Universidad Carlos III de Madrid
<u>ECONOMETRICS I</u>
Academic year 2006/07
FINAL EXAM
September 6, 2008

Exam Type: 1

TIME: 2 HOURS 30 MINUTES

1. **(Problem 1.)** The endogeneity problem of regressors in linear models causes that the OLS estimates,

   (i) Are generally biased.

   (ii) Need to be compared with corrected standard errors (White).

   (iii) Are not efficient, though in general are always consistent.

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

2. **(Problem 1.)** In a model of two simultaneous equations,

   (i) It is only necessary to estimate one estructural equation, since in general the dependent variable of the second equation show up as a regressor in the first one.

   (ii) None of the two equations can describe a causal relationship.

   (iii) If one estructural equation is identified, the other equation, and, therefore, the system are identified.

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

3. **(Problem 1.)** An instrumental variable for an equation,

   (i) It is an exogenous regressor in this equation.

   (ii) It is a dependent variable from other equation.

   (iii) Has not to be correlated with the error term.

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

4. **(Problem 1.)** The fact that the variable *open* is endogenoeus in the first equation,

   (i) Depends on whether γ_{21} is different from zero or not.

   (ii) Depends on whether the error terms u_1 and u_2 are correlated or not.

   (iii) Depends on whether δ_{22} is different from zero or not.

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

5. **(Problem 1.)** Assuming that *open* is endogenous, is the first equation identified?

   (a) It is identified, because $\log(land)$ is significative in the reduced form of *open*.
   (b) It is identified because the F is significative in Output 4.
   (c) It is identified because $\log(land)$ is significative in Output 2.
   (d) It is not identified because $\log(pcinc)$ is not significative in the reduced form of *open*.

6. **(Problem 1.)** Assuming that inf is endogenous, study the identification of the second equation:

   (i) It is identified because it has an additional exogenous variable included.

   (ii) It is not identified because $\log(pcinc)$ is not significative in Output 1.

   (iii) It is identified because $\log(land)$ is significative in Output 4.

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

7. **(Problem 1.)** We are interested in obtaining consistent estimates of the first equation, assuming that *open* is endogenous and that the variables $\log(pcinc)$ and $\log(land))$ are exogenous.

   (a) The estimates in Output 1 are consistent because the equation is identified.

   (b) The estimates in Output 5 are not consistent because $\log(pcinc)$ is not partially correlated with the explanatory variable *open* (Output 4).

   (c) The estimates in Output 5 are consistent because $\log(land)$ is partially correlated with the explanatory variable *open* (Output 4).

   (d) The estimates in Output 5 are consistent because $\log(pcinc)$ and $\log(land)$ are jointly significative in Output 4.

8. **(Problem 1.)** Given the empirical evidence, can be concluded that γ_{12} is different from zero?

   (i) Yes, attending to the estimation in Output 1, because OLS is always valid if the equation is identified.

   (ii) Yes, attending to the instrumental variables estimation.

   (iii) We do not know, because the first equation can not be identified in any case.

   (a) Only (ii) is true.

   (b) All are true.

   (c) Only (iii) is true.

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

9. **(Problem 1.)** Taking into account the result of Hausman test (Output 5) and any other relevant information, we can conclude that,

   (i) The estimates in Output 1 are consistent.

   (ii) The estimates in Output 5 are more efficient than those in Output 1.

   (iii) There is no empirical evidence to conclude that the variable *open* is endogenous in the first equation.

   (a) All are true.

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

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

   (d) Only (i) is true.

10. **(Problem 1.)** The result of Hausman test (Output 5):

   (i) Shows that the first equation is over-identified because the OLS estimates are consistent.

   (ii) Needs the equation to be over-identified to make sense.

   (iii) Shows that the estimation in Output 5 is not consistent.

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

11. **(Problem 1.)** Concerning the estimation of a system of two simultaneous equations by 2SLS,

   (i) The first step in the 2SLS to estimate one equation looks for an optimal instrumental variable when this equation is over-identified, but if the other equation is not identified, the estimation is not consistent.

   (ii) The two steps in 2SLS can be done by means of OLS regressions.

   (iii) The second step in the 2SLS method can be done as an instrumental variable regression, using as instrument only the optimal instrumental variable from the first step.

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

12. **(Problem 1.)** If we have available one additional valid instrument for *open* in the first equation,

   (i) The 2SLS estimation would be more efficient than that of Output 5 if we additionally assume that $\log(land)$ is also a valid instrument.

   (ii) We could test whether $\log(land)$ is uncorrelated with u_1.

   (iii) We would not need observations of $\log(land)$ to obtain consistent estimates of the parameters of the first equation.

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

13. **(Problem 1.)** If we include the residuals from Output 4 as an additional regressor in the OLS estimation of the first equation (Output 1), and assuming that this equation is identified:

   (i) The estimates of the coefficients of *open* and of *lpcinc* would be consistent.

   (ii) The estimates of the coefficients of *open* and of *lpcinc* would be consistent, but less efficient that those from Output 5.

   (iii) We could do an endogeneity test for *open* by means of a t test.

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

14. **(Problem 1.)** If we reject the null hypothesis of the endogeneity test,

   (i) We conclude that the regressor is endogenous, and therefore the OLS estimates are inconsistent.

   (ii) We conclude that the regressor is exogenous, and therefore the instrumental variable estimates are less efficient that the OLS estimates.

   (iii) We conclude that the regressor is exogenous, but that the OLS estimates could be inconsistent.

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

15. **(Problem 1.)** In case of over-identification,

   (i) The equation can not be estimated because the parameters take too many alternative values.

   (ii) We can not define an efficient instrumental variables estimate.

   (iii) We can not do an exogeneity test, because we do not have enough degrees of freedom.

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

16. **(Problem 1.)** The efficiency of the 2SLS,

(i) Increases with the correlation between the endogenous regressor and the instrumental variables, measured through an R^2 coefficient.

(ii) Decreases with the variability of the endogenous regressor, since instrumental variable methods are less efficient than OLS.

(iii) Does not depend on the variance of the error from the original equation, since 2SLS is based on two different regressions.

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

17. **(Problem 1.)** The weak instruments case,

(i) May arise only when we have an over-identified equation, because each instrument can only explain an small proportion of the variability of the endogenous regressor.

(ii) Arises when the instrument and the endogenous regressor have a low correlation, leading to inefficient instrumental variable estimates.

(iii) May cause serious problems of inconsistency of the instrumental variables estimate if the instrument is correlated, though minimally, with the error term.

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

18. **(Problem 2.)** In the problem discussed,

(i) It is not a good idea to compare two alternative models with the determination coefficient, R^2, since we know in advance that (2) will always fit better because it contains more regressors.

(ii) If all coefficients in the OLS regression are significant, the model is correctly specified.

(iii) Chose the right model is very important, since otherwise some of the necessary hypothesis to perform valid inference could be not valid.

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

19. **(Problem 2.)** To choose between the two models, the researcher estimates the following equation:

$$stndfnl = \beta_0 + \beta_1 atndrte + \beta_2 \log(priGPA) + \beta_3 \log(ACT) + \\ + \gamma_1 priGPA + \gamma_2 ACT + \gamma_3 priGPA^2 + \gamma_4 ACT^2 + u \quad (3)$$

If she wants to test whether model (2) is correct, which is the null hypothesis that she should test for in equation (3)?

(a) $H_0 : \beta_1 = \beta_2 = \beta_3 = 0.$

(b) $H_0 : \gamma_1 = \gamma_2 = 0.$

(c) $H_0 : \gamma_1 = \gamma_2 = \gamma_3 = \gamma_4 = 0.$

(d) $H_0 : \beta_2 = \beta_3 = 0.$

20. **(Problem 2.)** The F statistic to test H_0 in the previous question takes the value 1,00081 with a p-value of 0.368.

(a) The researcher rejects H_0 and decides that model (2) is correct.

(b) The researcher decides that both models are correct.

(c) The researcher rejects H_0 and decides that model (1) is correct.

(d) The researcher can not reject H_0 and decides that model (2) is correct.

21. **(Problem 2.)** To choose between both models, the researcher considers an alternative method estimating the following equation,

$$stndfnl = \beta_0 + \beta_1 atndrte + \beta_2 \log(priGPA) + \beta_3 \log(ACT) + \theta_1 \widehat{y} + u,$$

where $\widehat{y}$ are the adjusted values from the estimation of model (2).

(i) If we reject $H_0 : \theta_1 = 0$, then we conclude that model (1) has specification problems.

(ii) If we reject $H_0 : \theta_1 = 0$, then we conclude that model (2) has specification problems.

(iii) If we reject $H_0 : \theta_1 = 0$, then we conclude that model (2) is correctly specified.

(a) Only (i) is true.

(b) Only (ii) is true.

(c) All are false.

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

22. **(Problem 2.)** It seems that the researcher prefers model (2), so she continues testing model (2). She estimates model (2) by OLS and obtains $R^2 = 0.226$. Then, she omits from the model the explanatory variables $priGPA^2$ and ACT^2, and obtains $R^2 = 0.211$. Which of the following conclusions is true?

    (a) Model (2) will be much better than the model without the explanatory variables $priGPA^2$ and ACT^2 since it has a higher R^2.

    (b) The test is not valid to check the specification of the model because the previous hypotheses are not nested.

    (c) The F statistic for the joint significance test of the two variables $priGPA^2$ and ACT^2 is 11.081 with a p-value of 0.000018, so that there is no evidence of wrong functional specification in model (2).

    (d) The F statistic for the joint significance test of the two variables $priGPA^2$ and ACT^2 is 11.081 with a p-value of 0.000018, so that there is evidence of wrong functional specification in model (2).

23. **(Problem 2.)** The researcher considers an alternative method: Ramsey's RESET test.

    (i) RESET test uses the OLS residuals and its squares to check if they help to explain the dependent variable.

    (ii) If we reject the null hypothesis of the RESET test, we confirm the correct specification of the model.

    (iii) RESET test allows us also to test for the presence of heteroskedasticity.

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

    (b) Only (ii) is true.

    (c) All are false.

    (d) Only (iii) is true.

24. **(Problem 2.)** If Model (2) is correctly specified, and u satisfies all usual assumptions, but we do not include the variable ACT^2 in the OLS estimation,

    (i) The OLS estimates of the remaining parameters will be biased if $\beta_5 \neq 0$, because at least ACT and ACT^2 will be correlated.

    (ii) The OLS estimation of the remaining parameters will be consistent, but the new errors will be heteroskedastic and usual inference will not be valid, if $\beta_5 \neq 0$.

    (iii) OLS estimation of the remaining parameters, except β_3, will be consistent if $\beta_5 = 0$, since ACT and ACT^2 will be correlated.

    (a) Only (ii) is true.

    (b) Only (iii) is true.

    (c) All are false.

    (d) Only (i) is true.

25. **(Problem 2.)** In the next step of the empirical analysis, the researcher wishes to test for the presence of heteroskedasticity in the model using White's test. Assume that given the results of previous tests, she decides to focus on model (2). How is White's test implemented?

   (a) In model (2) we add the squares of the explanatory variables and then we calculate the F test for global significance.

   (b) We calculate the residuals of the OLS fit of model (2). We estimate the regression of the squares of the OLS residuals on the squares of all explanatory variables of model (2). We obtain the R^2 of this estimation and we compute the test statistic nR^2.

   (c) We calculate the residuals of the OLS fit of model (2). We estimate the regression of the squares of the OLS residuals on all explanatory variables of model (2), their squares, and all cross products. We obtain the R^2 of this estimation and we compute the test statistic nR^2.

   (d) In model (2) we add the squares of the explanatory variables and all cross products and then we calculate the F statistic for global significance.

26. **(Problem 2.)** The result of White's heteroskedasticity test is 13.993, with a p-value of 0.72. Which is the conclusion of the researcher on the null hypothesis of White's test, and therefore, on the presence of heteroskedasticity in model (2) given the empirical evidence?

   (a) She rejects H_0 and concludes that there is heteroskedasticity.

   (b) She rejects H_0 and concludes that there is not heteroskedasticity.

   (c) She can not reject H_0 and concludes that there is not heteroskedasticity.

   (d) She can not reject H_0 and concludes that there is heteroskedasticity.

27. **(Problem 2.)** In the presence of heteroskedasticity,

   (i) OLS estimates stop being consistent.

   (ii) To perform valid inference is always necessary to estimate the equation by means of generalized least squares.

   (iii) White's corrected standard errors are only valid if we know the form of the heteroskedasticity.

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

   (b) Only (ii) is true.

   (c) All are false.

   (d) Only (iii) is true.

28. **(Problem 2.)** Now assume that the researcher decides to include the interaction term $priGPA * atndrte$ in model (2):

$$\begin{aligned} stndfnl = &\beta_0 + \beta_1 atndrte + \beta_2 priGPA + \beta_3 ACT + \beta_4 priGPA^2 + \\ &+ \beta_5 ACT^2 + \beta_6 priGPA * atndrte + u. \end{aligned} \tag{4}$$

The results of the OLS estimation of this new model are as follows,

$$\begin{aligned} \widehat{stndfnl} = &\underset{(1.36)}{2.05} - \underset{(0.0102)}{0.0067} atndrte - \underset{(0.48)}{1.63} priGPA - \underset{(0.098)}{0.128} ACT + \underset{(0.101)}{0.296} priGPA^2 + \\ &+ \underset{(0.0022)}{0.0045} ACT^2 + \underset{(??)}{0.0056} priGPA * atndrte \\ &n = 680 \qquad R^2 = 0.228 \end{aligned}$$

(a) The effect of the interaction is without doubt significative, since the R^2 has increased.

(b) There is not enough information to test for the significance of the new estimate.

(c) The t test for the coefficient of the interaction is equal to 2.13, and therefore the estimate is significative.

(d) The t test for the coefficient of the interaction is equal to 1.32, and therefore the estimate is not significative.

29. **(Problem 2.)** The researcher wishes to calculate the partial effect of attendance to class (*atndrte*) on the result of the final exam (*stndfnl*). How would you estimate that effect?

(a) Is simply the estimated coefficient $\widehat{\beta}_1$.

(b) $\widehat{\beta}_1 + \widehat{\beta}_2 + \widehat{\beta}_6 priGPA$.

(c) $\widehat{\beta}_1 + \widehat{\beta}_6$.

(d) $\widehat{\beta}_1 + \widehat{\beta}_6 priGPA$.

30. **(Problem 2.)** If the averaged value of *priGPA* in the sample is 2.59, how would affect an increment of 10 percentage points of class attendance on the result of the final exam for a student with this value of *priGPA*?

(a) An increment of 10 percentage points in the class attendance leads to an increment of *stndfnl* of 0.078.

(b) An increment of 10 percentage points in the class attendance leads to an increment of *stndfnl* of 1.622.

(c) An increment of 10 percentage points in the class attendance leads to a decrement of *stndfnl* of 0.067.

(d) An increment of 10 percentage points in the class attendance leads to an increment of *stndfnl* of 0.0078.

31. **(Problem 2.)** The standard error of estimating the previous effect is

   (a) 0.15 if the OLS estimates of the coefficients were uncorrelated.
   (b) 0.102.
   (c) There is no enough available information to calculate it.
   (d) 0.25 if the OLS estimates of the coefficients were uncorrelated.

32. **(Problem 2.)** Using Model (4), which is the partial effect of the grade of previous exams (*priGPA*) over the final exam grading (*stndfnl*), when we fix $priGPA = 2.59$ and $atndrte = 0.82$?

   (a) -0.092.
   (b) 1.537.
   (c) -0.863.
   (d) -0.858.

33. **(Problem 2.)** What is like the effect of the average grade in the university access test (ACT) over the result of the final exam, ceteris paribus?

   (a) The effect would be negative because the estimated effect of the variable ACT, $\widehat{\beta}_3$, is negative.
   (b) If would depend on the values of the variable ACT.
   (c) The effect would be positive because the estimated coefficient of the variable ACT^2, $\widehat{\beta}_5$, is positive and since the variable ACT appears as a square, the sign of this variable will prevail.
   (d) The effect would be negative because, despite the estimated coefficient of the variable ACT^2 ($\widehat{\beta}_5$) being positive, it is smaller in comparison with the estimated coefficient $\widehat{\beta}_3$, which is negative.

34. **(Problem 2.)** From which value of the grading on, having an extra point in the access to university test, ACT, increases the result of the standardized final test?

