UNIVERSIDAD CARLOS III DE MADRID
ECONOMETRICS I
Academic year 2005/06
FINAL EXAM

January, 30, 2006

PROBLEM TEXT

PROBLEM 1: PROBABILITY OF DENYING A MORTGAGE

We consider three specifications to model the probability of NOT being granted a mortgage:

$$E(Y|X_1, X_2) = F(\beta_0 + \beta_1 X_1 + \beta_2 X_2)$$

where, for each applicant:
Y = binary variable which equals 1 if the mortgage application is denied and 0 if granted;
$X_1 = DEUDA$ = applicant's debt rate –ratio of debt to income– (excluding the mortgage under consideration);
$X_2 = NEG$ = binary variable which equals 1 if the applicant is black and 0 otherwise (if white).

Given a sample of 2380 applications in Boston (USA), three alternative models have been estimated:

OUTPUT 1

Dependent Variable: Y
Method: Least Squares
Sample: 1 2380
Included observations: 2380
White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	−0.091	0.029	−3.165	0.0016
DEUDA	0.559	0.089	6.307	0.0000
NEG	0.177	0.025	7.112	0.0000

R-squared	0.076003	Mean dependent var	0.119748
Adj. R-squared	0.075226	S.D. dependent var	0.324735
S.E. of regression	0.312282	Akaike info criterion	0.511438
Sum squared resid	231.8047	Schwarz criterion	0.518717
Log likelihood	−605.6107	F-statistic	97.76019
Durbin-Watson stat	1.517180	Prob (F-statistic)	0.00000

OUTPUT 2

Dependent Variable: Y
Method: ML - Binary Probit
Sample: 1 2380
Included observations: 2380
Convergence achieved after 4 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	−2.26	0.16	−14.225	0.0000
DEUDA	2.74	0.44	6.1738	0.0000
NEG	0.71	0.08	8.5146	0.0000
Mean dependent var	0.1197	S.D. dependent var		0.3247
S.E. of regression	0.3104	Akaike info criterion		0.6724
Sum squared resid	229.06	Schwarz criterion		0.6797
Log likelihood	−797.14	Hannan--Quinn criter.		0.6750
Restr. log likelihood	−872.09	Avg. log likelihood		−0.3349
LR statistic (2 df)	149.90	McFadden R-squared		0.0859
Probability (LR stat)	0.0000			
Obs. with Dep=0	2095	Total obs		2380
Obs. with Dep=1	285			

OUTPUT 3

Dependent Variable: Y
Method: ML - Binary Logit
Sample: 1 2380
Included observations: 2380
Convergence achieved after 5 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	−4.13	0.35	−11.93	0.0000
DEUDA	5.37	0.96	5.58	0.0000
NEG	1.27	0.15	8.71	0.0000
Mean dependent var	0.1197	S.D. dependent var		0.3247
S.E. of regression	0.3101	Akaike info criterion		0.6711
Sum squared resid	228.61	Schwarz criterion		0.6785
Log likelihood	−795.70	Hannan--Quinn criter.		0.6738
Restr. log likelihood	−872.09	Avg. log likelihood		−0.3343
LR statistic (2 df)	152.78	McFadden R-squared		0.0860
Probability (LR stat)	0.0000			
Obs. with Dep=0	2095	Total obs		2380
Obs. with Dep=1	285			

Nota: Remember that it is assumed in the probit model that, given (in our problem) $Z = \beta_0 + \beta_1 X_1 + \beta_2 X_2$, $F(Z) = \Phi(Z)$, where $\Phi(Z)$ is the cumulative distribution function of a $N(0,1)$ random variable, whose density function is $\phi(Z) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-Z^2}{2}\right) \simeq 0.3989 \times \exp\left(\frac{-Z^2}{2}\right)$, and in the logit model $F(Z) = \Lambda(Z)$, that is, the cumulative distribution function of a standard logistic random variable, $\Lambda(Z) = \frac{e^Z}{1+e^Z}$, whose density function is equal to $\lambda(Z) = \frac{e^Z}{(1+e^Z)^2}$.

Another researcher adds a measure of the applicant's risk of credit devolution (`MORC`) and a measure of the applicant's risk of mortgage devolution (`MORH`)

OUTPUT 4

```
Dependent Variable: Y
Method: ML - Binary Probit
Sample: 1 2380
Included observations: 2380
Convergence achieved after 5 iterations
```

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	−2.98	0.18	−16.30	0.0000
DEUDA	2.66	0.36	7.35	0.0000
NEG	0.51	0.09	5.74	0.0000
MORC	0.18	0.02	9.31	0.0000
MORH	0.19	0.07	2.82	0.0048

Mean dependent var	0.1197	S.D. dependent var	0.3247
S.E. of regression	0.3015	Akaike info criterion	0.6306
Sum squared resid	215.91	Schwarz criterion	0.6427
Log likelihood	−745.39	Hannan--Quinn criter.	0.6350
Restr. log likelihood	−872.09	Avg. log likelihood	−0.3132
LR statistic (4 df)	253.40	McFadden R-squared	0.1453
Probability (LR stat)	0.0000		
Obs. with Dep=0	2095	Total obs	2380
Obs. with Dep=1	285		

PROBLEMA 2: FERTILITY IMPACT ON WOMEN HOURS OF WORK

There is evidence that the labor market behavior of women is particularly determined by their fertility decisions. In order to evaluate the impact of fertility (measured by the number of children) on woman's usual weekly number of hours at work, we concentrate on the following specification:

$$\begin{aligned}\texttt{HRS} \;=\; & \beta_0 + \beta_1\texttt{WHITE} + \beta_2\texttt{BLACK} + \beta_3\texttt{HISPAN} \\ & + \beta_4\texttt{HIGHSC} + \beta_5\texttt{UNIV} + \beta_6\texttt{AGE} + \beta_7\texttt{AGE2} \\ & + \beta_8\texttt{SPOUSE} + \beta_9\left(\texttt{SPOUSE} \times \texttt{AGE}\right) + \beta_{10}\left(\texttt{SPOUSE} \times \texttt{AGE2}\right) \\ & + \beta_{11}\texttt{NCHILD} + \varepsilon\end{aligned} \tag{E.1}$$

where, for each woman:

HRS = number of hours worked per week;
WHITE = dummy variable that equals 1 if the woman is white and zero otherwise;
BLACK = dummy variable that equals 1 if the woman is black and zero otherwise;
HISPAN = dummy variable that equals 1 if the woman is hispanic and zero otherwise;
HIGHSC = dummy variable that equals 1 if the woman has only completed high school degree and zero otherwise;
UNIV = dummy variable that equals 1 if the woman has a university degree and zero otherwise.
AGE = age in years;
AGE2 = square of age;
SPOUSE = dummy variable that equals 1 if the woman has a spouse living at home and zero otherwise;
NCHILD = number of children younger than 18 years old that live at woman's home.

Note: There are four ethnic groups mutually excluyent: White, Black, Hispanic and Asian. There are only three (mutually excluyent) education degrees: primary or lower, high school, university.
To estimate this equation, we have household data for 69852 mothers from the 1980 US Census data.

Furthermore, we know that fertility decisions are correlated with unobserved characteristics that simultaneously affect labor decisions. For example, women with more market oriented abilities, not only might have in average higher wages (higher opportunity cost of staying at home) but also they might be the one with a higher cost of child bearing. Then we would expect that

$$C(\varepsilon, \texttt{NCHILD}) \neq 0.$$

whereas the rest of the right-hand-side variables in (E.1) **are not correlated with** other unobserved variables (ε).

Besides the variables mentioned above, we have information whether or not a woman has faced a multiple birth (MB). We understand that a woman had faced a multiple birth if she had had twins, triplet, quadruple or quintuple children in one birth. Then we can define the variable ***Multiple Births,*** MB:

MB = dummy variable that equals 1 if the woman has faced a multiple birth and zero otherwise.

