same interpretation and use as in the OLS case.

   (iii) The asymptotic variance of the 2SLS estimators is always greater or equal to that of the OLS estimators.

   (a) Only (ii) is true.
   [(b)] (i) and (ii) are false.
   (c) (ii) and (iii) are true.
   (d) Only (ii) is false.

12. **(Problem 3.)** Out of the four suggested specifications:

   [(a)] The only specification that cannot be estimated by OLS is (4).
   (b) Only (2) and (3) can be estimated by OLS.
   (c) All can be estimated by OLS, but only (2) provides consistent estimators.
   (d) None of the specifications can be estimated by OLS because they omit relevant variables.

13. **(Problem 3.)** Consider the following statements with respect to the previous models:

   (i) It is possible that assumption $E(u|x) = 0$ where x is a vector that contains all the explanatory variables included in each equation, is fulfilled in some of these specifications since they consider single and married working mothers separately.

   (ii) In specification (2), parameter β_3 is interpreted as the difference between the wage rate of married working mothers and single females.

   (iii) There are problems of perfect multicollinearity in Model (2).

   (a) (i) and (iii) are false.
   [(b)] All statements are false.
   (c) (i) and (ii) are false.
   (d) (ii) and (iii) are false.

14. **(Problem 3.)** In specification (2), the impact of being married on the wage rate is:

   (a) $\beta_2 + \beta_3 c$.
   (b) β_2.
   (c) $\beta_0 + \beta_2 + \beta_3 m$.
   [(d)] $\beta_2 + \beta_3 m$.

15. **(Problem 3.)** Select the correct answer with respect to specification (1):

(a) All OLS estimators can be biased with a bias equal to $\delta_j x_1$ where δ_j are the OLS estimators from the regression of the included variables on the omitted variable.

(b) OLS will only provide an unbiased estimator of the constant term.

(c) All OLS estimators can be biased with a bias equal to $\delta_1 x_1$ where δ_1 is the OLS estimator from the regression of the wage rate w on the omitted variable.

[(d)] All OLS estimators can be biased with a bias equal to $\delta_j \beta_1$ where δ_j are the OLS estimators from the regression of the excluded variable on the included variable, and β_1 is the coefficient of the excluded variable.

16. **(Problem 3.)** We assume that the decision to become a working mother depends on the level of family income, that is why we suggest another equation:

$$m = \gamma_0 + \gamma_1 w + \gamma_2 sm + \gamma_3 c + \varepsilon \qquad (6)$$

where sm is the wage rate of the partner with whom she lives. It is assumed that $E(\varepsilon | sm, c) = 0$. Considering specification (2) for w, and that sm and c are exogenous in the aforementioned equation, $E(u|sm, c) = 0$, select the correct response:

(a) Both equations satisfy the rank and order conditions.

(b) Both equations, (2) and (6), satisfy the order condition but fail to satisfy the rank condition.

(c) Equation (2) satisfies the order condition, but equation (6) does not satisfy the order condition.

[(d)] None of the two equations satisfies the rank nor the order conditions, since in general we expect that $m \times c$ would also be endogenous in equation (2).

17. **(Problem 3.)** Consider the following statements referring to equation (6) for m :

[(a)] First, we need to estimate the reduced form of w by OLS, so that we are able to estimate the structural form by OLS replacing w by the estimated values in the first regression, but this is not possible since we have not got enough instruments.

(b) It is equivalent to estimating a Linear Probability Model by OLS, and hence there are no problems with the bias and inconsistency of the estimators.

(c) It is impossible to estimate by OLS given the presence of the evident simultaneity problems. It can only be estimated by maximum likelihood i.e. using a logit or probit model according to the assumptions about the errors distribution.

(d) We can only obtain efficient estimators of the model if instead of OLS we use FGLS (Feasible Generalised Least Squares) in the 2SLS estimation.

18. **(Problem 3.)** What would occur if instead of using m as a dependent variable we used $50m$ in (6)?

   (i) The only effect on the OLS estimators will be that all are multiplied by 50 as well. The coefficient of determination R^2 will not be affected.

   (ii) The sum of the squared residuals will be multiplied by 2.500.

   (iii) The OLS estimators will now be biased upwards by 50.