   (a) From a grading of approximately 14.22 points.
   (b) From a grading of approximately 0.45 points.
   (c) From a grading of approximately 12.8 points.
   (d) From a grading of approximately 28.44 points.

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 1.)** The endogeneity problem of regressors in linear models causes that the OLS estimates,

   (i) Are generally biased.

   (ii) Need to be compared with corrected standard errors (White).

   (iii) Are not efficient, though in general are always consistent.

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

2. **(Problem 1.)** In a model of two simultaneous equations,

   (i) It is only necessary to estimate one estructural equation, since in general the dependent variable of the second equation show up as a regressor in the first one.

   (ii) None of the two equations can describe a causal relationship.

   (iii) If one estructural equation is identified, the other equation, and, therefore, the system are identified.

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

3. **(Problem 1.)** An instrumental variable for an equation,

   (i) It is an exogenous regressor in this equation.

   (ii) It is a dependent variable from other equation.

   (iii) Has not to be correlated with the error term.

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

4. **(Problem 1.)** The fact that the variable *open* is endogenoeus in the first equation,

   (i) Depends on whether γ_{21} is different from zero or not.

   (ii) Depends on whether the error terms u_1 and u_2 are correlated or not.

   (iii) Depends on whether δ_{22} is different from zero or not.

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

5. **(Problem 1.)** Assuming that *open* is endogenous, is the first equation identified?

   [(a)] It is identified, because $\log(land)$ is significative in the reduced form of *open*.
   (b) It is identified because the F is significative in Output 4.
   (c) It is identified because $\log(land)$ is significative in Output 2.
   (d) It is not identified because $\log(pcinc)$ is not significative in the reduced form of *open*.

6. **(Problem 1.)** Assuming that inf is endogenous, study the identification of the second equation:

   (i) It is identified because it has an additional exogenous variable included.

   (ii) It is not identified because $\log(pcinc)$ is not significative in Output 1.

   (iii) It is identified because $\log(land)$ is significative in Output 4.

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

7. **(Problem 1.)** We are interested in obtaining consistent estimates of the first equation, assuming that *open* is endogenous and that the variables $\log(pcinc)$ and $\log(land))$ are exogenous.

   (a) The estimates in Output 1 are consistent because the equation is identified.

   (b) The estimates in Output 5 are not consistent because $\log(pcinc)$ is not partially correlated with the explanatory variable *open* (Output 4).

   [(c)] The estimates in Output 5 are consistent because $\log(land)$ is partially correlated with the explanatory variable *open* (Output 4).

   (d) The estimates in Output 5 are consistent because $\log(pcinc)$ and $\log(land)$ are jointly significative in Output 4.

8. **(Problem 1.)** Given the empirical evidence, can be concluded that γ_{12} is different from zero?

   (i) Yes, attending to the estimation in Output 1, because OLS is always valid if the equation is identified.

   (ii) Yes, attending to the instrumental variables estimation.

   (iii) We do not know, because the first equation can not be identified in any case.

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

   (b) All are true.

   (c) Only (iii) is true.

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

9. **(Problem 1.)** Taking into account the result of Hausman test (Output 5) and any other relevant information, we can conclude that,

   (i) The estimates in Output 1 are consistent.

   (ii) The estimates in Output 5 are more efficient than those in Output 1.

   (iii) There is no empirical evidence to conclude that the variable *open* is endogenous in the first equation.

   (a) All are true.

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

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

   (d) Only (i) is true.

10. **(Problem 1.)** The result of Hausman test (Output 5):

    (i) Shows that the first equation is over-identified because the OLS estimates are consistent.

    (ii) Needs the equation to be over-identified to make sense.

    (iii) Shows that the estimation in Output 5 is not consistent.

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

11. **(Problem 1.)** Concerning the estimation of a system of two simultaneous equations by 2SLS,

    (i) The first step in the 2SLS to estimate one equation looks for an optimal instrumental variable when this equation is over-identified, but if the other equation is not identified, the estimation is not consistent.

    (ii) The two steps in 2SLS can be done by means of OLS regressions.

    (iii) The second step in the 2SLS method can be done as an instrumental variable regression, using as instrument only the optimal instrumental variable from the first step.

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

12. **(Problem 1.)** If we have available one additional valid instrument for *open* in the first equation,

    (i) The 2SLS estimation would be more efficient than that of Output 5 if we additionally assume that $\log(land)$ is also a valid instrument.

    (ii) We could test whether $\log(land)$ is uncorrelated with u_1.

    (iii) We would not need observations of $\log(land)$ to obtain consistent estimates of the parameters of the first equation.

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

13. **(Problem 1.)** If we include the residuals from Output 4 as an additional regressor in the OLS estimation of the first equation (Output 1), and assuming that this equation is identified:

(i) The estimates of the coefficients of *open* and of *lpcinc* would be consistent.

(ii) The estimates of the coefficients of *open* and of *lpcinc* would be consistent, but less efficient that those from Output 5.

(iii) We could do an endogeneity test for *open* by means of a t test.

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

14. **(Problem 1.)** If we reject the null hypothesis of the endogeneity test,

(i) We conclude that the regressor is endogenous, and therefore the OLS estimates are inconsistent.

(ii) We conclude that the regressor is exogenous, and therefore the instrumental variable estimates are less efficient that the OLS estimates.

(iii) We conclude that the regressor is exogenous, but that the OLS estimates could be inconsistent.

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

15. **(Problem 1.)** In case of over-identification,

(i) The equation can not be estimated because the parameters take too many alternative values.

(ii) We can not define an efficient instrumental variables estimate.

(iii) We can not do an exogeneity test, because we do not have enough degrees of freedom.

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

16. **(Problem 1.)** The efficiency of the 2SLS,

(i) Increases with the correlation between the endogenous regressor and the instrumental variables, measured through an R^2 coefficient.

(ii) Decreases with the variability of the endogenous regressor, since instrumental variable methods are less efficient than OLS.

(iii) Does not depend on the variance of the error from the original equation, since 2SLS is based on two different regressions.

(a) Only (ii) is true.

(b) All are true.

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

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

17. **(Problem 1.)** The weak instruments case,

(i) May arise only when we have an over-identified equation, because each instrument can only explain an small proportion of the variability of the endogenous regressor.

(ii) Arises when the instrument and the endogenous regressor have a low correlation, leading to inefficient instrumental variable estimates.

(iii) May cause serious problems of inconsistency of the instrumental variables estimate if the instrument is correlated, though minimally, with the error term.

(a) Only (ii) is true.

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

[(c)] (ii) and (iii) are true.

(d) Only (i) is true.

18. **(Problem 2.)** In the problem discussed,

(i) It is not a good idea to compare two alternative models with the determination coefficient, R^2, since we know in advance that (2) will always fit better because it contains more regressors.

(ii) If all coefficients in the OLS regression are significant, the model is correctly specified.

(iii) Chose the right model is very important, since otherwise some of the necessary hypothesis to perform valid inference could be not valid.

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

(b) All are true.

[(c)] Only (iii) is true.

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

19. **(Problem 2.)** To choose between the two models, the researcher estimates the following equation:

$$stndfnl = \beta_0 + \beta_1 atndrte + \beta_2 \log(priGPA) + \beta_3 \log(ACT) + \\ + \gamma_1 priGPA + \gamma_2 ACT + \gamma_3 priGPA^2 + \gamma_4 ACT^2 + u \quad (3)$$

If she wants to test whether model (2) is correct, which is the null hypothesis that she should test for in equation (3)?

(a) $H_0 : \beta_1 = \beta_2 = \beta_3 = 0.$

(b) $H_0 : \gamma_1 = \gamma_2 = 0.$

(c) $H_0 : \gamma_1 = \gamma_2 = \gamma_3 = \gamma_4 = 0.$

[(d)] $H_0 : \beta_2 = \beta_3 = 0.$

20. **(Problem 2.)** The F statistic to test H_0 in the previous question takes the value 1,00081 with a p-value of 0.368.

(a) The researcher rejects H_0 and decides that model (2) is correct.

(b) The researcher decides that both models are correct.

(c) The researcher rejects H_0 and decides that model (1) is correct.

[(d)] The researcher can not reject H_0 and decides that model (2) is correct.

21. **(Problem 2.)** To choose between both models, the researcher considers an alternative method estimating the following equation,

$$stndfnl = \beta_0 + \beta_1 atndrte + \beta_2 \log(priGPA) + \beta_3 \log(ACT) + \theta_1 \widehat{y} + u,$$

where $\widehat{y}$ are the adjusted values from the estimation of model (2).

(i) If we reject $H_0 : \theta_1 = 0$, then we conclude that model (1) has specification problems.

(ii) If we reject $H_0 : \theta_1 = 0$, then we conclude that model (2) has specification problems.

(iii) If we reject $H_0 : \theta_1 = 0$, then we conclude that model (2) is correctly specified.

[(a)] Only (i) is true.

(b) Only (ii) is true.

(c) All are false.

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

22. **(Problem 2.)** It seems that the researcher prefers model (2), so she continues testing model (2). She estimates model (2) by OLS and obtains $R^2 = 0.226$. Then, she omits from the model the explanatory variables $priGPA^2$ and ACT^2, and obtains $R^2 = 0.211$. Which of the following conclusions is true?

   (a) Model (2) will be much better than the model without the explanatory variables $priGPA^2$ and ACT^2 since it has a higher R^2.

   (b) The test is not valid to check the specification of the model because the previous hypotheses are not nested.

   [(c)] The F statistic for the joint significance test of the two variables $priGPA^2$ and ACT^2 is 11.081 with a p-value of 0.000018, so that there is no evidence of wrong functional specification in model (2).

   (d) The F statistic for the joint significance test of the two variables $priGPA^2$ and ACT^2 is 11.081 with a p-value of 0.000018, so that there is evidence of wrong functional specification in model (2).

23. **(Problem 2.)** The researcher considers an alternative method: Ramsey's RESET test.

   (i) RESET test uses the OLS residuals and its squares to check if they help to explain the dependent variable.

   (ii) If we reject the null hypothesis of the RESET test, we confirm the correct specification of the model.

   (iii) RESET test allows us also to test for the presence of heteroskedasticity.

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

   (b) Only (ii) is true.

   [(c)] All are false.

   (d) Only (iii) is true.

24. **(Problem 2.)** If Model (2) is correctly specified, and u satisfies all usual assumptions, but we do not include the variable ACT^2 in the OLS estimation,

   (i) The OLS estimates of the remaining parameters will be biased if $\beta_5 \neq 0$, because at least ACT and ACT^2 will be correlated.

   (ii) The OLS estimation of the remaining parameters will be consistent, but the new errors will be heteroskedastic and usual inference will not be valid, if $\beta_5 \neq 0$.

   (iii) OLS estimation of the remaining parameters, except β_3, will be consistent if $\beta_5 = 0$, since ACT and ACT^2 will be correlated.

   (a) Only (ii) is true.

   (b) Only (iii) is true.

   (c) All are false.

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

25. **(Problem 2.)** In the next step of the empirical analysis, the researcher wishes to test for the presence of heteroskedasticity in the model using White's test. Assume that given the results of previous tests, she decides to focus on model (2). How is White's test implemented?

   (a) In model (2) we add the squares of the explanatory variables and then we calculate the F test for global significance.

   (b) We calculate the residuals of the OLS fit of model (2). We estimate the regression of the squares of the OLS residuals on the squares of all explanatory variables of model (2). We obtain the R^2 of this estimation and we compute the test statistic nR^2.

   [(c)] We calculate the residuals of the OLS fit of model (2). We estimate the regression of the squares of the OLS residuals on all explanatory variables of model (2), their squares, and all cross products. We obtain the R^2 of this estimation and we compute the test statistic nR^2.

   (d) In model (2) we add the squares of the explanatory variables and all cross products and then we calculate the F statistic for global significance.

26. **(Problem 2.)** The result of White's heteroskedasticity test is 13.993, with a p-value of 0.72. Which is the conclusion of the researcher on the null hypothesis of White's test, and therefore, on the presence of heteroskedasticity in model (2) given the empirical evidence?

   (a) She rejects H_0 and concludes that there is heteroskedasticity.

   (b) She rejects H_0 and concludes that there is not heteroskedasticity.

   [(c)] She can not reject H_0 and concludes that there is not heteroskedasticity.

   (d) She can not reject H_0 and concludes that there is heteroskedasticity.

27. **(Problem 2.)** In the presence of heteroskedasticity,

   (i) OLS estimates stop being consistent.

   (ii) To perform valid inference is always necessary to estimate the equation by means of generalized least squares.

   (iii) White's corrected standard errors are only valid if we know the form of the heteroskedasticity.

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

   (b) Only (ii) is true.

   [(c)] All are false.

   (d) Only (iii) is true.

28. **(Problem 2.)** Now assume that the researcher decides to include the interaction term $priGPA * atndrte$ in model (2):

$$stndfnl = \beta_0 + \beta_1 atndrte + \beta_2 priGPA + \beta_3 ACT + \beta_4 priGPA^2 + \\ + \beta_5 ACT^2 + \beta_6 priGPA * atndrte + u. \quad (4)$$

The results of the OLS estimation of this new model are as follows,

$$\begin{aligned} \widehat{stndfnl} = & \underset{(1.36)}{2.05} - \underset{(0.0102)}{0.0067} atndrte - \underset{(0.48)}{1.63} priGPA - \underset{(0.098)}{0.128} ACT + \underset{(0.101)}{0.296} priGPA^2 + \\ & + \underset{(0.0022)}{0.0045} ACT^2 + \underset{(??)}{0.0056} priGPA * atndrte \\ & n = 680 \qquad R^2 = 0.228 \end{aligned}$$

(a) The effect of the interaction is without doubt significative, since the R^2 has increased.

(b) There is not enough information to test for the significance of the new estimate.

(c) The t test for the coefficient of the interaction is equal to 2.13, and therefore the estimate is significative.

[(d)] The t test for the coefficient of the interaction is equal to 1.32, and therefore the estimate is not significative.

29. **(Problem 2.)** The researcher wishes to calculate the partial effect of attendance to class (*atndrte*) on the result of the final exam (*stndfnl*). How would you estimate that effect?

(a) Is simply the estimated coefficient $\widehat{\beta}_1$.

(b) $\widehat{\beta}_1 + \widehat{\beta}_2 + \widehat{\beta}_6 priGPA$.

(c) $\widehat{\beta}_1 + \widehat{\beta}_6$.

[(d)] $\widehat{\beta}_1 + \widehat{\beta}_6 priGPA$.

30. **(Problem 2.)** If the averaged value of *priGPA* in the sample is 2.59, how would affect an increment of 10 percentage points of class attendance on the result of the final exam for a student with this value of *priGPA*?

[(a)] An increment of 10 percentage points in the class attendance leads to an increment of *stndfnl* of 0.078.

(b) An increment of 10 percentage points in the class attendance leads to an increment of *stndfnl* of 1.622.

(c) An increment of 10 percentage points in the class attendance leads to a decrement of *stndfnl* of 0.067.

(d) An increment of 10 percentage points in the class attendance leads to an increment of *stndfnl* of 0.0078.

31. **(Problem 2.)** The standard error of estimating the previous effect is

- [(a)] 0.15 if the OLS estimates of the coefficients were uncorrelated.
- (b) 0.102.
- (c) There is no enough available information to calculate it.
- (d) 0.25 if the OLS estimates of the coefficients were uncorrelated.

32. **(Problem 2.)** Using Model (4), which is the partial effect of the grade of previous exams (*priGPA*) over the final exam grading (*stndfnl*), when we fix $priGPA = 2.59$ and $atndrte = 0.82$?

- (a) -0.092.
- [(b)] 1.537.
- (c) -0.863.
- (d) -0.858.

33. **(Problem 2.)** What is like the effect of the average grade in the university access test (ACT) over the result of the final exam, ceteris paribus?

- (a) The effect would be negative because the estimated effect of the variable ACT, $\widehat{\beta_3}$, is negative.
- [(b)] If would depend on the values of the variable ACT.
- (c) The effect would be positive because the estimated coefficient of the variable ACT^2, $\widehat{\beta_5}$, is positive and since the variable ACT appears as a square, the sign of this variable will prevail.
- (d) The effect would be negative because, despite the estimated coefficient of the variable ACT^2 ($\widehat{\beta_5}$) being positive, it is smaller in comparison with the estimated coefficient $\widehat{\beta_3}$, which is negative.