Moreover, we know that $C(\mathtt{MB}, \varepsilon) = 0$.

We have obtained the following estimates:

OUTPUT 1

Dependent Variable: HRS
Method: Least Squares
Sample: 69852
Included observations: 69852

Variable	Coefficient	Std. Error	t-Statistic	Prob.
WHITE	1.3224	0.7901	1.67	0.094
BLACK	3.3607	0.8005	4.20	0.000
HISPAN	−0.5399	1.0007	−0.54	0.590
HIGHSC	4.4627	0.1770	25.22	0.000
UNIV	4.3868	0.2005	21.88	0.000
AGE	1.4505	0.0288	50.37	0.000
AGE2	−0.0207	0.0006	−36.04	0.000
SPOUSE	0.4258	0.4343	0.98	0.327
SPOUSE * AGE	−0.5668	0.0335	−16.90	0.000
SPOUSE * AGE2	0.0101	0.0006	15.42	0.000
$NCHILD$	−1.8506	0.0610	−30.32	0.000
C	0.2952	0.8568	0.34	0.730
R-squared		0.1344		
Adjusted R-squared		0.1343		
S.E. of regression		335.4738		

OUTPUT 2

Dependent Variable: $NCHILD$
Method: Least Squares
Sample: 69852
Included observations: 69852

Variable	Coefficient	Std. Error	t-Statistic	Prob.
WHITE	−0.3527	0.0482	−7.32	0.000
BLACK	0.2822	0.0488	5.78	0.000
HISPAN	0.0638	0.0610	1.05	0.296
HIGHSC	−0.4758	0.0106	−44.68	0.000
UNIV	−0.5453	0.0121	−45.22	0.000
AGE	0.0801	0.0017	46.26	0.000
AGE2	−0.0012	0.00003	−35.08	0.000
SPOUSE	0.2622	0.0265	9.90	0.000
SPOUSE * AGE	−0.0095	0.0020	−4.65	0.000
SPOUSE * AGE2	0.0001	0.00004	3.42	0.001
MB	1.2192	0.0255	47.79	0.000
C	1.5957	0.0519	30.73	0.000
R-squared		0.1568		
Adjusted R-squared		0.1566		

OUTPUT 3				
Dependent Variable: HRS				
Method: Two-Stage Least Squares				
Sample: 69852				
Included observations: 69852				
Instrument list: MB				
Variable Variable	Coefficient	Std. Error	t-Statistic	Prob.
WHITE	1.7045	0.8019	2.13	0.034
BLACK	3.0750	0.8077	3.81	0.000
HISPAN	−0.5925	1.0029	−0.59	0.555
HIGHSC	4.9467	0.2398	20.63	0.000
UNIV	4.9483	0.2746	18.02	0.000
AGE	1.3697	0.0395	34.69	0.000
AGE2	−0.0195	0.0007	−27.66	0.000
SPOUSE	0.1529	0.4446	0.34	0.731
SPOUSE ∗ AGE	−0.5575	0.0337	−16.52	0.000
SPOUSE ∗ AGE2	0.0100	0.0007	15.15	0.000
$NCHILD$	−0.8362	0.3437	−2.43	0.015
C	−1.3616	1.0209	−1.33	0.182
R-squared		0.1310		
Adjusted R-squared		0.1308		

OUTPUT 4				
Dependent Variable: HRS				
Method: Least Squares				
Sample: 69852				
Included observations: 69852				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
WHITE	1.7045	0.8002	2.13	0.033
BLACK	3.07502	0.8061	3.81	0.000
HISPAN	−0.5925	1.0008	−0.59	0.554
HIGHSC	4.9467	0.2393	20.67	0.000
UNIV	4.9483	0.2741	18.05	0.000
AGE	1.3697	0.0394	34.76	0.000
AGE2	−0.0195	0.0007	−27.71	0.000
SPOUSE	0.1529	0.4437	0.34	0.730
SPOUSE ∗ AGE	−0.5575	0.0337	−16.55	0.000
SPOUSE ∗ AGE2	0.0100	0.0007	15.19	0.000
$NCHILD$	−0.8362	0.3430	−2.44	0.015
RES	−1.0475	0.3486	−3.01	0.003
C	−1.3616	1.0188	−1.34	0.181
R-squared		0.1345		
Adj. R-squared		0.1344		

(**NOTE**: RES are the residuals from OUTPUT 2)

Universidad Carlos III de Madrid
ECONOMETRICS I
Academic year 2005/06
FINAL EXAM
January 30, 2006

Exam's type: $\boxed{1}$

TIEMPO: 2 HORAS 30 MINUTOS
TIME: 2 HOURS 30 MINUTES

Instrucciones:

- ANTES DE EMPEZAR A RESPONDER EL EXAMEN:
  BEFORE YOU START TO ANSWER THE EXAM:
  - Rellene sus datos personales en el **impreso de lectura óptica**, que será el único documento válido de respuesta. Recuerde que tiene que completar sus datos identificativos (Nombre y apellidos y NIE) tanto en letra como en las casillas correspondientes de lectura óptica.
    Muy importante: El número de identificación que debe rellenar es su **NIE** (NO el DNI o el Pasaporte), que tiene 9 dígitos y empieza siempre por 1000.
    Fill in your personal data in the optical reading form, which will be the only valid answering document. Remember that you must complete all your identifiying data (name and surname(s), and NIE) both in letters and in the corresponding optical reading boxes.
    Very important: The identification number that you must fill is your **NIE** (NOT the DNI or the Passport), which has 9 digit and always begins by 1000.
  - Rellene, tanto en letra como en las correspondientes casillas de lectura óptica, el código de la asignatura y su grupo, de acuerdo con la siguiente tabla:
    Fill in, both in letters and in the corresponding optical reading boxes, the course code and your group, according to the following table:

TITULACION	**GRUPOS**					**CODIGO DE ASIGNATURA**
Economía	61	62	63	64	65*	10188
ADE	71	72	73	74	75*	10188
ADE (Colmenarejo)	71					10188
Sim. Eco-Dcho.	69					42020
Sim. ADE-Dcho.	77	78				43020
Sim. ADE-Dcho (Colmenarejo)	17					43020

*Grupos bilingües

- Compruebe que este tiene 40 preguntas numeradas correlativamente.
  Check that this document contains 40 questions sequentially numbered.

- **Compruebe que el número de tipo de examen que aparece en el cuestionario de preguntas coincide con el señalado en el impreso de lectura óptica.**
  Check that the number of exam type that appears in the questionnaire matches the number indicated in the optical reading form.

- Lea las preguntas detenidamente.
  Cuando una pregunta se refiera a problemas del enunciado, el encabezado de la

pregunta incluirá entre paréntesis el número de problema.
Se recomienda leer atentamente el enunciado del problema **antes** de contestar las preguntas relacionadas.
Read the questions carefully.
Whenever a question is referred to the problem included in the enclosed document, the question will include within parentheses at the beginning of the question the letter P.
It is advised to read the text of the problem carefully before answering its corresponding questions.

- Para la fila correspondiente al número de cada una de las preguntas, rellene la casilla correspondiente a la respuesta escogida en el impreso de lectura óptica (A, B, C ó D).
For each row regarding the number of each question, fill the box which corresponds to your chosen option in the optical reading form (A, B, C or D).

- **Cada pregunta tiene una única respuesta correcta**.
Cualquier pregunta en la que se seleccione más de una opción será considerada nula y su puntuación será cero.
Each question only has one correct answer.
Any question in which more than one answer is selected will be considered incorrect and its score will be zero.

- Todas las preguntas respondidas correctamente tienen idéntica puntuación. Las respuestas incorrectas tendrán una puntuación de cero. Para aprobar el examen hay que responder correctamente un mínimo de 24 preguntas.
All the questions correctly answered have the same score. Any incorrect answer will score as zero. To pass the exam, you must correctly answer a minimum of 24 questions.