   [(a)] Only (i) and (ii) are true.
   (b) Only (i) is true.
   (c) Only (iii) is true.
   (d) All statements are true.

19. **(Problem 3.)** With respect to the assumption of heteroskedasticity in model (6) for m (and ignoring any **endogeneity** problems), consider the following statements:

   (i) We can estimate the equation by OLS, but for inference purposes we will use the robust t statistics, even though this estimations would not be efficient if heteroskedasticity is present.

   (ii) In this type of models there is no need to test for heteroskedasticity, so that it will no matter whether we use the usual standard errors or the corrected ones.

   (iii) We can estimate the model by Feasible Generalized LS, without the need to run in first place a regression for the squares of the OLS residuals.

   (a) Only (i) is true.
   (b) Only (i) and (ii) are true.
   [(c)] Only (i) and (iii) are true.
   (d) Only (ii) and(iii) are true.

20. **(Problem 3.)** Assuming that there are no simultaneity problems, but that w was measured with error in equation (6) such that:

$$w = w_{true} + e_1,$$

where e_1 denotes the measurement error and w_{true} is the true level of the wage rate.

   (i) The measurement error will always cause a decrease in the variance of the random disturbance.

   (ii) If $Cov\left(w, e_1\right) = 0$ holds, the OLS estimators will be unbiased and consistent.

   (iii) Under the CEV (Classical Errors in Variables) assumption, the OLS estimators will be unbiased, but with lower variance than without measurement error due to the attenuation bias.

   [(a)] Only (ii) is true.
   (b) Only (i) is true.
   (c) Only (ii) and (iii) are true.
   (d) Only (i) and (ii) are true.

21. **(Problem 3.)** We suspect that equation (6) is misspecified (assuming that there is no simultaneity problem). Consider the following statements:

   (i) If assumption $E\left(\varepsilon|x\right) = 0$ is satisfied, where x is a vector containing the explanatory variables, then none of the added non-linear functions of the $x's$ should be significant.

   (ii) The RESET test results indicate whether it is necessary to use heteroskedasticity robust standard errors.

   (iii) The RESET test does not inform us on how to proceed if the null hypothesis of correct specification is rejected.

   (a) Only (i) is true.
   [(b)] (i) and (iii) are true.
   (c) (ii) and (iii) are true.
   (d) All statements are true.

22. **(Problem 3.)** (*Continues*)

(i) An advantage of the RESET test is that it can compare models with different specifications of the dependent variable.

(ii) The disadvantage of the Mizon-Richard and Davidson-MacKinnon tests is that they can give inconclusive results regarding which model is preferable.

(iii) White's test has the capacity to detect more general specification errors compared to the Breusch-Pagan test.

[(a)] Only (*ii*) is true.
(b) Only (*i*) is true.
(c) Only (*i*) and (*iii*) are true.
(d) All statements are true.

23. **(Problem 2.)** In the Linear Probability Model,

(i) The error term would present heteroskedasticity (conditional on the explanatory variables).

(ii) The error term (conditional on the explanatory variables) would be normally distributed.

(iii) The predicted probabilities could be greater than one, but not lower than zero.

(a) Only (i) and (iii) are true.
(b) The three statements are true.
(c) Only (i) and (ii) are true.
[(d)] Only (i) is true.

24. **(Problem 2.)** Using the linear probability model estimates,

(a) The probability to participate in sports diminishes after the age of 39,5, approximately.
[(b)] The probability to participate in sports increases with age up to the age of 50, approximately.
(c) The probability to participate in sports diminishes with age after the age of 35, approximately.
(d) The probability to participate in sports always diminishes with age.

25. **(Problem 2.)** Using the linear probability model estimates, the probability that a 30 year old male with 10 years of education participates in sports is:

(a) 0.81.
[(b)] 0.52.
(c) 0.36.
(d) 0.94.

26. **(Problem 2.)** Using the linear probability model estimates, we can conclude that

(i) Females have a lower probability to participate in sports, holding all else equal.

(ii) Age does not affect the decision to participate in sports because the coefficient on $age2$ is not significant at the 5% level.

(iii) A male with 10 years of education has a higher probability to do sports than a female with 20 years of education, everything else being held equal.