34. **(Problem 2.)** From which value of the grading on, having an extra point in the access to university test, ACT, increases the result of the standardized final test?

- [(a)] From a grading of approximately 14.22 points.
- (b) From a grading of approximately 0.45 points.
- (c) From a grading of approximately 12.8 points.
- (d) From a grading of approximately 28.44 points.

Universidad Carlos III de Madrid
ECONOMETRICS I
Academic year 2006/07
FINAL EXAM
September 6, 2008

Exam Type: 2

TIME: 2 HOURS 30 MINUTES

1. **(Problem 1.)** The endogeneity problem of regressors in linear models causes that the OLS estimates,

   (i) Are generally biased.

   (ii) Need to be compared with corrected standard errors (White).

   (iii) Are not efficient, though in general are always consistent.

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

2. **(Problem 1.)** In a model of two simultaneous equations,

   (i) It is only necessary to estimate one estructural equation, since in general the dependent variable of the second equation show up as a regressor in the first one.

   (ii) None of the two equations can describe a causal relationship.

   (iii) If one estructural equation is identified, the other equation, and, therefore, the system are identified.

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

3. **(Problem 1.)** An instrumental variable for an equation,

   (i) It is an exogenous regressor in this equation.

   (ii) It is a dependent variable from other equation.

   (iii) Has not to be correlated with the error term.

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

4. **(Problem 1.)** The fact that the variable *open* is endogenoeus in the first equation,

   (i) Depends on whether γ_{21} is different from zero or not.

   (ii) Depends on whether the error terms u_1 and u_2 are correlated or not.

   (iii) Depends on whether δ_{22} is different from zero or not.

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

5. **(Problem 1.)** Assuming that *open* is endogenous, is the first equation identified?

   (a) It is not identified because $\log(pcinc)$ is not significative in the reduced form of *open*.
   (b) It is identified because $\log(land)$ is significative in Output 2.
   (c) It is identified because the F is significative in Output 4.
   (d) It is identified, because $\log(land)$ is significative in the reduced form of *open*.

6. **(Problem 1.)** Assuming that inf is endogenous, study the identification of the second equation:

   (i) It is identified because it has an additional exogenous variable included.

   (ii) It is not identified because $\log(pcinc)$ is not significative in Output 1.

   (iii) It is identified because $\log(land)$ is significative in Output 4.

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

7. **(Problem 1.)** We are interested in obtaining consistent estimates of the first equation, assuming that *open* is endogenous and that the variables $\log(pcinc)$ and $\log(land))$ are exogenous.

(a) The estimates in Output 5 are consistent because $\log(pcinc)$ and $\log(land)$ are jointly significative in Output 4.

(b) The estimates in Output 5 are consistent because $\log(land)$ is partially correlated with the explanatory variable *open* (Output 4).

(c) The estimates in Output 5 are not consistent because $\log(pcinc)$ is not partially correlated with the explanatory variable *open* (Output 4).

(d) The estimates in Output 1 are consistent because the equation is identified.

8. **(Problem 1.)** Given the empirical evidence, can be concluded that γ_{12} is different from zero?

(i) Yes, attending to the estimation in Output 1, because OLS is always valid if the equation is identified.

(ii) Yes, attending to the instrumental variables estimation.

(iii) We do not know, because the first equation can not be identified in any case.

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

(b) Only (iii) is true.

(c) All are true.

(d) Only (ii) is true.

9. **(Problem 1.)** Taking into account the result of Hausman test (Output 5) and any other relevant information, we can conclude that,

(i) The estimates in Output 1 are consistent.

(ii) The estimates in Output 5 are more efficient than those in Output 1.

(iii) There is no empirical evidence to conclude that the variable *open* is endogenous in the first equation.

(a) Only (i) is true.

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

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

(d) All are true.

10. **(Problem 1.)** The result of Hausman test (Output 5):

(i) Shows that the first equation is over-identified because the OLS estimates are consistent.

(ii) Needs the equation to be over-identified to make sense.

(iii) Shows that the estimation in Output 5 is not consistent.

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

11. **(Problem 1.)** Concerning the estimation of a system of two simultaneous equations by 2SLS,

(i) The first step in the 2SLS to estimate one equation looks for an optimal instrumental variable when this equation is over-identified, but if the other equation is not identified, the estimation is not consistent.

(ii) The two steps in 2SLS can be done by means of OLS regressions.

(iii) The second step in the 2SLS method can be done as an instrumental variable regression, using as instrument only the optimal instrumental variable from the first step.

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

12. **(Problem 1.)** If we have available one additional valid instrument for *open* in the first equation,

(i) The 2SLS estimation would be more efficient than that of Output 5 if we additionally assume that $\log(land)$ is also a valid instrument.

(ii) We could test whether $\log(land)$ is uncorrelated with u_1.

(iii) We would not need observations of $\log(land)$ to obtain consistent estimates of the parameters of the first equation.

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

13. **(Problem 1.)** If we include the residuals from Output 4 as an additional regressor in the OLS estimation of the first equation (Output 1), and assuming that this equation is identified:

(i) The estimates of the coefficients of *open* and of *lpcinc* would be consistent.

(ii) The estimates of the coefficients of *open* and of *lpcinc* would be consistent, but less efficient that those from Output 5.

(iii) We could do an endogeneity test for *open* by means of a t test.

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

14. **(Problem 1.)** If we reject the null hypothesis of the endogeneity test,

(i) We conclude that the regressor is endogenous, and therefore the OLS estimates are inconsistent.

(ii) We conclude that the regressor is exogenous, and therefore the instrumental variable estimates are less efficient that the OLS estimates.

(iii) We conclude that the regressor is exogenous, but that the OLS estimates could be inconsistent.

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

15. **(Problem 1.)** In case of over-identification,

(i) The equation can not be estimated because the parameters take too many alternative values.

(ii) We can not define an efficient instrumental variables estimate.

(iii) We can not do an exogeneity test, because we do not have enough degrees of freedom.

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

16. **(Problem 1.)** The efficiency of the 2SLS,

(i) Increases with the correlation between the endogenous regressor and the instrumental variables, measured through an R^2 coefficient.

(ii) Decreases with the variability of the endogenous regressor, since instrumental variable methods are less efficient than OLS.

(iii) Does not depend on the variance of the error from the original equation, since 2SLS is based on two different regressions.

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

17. **(Problem 1.)** The weak instruments case,

(i) May arise only when we have an over-identified equation, because each instrument can only explain an small proportion of the variability of the endogenous regressor.

(ii) Arises when the instrument and the endogenous regressor have a low correlation, leading to inefficient instrumental variable estimates.

(iii) May cause serious problems of inconsistency of the instrumental variables estimate if the instrument is correlated, though minimally, with the error term.

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

18. **(Problem 2.)** In the problem discussed,

(i) It is not a good idea to compare two alternative models with the determination coefficient, R^2, since we know in advance that (2) will always fit better because it contains more regressors.

(ii) If all coefficients in the OLS regression are significant, the model is correctly specified.

(iii) Chose the right model is very important, since otherwise some of the necessary hypothesis to perform valid inference could be not valid.

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

19. **(Problem 2.)** To choose between the two models, the researcher estimates the following equation:

$$stndfnl = \beta_0 + \beta_1 atndrte + \beta_2 \log(priGPA) + \beta_3 \log(ACT) + \\ + \gamma_1 priGPA + \gamma_2 ACT + \gamma_3 priGPA^2 + \gamma_4 ACT^2 + u \qquad (3)$$

If she wants to test whether model (2) is correct, which is the null hypothesis that she should test for in equation (3)?

(a) $H_0 : \beta_2 = \beta_3 = 0.$

(b) $H_0 : \gamma_1 = \gamma_2 = \gamma_3 = \gamma_4 = 0.$

(c) $H_0 : \gamma_1 = \gamma_2 = 0.$

(d) $H_0 : \beta_1 = \beta_2 = \beta_3 = 0.$

20. **(Problem 2.)** The F statistic to test H_0 in the previous question takes the value 1,00081 with a p-value of 0.368.

(a) The researcher can not reject H_0 and decides that model (2) is correct.

(b) The researcher rejects H_0 and decides that model (1) is correct.

(c) The researcher decides that both models are correct.

(d) The researcher rejects H_0 and decides that model (2) is correct.

21. **(Problem 2.)** To choose between both models, the researcher considers an alternative method estimating the following equation,

$$stndfnl = \beta_0 + \beta_1 atndrte + \beta_2 \log(priGPA) + \beta_3 \log(ACT) + \theta_1 \widehat{y} + u,$$

where $\widehat{y}$ are the adjusted values from the estimation of model (2).

(i) If we reject $H_0 : \theta_1 = 0$, then we conclude that model (1) has specification problems.

(ii) If we reject $H_0 : \theta_1 = 0$, then we conclude that model (2) has specification problems.

(iii) If we reject $H_0 : \theta_1 = 0$, then we conclude that model (2) is correctly specified.

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

(b) All are false.

(c) Only (ii) is true.

(d) Only (i) is true.

22. **(Problem 2.)** It seems that the researcher prefers model (2), so she continues testing model (2). She estimates model (2) by OLS and obtains $R^2 = 0.226$. Then, she omits from the model the explanatory variables $priGPA^2$ and ACT^2, and obtains $R^2 = 0.211$. Which of the following conclusions is true?

   (a) The F statistic for the joint significance test of the two variables $priGPA^2$ and ACT^2 is 11.081 with a p-value of 0.000018, so that there is evidence of wrong functional specification in model (2).

   (b) The F statistic for the joint significance test of the two variables $priGPA^2$ and ACT^2 is 11.081 with a p-value of 0.000018, so that there is no evidence of wrong functional specification in model (2).

   (c) The test is not valid to check the specification of the model because the previous hypotheses are not nested.

   (d) Model (2) will be much better than the model without the explanatory variables $priGPA^2$ and ACT^2 since it has a higher R^2.

23. **(Problem 2.)** The researcher considers an alternative method: Ramsey's RESET test.

   (i) RESET test uses the OLS residuals and its squares to check if they help to explain the dependent variable.

   (ii) If we reject the null hypothesis of the RESET test, we confirm the correct specification of the model.

   (iii) RESET test allows us also to test for the presence of heteroskedasticity.

   (a) Only (iii) is true.

   (b) All are false.

   (c) Only (ii) is true.

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

24. **(Problem 2.)** If Model (2) is correctly specified, and u satisfies all usual assumptions, but we do not include the variable ACT^2 in the OLS estimation,

   (i) The OLS estimates of the remaining parameters will be biased if $\beta_5 \neq 0$, because at least ACT and ACT^2 will be correlated.

   (ii) The OLS estimation of the remaining parameters will be consistent, but the new errors will be heteroskedastic and usual inference will not be valid, if $\beta_5 \neq 0$.

   (iii) OLS estimation of the remaining parameters, except β_3, will be consistent if $\beta_5 = 0$, since ACT and ACT^2 will be correlated.

   (a) Only (i) is true.

   (b) All are false.

   (c) Only (iii) is true.

   (d) Only (ii) is true.

25. **(Problem 2.)** In the next step of the empirical analysis, the researcher wishes to test for the presence of heteroskedasticity in the model using White's test. Assume that given the results of previous tests, she decides to focus on model (2). How is White's test implemented?

   (a) In model (2) we add the squares of the explanatory variables and all cross products and then we calculate the F statistic for global significance.

   (b) We calculate the residuals of the OLS fit of model (2). We estimate the regression of the squares of the OLS residuals on all explanatory variables of model (2), their squares, and all cross products. We obtain the R^2 of this estimation and we compute the test statistic nR^2.

   (c) We calculate the residuals of the OLS fit of model (2). We estimate the regression of the squares of the OLS residuals on the squares of all explanatory variables of model (2). We obtain the R^2 of this estimation and we compute the test statistic nR^2.

   (d) In model (2) we add the squares of the explanatory variables and then we calculate the F test for global significance.

26. **(Problem 2.)** The result of White's heteroskedasticity test is 13.993, with a p-value of 0.72. Which is the conclusion of the researcher on the null hypothesis of White's test, and therefore, on the presence of heteroskedasticity in model (2) given the empirical evidence?

   (a) She can not reject H_0 and concludes that there is heteroskedasticity.

   (b) She can not reject H_0 and concludes that there is not heteroskedasticity.

   (c) She rejects H_0 and concludes that there is not heteroskedasticity.

   (d) She rejects H_0 and concludes that there is heteroskedasticity.

27. **(Problem 2.)** In the presence of heteroskedasticity,

   (i) OLS estimates stop being consistent.

   (ii) To perform valid inference is always necessary to estimate the equation by means of generalized least squares.

   (iii) White's corrected standard errors are only valid if we know the form of the heteroskedasticity.

   (a) Only (iii) is true.

   (b) All are false.

   (c) Only (ii) is true.

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

28. **(Problem 2.)** Now assume that the researcher decides to include the interaction term $priGPA * atndrte$ in model (2):

$$stndfnl = \beta_0 + \beta_1 atndrte + \beta_2 priGPA + \beta_3 ACT + \beta_4 priGPA^2 + \\ + \beta_5 ACT^2 + \beta_6 priGPA * atndrte + u. \tag{4}$$

The results of the OLS estimation of this new model are as follows,

$$\begin{aligned} \widehat{stndfnl} &= \underset{(1.36)}{2.05} - \underset{(0.0102)}{0.0067} atndrte - \underset{(0.48)}{1.63} priGPA - \underset{(0.098)}{0.128} ACT + \underset{(0.101)}{0.296} priGPA^2 + \\ &+ \underset{(0.0022)}{0.0045} ACT^2 + \underset{(??)}{0.0056} priGPA * atndrte \\ n &= 680 \qquad R^2 = 0.228 \end{aligned}$$

(a) The t test for the coefficient of the interaction is equal to 1.32, and therefore the estimate is not significative.

(b) The t test for the coefficient of the interaction is equal to 2.13, and therefore the estimate is significative.

(c) There is not enough information to test for the significance of the new estimate.

(d) The effect of the interaction is without doubt significative, since the R^2 has increased.

29. **(Problem 2.)** The researcher wishes to calculate the partial effect of attendance to class (*atndrte*) on the result of the final exam (*stndfnl*). How would you estimate that effect?

(a) $\widehat{\beta}_1 + \widehat{\beta}_6 priGPA$.

(b) $\widehat{\beta}_1 + \widehat{\beta}_6$.

(c) $\widehat{\beta}_1 + \widehat{\beta}_2 + \widehat{\beta}_6 priGPA$.

(d) Is simply the estimated coefficient $\widehat{\beta}_1$.

30. **(Problem 2.)** If the averaged value of *priGPA* in the sample is 2.59, how would affect an increment of 10 percentage points of class attendance on the result of the final exam for a student with this value of *priGPA*?

(a) An increment of 10 percentage points in the class attendance leads to an increment of *stndfnl* of 0.0078.

(b) An increment of 10 percentage points in the class attendance leads to a decrement of *stndfnl* of 0.067.

(c) An increment of 10 percentage points in the class attendance leads to an increment of *stndfnl* of 1.622.

(d) An increment of 10 percentage points in the class attendance leads to an increment of *stndfnl* of 0.078.

31. **(Problem 2.)** The standard error of estimating the previous effect is

(a) 0.25 if the OLS estimates of the coefficients were uncorrelated.

(b) There is no enough available information to calculate it.

(c) 0.102.

(d) 0.15 if the OLS estimates of the coefficients were uncorrelated.

32. **(Problem 2.)** Using Model (4), which is the partial effect of the grade of previous exams (*priGPA*) over the final exam grading (*stndfnl*), when we fix $priGPA = 2.59$ and $atndrte = 0.82$?

(a) -0.858.

(b) -0.863.

(c) 1.537.

(d) -0.092.

33. **(Problem 2.)** What is like the effect of the average grade in the university access test (ACT) over the result of the final exam, ceteris paribus?

(a) The effect would be negative because, despite the estimated coefficient of the variable ACT^2 ($\widehat{\beta_5}$) being positive, it is smaller in comparison with the estimated coefficient $\widehat{\beta_3}$, which is negative.

(b) The effect would be positive because the estimated coefficient of the variable ACT^2, $\widehat{\beta_5}$, is positive and since the variable ACT appears as a square, the sign of this variable will prevail.