- Si lo desea, puede utilizar la plantilla de respuestas que aparece a continuación como borrador, si bien dicha plantilla carece por completo de validez oficial.
If you want, you may use the answer table as a draft, although such table will not have any official validity.

- Puede utilizar el reverso de las hojas como borrador (no se facilitará más papel).
You can use the back side of the sheets as a draft (no additional sheets will be handed out).

- Al final de este documento, se adjuntan tablas estadísticas.
Statistical tables are enclosed at the end of the document.

- **Cualquier alumno que sea sorprendido hablando o intercambiando cualquier tipo de material en el examen será expulsado en el acto y su calificación será de cero, sin perjuicio de otras medidas que se puedan adoptar.**
Any student who was found talking or sharing any sort of material during the exam will be expelled immediately and his/her overall score will be zero, independently of any other measure that could be undertaken.

- **Fechas de publicación de calificaciones:** Miércoles 1 de Febrero.
Date of grades publication: Wednesday, February, 1st.

- **Fecha de revisión:**
Date of exam revision:

  - Lunes 6 de Febrero a las 15 h en las AULAS 15.0.04, 15.0.05 y 15.0.06.
  Monday, February, 6th, at 15 h in classrooms 15.0.04, 15.0.05 y 15.0.06.

- **Normas para la revisión:**
  Rules for exam revision:
  - La revisión sólo tendrá por objeto comprobar el número de respuestas correctas del examen.
    Its only purpose will be to check that the number of correct answers is right.
  - Para tener derecho a revisión, el alumno deberá:
    To be entitled for revision, the student should:
    * **Solicitarlo por escrito**, apuntándose en la lista situada en el Tablón de Información del departamento de Economía (junto al despacho 15.2.22), indicando titulación y grupo.
      Apply in writing, enrolling in a list located in the Tablón de Información of the Department of Economics (close to room 15.2.22), establishing your titulation and group.
    * **Acudir a la revisión con una copia impresa de las soluciones** del examen, que estarán disponibles en Aula Global desde el día de publicación de las calificaciones.
      Bring a printed copy of the exam solutions, which will be available in Aula Global from the date of grades publication..

Draft of ANSWERS									
QUESTION	**(a)**	**(b)**	**(c)**	**(d)**	**QUESTION**	**(a)**	**(b)**	**(c)**	**(d)**
1.					**21.**				
2.					**22.**				
3.					**23.**				
4.					**24.**				
5.					**25.**				
6.					**26.**				
7.					**27.**				
8.					**28.**				
9.					**29.**				
10.					**30.**				
11.					**31.**				
12.					**32.**				
13.					**33.**				
14.					**34.**				
15.					**35.**				
16.					**36.**				
17.					**37.**				
18.					**38.**				
19.					**39.**				
20.					**40.**				

1. Define in which of the following models the parameters α and β can be correctly estimated by OLS, where $A = e^{\alpha}$, and u is an unobservable error term which is independent of X:

   (a) $Y = AX^{\beta} + u$

   (b) It is not possible in any of these models.

   (c) $Y = AX^{\beta}u$

   (d) $\Pr(Y = 1|X) = \Phi(\alpha + \beta X)$

2. Given the OLS estimate for the slope in a linear regression model of Y on X, What Can you say about the OLS estimate for the slope in the liner regression model of X on Y?

   (a) The slope in the first regression is the inverse of the the slope in the second regression.

   (b) The estimated slopes in both regressions are the same.

   (c) The estimated slopes in both regression are equal but with different sign.

   (d) None of the other statements is correct.

3. Given a model where all the assumptions of the classical regression hold but conditional homoscedasticity:

   (a) A proper measure for the goodness of fit is either the regression's standard error or the R^2.

   (b) None of the previous statements is correct.

   (c) A proper measure for the goodness of fit is the R^2 but the regression's standard error is not longer a proper measure.

   (d) A proper measure for the goodness of fit is the regression's standard error but the R^2 is not longer a proper measure.

4. **(Problem 1)** Under the estimated LOGIT model, the predicted probability of not being granted a mortgage for a White applicant with 30% debt rate is approximately:

   (a) 0.0745.

   (b) 0.2229.

   (c) $(0.3 \times 5.37)/100 \simeq 0.016\,1$

   (d) 0.0690.

5. **(Problem 1)** Ignoring the problems in the linear probability model in OUTPUT1 that describes the probability of rejection of a mortgage, the predicted probability for an applicant with the same debt rate than a White applicant whose debt rate is 30% but being Black, is equal to:

   (a) 0.2537.

   (b) 0.0767.

   (c) $0.559 \times 0.30 = 0.1677$.

   (d) 0.559.

6. **(Problem 1)** Using the estimated LOGIT model, the expected differences in probability of mortgage rejection between a White applicant in respect to a Black one when both have a debt rate of 30% is equal to:

   (a) $-[0.0745 - 0.2229] = 0.1483$.

   (b) $1.27 \times 0.1732 = 0.2196$.

   (c) $0.0745 - 0.2229 = -0.1483$.

   (d) $1.27 \times 0.0690 \simeq 0.0876$.

7. **(Problem 1)** Ignoring the problems in the linear probability model, if we use the linear model in OUTPUT1 to describe the probability of mortgage rejection, the expected difference in this previous probability between a White applicant and a Black one both with the same debt rate is equal to:

   (a) 0.177 (and independent of the value of the variable `DEUDA`).

   (b) $0.559 \times \texttt{DEUDA} + 0.177$.

   (c) -0.177 (and independent of the value of the variable `DEUDA`).

   (d) $-0.091 + 0.559 \times \texttt{DEUDA} + 0.177$.

8. **(Problem 1)** According to the estimated PROBIT model, the marginal effect associated to the debt rate for a white applicant with a debt rate of 30% on the probability of mortgage rejection is:

   (a) 0.8386.

   (b) 0.622.

   (c) 0.556.

   (d) 0.3887.

9. **(Problem 1)** A comparison between the estimates from a linear probability model and a LOGIT one, allows us to conclude that the expected value of a change in the debt rate on the probability of mortgage rejection for a White applicant:

   (a) is higher in the linear probability model.

   (b) None of the other statements is correct.

   (c) in both models, the estimated effect is not constant, as it depends on the values that the variable `DEUDA` takes.

   (d) is higher in the LOGIT model.

10. **(Problem 1)** Based on the estimated models, the following statements are made:
    (i) Before estimating the linear probability model with robust standard errors, we should have analyzed the residual because it might be the case that there is no heteroscedasticity in which case we should use the usual standard errors.
    (ii) The probability of mortgage rejection for a White applicant with a debt rate of $\texttt{DEUDA} = 0.3$ according to the linear probability model is lower to the rejection rate in the sample.
    (iii) A direct comparison between the coefficient in OUTPUT1 and OUTPUT2 allows us to know in which of these two models the impact of a variable on the probability is higher.

    (a) Only (i) is correct.

    (b) Only (ii) is correct.

    (c) Only (i) y (iii) are correct.

    (d) All are correct.

11. **(Problem 1)** We want to test the joint significance of the two added variables in OUTPUT4.

    (a) The value of the statistics, whose asymptotic distribution follows a χ^2_2, is equal to $2 \times (-745.39 + 797.14) = 103.50$.

    (b) Under the information provided for the "Probability(LR stat)" in OUTPUT4, we reject the null hypothesis.

    (c) This test can not be performed because we need additional information about the covariance of the estimated coefficients.

    (d) The null hypothesis can not be rejected at 5% significance level.

12. Assume that you have estimated a multiple regression model with time series data and you got a Durbin-Watson of 0.20. Define which of the following answers is correct:

    (a) It is a clear sign of including an irrelevant variable in the model.