[(a)] (i) and (iii) are true.
(b) Only (i) is true.
(c) (ii) is true.
(d) (ii) and (iii) are true.

27. **(Problem 2.)** The estimated linear probability model specification (Output 1),

(i) Permits measuring the effect of sex on the probability to do sports.

(ii) Permits that the partial effect of age varies with its value.

(iii) Permits that the effect of education on the probability to do sports varies by gender.

(a) Only (i) and (iii) are true.
(b) Only (i) is true.
[(c)] Only (i) and (ii) are true.
(d) Only (iii) is true.

28. **(Problem 2.)** In a Logit model,

(i) The partial effect of variable x_j on the probability that $y = 1$ (conditional on the explanatory variables), depends on β_j and x_j, but not on the rest of the explanatory variables or parameters.

(ii) The partial effect of variable x_j on the probability that $y = 1$ (conditional on the explanatory variables), depends on $\beta_0, \beta_1, \ldots, \beta_k$ and on x_j, but not on the rest of the explanatory variables.

(iii) The partial effect of variable x_j on the probability that $y = 1$ (conditional on the explanatory variables), depends on β_j and on $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_k$, but not on the rest of the parameters.

(a) Only (ii) is true.
[(b)] The three statements are false.
(c) Only (i) is true.
(d) Only (iii) is true.

29. **(Problem 2.)** Using the Logit model estimates, the effect on the probability to participate in sports, of a 10 year increase in education (ceteris paribus) for an individual with the mean characteristics in the sample, is

(a) 0.074.

(b) 0.74.

[(c)] 0.17.

(d) 0.017.

30. **(Problem 2.)** Using the Logit model estimates,

(i) For some females it is possible that the predicted probability to do sports is negative.

(ii) The difference in the probability to do sports between a male and a female is always 0.29 in favour of the male, independently of his age.

(iii) The magnitude of the partial effect of years of education on the probability to do sports, diminishes when *yedu* increases if *yedu* is sufficiently high, all else held equal.

(a) Only (*i*) and (*ii*) are true.

(b) Only (*ii*) and (*iii*) are true.

(c) Only (*i*) is true.

[(d)] Only (*iii*) is true.

31. **(Problem 2.)** Using the Logit model estimates, to test that the effect of education on the probability to do sports does not depend on gender:

(a) It is necessary to estimate a model that is inclusive of interactions of *yedu* with the variable $female$.

(b) I would have tested if the coefficients of $female$ and *yedu* are jointly significant with a likelihood ratio test.

(c) I would have used the log-likelihoods of this model and another which is not inclusive of variables $female$ and *yedu*.

[(d)] I would have tested if the coefficient on $female$ is significant using a t test.

32. **(Problem 2.)** Using the Logit model estimates,

(i) We can conclude that the explanatory variables aid in predicting the probability to do sports because the percent correctly predicted is fairly high.

(ii) We can reject the hypothesis that the coefficients on $female$, age, $age2$ and $yedu$ are equal to zero at any usual significance level.

(iii) The model is not properly specified since the pseudo-R^2 is very low.

(a) Only (i) and (iii) are true.
(b) Only (i) and (ii) are true.
[(c)] Only (ii) is true.
(d) Only (i) is true.

33. **(Problem 2.)** Using the Logit model estimates, and assuming that the coefficient on $age2$ is exactly zero,

(a) The partial effect on the probability to do sports of an additional year of education is equivalent to being 0,88 years younger, but only for females.
[(b)] The partial effect on the probability to do sports of an additional year of education is equivalent to being 0,88 years younger.
(c) The partial effect on the probability to do sports of an additional year of education is equivalent to being 0,88 years younger, but only for males.
(d) The partial effect on the probability to do sports of an additional year of education is equivalent to being 0,88 years younger, but only if all explanatory variables are evaluated at the sample means.

34. **(Problem 2.)** Using the Logit model estimates, the probability that a 30 year old female 30 with 20 years of education participates in sports is

(a) 0.29 less than that of a male with the mean characteristics in the sample.
(b) 0.29 less than that of a male with equal characteristics.
[(c)] 0.31 less than that of a male with equal characteristics.
(d) 0.27 less than that of a male with equal characteristics.