(c) If would depend on the values of the variable ACT.

(d) The effect would be negative because the estimated effect of the variable *ACT*, $\widehat{\beta_3}$, is negative.

34. **(Problem 2.)** From which value of the grading on, having an extra point in the access to university test, ACT, increases the result of the standardized final test?

(a) From a grading of approximately 28.44 points.

(b) From a grading of approximately 12.8 points.

(c) From a grading of approximately 0.45 points.

(d) From a grading of approximately 14.22 points.

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 1.)** The endogeneity problem of regressors in linear models causes that the OLS estimates,

   (i) Are generally biased.

   (ii) Need to be compared with corrected standard errors (White).

   (iii) Are not efficient, though in general are always consistent.

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

2. **(Problem 1.)** In a model of two simultaneous equations,

   (i) It is only necessary to estimate one estructural equation, since in general the dependent variable of the second equation show up as a regressor in the first one.

   (ii) None of the two equations can describe a causal relationship.

   (iii) If one estructural equation is identified, the other equation, and, therefore, the system are identified.

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

3. **(Problem 1.)** An instrumental variable for an equation,

   (i) It is an exogenous regressor in this equation.

   (ii) It is a dependent variable from other equation.

   (iii) Has not to be correlated with the error term.

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

4. **(Problem 1.)** The fact that the variable *open* is endogenoeus in the first equation,

   (i) Depends on whether γ_{21} is different from zero or not.

   (ii) Depends on whether the error terms u_1 and u_2 are correlated or not.

   (iii) Depends on whether δ_{22} is different from zero or not.

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

5. **(Problem 1.)** Assuming that *open* is endogenous, is the first equation identified?

   (a) It is not identified because $\log(pcinc)$ is not significative in the reduced form of *open*.
   (b) It is identified because $\log(land)$ is significative in Output 2.
   (c) It is identified because the F is significative in Output 4.
   [(d)] It is identified, because $\log(land)$ is significative in the reduced form of *open*.

6. **(Problem 1.)** Assuming that inf is endogenous, study the identification of the second equation:

   (i) It is identified because it has an additional exogenous variable included.

   (ii) It is not identified because $\log(pcinc)$ is not significative in Output 1.

   (iii) It is identified because $\log(land)$ is significative in Output 4.

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

7. **(Problem 1.)** We are interested in obtaining consistent estimates of the first equation, assuming that *open* is endogenous and that the variables $\log(pcinc)$ and $\log(land))$ are exogenous.

   (a) The estimates in Output 5 are consistent because $\log(pcinc)$ and $\log(land)$ are jointly significative in Output 4.

   [(b)] The estimates in Output 5 are consistent because $\log(land)$ is partially correlated with the explanatory variable *open* (Output 4).

   (c) The estimates in Output 5 are not consistent because $\log(pcinc)$ is not partially correlated with the explanatory variable *open* (Output 4).

   (d) The estimates in Output 1 are consistent because the equation is identified.

8. **(Problem 1.)** Given the empirical evidence, can be concluded that γ_{12} is different from zero?

   (i) Yes, attending to the estimation in Output 1, because OLS is always valid if the equation is identified.

   (ii) Yes, attending to the instrumental variables estimation.

   (iii) We do not know, because the first equation can not be identified in any case.

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

   (b) Only (iii) is true.

   (c) All are true.

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

9. **(Problem 1.)** Taking into account the result of Hausman test (Output 5) and any other relevant information, we can conclude that,

   (i) The estimates in Output 1 are consistent.

   (ii) The estimates in Output 5 are more efficient than those in Output 1.

   (iii) There is no empirical evidence to conclude that the variable *open* is endogenous in the first equation.

   (a) Only (i) is true.

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

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

   (d) All are true.

10. **(Problem 1.)** The result of Hausman test (Output 5):

   (i) Shows that the first equation is over-identified because the OLS estimates are consistent.

   (ii) Needs the equation to be over-identified to make sense.

   (iii) Shows that the estimation in Output 5 is not consistent.

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

11. **(Problem 1.)** Concerning the estimation of a system of two simultaneous equations by 2SLS,

   (i) The first step in the 2SLS to estimate one equation looks for an optimal instrumental variable when this equation is over-identified, but if the other equation is not identified, the estimation is not consistent.

   (ii) The two steps in 2SLS can be done by means of OLS regressions.

   (iii) The second step in the 2SLS method can be done as an instrumental variable regression, using as instrument only the optimal instrumental variable from the first step.

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

12. **(Problem 1.)** If we have available one additional valid instrument for *open* in the first equation,

   (i) The 2SLS estimation would be more efficient than that of Output 5 if we additionally assume that $\log(land)$ is also a valid instrument.

   (ii) We could test whether $\log(land)$ is uncorrelated with u_1.

   (iii) We would not need observations of $\log(land)$ to obtain consistent estimates of the parameters of the first equation.

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

13. **(Problem 1.)** If we include the residuals from Output 4 as an additional regressor in the OLS estimation of the first equation (Output 1), and assuming that this equation is identified:

   (i) The estimates of the coefficients of *open* and of *lpcinc* would be consistent.

   (ii) The estimates of the coefficients of *open* and of *lpcinc* would be consistent, but less efficient that those from Output 5.

   (iii) We could do an endogeneity test for *open* by means of a t test.

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

14. **(Problem 1.)** If we reject the null hypothesis of the endogeneity test,

   (i) We conclude that the regressor is endogenous, and therefore the OLS estimates are inconsistent.

   (ii) We conclude that the regressor is exogenous, and therefore the instrumental variable estimates are less efficient that the OLS estimates.

   (iii) We conclude that the regressor is exogenous, but that the OLS estimates could be inconsistent.

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

15. **(Problem 1.)** In case of over-identification,

   (i) The equation can not be estimated because the parameters take too many alternative values.

   (ii) We can not define an efficient instrumental variables estimate.

   (iii) We can not do an exogeneity test, because we do not have enough degrees of freedom.

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

16. **(Problem 1.)** The efficiency of the 2SLS,

(i) Increases with the correlation between the endogenous regressor and the instrumental variables, measured through an R^2 coefficient.

(ii) Decreases with the variability of the endogenous regressor, since instrumental variable methods are less efficient than OLS.

(iii) Does not depend on the variance of the error from the original equation, since 2SLS is based on two different regressions.

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

17. **(Problem 1.)** The weak instruments case,

(i) May arise only when we have an over-identified equation, because each instrument can only explain an small proportion of the variability of the endogenous regressor.

(ii) Arises when the instrument and the endogenous regressor have a low correlation, leading to inefficient instrumental variable estimates.

(iii) May cause serious problems of inconsistency of the instrumental variables estimate if the instrument is correlated, though minimally, with the error term.

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

18. **(Problem 2.)** In the problem discussed,

(i) It is not a good idea to compare two alternative models with the determination coefficient, R^2, since we know in advance that (2) will always fit better because it contains more regressors.

(ii) If all coefficients in the OLS regression are significant, the model is correctly specified.

(iii) Chose the right model is very important, since otherwise some of the necessary hypothesis to perform valid inference could be not valid.

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

19. **(Problem 2.)** To choose between the two models, the researcher estimates the following equation:

$$stndfnl = \beta_0 + \beta_1 atndrte + \beta_2 \log(priGPA) + \beta_3 \log(ACT) + \\ + \gamma_1 priGPA + \gamma_2 ACT + \gamma_3 priGPA^2 + \gamma_4 ACT^2 + u \quad (3)$$

If she wants to test whether model (2) is correct, which is the null hypothesis that she should test for in equation (3)?

[(a)] $H_0 : \beta_2 = \beta_3 = 0.$

(b) $H_0 : \gamma_1 = \gamma_2 = \gamma_3 = \gamma_4 = 0.$

(c) $H_0 : \gamma_1 = \gamma_2 = 0.$

(d) $H_0 : \beta_1 = \beta_2 = \beta_3 = 0.$

20. **(Problem 2.)** The F statistic to test H_0 in the previous question takes the value 1,00081 with a p-value of 0.368.

[(a)] The researcher can not reject H_0 and decides that model (2) is correct.

(b) The researcher rejects H_0 and decides that model (1) is correct.

(c) The researcher decides that both models are correct.

(d) The researcher rejects H_0 and decides that model (2) is correct.

21. **(Problem 2.)** To choose between both models, the researcher considers an alternative method estimating the following equation,

$$stndfnl = \beta_0 + \beta_1 atndrte + \beta_2 \log(priGPA) + \beta_3 \log(ACT) + \theta_1 \widehat{\hat{y}} + u,$$

where $\widehat{\hat{y}}$ are the adjusted values from the estimation of model (2).

(i) If we reject $H_0 : \theta_1 = 0$, then we conclude that model (1) has specification problems.

(ii) If we reject $H_0 : \theta_1 = 0$, then we conclude that model (2) has specification problems.

(iii) If we reject $H_0 : \theta_1 = 0$, then we conclude that model (2) is correctly specified.

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

(b) All are false.

(c) Only (ii) is true.

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

22. **(Problem 2.)** It seems that the researcher prefers model (2), so she continues testing model (2). She estimates model (2) by OLS and obtains $R^2 = 0.226$. Then, she omits from the model the explanatory variables $priGPA^2$ and ACT^2, and obtains $R^2 = 0.211$. Which of the following conclusions is true?

    (a) The F statistic for the joint significance test of the two variables $priGPA^2$ and ACT^2 is 11.081 with a p-value of 0.000018, so that there is evidence of wrong functional specification in model (2).

    [(b)] The F statistic for the joint significance test of the two variables $priGPA^2$ and ACT^2 is 11.081 with a p-value of 0.000018, so that there is no evidence of wrong functional specification in model (2).

    (c) The test is not valid to check the specification of the model because the previous hypotheses are not nested.

    (d) Model (2) will be much better than the model without the explanatory variables $priGPA^2$ and ACT^2 since it has a higher R^2.

23. **(Problem 2.)** The researcher considers an alternative method: Ramsey's RESET test.

    (i) RESET test uses the OLS residuals and its squares to check if they help to explain the dependent variable.

    (ii) If we reject the null hypothesis of the RESET test, we confirm the correct specification of the model.

    (iii) RESET test allows us also to test for the presence of heteroskedasticity.

    (a) Only (iii) is true.

    [(b)] All are false.

    (c) Only (ii) is true.

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

24. **(Problem 2.)** If Model (2) is correctly specified, and u satisfies all usual assumptions, but we do not include the variable ACT^2 in the OLS estimation,

    (i) The OLS estimates of the remaining parameters will be biased if $\beta_5 \neq 0$, because at least ACT and ACT^2 will be correlated.

    (ii) The OLS estimation of the remaining parameters will be consistent, but the new errors will be heteroskedastic and usual inference will not be valid, if $\beta_5 \neq 0$.

    (iii) OLS estimation of the remaining parameters, except β_3, will be consistent if $\beta_5 = 0$, since ACT and ACT^2 will be correlated.

    [(a)] Only (i) is true.

    (b) All are false.

    (c) Only (iii) is true.

    (d) Only (ii) is true.

25. **(Problem 2.)** In the next step of the empirical analysis, the researcher wishes to test for the presence of heteroskedasticity in the model using White's test. Assume that given the results of previous tests, she decides to focus on model (2). How is White's test implemented?

(a) In model (2) we add the squares of the explanatory variables and all cross products and then we calculate the F statistic for global significance.

[(b)] We calculate the residuals of the OLS fit of model (2). We estimate the regression of the squares of the OLS residuals on all explanatory variables of model (2), their squares, and all cross products. We obtain the R^2 of this estimation and we compute the test statistic nR^2.

(c) We calculate the residuals of the OLS fit of model (2). We estimate the regression of the squares of the OLS residuals on the squares of all explanatory variables of model (2). We obtain the R^2 of this estimation and we compute the test statistic nR^2.

(d) In model (2) we add the squares of the explanatory variables and then we calculate the F test for global significance.

26. **(Problem 2.)** The result of White's heteroskedasticity test is 13.993, with a p-value of 0.72. Which is the conclusion of the researcher on the null hypothesis of White's test, and therefore, on the presence of heteroskedasticity in model (2) given the empirical evidence?

(a) She can not reject H_0 and concludes that there is heteroskedasticity.

[(b)] She can not reject H_0 and concludes that there is not heteroskedasticity.

(c) She rejects H_0 and concludes that there is not heteroskedasticity.

(d) She rejects H_0 and concludes that there is heteroskedasticity.

27. **(Problem 2.)** In the presence of heteroskedasticity,

(i) OLS estimates stop being consistent.

(ii) To perform valid inference is always necessary to estimate the equation by means of generalized least squares.

(iii) White's corrected standard errors are only valid if we know the form of the heteroskedasticity.

(a) Only (iii) is true.

[(b)] All are false.

(c) Only (ii) is true.

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

28. **(Problem 2.)** Now assume that the researcher decides to include the interaction term $priGPA * atndrte$ in model (2):

$$stndfnl = \beta_0 + \beta_1 atndrte + \beta_2 priGPA + \beta_3 ACT + \beta_4 priGPA^2 + \\ + \beta_5 ACT^2 + \beta_6 priGPA * atndrte + u. \quad (4)$$

The results of the OLS estimation of this new model are as follows,

$$\underset{}{\widehat{stndfnl}} = \underset{(1.36)}{2.05} - \underset{(0.0102)}{0.0067} atndrte - \underset{(0.48)}{1.63} priGPA - \underset{(0.098)}{0.128} ACT + \underset{(0.101)}{0.296} priGPA^2 + \\ + \underset{(0.0022)}{0.0045} ACT^2 + \underset{(??)}{0.0056} priGPA * atndrte \\ n = 680 \qquad R^2 = 0.228$$

(a) The t test for the coefficient of the interaction is equal to 1.32, and therefore the estimate is not significative.
(b) The t test for the coefficient of the interaction is equal to 2.13, and therefore the estimate is significative.
(c) There is not enough information to test for the significance of the new estimate.
(d) The effect of the interaction is without doubt significative, since the R^2 has increased.

29. **(Problem 2.)** The researcher wishes to calculate the partial effect of attendance to class (*atndrte*) on the result of the final exam (*stndfnl*). How would you estimate that effect?

(a) $\widehat{\beta}_1 + \widehat{\beta}_6 priGPA$.
(b) $\widehat{\beta}_1 + \widehat{\beta}_6$.
(c) $\widehat{\beta}_1 + \widehat{\beta}_2 + \widehat{\beta}_6 priGPA$.
(d) Is simply the estimated coefficient $\widehat{\beta}_1$.

30. **(Problem 2.)** If the averaged value of *priGPA* in the sample is 2.59, how would affect an increment of 10 percentage points of class attendance on the result of the final exam for a student with this value of $priGPA$?

(a) An increment of 10 percentage points in the class attendance leads to an increment of *stndfnl* of 0.0078.
(b) An increment of 10 percentage points in the class attendance leads to a decrement of *stndfnl* of 0.067.
(c) An increment of 10 percentage points in the class attendance leads to an increment of *stndfnl* of 1.622.
(d) An increment of 10 percentage points in the class attendance leads to an increment of *stndfnl* of 0.078.

31. **(Problem 2.)** The standard error of estimating the previous effect is

(a) 0.25 if the OLS estimates of the coefficients were uncorrelated.

(b) There is no enough available information to calculate it.

(c) 0.102.

[(d)] 0.15 if the OLS estimates of the coefficients were uncorrelated.

32. **(Problem 2.)** Using Model (4), which is the partial effect of the grade of previous exams (*priGPA*) over the final exam grading (*stndfnl*), when we fix $priGPA = 2.59$ and $atndrte = 0.82$?

(a) -0.858.

(b) -0.863.

[(c)] 1.537.

(d) -0.092.

33. **(Problem 2.)** What is like the effect of the average grade in the university access test (ACT) over the result of the final exam, ceteris paribus?

(a) The effect would be negative because, despite the estimated coefficient of the variable ACT^2 ($\widehat{\beta}_5$) being positive, it is smaller in comparison with the estimated coefficient $\widehat{\beta}_3$, which is negative.

(b) The effect would be positive because the estimated coefficient of the variable ACT^2, $\widehat{\beta}_5$, is positive and since the variable ACT appears as a square, the sign of this variable will prevail.