    (b) It provides information that there might be first order autocorrelation of 0.90.

    (c) It provides information that there might be first order autocorrelation of 0.10.

    (d) It is a clear sign that the dependent variable should be in logs.

13. Consider the simple regression model studied this semester. Assume that we have time series data and all the usual assumptions in the classical regression model but autocorrelation hold. Define which of the following statements is correct :

    (a) The OLS estimators for the constant and the slope are NOT consistent.

    (b) To make valid inference, we must use the Eicker-White robust standard errors.

    (c) The OLS estimators for the constant and the slope are efficient.

    (d) The Newey-West estimator for the standard errors of the estimated parameters are consistent.

14. Define which of the following statements is correct:

    (a) A test to detect autocorrelation in the error term can be based on the regression residuals.

    (b) Moving from quarterly to monthly data this will imply that the maximum relevant autocorrelation order, J, that is used to compute the Newey-West estimator of the standard errors should necessarily decrease.

    (c) The estimated coefficients by OLS and the estimated coefficients by OLS with standard errors robust to autocorrelation are quite different, in spite of both being consistent estimates.

    (d) With annual data, the autocorrelation order in the error term can not be higher than one.

15. In order to ensure the consistency of OLS estimates in a multiple regression model, define which of the following assumptions are NOT needed:

    (a) The conditional expectation of the error term (conditioning on the explanatory variables) is equal to zero.

    (b) Linearity in parameters.

    (c) Non correlation between the regressors and the error term.

    (d) Conditional Homoscedasticity (conditioning on the explanatory variables).

16. Given a random sample of the variables (Y, X_1, X_2) with a relationship

$$E(Y|X_1, X_2) = \beta_0 + \beta_1 X_1 + \beta_2 X_2,$$

the following models have been estimated:

$$\begin{aligned} Y &= b_0 + b_1 X_1 + b_2 X_2 + e \\ Y &= c_0 + c_1 X_1 + u \\ X_1 &= a_0 + a_1 X_2 + v \\ Y &= d_0 + d_1 v + w \end{aligned}$$

with b_0, b_1, b_2, c_0, c_1, a_0, a_1, d_0 y d_1 representing the OLS estimators for the respectively parameters and e, u, v y w being the OLS residuals. Consider the following statements :
(i) $V(c_1)$ is lower or equal than $V(b_1)$.
(ii) $d_1 = b_1$
(iii) The estimated variance of c_1 can be higher than the estimated variance of b_1.

(a) Only (i) is correct.
(b) The three statements are correct.
(c) Only (ii) is correct.
(d) Only (i) y (ii) are correct.

17. In a regression, the most general form of the Newey-West estimator solves:

(a) Only a heteroskedasticity problem.
(b) Only a second order autocorrelation problem.
(c) Simultaneously, the autocorrelation and heteroskedasticity problems.
(d) Only a first order autocorrelation problem.

18. **(Problem 2)** Assume that the exogenous variables of equation (E.1) (`WHITE`, `BLACK`, `HISPAN`, `HIGSCH`, `UNIV`, `AGE`, `AGE2`, `SPOUSE`, `AGE` $\times$ `SPOUSE`, `AGE2` $\times$ `SPOUSE`) are uncorrelated with `NCHILD`. Suppose that we estimate by OLS an equation for `HRS` just with these explanatory variables (and thus omitting `NCHILD`). Consider the following assertions.:
(i) The estimated coefficients will be inconsistent. In particular, the coefficient of WHITE will tend to be underestimated.
(ii) The estimated coefficients will be inconsistent. In particular, the coefficient of AGE will tend to be underestimated.
(iii) The estimated coefficients will be inconsistent. In particular, the coefficient of SPOUSE will tend to be underestimated.

(a) Only (i) and (ii) are true.
(b) None of the above assertions are true.
(c) Both (i), (ii) and (iii) are true.
(d) Only (ii) and (iii) are true.

19. **(Problem 2)** Assume that $C(\texttt{NCHILD}, \varepsilon) = 0$, and the model (E.1) satisfies all the further assumptions of the classical linear regression model. Consider the following assertions:
(i) The estimate of $V(\texttt{HRS})$ is equal to $(335.4738)^2$.
(ii) The estimate of

$$V(\texttt{HRS} \mid \texttt{WHITE}, \texttt{BLACK}, \texttt{HISPAN}, \texttt{HIGSCH}, \texttt{UNIV}, \texttt{AGE}, \texttt{AGE2}, \texttt{SPOUSE}, \texttt{NCHILD})$$

is equal to $(335.4738)^2$.
(iii) The estimate of

$$V\left(\texttt{HRS} \;\middle|\; \begin{array}{l} \texttt{WHITE}, \texttt{BLACK}, \texttt{HISPAN}, \texttt{HIGSCH}, \texttt{UNIV}, \texttt{AGE}, \\ \texttt{SPOUSE}, \texttt{AGE} \times \texttt{SPOUSE}, \texttt{AGE2} \times \texttt{SPOUSE}, \texttt{NCHILD} \end{array}\right)$$

is equal to $(335.4738)^2$.

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

20. **(Problem 2)** Assume that $C(\texttt{NCHILD}, \varepsilon) = 0$, and the model (E.1) satisfies all the further assumptions of the classical linear regression model except the assumption of conditional homoscedasticity. Consider the following assertions:
(i) The parameter estimates of OUTPUT 1 are not consistent.
(ii) The standard errors of the parameters of OUTPUT 1 are not consistent.
(iii) The R^2 of the model has no sense.

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

21. **(Problem 2)** Suppose that from model (E.1) we want to test the null hypothesis that hours worked per week are independent of age.

(a) The null hypothesis would be $H_0 : \beta_6 = 0$.
(b) The null hypothesis would be $H_0 : \beta_6 = \beta_7 = \beta_9 = \beta_{10} = 0$.
(c) The null hypothesis would be $H_0 : \beta_6 = \beta_7 = \beta_8 = \beta_9 = \beta_{10} = 0$.
(d) The null hypothesis would be $H_0 : \beta_6 = \beta_7 = 0$.

22. **(Problem 2)** Suppose that from model (E.1) we want to test the null hypothesis that hours worked per week are independent on whether the woman has a spouse living at her home.

(a) The null hypothesis would be $H_0 : \beta_8 = 0$.
(b) The null hypothesis would be $H_0 : \beta_0 = \beta_8 = \beta_9 = \beta_{10} = 0$.
(c) The null hypothesis would be $H_0 : \beta_8 = \beta_9 = \beta_{10} = 0$.
(d) The null hypothesis would be $H_0 : \beta_8 = \beta_9 = 0$.

23. **(Problem 2)** Suppose that from model (E.1) we want to test the null hypothesis that hours worked per week are independent on woman's education.

    (a) The null hypothesis would be $H_0 : \beta_0 = 0$.

    (b) The null hypothesis would be $H_0 : \beta_4 = \beta_5 = 0$.

    (c) The null hypothesis would be $H_0 : \beta_4 = \beta_5 = \beta_0 = 0$.

    (d) The null hypothesis would be $H_0 : \beta_4 - \beta_5 = 0$.

24. **(Problem 2)** We are concerned with obtaining consistent estimators for all the coefficients of equation (E.1).

    (a) The estimators of OUTPUT 1 are consistent.

    (b) The first stage regression for `NCHILD` in OUTPUT 2 is incorrect, since it should include only the instrument.

    (c) The estimators of OUTPUT3 are consistent, because the instrument (`MB`) fulfills the two required conditions to be a valid instrument: being uncorrelated with ε (as it is pointed out in the question statement) and being correlated with the endogenous variable `NCHILD` (as it can be seen in the first stage regression OUTPUT 2).

    (d) The estimators of OUTPUT 3 are not consistent, because we would need that the instrument (`MB`) was not correlated with the endogenous variable `NCHILD`, what does not appear to be the case given OUTPUT 2.