[(c)] If would depend on the values of the variable ACT.

(d) The effect would be negative because the estimated effect of the variable ACT, $\widehat{\beta}_3$, is negative.

34. **(Problem 2.)** From which value of the grading on, having an extra point in the access to university test, ACT, increases the result of the standardized final test?

(a) From a grading of approximately 28.44 points.

(b) From a grading of approximately 12.8 points.

(c) From a grading of approximately 0.45 points.

[(d)] From a grading of approximately 14.22 points.

Universidad Carlos III de Madrid
<u>ECONOMETRICS I</u>
Academic year 2006/07
FINAL EXAM
September 6, 2008

Exam Type: 3

TIME: 2 HOURS 30 MINUTES

1. **(Problem 2.)** In the problem discussed,

   (i) It is not a good idea to compare two alternative models with the determination coefficient, R^2, since we know in advance that (2) will always fit better because it contains more regressors.

   (ii) If all coefficients in the OLS regression are significant, the model is correctly specified.

   (iii) Chose the right model is very important, since otherwise some of the necessary hypothesis to perform valid inference could be not valid.

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

2. **(Problem 2.)** To choose between the two models, the researcher estimates the following equation:

$$stndfnl = \beta_0 + \beta_1 atndrte + \beta_2 \log(priGPA) + \beta_3 \log(ACT) + \\ + \gamma_1 priGPA + \gamma_2 ACT + \gamma_3 priGPA^2 + \gamma_4 ACT^2 + u \quad (1)$$

   If she wants to test whether model (2) is correct, which is the null hypothesis that she should test for in equation (3)?

   (a) $H_0 : \gamma_1 = \gamma_2 = 0.$
   (b) $H_0 : \beta_2 = \beta_3 = 0.$
   (c) $H_0 : \beta_1 = \beta_2 = \beta_3 = 0.$
   (d) $H_0 : \gamma_1 = \gamma_2 = \gamma_3 = \gamma_4 = 0.$

3. **(Problem 2.)** The F statistic to test H_0 in the previous question takes the value 1,00081 with a p-value of 0.368.

   (a) The researcher decides that both models are correct.
   (b) The researcher can not reject H_0 and decides that model (2) is correct.
   (c) The researcher rejects H_0 and decides that model (2) is correct.
   (d) The researcher rejects H_0 and decides that model (1) is correct.

4. **(Problem 2.)** To choose between both models, the researcher considers an alternative method estimating the following equation,

$$stndfnl = \beta_0 + \beta_1 atndrte + \beta_2 \log(priGPA) + \beta_3 \log(ACT) + \theta_1 \widehat{y} + u,$$

where $\widehat{y}$ are the adjusted values from the estimation of model (2).

(i) If we reject $H_0 : \theta_1 = 0$, then we conclude that model (1) has specification problems.

(ii) If we reject $H_0 : \theta_1 = 0$, then we conclude that model (2) has specification problems.

(iii) If we reject $H_0 : \theta_1 = 0$, then we conclude that model (2) is correctly specified.

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

5. **(Problem 2.)** It seems that the researcher prefers model (2), so she continues testing model (2). She estimates model (2) by OLS and obtains $R^2 = 0.226$. Then, she omits from the model the explanatory variables $priGPA^2$ and ACT^2, and obtains $R^2 = 0.211$. Which of the following conclusions is true?

(a) The test is not valid to check the specification of the model because the previous hypotheses are not nested.

(b) The F statistic for the joint significance test of the two variables $priGPA^2$ and ACT^2 is 11.081 with a p-value of 0.000018, so that there is evidence of wrong functional specification in model (2).

(c) Model (2) will be much better than the model without the explanatory variables $priGPA^2$ and ACT^2 since it has a higher R^2.

(d) The F statistic for the joint significance test of the two variables $priGPA^2$ and ACT^2 is 11.081 with a p-value of 0.000018, so that there is no evidence of wrong functional specification in model (2).

6. **(Problem 2.)** The researcher considers an alternative method: Ramsey's RESET test.

   (i) RESET test uses the OLS residuals and its squares to check if they help to explain the dependent variable.

   (ii) If we reject the null hypothesis of the RESET test, we confirm the correct specification of the model.

   (iii) RESET test allows us also to test for the presence of heteroskedasticity.

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

7. **(Problem 2.)** If Model (2) is correctly specified, and u satisfies all usual assumptions, but we do not include the variable ACT^2 in the OLS estimation,

   (i) The OLS estimates of the remaining parameters will be biased if $\beta_5 \neq 0$, because at least ACT and ACT^2 will be correlated.

   (ii) The OLS estimation of the remaining parameters will be consistent, but the new errors will be heteroskedastic and usual inference will not be valid, if $\beta_5 \neq 0$.

   (iii) OLS estimation of the remaining parameters, except β_3, will be consistent if $\beta_5 = 0$, since ACT and ACT^2 will be correlated.

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

8. **(Problem 2.)** In the next step of the empirical analysis, the researcher wishes to test for the presence of heteroskedasticity in the model using White's test. Assume that given the results of previous tests, she decides to focus on model (2). How is White's test implemented?

   (a) We calculate the residuals of the OLS fit of model (2). We estimate the regression of the squares of the OLS residuals on the squares of all explanatory variables of model (2). We obtain the R^2 of this estimation and we compute the test statistic nR^2.

   (b) In model (2) we add the squares of the explanatory variables and all cross products and then we calculate the F statistic for global significance.

   (c) In model (2) we add the squares of the explanatory variables and then we calculate the F test for global significance.

   (d) We calculate the residuals of the OLS fit of model (2). We estimate the regression of the squares of the OLS residuals on all explanatory variables of model (2), their squares, and all cross products. We obtain the R^2 of this estimation and we compute the test statistic nR^2.

9. **(Problem 2.)** The result of White's heteroskedasticity test is 13.993, with a p-value of 0.72. Which is the conclusion of the researcher on the null hypothesis of White's test, and therefore, on the presence of heteroskedasticity in model (2) given the empirical evidence?

   (a) She rejects H_0 and concludes that there is not heteroskedasticity.

   (b) She can not reject H_0 and concludes that there is heteroskedasticity.

   (c) She rejects H_0 and concludes that there is heteroskedasticity.

   (d) She can not reject H_0 and concludes that there is not heteroskedasticity.

10. **(Problem 2.)** In the presence of heteroskedasticity,

    (i) OLS estimates stop being consistent.

    (ii) To perform valid inference is always necessary to estimate the equation by means of generalized least squares.

    (iii) White's corrected standard errors are only valid if we know the form of the heteroskedasticity.

    (a) Only (ii) is true.

    (b) Only (iii) is true.

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

    (d) All are false.

11. **(Problem 2.)** Now assume that the researcher decides to include the interaction term $priGPA * atndrte$ in model (2):

$$\begin{aligned} stndfnl = \beta_0 + \beta_1 atndrte + \beta_2 priGPA + \beta_3 ACT + \beta_4 priGPA^2 + \\ + \beta_5 ACT^2 + \beta_6 priGPA * atndrte + u. \end{aligned} \tag{2}$$

The results of the OLS estimation of this new model are as follows,

$$\begin{aligned} \widehat{stndfnl} &= \underset{(1.36)}{2.05} - \underset{(0.0102)}{0.0067} atndrte - \underset{(0.48)}{1.63} priGPA - \underset{(0.098)}{0.128} ACT + \underset{(0.101)}{0.296} priGPA^2 + \\ &+ \underset{(0.0022)}{0.0045} ACT^2 + \underset{(??)}{0.0056} priGPA * atndrte \\ n &= 680 \qquad R^2 = 0.228 \end{aligned}$$

(a) There is not enough information to test for the significance of the new estimate.

(b) The t test for the coefficient of the interaction is equal to 1.32, and therefore the estimate is not significative.

(c) The effect of the interaction is without doubt significative, since the R^2 has increased.

(d) The t test for the coefficient of the interaction is equal to 2.13, and therefore the estimate is significative.

12. **(Problem 2.)** The researcher wishes to calculate the partial effect of attendance to class (*atndrte*) on the result of the final exam (*stndfnl*). How would you estimate that effect?

(a) $\widehat{\beta}_1 + \widehat{\beta}_2 + \widehat{\beta}_6 priGPA$.

(b) $\widehat{\beta}_1 + \widehat{\beta}_6 priGPA$.

(c) Is simply the estimated coefficient $\widehat{\beta}_1$.

(d) $\widehat{\beta}_1 + \widehat{\beta}_6$.

13. **(Problem 2.)** If the averaged value of *priGPA* in the sample is 2.59, how would affect an increment of 10 percentage points of class attendance on the result of the final exam for a student with this value of $priGPA$?

(a) An increment of 10 percentage points in the class attendance leads to an increment of *stndfnl* of 1.622.

(b) An increment of 10 percentage points in the class attendance leads to an increment of *stndfnl* of 0.0078.

(c) An increment of 10 percentage points in the class attendance leads to an increment of *stndfnl* of 0.078.

(d) An increment of 10 percentage points in the class attendance leads to a decrement of *stndfnl* of 0.067.

14. **(Problem 2.)** The standard error of estimating the previous effect is

 (a) 0.102.
 (b) 0.25 if the OLS estimates of the coefficients were uncorrelated.
 (c) 0.15 if the OLS estimates of the coefficients were uncorrelated.
 (d) There is no enough available information to calculate it.

15. **(Problem 2.)** Using Model (4), which is the partial effect of the grade of previous exams (*priGPA*) over the final exam grading (*stndfnl*), when we fix $priGPA = 2.59$ and $atndrte = 0.82$?

 (a) 1.537.
 (b) -0.858.
 (c) -0.092.
 (d) -0.863.

16. **(Problem 2.)** What is like the effect of the average grade in the university access test (ACT) over the result of the final exam, ceteris paribus?

 (a) If would depend on the values of the variable ACT.
 (b) The effect would be negative because, despite the estimated coefficient of the variable ACT^2 ($\widehat{\beta}_5$) being positive, it is smaller in comparison with the estimated coefficient $\widehat{\beta}_3$, which is negative.
 (c) The effect would be negative because the estimated effect of the variable ACT, $\widehat{\beta}_3$, is negative.
 (d) The effect would be positive because the estimated coefficient of the variable ACT^2, $\widehat{\beta}_5$, is positive and since the variable ACT appears as a square, the sign of this variable will prevail.

17. **(Problem 2.)** From which value of the grading on, having an extra point in the access to university test, ACT, increases the result of the standardized final test?

 (a) From a grading of approximately 0.45 points.
 (b) From a grading of approximately 28.44 points.
 (c) From a grading of approximately 14.22 points.
 (d) From a grading of approximately 12.8 points.

18. **(Problem 1.)** The endogeneity problem of regressors in linear models causes that the OLS estimates,

   (i) Are generally biased.

   (ii) Need to be compared with corrected standard errors (White).

   (iii) Are not efficient, though in general are always consistent.

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

19. **(Problem 1.)** In a model of two simultaneous equations,

   (i) It is only necessary to estimate one estructural equation, since in general the dependent variable of the second equation show up as a regressor in the first one.

   (ii) None of the two equations can describe a causal relationship.

   (iii) If one estructural equation is identified, the other equation, and, therefore, the system are identified.

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

20. **(Problem 1.)** An instrumental variable for an equation,

   (i) It is an exogenous regressor in this equation.

   (ii) It is a dependent variable from other equation.

   (iii) Has not to be correlated with the error term.

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

21. **(Problem 1.)** The fact that the variable *open* is endogenoeus in the first equation,

   (i) Depends on whether γ_{21} is different from zero or not.

   (ii) Depends on whether the error terms u_1 and u_2 are correlated or not.

   (iii) Depends on whether δ_{22} is different from zero or not.

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

22. **(Problem 1.)** Assuming that *open* is endogenous, is the first equation identified?

   (a) It is identified because the F is significative in Output 4.
   (b) It is not identified because $\log(pcinc)$ is not significative in the reduced form of *open*.
   (c) It is identified, because $\log(land)$ is significative in the reduced form of *open*.
   (d) It is identified because $\log(land)$ is significative in Output 2.

23. **(Problem 1.)** Assuming that inf is endogenous, study the identification of the second equation:

   (i) It is identified because it has an additional exogenous variable included.

   (ii) It is not identified because $\log(pcinc)$ is not significative in Output 1.

   (iii) It is identified because $\log(land)$ is significative in Output 4.

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

24. **(Problem 1.)** We are interested in obtaining consistent estimates of the first equation, assuming that *open* is endogenous and that the variables $\log(pcinc)$ and $\log(land)$) are exogenous.

   (a) The estimates in Output 5 are not consistent because $\log(pcinc)$ is not partially correlated with the explanatory variable *open* (Output 4).
   (b) The estimates in Output 5 are consistent because $\log(pcinc)$ and $\log(land)$ are jointly significative in Output 4.
   (c) The estimates in Output 1 are consistent because the equation is identified.
   (d) The estimates in Output 5 are consistent because $\log(land)$ is partially correlated with the explanatory variable *open* (Output 4).

25. **(Problem 1.)** Given the empirical evidence, can be concluded that γ_{12} is different from zero?

(i) Yes, attending to the estimation in Output 1, because OLS is always valid if the equation is identified.

(ii) Yes, attending to the instrumental variables estimation.

(iii) We do not know, because the first equation can not be identified in any case.

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

26. **(Problem 1.)** Taking into account the result of Hausman test (Output 5) and any other relevant information, we can conclude that,

(i) The estimates in Output 1 are consistent.

(ii) The estimates in Output 5 are more efficient than those in Output 1.

(iii) There is no empirical evidence to conclude that the variable *open* is endogenous in the first equation.

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

27. **(Problem 1.)** The result of Hausman test (Output 5):

(i) Shows that the first equation is over-identified because the OLS estimates are consistent.

(ii) Needs the equation to be over-identified to make sense.

(iii) Shows that the estimation in Output 5 is not consistent.

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

28. **(Problem 1.)** Concerning the estimation of a system of two simultaneous equations by 2SLS,

   (i) The first step in the 2SLS to estimate one equation looks for an optimal instrumental variable when this equation is over-identified, but if the other equation is not identified, the estimation is not consistent.

   (ii) The two steps in 2SLS can be done by means of OLS regressions.

   (iii) The second step in the 2SLS method can be done as an instrumental variable regression, using as instrument only the optimal instrumental variable from the first step.

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

29. **(Problem 1.)** If we have available one additional valid instrument for *open* in the first equation,

   (i) The 2SLS estimation would be more efficient than that of Output 5 if we additionally assume that $\log(land)$ is also a valid instrument.

   (ii) We could test whether $\log(land)$ is uncorrelated with u_1.

   (iii) We would not need observations of $\log(land)$ to obtain consistent estimates of the parameters of the first equation.

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

30. **(Problem 1.)** If we include the residuals from Output 4 as an additional regressor in the OLS estimation of the first equation (Output 1), and assuming that this equation is identified:

   (i) The estimates of the coefficients of *open* and of *lpcinc* would be consistent.

   (ii) The estimates of the coefficients of *open* and of *lpcinc* would be consistent, but less efficient that those from Output 5.

   (iii) We could do an endogeneity test for *open* by means of a t test.

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

31. **(Problem 1.)** If we reject the null hypothesis of the endogeneity test,

(i) We conclude that the regressor is endogenous, and therefore the OLS estimates are inconsistent.

(ii) We conclude that the regressor is exogenous, and therefore the instrumental variable estimates are less efficient that the OLS estimates.

(iii) We conclude that the regressor is exogenous, but that the OLS estimates could be inconsistent.

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

32. **(Problem 1.)** In case of over-identification,

(i) The equation can not be estimated because the parameters take too many alternative values.

(ii) We can not define an efficient instrumental variables estimate.

(iii) We can not do an exogeneity test, because we do not have enough degrees of freedom.

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

33. **(Problem 1.)** The efficiency of the 2SLS,

(i) Increases with the correlation between the endogenous regressor and the instrumental variables, measured through an R^2 coefficient.

(ii) Decreases with the variability of the endogenous regressor, since instrumental variable methods are less efficient than OLS.

(iii) Does not depend on the variance of the error from the original equation, since 2SLS is based on two different regressions.

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

34. **(Problem 1.)** The weak instruments case,

(i) May arise only when we have an over-identified equation, because each instrument can only explain an small proportion of the variability of the endogenous regressor.