25. **(Problem 2)** Assume that $C(\texttt{NCHILD}, \varepsilon) = 0$, so that `NCHILD` is exogenous. A woman 20 years old without children, without spouse, Asian and with less than high school will work on average, approximately:

    (a) 21 hours per week.

    (b) It can not be determined.

    (c) 3 hours per week.

    (d) 18 hours per week.

26. **(Problem 2)** Assume that $C(\texttt{NCHILD}, \varepsilon) = 0$, so that `NCHILD` is exogenous. For a given race, education level, age and spouse status, an additional child entails on average an estimated decrease of hours of work:

    (a) 1.85% hours per month.

    (b) 1.85 hours per week.

    (c) It depends on the age of the individual.

    (d) 0.83 hours per month.

27. **(Problem 2)** Assume for this question that $C(\texttt{MB}, \varepsilon) \neq 0$, then

    (a) The number of children coefficient in OUTPUT 1 is a consistent estimate of β_{11}.

    (b) None of the other answers is correct.

    (c) The Multiple birth coefficient in OUTPUT 2 is inconsistent.

    (d) number of children coefficient in OUTPUT 3 is a consistent estimate of β_{11}.

28. **(Problem 2)** If we want to evaluate whether the variable `NCHILD` is an endogenous variable:

    (a) We will test whether `NCHILD` is endogenous in the first stage equation by means of a t-test for the coefficient of `MB`.

    (b) We will test the joint significance of all the regressors in OUTPUT 2 (test of joint significance, or regression test).

    (c) We will test whether `NCHILD` is endogenous in the hours equation by means of a Hausman test.

    (d) We cannot test that hypothesis because we need at least two instruments.

29. **(Problem 2)** Given the results:

    (a) Since `RES` is statistically significant in the OLS estimation of the augmented hours equation (OUTPUT 4), we do NOT reject that `NCHILD` is exogenous.

    (b) None of the other answers is correct.

    (c) The reported test for endogeneity is incorrect, since the reduced form in which the residuals are based does incorrectly include the exogenous variables of the hours equation (`WHITE`, `BLACK`, `HISPAN`, `HIGSCH`, `UNIV`, `AGE`, `AGE2`, `SPOUSE`, `AGE` × `SPOUSE`, `AGE2` × `SPOUSE`).

    (d) Since `RES` is statistically significant in the OLS estimation of the augmented hours equation (OUTPUT 4), we reject that `NCHILD` is exogenous.

30. **(Problem 2)** Using the appropriate estimates, we can conclude that for a given race, education level, age and spouse status, an additional child will approximately decrease on average the estimated hour supplied by:

    (a) 0.84% per week.

    (b) 1.85 hours per week.

    (c) 1.21 hours per week.

    (d) 0.84 hours per week.

31. **(Problem 2)** Using the appropriate estimates, we can conclude that for a given number of children, age, and spouse status a White woman with university education, respect to a black woman with high school will approximately work on average:

    (a) 1.37 hours less per week.

    (b) 1.37 hours more per week.

    (c) 1.7% hours more per week.

    (d) 3.07 hours more per week.

32. **(Problem 2)** Using the appropriate estimates, we can conclude that for a given number of children, race, and education level, getting one year older for women with 46 years of age who does not have a spouse implies, approximately, that she will work on average:

    (a) 1.33 hours per week more.

    (b) 0.79 hours per week more.

    (c) 1.33 hours per week less.

    (d) 0.42 hours per week less.

33. **(Problem 2)** Using the appropriate estimates, we can conclude that for a given number of children, age, education level and race a woman with a spouse, in respect to a woman without spouse will work on average:

(a) $[0.4258 - 0.5668 \times \texttt{AGE} + 0.0101 \times \texttt{AGE2}]$ hours more per week.

(b) $[0.1529 - 0.5575 \times \texttt{AGE} + 2 \times 0.01 \times \texttt{AGE2}]$ hours more per week.

(c) $[0.1529 - 0.5575 \times \texttt{AGE} + 0.01 \times \texttt{AGE2}]$ hours more per week.

(d) $[0.4258 - 0.5668 \times \texttt{AGE} + 2 \times 0.0101 \times \texttt{AGE2}]$ hours more per week.

34. **(Problem 2)** Using the appropriate estimates, we can conclude that for a given number of children and age, a white woman with university education and without spouse, in respect to a black woman with high school and with spouse will work on average:

(a) $-1.37 + [0.1529 - 0.5575 \times \texttt{AGE} + 0.01 \times \texttt{AGE2}]$ hours more per week.

(b) $-1.37 - [0.1529 - 0.5575 \times \texttt{AGE} + 0.01 \times \texttt{AGE2}]$ hours more per week.

(c) $-2.11 - [0.4258 - 0.5668 \times \texttt{AGE} + 2 \times 0.0101 \times \texttt{AGE2}]$ hours more per week.

(d) $-2.11 + [0.4258 - 0.5668 \times \texttt{AGE} + 2 \times 0.0101 \times \texttt{AGE2}]$ hours more per week.

35. **(Problem 2)** Using the appropriate estimates, the age at which a woman who has a spouse will work, on average, the highest number of hours will be approximately:

(a) 42.7 years old.

(b) 35.9 years old.

(c) 45 years old.

(d) 35 years old.

36. **(Problem 2)** Using the appropriate estimates, if we compare two women with the same race, education level, and number of children but the first one has a spouse, while the second one has not, the difference, in absolute years, between the ages at which they work the highest number of hours will be approximately:

(a) 2.2 years.

(b) 15.5 years.

(c) 10.1 years.

(d) 7.6 years.

37. **(Problem 2)** Suppose that from model (E.1) we want to test the null hypothesis that hours worked per week are independent of age. Using the appropriate estimates:

(a) The associated t-statistic, which is asymptotically distributed as a $N(0,1)$, is 34.69.

(b) It is not possible to evaluate such hypothesis with the available information.

(c) The associated t-statistic, which is asymptotically distributed as a $N(0,1)$, is 46.26.

(d) The associated t-statistic, which is asymptotically distributed as a $N(0,1)$, is 50.37.

38. **(Problem 2)** Suppose that both in OUTPUT 1 and OUTPUT 3, the estimation of the correlation coefficient between the estimated coefficients of `HIGSCH` and `UNIV` is $\widehat{corr}\,(b_4, b_5) = 0.6$. Using the appropriate estimates, if we want to test the null hypothesis that having high school education has the same effect on hours worked per week than having a university degree:

(a) The appropriate statistic is

$$t = \left| \frac{-0.0016}{\sqrt{0.2398^2 + 0.2746^2 - 2 \times 0.6 \times 0.2398 \times 0.2746}} \right| = 6.892\,3 \times 10^{-2}.$$

(b) The appropriate statistic is

$$t = \left| \frac{-0.0016}{\sqrt{0.2398^2 + 0.2746^2 + 2 \times 0.6 \times 0.2398 \times 0.2746}} \right| = 3.475\,6 \times 10^{-2}.$$

(c) We do not have information enough to test such hypothesis.

(d) The appropriate statistic is

$$t = \left| \frac{0.0759}{\sqrt{0.1770^2 + 0.2005^2 - 2 \times 0.6 \times 0.1770 \times 0.2005}} \right| = 0.44614.$$

39. Given the model

$$E\,(Y|\,X) = F\,(\beta_0 + \beta_1 X)$$

where Y is a binary variable. Consider the following statements:
(i) $E\,(Y|\,X) = PLO\,(Y|\,X)$.
(ii) $\Pr\,(Y = 1|\,X) = F\,(\beta_0 + \beta_1 X)$, where $F(\beta_0 + \beta_1 X) \in [0, 1]$ for any value of X.
(iii) $F\,(\beta_0 + \beta_1 X) = \beta_0 + \beta_1 X$.

(a) The three statement are correct.

(b) Only (ii) is correct.

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

(d) Only (i) and (ii) are correct.

40. Given the following joint distribution for the discrete random variables (X, Y):

$Pr(X = x, Y = y)$	$X = 1$	$X = 2$	$X = 3$
$Y = 1$	0.15	0.10	0.15
$Y = 2$	0.15	0.30	0.15

(a) $E(Y|X) = 1.60$.

(b) $E(Y|X) = 1.75 - 0.25 \times \sqrt{(X-2)^2}$.

(c) $E(Y|X) = PLO(Y|X)$.

(d) $E(Y|X) = 0$ because they are independent.

Solution to Exam Type: $\boxed{1}$

Universidad Carlos III de Madrid

ECONOMETRICS I

Academic year 2005/06
FINAL EXAM

January 30, 2006

TIEMPO: 2 HORAS 30 MINUTOS
TIME: 2 HOURS 30 MINUTES

1. Define in which of the following models the parameters α and β can be correctly estimated by OLS, where $A = e^{\alpha}$, and u is an unobservable error term which is independent of X:

   (a) $Y = AX^{\beta} + u$

   (b) It is not possible in any of these models.

   $\boxed{\text{(c)}}$ $Y = AX^{\beta}u$

   (d) $\Pr(Y = 1|X) = \Phi(\alpha + \beta X)$

2. Given the OLS estimate for the slope in a linear regression model of Y on X, What Can you say about the OLS estimate for the slope in the liner regression model of X on Y?

   (a) The slope in the first regression is the inverse of the the slope in the second regression.

   (b) The estimated slopes in both regressions are the same.

   (c) The estimated slopes in both regression are equal but with different sign.

   $\boxed{\text{(d)}}$ None of the other statements is correct.

3. Given a model where all the assumptions of the classical regression hold but conditional homoscedasticity:

   (a) A proper measure for the goodness of fit is either the regression's standard error or the R^2.

   $\boxed{\text{(b)}}$ None of the previous statements is correct.

   (c) A proper measure for the goodness of fit is the R^2 but the regression's standard error is not longer a proper measure.

   (d) A proper measure for the goodness of fit is the regression's standard error but the R^2 is not longer a proper measure.

4. **(Problem 1)** Under the estimated LOGIT model, the predicted probability of not being granted a mortgage for a White applicant with 30% debt rate is approximately:

   $\boxed{\text{(a)}}$ 0.0745.

   (b) 0.2229.

   (c) $(0.3 \times 5.37)/100 \simeq 0.0161$

   (d) 0.0690.

5. **(Problem 1)** Ignoring the problems in the linear probability model in OUTPUT1 that describes the probability of rejection of a mortgage, the predicted probability for an applicant with the same debt rate than a White applicant whose debt rate is 30% but being Black, is equal to:

   (a) 0.2537.
   (b) 0.0767.
   (c) $0.559 \times 0.30 = 0.1677$.
   (d) 0.559.

6. **(Problem 1)** Using the estimated LOGIT model, the expected differences in probability of mortgage rejection between a White applicant in respect to a Black one when both have a debt rate of 30% is equal to:

   (a) $-[0.0745 - 0.2229] = 0.1483$.
   (b) $1.27 \times 0.1732 = 0.2196$.
   (c) $0.0745 - 0.2229 = -0.1483$.
   (d) $1.27 \times 0.0690 \simeq 0.0876$.

7. **(Problem 1)** Ignoring the problems in the linear probability model, if we use the linear model in OUTPUT1 to describe the probability of mortgage rejection, the expected difference in this previous probability between a White applicant and a Black one both with the same debt rate is equal to:

   (a) 0.177 (and independent of the value of the variable `DEUDA`).
   (b) $0.559 \times$ `DEUDA` $+ 0.177$.
   (c) -0.177 (and independent of the value of the variable `DEUDA`).
   (d) $-0.091 + 0.559 \times$ `DEUDA` $+ 0.177$.

8. **(Problem 1)** According to the estimated PROBIT model, the marginal effect associated to the debt rate for a white applicant with a debt rate of 30% on the probability of mortgage rejection is:

   (a) 0.8386.
   (b) 0.622.
   (c) 0.556.
   (d) 0.3887.

9. **(Problem 1)** A comparison between the estimates from a linear probability model and a LOGIT one, allows us to conclude that the expected value of a change in the debt rate on the probability of mortgage rejection for a White applicant:

   (a) is higher in the linear probability model.
   (b) None of the other statements is correct.
   (c) in both models, the estimated effect is not constant, as it depends on the values that the variable `DEUDA` takes.
   (d) is higher in the LOGIT model.

10. **(Problem 1)** Based on the estimated models, the following statements are made:
(i) Before estimating the linear probability model with robust standard errors, we should have analyzed the residual because it might be the case that there is no heteroscedasticity in which case we should use the usual standard errors.
(ii) The probability of mortgage rejection for a White applicant with a debt rate of `DEUDA` $= 0.3$ according to the linear probability model is lower to the rejection rate in the sample.
(iii) A direct comparison between the coefficient in OUTPUT1 and OUTPUT2 allows us to know in which of these two models the impact of a variable on the probability is higher.

   (a) Only (i) is correct.
   $\boxed{\text{(b)}}$ Only (ii) is correct.
   (c) Only (i) y (iii) are correct.
   (d) All are correct.

11. **(Problem 1)** We want to test the joint significance of the two added variables in OUTPUT4.

   $\boxed{\text{(a)}}$ The value of the statistics, whose asymptotic distribution follows a χ^2_2, is equal to $2 \times (-745.39 + 797.14) = 103.50$.
   (b) Under the information provided for the "Probability(LR stat)" in OUTPUT4, we reject the null hypothesis.
   (c) This test can not be performed because we need additional information about the covariance of the estimated coefficients.
   (d) The null hypothesis can not be rejected at 5% significance level.

12. Assume that you have estimated a multiple regression model with time series data and you got a Durbin-Watson of 0.20. Define which of the following answers is correct:

   (a) It is a clear sign of including an irrelevant variable in the model.
   $\boxed{\text{(b)}}$ It provides information that there might be first order autocorrelation of 0.90.
   (c) It provides information that there might be first order autocorrelation of 0.10.
   (d) It is a clear sign that the dependent variable should be in logs.

13. Consider the simple regression model studied this semester. Assume that we have time series data and all the usual assumptions in the classical regression model but autocorrelation hold. Define which of the following statements is correct :

   (a) The OLS estimators for the constant and the slope are NOT consistent.
   (b) To make valid inference, we must use the Eicker-White robust standard errors.
   (c) The OLS estimators for the constant and the slope are efficient.
   $\boxed{\text{(d)}}$ The Newey-West estimator for the standard errors of the estimated parameters are consistent.

14. Define which of the following statements is correct:

   $\boxed{\text{(a)}}$ A test to detect autocorrelation in the error term can be based on the regression residuals.

   (b) Moving from quarterly to monthly data this will imply that the maximum relevant autocorrelation order, J, that is used to compute the Newey-West estimator of the standard errors should necessarily decrease.

   (c) The estimated coefficients by OLS and the estimated coefficients by OLS with standard errors robust to autocorrelation are quite different, in spite of both being consistent estimates.

   (d) With annual data, the autocorrelation order in the error term can not be higher than one.

15. In order to ensure the consistency of OLS estimates in a multiple regression model, define which of the following assumptions are NOT needed:

   (a) The conditional expectation of the error term (conditioning on the explanatory variables) is equal to zero.

   (b) Linearity in parameters.

   (c) Non correlation between the regressors and the error term.

   $\boxed{\text{(d)}}$ Conditional Homoscedasticity (conditioning on the explanatory variables).