(ii) Arises when the instrument and the endogenous regressor have a low correlation, leading to inefficient instrumental variable estimates.

(iii) May cause serious problems of inconsistency of the instrumental variables estimate if the instrument is correlated, though minimally, with the error term.

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

(b) Only (i) is true.

(c) Only (ii) is true.

(d) (ii) and (iii) are 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 2.)** In the problem discussed,

   (i) It is not a good idea to compare two alternative models with the determination coefficient, R^2, since we know in advance that (2) will always fit better because it contains more regressors.

   (ii) If all coefficients in the OLS regression are significant, the model is correctly specified.

   (iii) Chose the right model is very important, since otherwise some of the necessary hypothesis to perform valid inference could be not valid.

   (a) All are true.

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

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

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

2. **(Problem 2.)** To choose between the two models, the researcher estimates the following equation:

$$stndfnl = \beta_0 + \beta_1 atndrte + \beta_2 \log(priGPA) + \beta_3 \log(ACT) + \\ + \gamma_1 priGPA + \gamma_2 ACT + \gamma_3 priGPA^2 + \gamma_4 ACT^2 + u \quad (1)$$

   If she wants to test whether model (2) is correct, which is the null hypothesis that she should test for in equation (3)?

   (a) $H_0 : \gamma_1 = \gamma_2 = 0.$

   [(b)] $H_0 : \beta_2 = \beta_3 = 0.$

   (c) $H_0 : \beta_1 = \beta_2 = \beta_3 = 0.$

   (d) $H_0 : \gamma_1 = \gamma_2 = \gamma_3 = \gamma_4 = 0.$

3. **(Problem 2.)** The F statistic to test H_0 in the previous question takes the value 1,00081 with a p-value of 0.368.

(a) The researcher decides that both models are correct.

[(b)] The researcher can not reject H_0 and decides that model (2) is correct.

(c) The researcher rejects H_0 and decides that model (2) is correct.

(d) The researcher rejects H_0 and decides that model (1) is correct.

4. **(Problem 2.)** To choose between both models, the researcher considers an alternative method estimating the following equation,

$$stndfnl = \beta_0 + \beta_1 atndrte + \beta_2 \log(priGPA) + \beta_3 \log(ACT) + \theta_1 \widehat{y} + u,$$

where $\widehat{y}$ are the adjusted values from the estimation of model (2).

(i) If we reject $H_0 : \theta_1 = 0$, then we conclude that model (1) has specification problems.

(ii) If we reject $H_0 : \theta_1 = 0$, then we conclude that model (2) has specification problems.

(iii) If we reject $H_0 : \theta_1 = 0$, then we conclude that model (2) is correctly specified.

(a) Only (ii) is true.

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

[(c)] Only (i) is true.

(d) All are false.

5. **(Problem 2.)** It seems that the researcher prefers model (2), so she continues testing model (2). She estimates model (2) by OLS and obtains $R^2 = 0.226$. Then, she omits from the model the explanatory variables $priGPA^2$ and ACT^2, and obtains $R^2 = 0.211$. Which of the following conclusions is true?

(a) The test is not valid to check the specification of the model because the previous hypotheses are not nested.

(b) The F statistic for the joint significance test of the two variables $priGPA^2$ and ACT^2 is 11.081 with a p-value of 0.000018, so that there is evidence of wrong functional specification in model (2).

(c) Model (2) will be much better than the model without the explanatory variables $priGPA^2$ and ACT^2 since it has a higher R^2.

[(d)] The F statistic for the joint significance test of the two variables $priGPA^2$ and ACT^2 is 11.081 with a p-value of 0.000018, so that there is no evidence of wrong functional specification in model (2).

6. **(Problem 2.)** The researcher considers an alternative method: Ramsey's RESET test.

   (i) RESET test uses the OLS residuals and its squares to check if they help to explain the dependent variable.

   (ii) If we reject the null hypothesis of the RESET test, we confirm the correct specification of the model.

   (iii) RESET test allows us also to test for the presence of heteroskedasticity.

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

7. **(Problem 2.)** If Model (2) is correctly specified, and u satisfies all usual assumptions, but we do not include the variable ACT^2 in the OLS estimation,

   (i) The OLS estimates of the remaining parameters will be biased if $\beta_5 \neq 0$, because at least ACT and ACT^2 will be correlated.

   (ii) The OLS estimation of the remaining parameters will be consistent, but the new errors will be heteroskedastic and usual inference will not be valid, if $\beta_5 \neq 0$.

   (iii) OLS estimation of the remaining parameters, except β_3, will be consistent if $\beta_5 = 0$, since ACT and ACT^2 will be correlated.

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

8. **(Problem 2.)** In the next step of the empirical analysis, the researcher wishes to test for the presence of heteroskedasticity in the model using White's test. Assume that given the results of previous tests, she decides to focus on model (2). How is White's test implemented?

   (a) We calculate the residuals of the OLS fit of model (2). We estimate the regression of the squares of the OLS residuals on the squares of all explanatory variables of model (2). We obtain the R^2 of this estimation and we compute the test statistic nR^2.

   (b) In model (2) we add the squares of the explanatory variables and all cross products and then we calculate the F statistic for global significance.

   (c) In model (2) we add the squares of the explanatory variables and then we calculate the F test for global significance.

   [(d)] We calculate the residuals of the OLS fit of model (2). We estimate the regression of the squares of the OLS residuals on all explanatory variables of model (2), their squares, and all cross products. We obtain the R^2 of this estimation and we compute the test statistic nR^2.

9. **(Problem 2.)** The result of White's heteroskedasticity test is 13.993, with a p-value of 0.72. Which is the conclusion of the researcher on the null hypothesis of White's test, and therefore, on the presence of heteroskedasticity in model (2) given the empirical evidence?

   (a) She rejects H_0 and concludes that there is not heteroskedasticity.

   (b) She can not reject H_0 and concludes that there is heteroskedasticity.

   (c) She rejects H_0 and concludes that there is heteroskedasticity.

   [(d)] She can not reject H_0 and concludes that there is not heteroskedasticity.

10. **(Problem 2.)** In the presence of heteroskedasticity,

    (i) OLS estimates stop being consistent.

    (ii) To perform valid inference is always necessary to estimate the equation by means of generalized least squares.

    (iii) White's corrected standard errors are only valid if we know the form of the heteroskedasticity.

    (a) Only (ii) is true.

    (b) Only (iii) is true.

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

    [(d)] All are false.

11. **(Problem 2.)** Now assume that the researcher decides to include the interaction term $priGPA * atndrte$ in model (2):

$$stndfnl = \beta_0 + \beta_1 atndrte + \beta_2 priGPA + \beta_3 ACT + \beta_4 priGPA^2 + + \beta_5 ACT^2 + \beta_6 priGPA * atndrte + u. \tag{2}$$

The results of the OLS estimation of this new model are as follows,

$$\begin{aligned}\widehat{stndfnl} = &\underset{(1.36)}{2.05} - \underset{(0.0102)}{0.0067} atndrte - \underset{(0.48)}{1.63} priGPA - \underset{(0.098)}{0.128} ACT + \underset{(0.101)}{0.296} priGPA^2 + \\ &+ \underset{(0.0022)}{0.0045} ACT^2 + \underset{(??)}{0.0056} priGPA * atndrte \\ &n = 680 \qquad R^2 = 0.228\end{aligned}$$

(a) There is not enough information to test for the significance of the new estimate.

[(b)] The t test for the coefficient of the interaction is equal to 1.32, and therefore the estimate is not significative.

(c) The effect of the interaction is without doubt significative, since the R^2 has increased.

(d) The t test for the coefficient of the interaction is equal to 2.13, and therefore the estimate is significative.

12. **(Problem 2.)** The researcher wishes to calculate the partial effect of attendance to class (*atndrte*) on the result of the final exam (*stndfnl*). How would you estimate that effect?

(a) $\widehat{\beta}_1 + \widehat{\beta}_2 + \widehat{\beta}_6 priGPA$.

[(b)] $\widehat{\beta}_1 + \widehat{\beta}_6 priGPA$.

(c) Is simply the estimated coefficient $\widehat{\beta}_1$.

(d) $\widehat{\beta}_1 + \widehat{\beta}_6$.

13. **(Problem 2.)** If the averaged value of *priGPA* in the sample is 2.59, how would affect an increment of 10 percentage points of class attendance on the result of the final exam for a student with this value of $priGPA$?

(a) An increment of 10 percentage points in the class attendance leads to an increment of *stndfnl* of 1.622.

(b) An increment of 10 percentage points in the class attendance leads to an increment of *stndfnl* of 0.0078.

[(c)] An increment of 10 percentage points in the class attendance leads to an increment of *stndfnl* of 0.078.

(d) An increment of 10 percentage points in the class attendance leads to a decrement of *stndfnl* of 0.067.

14. **(Problem 2.)** The standard error of estimating the previous effect is

(a) 0.102.

(b) 0.25 if the OLS estimates of the coefficients were uncorrelated.

[(c)] 0.15 if the OLS estimates of the coefficients were uncorrelated.

(d) There is no enough available information to calculate it.

15. **(Problem 2.)** Using Model (4), which is the partial effect of the grade of previous exams (*priGPA*) over the final exam grading (*stndfnl*), when we fix $priGPA = 2.59$ and $atndrte = 0.82$?

[(a)] 1.537.

(b) -0.858.

(c) -0.092.

(d) -0.863.

16. **(Problem 2.)** What is like the effect of the average grade in the university access test (ACT) over the result of the final exam, ceteris paribus?

[(a)] If would depend on the values of the variable ACT.

(b) The effect would be negative because, despite the estimated coefficient of the variable ACT^2 ($\widehat{\beta}_5$) being positive, it is smaller in comparison with the estimated coefficient $\widehat{\beta}_3$, which is negative.

(c) The effect would be negative because the estimated effect of the variable ACT, $\widehat{\beta}_3$, is negative.

(d) The effect would be positive because the estimated coefficient of the variable ACT^2, $\widehat{\beta}_5$, is positive and since the variable ACT appears as a square, the sign of this variable will prevail.

17. **(Problem 2.)** From which value of the grading on, having an extra point in the access to university test, ACT, increases the result of the standardized final test?

(a) From a grading of approximately 0.45 points.

(b) From a grading of approximately 28.44 points.

[(c)] From a grading of approximately 14.22 points.

(d) From a grading of approximately 12.8 points.

18. **(Problem 1.)** The endogeneity problem of regressors in linear models causes that the OLS estimates,

   (i) Are generally biased.

   (ii) Need to be compared with corrected standard errors (White).

   (iii) Are not efficient, though in general are always consistent.

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

19. **(Problem 1.)** In a model of two simultaneous equations,

   (i) It is only necessary to estimate one estructural equation, since in general the dependent variable of the second equation show up as a regressor in the first one.

   (ii) None of the two equations can describe a causal relationship.

   (iii) If one estructural equation is identified, the other equation, and, therefore, the system are identified.

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

20. **(Problem 1.)** An instrumental variable for an equation,

   (i) It is an exogenous regressor in this equation.

   (ii) It is a dependent variable from other equation.

   (iii) Has not to be correlated with the error term.

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

21. **(Problem 1.)** The fact that the variable *open* is endogenoeus in the first equation,

(i) Depends on whether γ_{21} is different from zero or not.

(ii) Depends on whether the error terms u_1 and u_2 are correlated or not.

(iii) Depends on whether δ_{22} is different from zero or not.

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

22. **(Problem 1.)** Assuming that *open* is endogenous, is the first equation identified?

(a) It is identified because the F is significative in Output 4.
(b) It is not identified because log(*pcinc*) is not significative in the reduced form of *open*.
[(c)] It is identified, because log (*land*) is significative in the reduced form of *open*.
(d) It is identified because log(*land*) is significative in Output 2.

23. **(Problem 1.)** Assuming that inf is endogenous, study the identification of the second equation:

(i) It is identified because it has an additional exogenous variable included.

(ii) It is not identified because log (*pcinc*) is not significative in Output 1.

(iii) It is identified because log(*land*) is significative in Output 4.

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

24. **(Problem 1.)** We are interested in obtaining consistent estimates of the first equation, assuming that *open* is endogenous and that the variables log(*pcinc*) and log (*land*)) are exogenous.

(a) The estimates in Output 5 are not consistent because log (*pcinc*) is not partially correlated with the explanatory variable *open* (Output 4).
(b) The estimates in Output 5 are consistent because log(*pcinc*) and log (*land*) are jointly significative in Output 4.
(c) The estimates in Output 1 are consistent because the equation is identified.
[(d)] The estimates in Output 5 are consistent because log (*land*) is partially correlated with the explanatory variable *open* (Output 4).

25. **(Problem 1.)** Given the empirical evidence, can be concluded that γ_{12} is different from zero?

(i) Yes, attending to the estimation in Output 1, because OLS is always valid if the equation is identified.

(ii) Yes, attending to the instrumental variables estimation.

(iii) We do not know, because the first equation can not be identified in any case.

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

26. **(Problem 1.)** Taking into account the result of Hausman test (Output 5) and any other relevant information, we can conclude that,

(i) The estimates in Output 1 are consistent.

(ii) The estimates in Output 5 are more efficient than those in Output 1.

(iii) There is no empirical evidence to conclude that the variable *open* is endogenous in the first equation.

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

27. **(Problem 1.)** The result of Hausman test (Output 5):

(i) Shows that the first equation is over-identified because the OLS estimates are consistent.

(ii) Needs the equation to be over-identified to make sense.

(iii) Shows that the estimation in Output 5 is not consistent.

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

28. **(Problem 1.)** Concerning the estimation of a system of two simultaneous equations by 2SLS,

   (i) The first step in the 2SLS to estimate one equation looks for an optimal instrumental variable when this equation is over-identified, but if the other equation is not identified, the estimation is not consistent.

   (ii) The two steps in 2SLS can be done by means of OLS regressions.

   (iii) The second step in the 2SLS method can be done as an instrumental variable regression, using as instrument only the optimal instrumental variable from the first step.

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

29. **(Problem 1.)** If we have available one additional valid instrument for *open* in the first equation,

   (i) The 2SLS estimation would be more efficient than that of Output 5 if we additionally assume that $\log(land)$ is also a valid instrument.

   (ii) We could test whether $\log(land)$ is uncorrelated with u_1.

   (iii) We would not need observations of $\log(land)$ to obtain consistent estimates of the parameters of the first equation.

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

30. **(Problem 1.)** If we include the residuals from Output 4 as an additional regressor in the OLS estimation of the first equation (Output 1), and assuming that this equation is identified:

   (i) The estimates of the coefficients of *open* and of *lpcinc* would be consistent.

   (ii) The estimates of the coefficients of *open* and of *lpcinc* would be consistent, but less efficient that those from Output 5.

   (iii) We could do an endogeneity test for *open* by means of a t test.

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

31. **(Problem 1.)** If we reject the null hypothesis of the endogeneity test,

(i) We conclude that the regressor is endogenous, and therefore the OLS estimates are inconsistent.

(ii) We conclude that the regressor is exogenous, and therefore the instrumental variable estimates are less efficient that the OLS estimates.

(iii) We conclude that the regressor is exogenous, but that the OLS estimates could be inconsistent.

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

32. **(Problem 1.)** In case of over-identification,

(i) The equation can not be estimated because the parameters take too many alternative values.

(ii) We can not define an efficient instrumental variables estimate.

(iii) We can not do an exogeneity test, because we do not have enough degrees of freedom.

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

33. **(Problem 1.)** The efficiency of the 2SLS,

(i) Increases with the correlation between the endogenous regressor and the instrumental variables, measured through an R^2 coefficient.

(ii) Decreases with the variability of the endogenous regressor, since instrumental variable methods are less efficient than OLS.

(iii) Does not depend on the variance of the error from the original equation, since 2SLS is based on two different regressions.

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

34. **(Problem 1.)** The weak instruments case,

(i) May arise only when we have an over-identified equation, because each instrument can only explain an small proportion of the variability of the endogenous regressor.

(ii) Arises when the instrument and the endogenous regressor have a low correlation, leading to inefficient instrumental variable estimates.

(iii) May cause serious problems of inconsistency of the instrumental variables estimate if the instrument is correlated, though minimally, with the error term.

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

(b) Only (i) is true.

(c) Only (ii) is true.