16. Given a random sample of the variables (Y, X_1, X_2) with a relationship

$$E(Y|X_1, X_2) = \beta_0 + \beta_1 X_1 + \beta_2 X_2,$$

the following models have been estimated:

$$\begin{aligned} Y &= b_0 + b_1 X_1 + b_2 X_2 + e \\ Y &= c_0 + c_1 X_1 + u \\ X_1 &= a_0 + a_1 X_2 + v \\ Y &= d_0 + d_1 v + w \end{aligned}$$

with b_0, b_1, b_2, c_0, c_1, a_0, a_1, d_0 y d_1 representing the OLS estimators for the respectively parameters and e, u, v y w being the OLS residuals. Consider the following statements :
(i) $V(c_1)$ is lower or equal than $V(b_1)$.
(ii) $d_1 = b_1$
(iii) The estimated variance of c_1 can be higher than the estimated variance of b_1.

   (a) Only (i) is correct.

   $\boxed{\text{(b)}}$ The three statements are correct.

   (c) Only (ii) is correct.

   (d) Only (i) y (ii) are correct.

17. In a regression, the most general form of the Newey-West estimator solves:

   (a) Only a heteroskedasticity problem.

   (b) Only a second order autocorrelation problem.

   $\boxed{\text{(c)}}$ Simultaneously, the autocorrelation and heteroskedasticity problems.

   (d) Only a first order autocorrelation problem.

18. **(Problem 2)** Assume that the exogenous variables of equation (E.1) ($\texttt{WHITE}$, $\texttt{BLACK}$, $\texttt{HISPAN}$, $\texttt{HIGSCH}$, $\texttt{UNIV}$, $\texttt{AGE}$, $\texttt{AGE2}$, $\texttt{SPOUSE}$, $\texttt{AGE} \times \texttt{SPOUSE}$, $\texttt{AGE2} \times \texttt{SPOUSE}$) are uncorrelated with $\texttt{NCHILD}$. Suppose that we estimate by OLS an equation for $\texttt{HRS}$ just with these explanatory variables (and thus omitting $\texttt{NCHILD}$). Consider the following assertions.:
(i) The estimated coefficients will be inconsistent. In particular, the coefficient of WHITE will tend to be underestimated.
(ii) The estimated coefficients will be inconsistent. In particular, the coefficient of AGE will tend to be underestimated.
(iii) The estimated coefficients will be inconsistent. In particular, the coefficient of SPOUSE will tend to be underestimated.

   (a) Only (i) and (ii) are true.
   $\boxed{\text{(b)}}$ None of the above assertions are true.
   (c) Both (i), (ii) and (iii) are true.
   (d) Only (ii) and (iii) are true.

19. **(Problem 2)** Assume that $C(\texttt{NCHILD}, \varepsilon) = 0$, and the model (E.1) satisfies all the further assumptions of the classical linear regression model. Consider the following assertions:
(i) The estimate of $V(\texttt{HRS})$ is equal to $(335.4738)^2$.
(ii) The estimate of

$$V\left(\texttt{HRS} \mid \texttt{WHITE}, \texttt{BLACK}, \texttt{HISPAN}, \texttt{HIGSCH}, \texttt{UNIV}, \texttt{AGE}, \texttt{AGE2}, \texttt{SPOUSE}, \texttt{NCHILD}\right)$$

is equal to $(335.4738)^2$.
(iii) The estimate of

$$V\left(\texttt{HRS} \,\middle|\, \begin{array}{l} \texttt{WHITE}, \texttt{BLACK}, \texttt{HISPAN}, \texttt{HIGSCH}, \texttt{UNIV}, \texttt{AGE}, \\ \texttt{SPOUSE}, \texttt{AGE} \times \texttt{SPOUSE}, \texttt{AGE2} \times \texttt{SPOUSE}, \texttt{NCHILD} \end{array}\right)$$

is equal to $(335.4738)^2$.

   $\boxed{\text{(a)}}$ Only (ii) and (iii) are true.
   (b) Only (i) is true.
   (c) Only (iii) is true.
   (d) Only (i) and (iii) are true.

20. **(Problem 2)** Assume that $C(\texttt{NCHILD}, \varepsilon) = 0$, and the model (E.1) satisfies all the further assumptions of the classical linear regression model except the assumption of conditional homoscedasticity. Consider the following assertions:
(i) The parameter estimates of OUTPUT 1 are not consistent.
(ii) The standard errors of the parameters of OUTPUT 1 are not consistent.
(iii) The R^2 of the model has no sense.

   (a) Only (ii) is true.
   $\boxed{\text{(b)}}$ Only (ii) and (iii) are true.
   (c) The three assertions are true.
   (d) Only (i) and (ii) are true.

21. **(Problem 2)** Suppose that from model (E.1) we want to test the null hypothesis that hours worked per week are independent of age.

(a) The null hypothesis would be $H_0 : \beta_6 = 0$.

$\boxed{\text{(b)}}$ The null hypothesis would be $H_0 : \beta_6 = \beta_7 = \beta_9 = \beta_{10} = 0$.

(c) The null hypothesis would be $H_0 : \beta_6 = \beta_7 = \beta_8 = \beta_9 = \beta_{10} = 0$.

(d) The null hypothesis would be $H_0 : \beta_6 = \beta_7 = 0$.

22. **(Problem 2)** Suppose that from model (E.1) we want to test the null hypothesis that hours worked per week are independent on whether the woman has a spouse living at her home.

(a) The null hypothesis would be $H_0 : \beta_8 = 0$.

(b) The null hypothesis would be $H_0 : \beta_0 = \beta_8 = \beta_9 = \beta_{10} = 0$.

$\boxed{\text{(c)}}$ The null hypothesis would be $H_0 : \beta_8 = \beta_9 = \beta_{10} = 0$.

(d) The null hypothesis would be $H_0 : \beta_8 = \beta_9 = 0$.

23. **(Problem 2)** Suppose that from model (E.1) we want to test the null hypothesis that hours worked per week are independent on woman's education.

(a) The null hypothesis would be $H_0 : \beta_0 = 0$.

$\boxed{\text{(b)}}$ The null hypothesis would be $H_0 : \beta_4 = \beta_5 = 0$.

(c) The null hypothesis would be $H_0 : \beta_4 = \beta_5 = \beta_0 = 0$.

(d) The null hypothesis would be $H_0 : \beta_4 - \beta_5 = 0$.

24. **(Problem 2)** We are concerned with obtaining consistent estimators for all the coefficients of equation (E.1).

(a) The estimators of OUTPUT 1 are consistent.

(b) The first stage regression for `NCHILD` in OUTPUT 2 is incorrect, since it should include only the instrument.

$\boxed{\text{(c)}}$ The estimators of OUTPUT3 are consistent, because the instrument (`MB`) fulfills the two required conditions to be a valid instrument: being uncorrelated with ε (as it is pointed out in the question statement) and being correlated with the endogenous variable `NCHILD` (as it can be seen in the first stage regression OUTPUT 2).

(d) The estimators of OUTPUT 3 are not consistent, because we would need that the instrument (`MB`) was not correlated with the endogenous variable `NCHILD`, what does not appear to be the case given OUTPUT 2.

25. **(Problem 2)** Assume that $C(\texttt{NCHILD}, \varepsilon) = 0$, so that `NCHILD` is exogenous. A woman 20 years old without children, without spouse, Asian and with less than high school will work on average, approximately:

$\boxed{\text{(a)}}$ 21 hours per week.

(b) It can not be determined.

(c) 3 hours per week.

(d) 18 hours per week.

26. **(Problem 2)** Assume that $C(\texttt{NCHILD}, \varepsilon) = 0$, so that NCHILD is exogenous. For a given race, education level, age and spouse status, an additional child entails on average an estimated decrease of hours of work:

(a) 1.85% hours per month.

$\boxed{\text{(b)}}$ 1.85 hours per week.

(c) It depends on the age of the individual.

(d) 0.83 hours per month.