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

Universidad Carlos III de Madrid
ECONOMETRICS I
Academic year 2006/07
FINAL EXAM
September 6, 2008

Exam Type: 4

TIME: 2 HOURS 30 MINUTES

1. **(Problem 1.)** The endogeneity problem of regressors in linear models causes that the OLS estimates,

   (i) Are generally biased.

   (ii) Need to be compared with corrected standard errors (White).

   (iii) Are not efficient, though in general are always consistent.

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

2. **(Problem 1.)** In a model of two simultaneous equations,

   (i) It is only necessary to estimate one estructural equation, since in general the dependent variable of the second equation show up as a regressor in the first one.

   (ii) None of the two equations can describe a causal relationship.

   (iii) If one estructural equation is identified, the other equation, and, therefore, the system are identified.

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

3. **(Problem 1.)** An instrumental variable for an equation,

   (i) It is an exogenous regressor in this equation.

   (ii) It is a dependent variable from other equation.

   (iii) Has not to be correlated with the error term.

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

4. **(Problem 1.)** The fact that the variable *open* is endogenoeus in the first equation,

   (i) Depends on whether γ_{21} is different from zero or not.

   (ii) Depends on whether the error terms u_1 and u_2 are correlated or not.

   (iii) Depends on whether δ_{22} is different from zero or not.

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

5. **(Problem 1.)** Assuming that *open* is endogenous, is the first equation identified?

   (a) It is identified because $\log(land)$ is significative in Output 2.
   (b) It is identified, because $\log(land)$ is significative in the reduced form of *open*.
   (c) It is not identified because $\log(pcinc)$ is not significative in the reduced form of *open*.
   (d) It is identified because the F is significative in Output 4.

6. **(Problem 1.)** Assuming that inf is endogenous, study the identification of the second equation:

   (i) It is identified because it has an additional exogenous variable included.

   (ii) It is not identified because $\log(pcinc)$ is not significative in Output 1.

   (iii) It is identified because $\log(land)$ is significative in Output 4.

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

7. **(Problem 1.)** We are interested in obtaining consistent estimates of the first equation, assuming that *open* is endogenous and that the variables $\log(pcinc)$ and $\log(land))$ are exogenous.

   (a) The estimates in Output 5 are consistent because $\log(land)$ is partially correlated with the explanatory variable *open* (Output 4).

   (b) The estimates in Output 1 are consistent because the equation is identified.

   (c) The estimates in Output 5 are consistent because $\log(pcinc)$ and $\log(land)$ are jointly significative in Output 4.

   (d) The estimates in Output 5 are not consistent because $\log(pcinc)$ is not partially correlated with the explanatory variable *open* (Output 4).

8. **(Problem 1.)** Given the empirical evidence, can be concluded that γ_{12} is different from zero?

   (i) Yes, attending to the estimation in Output 1, because OLS is always valid if the equation is identified.

   (ii) Yes, attending to the instrumental variables estimation.

   (iii) We do not know, because the first equation can not be identified in any case.

   (a) Only (iii) is true.

   (b) Only (ii) is true.

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

   (d) All are true.

9. **(Problem 1.)** Taking into account the result of Hausman test (Output 5) and any other relevant information, we can conclude that,

   (i) The estimates in Output 1 are consistent.

   (ii) The estimates in Output 5 are more efficient than those in Output 1.

   (iii) There is no empirical evidence to conclude that the variable *open* is endogenous in the first equation.

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

   (b) All are true.

   (c) Only (i) is true.

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

10. **(Problem 1.)** The result of Hausman test (Output 5):

   (i) Shows that the first equation is over-identified because the OLS estimates are consistent.

   (ii) Needs the equation to be over-identified to make sense.

   (iii) Shows that the estimation in Output 5 is not consistent.

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

11. **(Problem 1.)** Concerning the estimation of a system of two simultaneous equations by 2SLS,

   (i) The first step in the 2SLS to estimate one equation looks for an optimal instrumental variable when this equation is over-identified, but if the other equation is not identified, the estimation is not consistent.

   (ii) The two steps in 2SLS can be done by means of OLS regressions.

   (iii) The second step in the 2SLS method can be done as an instrumental variable regression, using as instrument only the optimal instrumental variable from the first step.

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

12. **(Problem 1.)** If we have available one additional valid instrument for *open* in the first equation,

   (i) The 2SLS estimation would be more efficient than that of Output 5 if we additionally assume that $\log(land)$ is also a valid instrument.

   (ii) We could test whether $\log(land)$ is uncorrelated with u_1.

   (iii) We would not need observations of $\log(land)$ to obtain consistent estimates of the parameters of the first equation.

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

13. **(Problem 1.)** If we include the residuals from Output 4 as an additional regressor in the OLS estimation of the first equation (Output 1), and assuming that this equation is identified:

(i) The estimates of the coefficients of *open* and of *lpcinc* would be consistent.

(ii) The estimates of the coefficients of *open* and of *lpcinc* would be consistent, but less efficient that those from Output 5.

(iii) We could do an endogeneity test for *open* by means of a t test.

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

(b) Only (iii) is true.

(c) Only (i) is true.

(d) All are true.

14. **(Problem 1.)** If we reject the null hypothesis of the endogeneity test,

(i) We conclude that the regressor is endogenous, and therefore the OLS estimates are inconsistent.

(ii) We conclude that the regressor is exogenous, and therefore the instrumental variable estimates are less efficient that the OLS estimates.

(iii) We conclude that the regressor is exogenous, but that the OLS estimates could be inconsistent.

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

(b) Only (i) is true.

(c) Only (iii) is true.

(d) All are false.

15. **(Problem 1.)** In case of over-identification,

(i) The equation can not be estimated because the parameters take too many alternative values.

(ii) We can not define an efficient instrumental variables estimate.

(iii) We can not do an exogeneity test, because we do not have enough degrees of freedom.

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

(b) All are false.

(c) Only (i) is true.

(d) Only (ii) is true.

16. **(Problem 1.)** The efficiency of the 2SLS,

(i) Increases with the correlation between the endogenous regressor and the instrumental variables, measured through an R^2 coefficient.

(ii) Decreases with the variability of the endogenous regressor, since instrumental variable methods are less efficient than OLS.

(iii) Does not depend on the variance of the error from the original equation, since 2SLS is based on two different regressions.

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

17. **(Problem 1.)** The weak instruments case,

(i) May arise only when we have an over-identified equation, because each instrument can only explain an small proportion of the variability of the endogenous regressor.

(ii) Arises when the instrument and the endogenous regressor have a low correlation, leading to inefficient instrumental variable estimates.

(iii) May cause serious problems of inconsistency of the instrumental variables estimate if the instrument is correlated, though minimally, with the error term.

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

18. **(Problem 2.)** In the problem discussed,

(i) It is not a good idea to compare two alternative models with the determination coefficient, R^2, since we know in advance that (2) will always fit better because it contains more regressors.

(ii) If all coefficients in the OLS regression are significant, the model is correctly specified.

(iii) Chose the right model is very important, since otherwise some of the necessary hypothesis to perform valid inference could be not valid.

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

19. **(Problem 2.)** To choose between the two models, the researcher estimates the following equation:

$$stndfnl = \beta_0 + \beta_1 atndrte + \beta_2 \log(priGPA) + \beta_3 \log(ACT) + \\ + \gamma_1 priGPA + \gamma_2 ACT + \gamma_3 priGPA^2 + \gamma_4 ACT^2 + u \quad (3)$$

If she wants to test whether model (2) is correct, which is the null hypothesis that she should test for in equation (3)?

(a) $H_0 : \gamma_1 = \gamma_2 = \gamma_3 = \gamma_4 = 0.$

(b) $H_0 : \beta_1 = \beta_2 = \beta_3 = 0.$

(c) $H_0 : \beta_2 = \beta_3 = 0.$

(d) $H_0 : \gamma_1 = \gamma_2 = 0.$

20. **(Problem 2.)** The F statistic to test H_0 in the previous question takes the value 1,00081 with a p-value of 0.368.

(a) The researcher rejects H_0 and decides that model (1) is correct.

(b) The researcher rejects H_0 and decides that model (2) is correct.

(c) The researcher can not reject H_0 and decides that model (2) is correct.

(d) The researcher decides that both models are correct.

21. **(Problem 2.)** To choose between both models, the researcher considers an alternative method estimating the following equation,

$$stndfnl = \beta_0 + \beta_1 atndrte + \beta_2 \log(priGPA) + \beta_3 \log(ACT) + \theta_1 \widehat{y} + u,$$

where $\widehat{y}$ are the adjusted values from the estimation of model (2).

(i) If we reject $H_0 : \theta_1 = 0$, then we conclude that model (1) has specification problems.

(ii) If we reject $H_0 : \theta_1 = 0$, then we conclude that model (2) has specification problems.

(iii) If we reject $H_0 : \theta_1 = 0$, then we conclude that model (2) is correctly specified.

(a) All are false.

(b) Only (i) is true.

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

(d) Only (ii) is true.

22. **(Problem 2.)** It seems that the researcher prefers model (2), so she continues testing model (2). She estimates model (2) by OLS and obtains $R^2 = 0.226$. Then, she omits from the model the explanatory variables $priGPA^2$ and ACT^2, and obtains $R^2 = 0.211$. Which of the following conclusions is true?

   (a) The F statistic for the joint significance test of the two variables $priGPA^2$ and ACT^2 is 11.081 with a p-value of 0.000018, so that there is no evidence of wrong functional specification in model (2).

   (b) Model (2) will be much better than the model without the explanatory variables $priGPA^2$ and ACT^2 since it has a higher R^2.

   (c) The F statistic for the joint significance test of the two variables $priGPA^2$ and ACT^2 is 11.081 with a p-value of 0.000018, so that there is evidence of wrong functional specification in model (2).

   (d) The test is not valid to check the specification of the model because the previous hypotheses are not nested.

23. **(Problem 2.)** The researcher considers an alternative method: Ramsey's RESET test.

   (i) RESET test uses the OLS residuals and its squares to check if they help to explain the dependent variable.

   (ii) If we reject the null hypothesis of the RESET test, we confirm the correct specification of the model.

   (iii) RESET test allows us also to test for the presence of heteroskedasticity.

   (a) All are false.

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

   (c) Only (iii) is true.

   (d) Only (ii) is true.

24. **(Problem 2.)** If Model (2) is correctly specified, and u satisfies all usual assumptions, but we do not include the variable ACT^2 in the OLS estimation,

   (i) The OLS estimates of the remaining parameters will be biased if $\beta_5 \neq 0$, because at least ACT and ACT^2 will be correlated.

   (ii) The OLS estimation of the remaining parameters will be consistent, but the new errors will be heteroskedastic and usual inference will not be valid, if $\beta_5 \neq 0$.

   (iii) OLS estimation of the remaining parameters, except β_3, will be consistent if $\beta_5 = 0$, since ACT and ACT^2 will be correlated.

   (a) All are false.

   (b) Only (ii) is true.

   (c) Only (i) is true.

   (d) Only (iii) is true.

25. **(Problem 2.)** In the next step of the empirical analysis, the researcher wishes to test for the presence of heteroskedasticity in the model using White's test. Assume that given the results of previous tests, she decides to focus on model (2). How is White's test implemented?

    (a) We calculate the residuals of the OLS fit of model (2). We estimate the regression of the squares of the OLS residuals on all explanatory variables of model (2), their squares, and all cross products. We obtain the R^2 of this estimation and we compute the test statistic nR^2.

    (b) In model (2) we add the squares of the explanatory variables and then we calculate the F test for global significance.

    (c) In model (2) we add the squares of the explanatory variables and all cross products and then we calculate the F statistic for global significance.

    (d) We calculate the residuals of the OLS fit of model (2). We estimate the regression of the squares of the OLS residuals on the squares of all explanatory variables of model (2). We obtain the R^2 of this estimation and we compute the test statistic nR^2.

26. **(Problem 2.)** The result of White's heteroskedasticity test is 13.993, with a p-value of 0.72. Which is the conclusion of the researcher on the null hypothesis of White's test, and therefore, on the presence of heteroskedasticity in model (2) given the empirical evidence?

    (a) She can not reject H_0 and concludes that there is not heteroskedasticity.

    (b) She rejects H_0 and concludes that there is heteroskedasticity.

    (c) She can not reject H_0 and concludes that there is heteroskedasticity.

    (d) She rejects H_0 and concludes that there is not heteroskedasticity.

27. **(Problem 2.)** In the presence of heteroskedasticity,

    (i) OLS estimates stop being consistent.

    (ii) To perform valid inference is always necessary to estimate the equation by means of generalized least squares.

    (iii) White's corrected standard errors are only valid if we know the form of the heteroskedasticity.

    (a) All are false.

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

    (c) Only (iii) is true.

    (d) Only (ii) is true.

28. **(Problem 2.)** Now assume that the researcher decides to include the interaction term $priGPA * atndrte$ in model (2):

$$\begin{aligned} stndfnl = {} & \beta_0 + \beta_1 atndrte + \beta_2 priGPA + \beta_3 ACT + \beta_4 priGPA^2 + \\ & + \beta_5 ACT^2 + \beta_6 priGPA * atndrte + u. \end{aligned} \tag{4}$$

The results of the OLS estimation of this new model are as follows,

$$\begin{aligned} \widehat{stndfnl} = {} & \underset{(1.36)}{2.05} - \underset{(0.0102)}{0.0067} atndrte - \underset{(0.48)}{1.63} priGPA - \underset{(0.098)}{0.128} ACT + \underset{(0.101)}{0.296} priGPA^2 + \\ & + \underset{(0.0022)}{0.0045} ACT^2 + \underset{(??)}{0.0056} priGPA * atndrte \\ & n = 680 \qquad R^2 = 0.228 \end{aligned}$$

(a) The t test for the coefficient of the interaction is equal to 2.13, and therefore the estimate is significative.

(b) The effect of the interaction is without doubt significative, since the R^2 has increased.

(c) The t test for the coefficient of the interaction is equal to 1.32, and therefore the estimate is not significative.

(d) There is not enough information to test for the significance of the new estimate.

29. **(Problem 2.)** The researcher wishes to calculate the partial effect of attendance to class (*atndrte*) on the result of the final exam (*stndfnl*). How would you estimate that effect?

(a) $\widehat{\beta}_1 + \widehat{\beta}_6$.

(b) Is simply the estimated coefficient $\widehat{\beta}_1$.

(c) $\widehat{\beta}_1 + \widehat{\beta}_6 priGPA$.

(d) $\widehat{\beta}_1 + \widehat{\beta}_2 + \widehat{\beta}_6 priGPA$.

30. **(Problem 2.)** If the averaged value of *priGPA* in the sample is 2.59, how would affect an increment of 10 percentage points of class attendance on the result of the final exam for a student with this value of *priGPA*?

(a) An increment of 10 percentage points in the class attendance leads to a decrement of *stndfnl* of 0.067.

(b) An increment of 10 percentage points in the class attendance leads to an increment of *stndfnl* of 0.078.

(c) An increment of 10 percentage points in the class attendance leads to an increment of *stndfnl* of 0.0078.

(d) An increment of 10 percentage points in the class attendance leads to an increment of *stndfnl* of 1.622.

31. **(Problem 2.)** The standard error of estimating the previous effect is

(a) There is no enough available information to calculate it.

(b) 0.15 if the OLS estimates of the coefficients were uncorrelated.

(c) 0.25 if the OLS estimates of the coefficients were uncorrelated.

(d) 0.102.

32. **(Problem 2.)** Using Model (4), which is the partial effect of the grade of previous exams (*priGPA*) over the final exam grading (*stndfnl*), when we fix $priGPA = 2.59$ and $atndrte = 0.82$?

(a) -0.863.

(b) -0.092.

(c) -0.858.

(d) 1.537.

33. **(Problem 2.)** What is like the effect of the average grade in the university access test (ACT) over the result of the final exam, ceteris paribus?

(a) The effect would be positive because the estimated coefficient of the variable ACT^2, $\widehat{\beta}_5$, is positive and since the variable ACT appears as a square, the sign of this variable will prevail.

(b) The effect would be negative because the estimated effect of the variable *ACT*, $\widehat{\beta}_3$, is negative.

(c) The effect would be negative because, despite the estimated coefficient of the variable ACT^2 ($\widehat{\beta}_5$) being positive, it is smaller in comparison with the estimated coefficient $\widehat{\beta}_3$, which is negative.

(d) If would depend on the values of the variable ACT.

34. **(Problem 2.)** From which value of the grading on, having an extra point in the access to university test, ACT, increases the result of the standardized final test?

(a) From a grading of approximately 12.8 points.

(b) From a grading of approximately 14.22 points.

(c) From a grading of approximately 28.44 points.

(d) From a grading of approximately 0.45 points.

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.)** The endogeneity problem of regressors in linear models causes that the OLS estimates,

   (i) Are generally biased.

   (ii) Need to be compared with corrected standard errors (White).

   (iii) Are not efficient, though in general are always consistent.

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

2. **(Problem 1.)** In a model of two simultaneous equations,

   (i) It is only necessary to estimate one estructural equation, since in general the dependent variable of the second equation show up as a regressor in the first one.

   (ii) None of the two equations can describe a causal relationship.

   (iii) If one estructural equation is identified, the other equation, and, therefore, the system are identified.

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

3. **(Problem 1.)** An instrumental variable for an equation,

(i) It is an exogenous regressor in this equation.

(ii) It is a dependent variable from other equation.

(iii) Has not to be correlated with the error term.

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

4. **(Problem 1.)** The fact that the variable *open* is endogenoeus in the first equation,

(i) Depends on whether γ_{21} is different from zero or not.

(ii) Depends on whether the error terms u_1 and u_2 are correlated or not.

(iii) Depends on whether δ_{22} is different from zero or not.

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

5. **(Problem 1.)** Assuming that *open* is endogenous, is the first equation identified?

(a) It is identified because $\log(land)$ is significative in Output 2.
[(b)] It is identified, because $\log(land)$ is significative in the reduced form of *open*.
(c) It is not identified because $\log(pcinc)$ is not significative in the reduced form of *open*.
(d) It is identified because the F is significative in Output 4.

6. **(Problem 1.)** Assuming that inf is endogenous, study the identification of the second equation:

(i) It is identified because it has an additional exogenous variable included.

(ii) It is not identified because $\log(pcinc)$ is not significative in Output 1.

(iii) It is identified because $\log(land)$ is significative in Output 4.

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

7. **(Problem 1.)** We are interested in obtaining consistent estimates of the first equation, assuming that *open* is endogenous and that the variables $\log(pcinc)$ and $\log(land)$) are exogenous.

   - [(a)] The estimates in Output 5 are consistent because $\log(land)$ is partially correlated with the explanatory variable *open* (Output 4).
   - (b) The estimates in Output 1 are consistent because the equation is identified.
   - (c) The estimates in Output 5 are consistent because $\log(pcinc)$ and $\log(land)$ are jointly significative in Output 4.
   - (d) The estimates in Output 5 are not consistent because $\log(pcinc)$ is not partially correlated with the explanatory variable *open* (Output 4).

8. **(Problem 1.)** Given the empirical evidence, can be concluded that γ_{12} is different from zero?

   (i) Yes, attending to the estimation in Output 1, because OLS is always valid if the equation is identified.

   (ii) Yes, attending to the instrumental variables estimation.

   (iii) We do not know, because the first equation can not be identified in any case.

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

9. **(Problem 1.)** Taking into account the result of Hausman test (Output 5) and any other relevant information, we can conclude that,

   (i) The estimates in Output 1 are consistent.

   (ii) The estimates in Output 5 are more efficient than those in Output 1.

   (iii) There is no empirical evidence to conclude that the variable *open* is endogenous in the first equation.

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

10. **(Problem 1.)** The result of Hausman test (Output 5):

(i) Shows that the first equation is over-identified because the OLS estimates are consistent.

(ii) Needs the equation to be over-identified to make sense.

(iii) Shows that the estimation in Output 5 is not consistent.

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

11. **(Problem 1.)** Concerning the estimation of a system of two simultaneous equations by 2SLS,

(i) The first step in the 2SLS to estimate one equation looks for an optimal instrumental variable when this equation is over-identified, but if the other equation is not identified, the estimation is not consistent.

(ii) The two steps in 2SLS can be done by means of OLS regressions.

(iii) The second step in the 2SLS method can be done as an instrumental variable regression, using as instrument only the optimal instrumental variable from the first step.

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

12. **(Problem 1.)** If we have available one additional valid instrument for *open* in the first equation,

(i) The 2SLS estimation would be more efficient than that of Output 5 if we additionally assume that $\log(land)$ is also a valid instrument.

(ii) We could test whether $\log(land)$ is uncorrelated with u_1.

(iii) We would not need observations of $\log(land)$ to obtain consistent estimates of the parameters of the first equation.

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

13. **(Problem 1.)** If we include the residuals from Output 4 as an additional regressor in the OLS estimation of the first equation (Output 1), and assuming that this equation is identified:

(i) The estimates of the coefficients of *open* and of *lpcinc* would be consistent.

(ii) The estimates of the coefficients of *open* and of *lpcinc* would be consistent, but less efficient that those from Output 5.

(iii) We could do an endogeneity test for *open* by means of a t test.

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

14. **(Problem 1.)** If we reject the null hypothesis of the endogeneity test,

(i) We conclude that the regressor is endogenous, and therefore the OLS estimates are inconsistent.

(ii) We conclude that the regressor is exogenous, and therefore the instrumental variable estimates are less efficient that the OLS estimates.

(iii) We conclude that the regressor is exogenous, but that the OLS estimates could be inconsistent.

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

15. **(Problem 1.)** In case of over-identification,

(i) The equation can not be estimated because the parameters take too many alternative values.

(ii) We can not define an efficient instrumental variables estimate.

(iii) We can not do an exogeneity test, because we do not have enough degrees of freedom.

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

16. **(Problem 1.)** The efficiency of the 2SLS,

(i) Increases with the correlation between the endogenous regressor and the instrumental variables, measured through an R^2 coefficient.

(ii) Decreases with the variability of the endogenous regressor, since instrumental variable methods are less efficient than OLS.

(iii) Does not depend on the variance of the error from the original equation, since 2SLS is based on two different regressions.

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

(b) Only (ii) is true.

[(c)] Only (i) is true.

(d) All are true.

17. **(Problem 1.)** The weak instruments case,

(i) May arise only when we have an over-identified equation, because each instrument can only explain an small proportion of the variability of the endogenous regressor.

(ii) Arises when the instrument and the endogenous regressor have a low correlation, leading to inefficient instrumental variable estimates.

(iii) May cause serious problems of inconsistency of the instrumental variables estimate if the instrument is correlated, though minimally, with the error term.

[(a)] (ii) and (iii) are true.

(b) Only (ii) is true.

(c) Only (i) is true.

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

18. **(Problem 2.)** In the problem discussed,

   (i) It is not a good idea to compare two alternative models with the determination coefficient, R^2, since we know in advance that (2) will always fit better because it contains more regressors.

   (ii) If all coefficients in the OLS regression are significant, the model is correctly specified.

   (iii) Chose the right model is very important, since otherwise some of the necessary hypothesis to perform valid inference could be not valid.

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

19. **(Problem 2.)** To choose between the two models, the researcher estimates the following equation:

$$stndfnl = \beta_0 + \beta_1 atndrte + \beta_2 \log(priGPA) + \beta_3 \log(ACT) + \\ + \gamma_1 priGPA + \gamma_2 ACT + \gamma_3 priGPA^2 + \gamma_4 ACT^2 + u \tag{3}$$

   If she wants to test whether model (2) is correct, which is the null hypothesis that she should test for in equation (3)?

   (a) $H_0 : \gamma_1 = \gamma_2 = \gamma_3 = \gamma_4 = 0.$
   (b) $H_0 : \beta_1 = \beta_2 = \beta_3 = 0.$
   [(c)] $H_0 : \beta_2 = \beta_3 = 0.$
   (d) $H_0 : \gamma_1 = \gamma_2 = 0.$

20. **(Problem 2.)** The F statistic to test H_0 in the previous question takes the value 1,00081 with a p-value of 0.368.

   (a) The researcher rejects H_0 and decides that model (1) is correct.
   (b) The researcher rejects H_0 and decides that model (2) is correct.
   [(c)] The researcher can not reject H_0 and decides that model (2) is correct.
   (d) The researcher decides that both models are correct.

21. **(Problem 2.)** To choose between both models, the researcher considers an alternative method estimating the following equation,

$$stndfnl = \beta_0 + \beta_1 atndrte + \beta_2 \log(priGPA) + \beta_3 \log(ACT) + \theta_1 \widehat{y} + u,$$

where $\widehat{y}$ are the adjusted values from the estimation of model (2).

(i) If we reject $H_0 : \theta_1 = 0$, then we conclude that model (1) has specification problems.

(ii) If we reject $H_0 : \theta_1 = 0$, then we conclude that model (2) has specification problems.

(iii) If we reject $H_0 : \theta_1 = 0$, then we conclude that model (2) is correctly specified.

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

22. **(Problem 2.)** It seems that the researcher prefers model (2), so she continues testing model (2). She estimates model (2) by OLS and obtains $R^2 = 0.226$. Then, she omits from the model the explanatory variables $priGPA^2$ and ACT^2, and obtains $R^2 = 0.211$. Which of the following conclusions is true?

[(a)] The F statistic for the joint significance test of the two variables $priGPA^2$ and ACT^2 is 11.081 with a p-value of 0.000018, so that there is no evidence of wrong functional specification in model (2).

(b) Model (2) will be much better than the model without the explanatory variables $priGPA^2$ and ACT^2 since it has a higher R^2.

(c) The F statistic for the joint significance test of the two variables $priGPA^2$ and ACT^2 is 11.081 with a p-value of 0.000018, so that there is evidence of wrong functional specification in model (2).

(d) The test is not valid to check the specification of the model because the previous hypotheses are not nested.

23. **(Problem 2.)** The researcher considers an alternative method: Ramsey's RESET test.

(i) RESET test uses the OLS residuals and its squares to check if they help to explain the dependent variable.

(ii) If we reject the null hypothesis of the RESET test, we confirm the correct specification of the model.

(iii) RESET test allows us also to test for the presence of heteroskedasticity.

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

24. **(Problem 2.)** If Model (2) is correctly specified, and u satisfies all usual assumptions, but we do not include the variable ACT^2 in the OLS estimation,

(i) The OLS estimates of the remaining parameters will be biased if $\beta_5 \neq 0$, because at least ACT and ACT^2 will be correlated.

(ii) The OLS estimation of the remaining parameters will be consistent, but the new errors will be heteroskedastic and usual inference will not be valid, if $\beta_5 \neq 0$.

(iii) OLS estimation of the remaining parameters, except β_3, will be consistent if $\beta_5 = 0$, since ACT and ACT^2 will be correlated.

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

25. **(Problem 2.)** In the next step of the empirical analysis, the researcher wishes to test for the presence of heteroskedasticity in the model using White's test. Assume that given the results of previous tests, she decides to focus on model (2). How is White's test implemented?

   - [(a)] We calculate the residuals of the OLS fit of model (2). We estimate the regression of the squares of the OLS residuals on all explanatory variables of model (2), their squares, and all cross products. We obtain the R^2 of this estimation and we compute the test statistic nR^2.
   - (b) In model (2) we add the squares of the explanatory variables and then we calculate the F test for global significance.
   - (c) In model (2) we add the squares of the explanatory variables and all cross products and then we calculate the F statistic for global significance.
   - (d) We calculate the residuals of the OLS fit of model (2). We estimate the regression of the squares of the OLS residuals on the squares of all explanatory variables of model (2). We obtain the R^2 of this estimation and we compute the test statistic nR^2.

26. **(Problem 2.)** The result of White's heteroskedasticity test is 13.993, with a p-value of 0.72. Which is the conclusion of the researcher on the null hypothesis of White's test, and therefore, on the presence of heteroskedasticity in model (2) given the empirical evidence?

   - [(a)] She can not reject H_0 and concludes that there is not heteroskedasticity.
   - (b) She rejects H_0 and concludes that there is heteroskedasticity.
   - (c) She can not reject H_0 and concludes that there is heteroskedasticity.
   - (d) She rejects H_0 and concludes that there is not heteroskedasticity.

27. **(Problem 2.)** In the presence of heteroskedasticity,

   (i) OLS estimates stop being consistent.

   (ii) To perform valid inference is always necessary to estimate the equation by means of generalized least squares.

   (iii) White's corrected standard errors are only valid if we know the form of the heteroskedasticity.

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

28. **(Problem 2.)** Now assume that the researcher decides to include the interaction term $priGPA * atndrte$ in model (2):

$$stndfnl = \beta_0 + \beta_1 atndrte + \beta_2 priGPA + \beta_3 ACT + \beta_4 priGPA^2 + \\ + \beta_5 ACT^2 + \beta_6 priGPA * atndrte + u. \quad (4)$$

The results of the OLS estimation of this new model are as follows,

$$\widehat{stndfnl} = \underset{(1.36)}{2.05} - \underset{(0.0102)}{0.0067} atndrte - \underset{(0.48)}{1.63} priGPA - \underset{(0.098)}{0.128} ACT + \underset{(0.101)}{0.296} priGPA^2 + \\ + \underset{(0.0022)}{0.0045} ACT^2 + \underset{(??)}{0.0056} priGPA * atndrte$$

$$n = 680 \qquad R^2 = 0.228$$

(a) The t test for the coefficient of the interaction is equal to 2.13, and therefore the estimate is significative.

(b) The effect of the interaction is without doubt significative, since the R^2 has increased.

[(c)] The t test for the coefficient of the interaction is equal to 1.32, and therefore the estimate is not significative.

(d) There is not enough information to test for the significance of the new estimate.

29. **(Problem 2.)** The researcher wishes to calculate the partial effect of attendance to class (*atndrte*) on the result of the final exam (*stndfnl*). How would you estimate that effect?

(a) $\widehat{\beta}_1 + \widehat{\beta}_6$.

(b) Is simply the estimated coefficient $\widehat{\beta}_1$.

[(c)] $\widehat{\beta}_1 + \widehat{\beta}_6 priGPA$.

(d) $\widehat{\beta}_1 + \widehat{\beta}_2 + \widehat{\beta}_6 priGPA$.

30. **(Problem 2.)** If the averaged value of *priGPA* in the sample is 2.59, how would affect an increment of 10 percentage points of class attendance on the result of the final exam for a student with this value of $priGPA$?

(a) An increment of 10 percentage points in the class attendance leads to a decrement of *stndfnl* of 0.067.

[(b)] An increment of 10 percentage points in the class attendance leads to an increment of *stndfnl* of 0.078.

(c) An increment of 10 percentage points in the class attendance leads to an increment of *stndfnl* of 0.0078.

(d) An increment of 10 percentage points in the class attendance leads to an increment of *stndfnl* of 1.622.

31. **(Problem 2.)** The standard error of estimating the previous effect is

(a) There is no enough available information to calculate it.

[(b)] 0.15 if the OLS estimates of the coefficients were uncorrelated.

(c) 0.25 if the OLS estimates of the coefficients were uncorrelated.

(d) 0.102.

32. **(Problem 2.)** Using Model (4), which is the partial effect of the grade of previous exams (*priGPA*) over the final exam grading (*stndfnl*), when we fix $priGPA = 2.59$ and $atndrte = 0.82$?

(a) -0.863.

(b) -0.092.

(c) -0.858.

[(d)] 1.537.

33. **(Problem 2.)** What is like the effect of the average grade in the university access test (ACT) over the result of the final exam, ceteris paribus?

(a) The effect would be positive because the estimated coefficient of the variable ACT^2, $\widehat{\beta}_5$, is positive and since the variable ACT appears as a square, the sign of this variable will prevail.

(b) The effect would be negative because the estimated effect of the variable *ACT*, $\widehat{\beta}_3$, is negative.

(c) The effect would be negative because, despite the estimated coefficient of the variable ACT^2 ($\widehat{\beta}_5$) being positive, it is smaller in comparison with the estimated coefficient $\widehat{\beta}_3$, which is negative.

[(d)] If would depend on the values of the variable ACT.

34. **(Problem 2.)** From which value of the grading on, having an extra point in the access to university test, ACT, increases the result of the standardized final test?

(a) From a grading of approximately 12.8 points.

[(b)] From a grading of approximately 14.22 points.

(c) From a grading of approximately 28.44 points.

(d) From a grading of approximately 0.45 points.