27. **(Problem 2)** Assume for this question that $C(\texttt{MB}, \varepsilon) \neq 0$, then

(a) The number of children coefficient in OUTPUT 1 is a consistent estimate of β_{11}.

$\boxed{\text{(b)}}$ None of the other answers is correct.

(c) The Multiple birth coefficient in OUTPUT 2 is inconsistent.

(d) number of children coefficient in OUTPUT 3 is a consistent estimate of β_{11}.

28. **(Problem 2)** If we want to evaluate whether the variable NCHILD is an endogenous variable:

(a) We will test whether NCHILD is endogenous in the first stage equation by means of a t-test for the coefficient of MB.

(b) We will test the joint significance of all the regressors in OUTPUT 2 (test of joint significance, or regression test).

$\boxed{\text{(c)}}$ We will test whether NCHILD is endogenous in the hours equation by means of a Hausman test.

(d) We cannot test that hypothesis because we need at least two instruments.

29. **(Problem 2)** Given the results:

(a) Since RES is statistically significant in the OLS estimation of the augmented hours equation (OUTPUT 4), we do NOT reject that NCHILD is exogenous.

(b) None of the other answers is correct.

(c) The reported test for endogeneity is incorrect, since the reduced form in which the residuals are based does incorrectly include the exogenous variables of the hours equation (WHITE, BLACK, HISPAN, HIGSCH, UNIV, AGE, AGE2, SPOUSE, AGE $\times$ SPOUSE, AGE2 $\times$ SPOUSE).

$\boxed{\text{(d)}}$ Since RES is statistically significant in the OLS estimation of the augmented hours equation (OUTPUT 4), we reject that NCHILD is exogenous.

30. **(Problem 2)** Using the appropriate estimates, we can conclude that for a given race, education level, age and spouse status, an additional child will approximately decrease on average the estimated hour supplied by:

(a) 0.84% per week.

(b) 1.85 hours per week.

(c) 1.21 hours per week.

$\boxed{\text{(d)}}$ 0.84 hours per week.

31. **(Problem 2)** Using the appropriate estimates, we can conclude that for a given number of children, age, and spouse status a White woman with university education, respect to a black woman with high school will approximately work on average:

    (a) 1.37 hours less per week.
    (b) 1.37 hours more per week.
    (c) 1.7% hours more per week.
    (d) 3.07 hours more per week.

32. **(Problem 2)** Using the appropriate estimates, we can conclude that for a given number of children, race, and education level, getting one year older for women with 46 years of age who does not have a spouse implies, approximately, that she will work on average:

    (a) 1.33 hours per week more.
    (b) 0.79 hours per week more.
    (c) 1.33 hours per week less.
    (d) 0.42 hours per week less.

33. **(Problem 2)** Using the appropriate estimates, we can conclude that for a given number of children, age, education level and race a woman with a spouse, in respect to a woman without spouse will work on average:

    (a) $[0.4258 - 0.5668 \times \texttt{AGE} + 0.0101 \times \texttt{AGE2}]$ hours more per week.
    (b) $[0.1529 - 0.5575 \times \texttt{AGE} + 2 \times 0.01 \times \texttt{AGE2}]$ hours more per week.
    (c) $[0.1529 - 0.5575 \times \texttt{AGE} + 0.01 \times \texttt{AGE2}]$ hours more per week.
    (d) $[0.4258 - 0.5668 \times \texttt{AGE} + 2 \times 0.0101 \times \texttt{AGE2}]$ hours more per week.

34. **(Problem 2)** Using the appropriate estimates, we can conclude that for a given number of children and age, a white woman with university education and without spouse, in respect to a black woman with high school and with spouse will work on average:

    (a) $-1.37 + [0.1529 - 0.5575 \times \texttt{AGE} + 0.01 \times \texttt{AGE2}]$ hours more per week.
    (b) $-1.37 - [0.1529 - 0.5575 \times \texttt{AGE} + 0.01 \times \texttt{AGE2}]$ hours more per week.
    (c) $-2.11 - [0.4258 - 0.5668 \times \texttt{AGE} + 2 \times 0.0101 \times \texttt{AGE2}]$ hours more per week.
    (d) $-2.11 + [0.4258 - 0.5668 \times \texttt{AGE} + 2 \times 0.0101 \times \texttt{AGE2}]$ hours more per week.

35. **(Problem 2)** Using the appropriate estimates, the age at which a woman who has a spouse will work, on average, the highest number of hours will be approximately:

    (a) 42.7 years old.
    (b) 35.9 years old.
    (c) 45 years old.
    (d) 35 years old.

36. **(Problem 2)** Using the appropriate estimates, if we compare two women with the same race, education level, and number of children but the first one has a spouse, while the second one has not, the difference, in absolute years, between the ages at which they work the highest number of hours will be approximately:

(a) 2.2 years.

(b) 15.5 years.

(c) 10.1 years.

$\boxed{\text{(d)}}$ 7.6 years.

37. **(Problem 2)** Suppose that from model (E.1) we want to test the null hypothesis that hours worked per week are independent of age. Using the appropriate estimates:

(a) The associated t-statistic, which is asymptotically distributed as a $N(0,1)$, is 34.69.

$\boxed{\text{(b)}}$ It is not possible to evaluate such hypothesis with the available information.

(c) The associated t-statistic, which is asymptotically distributed as a $N(0,1)$, is 46.26.

(d) The associated t-statistic, which is asymptotically distributed as a $N(0,1)$, is 50.37.

38. **(Problem 2)** Suppose that both in OUTPUT 1 and OUTPUT 3, the estimation of the correlation coefficient between the estimated coefficients of `HIGSCH` and `UNIV` is $\widehat{corr}(b_4, b_5) = 0.6$. Using the appropriate estimates, if we want to test the null hypothesis that having high school education has the same effect on hours worked per week than having a university degree:

$\boxed{\text{(a)}}$ The appropriate statistic is

$$t = \left| \frac{-0.0016}{\sqrt{0.2398^2 + 0.2746^2 - 2 \times 0.6 \times 0.2398 \times 0.2746}} \right| = 6.892\,3 \times 10^{-2}.$$

(b) The appropriate statistic is

$$t = \left| \frac{-0.0016}{\sqrt{0.2398^2 + 0.2746^2 + 2 \times 0.6 \times 0.2398 \times 0.2746}} \right| = 3.475\,6 \times 10^{-2}.$$

(c) We do not have information enough to test such hypothesis.

(d) The appropriate statistic is

$$t = \left| \frac{0.0759}{\sqrt{0.1770^2 + 0.2005^2 - 2 \times 0.6 \times 0.1770 \times 0.2005}} \right| = 0.44614.$$

39. Given the model

$$E(Y|X) = F(\beta_0 + \beta_1 X)$$

where Y is a binary variable. Consider the following statements:
(i) $E(Y|X) = PLO(Y|X)$.
(ii) $\Pr(Y = 1|X) = F(\beta_0 + \beta_1 X)$, where $F(\beta_0 + \beta_1 X) \in [0,1]$ for any value of X.
(iii) $F(\beta_0 + \beta_1 X) = \beta_0 + \beta_1 X$.

(a) The three statement are correct.

$\boxed{\text{(b)}}$ Only (ii) is correct.

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

(d) Only (i) and (ii) are correct.

40. Given the following joint distribution for the discrete random variables (X, Y):

$Pr(X = x, Y = y)$	$X = 1$	$X = 2$	$X = 3$
$Y = 1$	0.15	0.10	0.15
$Y = 2$	0.15	0.30	0.15

(a) $E(Y|X) = 1.60$.

$\boxed{\text{(b)}}$ $E(Y|X) = 1.75 - 0.25 \times \sqrt{(X-2)^2}$.

(c) $E(Y|X) = PLO(Y|X)$.

(d) $E(Y|X) = 0$ because they are independent.