UNIVERSIDAD CARLOS III DE MADRID
ECONOMETRICS
FINAL EXAM (Type B)

DURATION: 125 MINUTES

Directions:

1. This is an example of a exam that you can use to self-evaluate about the contents of the course Econometrics in OCW, Universidad Carlos III de Madrid, except the two last topics (Heterokedasticity and Autocorrelation).
2. This document is self contained. Your are not allowed to use any other material.
3. Read the problem text and the questions carefully. Each question, unless otherwise stated, requires a complete analysis of all the outputs shown in the corresponding problem.
   For example, to answer those questions referring to "appropriate estimates", or "given the estimates" or "given the problem conditions", the results based on the consistent and more efficient among outputs, must be used.
4. Each output includes all the explanatory variables used in the corresponding estimation.
5. Some results in the output shown may have been omitted.
6. The dependent variable can vary among outputs within the same problem.
7. For the sake of brevity, we will say that a model is well specified if it is linear in the conditioning variables (as they appear in the model) and its error term is mean-independent of such variables.
8. OLS, and 2SLS or TSLS, are the corresponding abbreviations of ordinary least squares and two stage least squares, respectively.
9. Statistical tables are included at the end of the problem, before the questions.
10. Each question only has one correct answer.
11. At the end of this document you will find the answer keys of this exam type. Do the exam as if you were sitting it for grade. After that check your answers with the keys given at the end. Do obtain your grade use the following formula:

$$[0{,}20 \times (\#\text{ correct answers}) - 0{,}06 \times (\#\text{ incorrect answers})] + 0{,}4$$

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

Problem: Determinants of fertility.

We would like to study the determinants of the total woman's number of children ($KIDS$). We are interested in knowing if fertility rates (meaning the average number of children per woman) have changed over time. We have a sample of 476 women from the *General Social Survey* of the National Opinion Research Center (NORC) in the USA for the years 1972, 1978 and 1984.

The characteristics of the woman we are interested in are $EDUC$ (Years of education), AGE (Age, in years), AGE^2 (Age squared), $BLACK$ (Binary variable that takes the value of 1 if the woman is black and 0 otherwise).

Moreover, in order to consider the possibility that fertility rates change over time, we have the variables $YEAR$ (The year corresponding to the observation; this variable takes three possible values: 72, 78 or 84); $Y72$ (Binary variable that takes the value of 1 if the observation corresponds to the year 1972 and 0 otherwise); $Y78$ (Binary variable that takes the value of 1 if the observation corresponds to the year 1978 and 0 otherwise); $Y84$ (Binary variable that takes the value of 1 if the observation corresponds to the year 1984 and 0 otherwise).

Also, we take into account the possibility of interacting $YEAR$ with education, ($YEAR \times EDUC$).

The following models are considered to analyze the determinants of the number of children women have:

$$KIDS = \beta_0 + \beta_1 AGE + \beta_2 AGE^2 + \beta_3 BLACK + \beta_4 EDUC + \beta_5 YEAR + \varepsilon_1 \tag{I}$$

$$KIDS = \delta_0 + \delta_1 AGE + \delta_2 AGE^2 + \delta_3 BLACK + \delta_4 EDUC + \delta_5 Y78 + \delta_6 Y84 + \varepsilon_2 \tag{II}$$

$$\begin{aligned} KIDS = {} & \gamma_0 + \gamma_1 AGE + \gamma_2 AGE^2 + \gamma_3 BLACK + \gamma_4 EDUC + \gamma_5 YEAR \\ & + \gamma_6 (YEAR \times EDUC) + \varepsilon_3 \end{aligned} \tag{III}$$

We also observe two further binary variables, $RURAL$ (that takes the value of 1 if the woman lived in a rural area in her teens and 0 otherwise) and $LPOP$ (that takes the value of 1 if the woman lived in a highly populated area in her teens and 0 otherwise). Of course, interactions of $YEAR$ with these two variables, ($YEAR \times RURAL$) y ($YEAR \times LPOP$), can also be considered.

The results of the various estimations are presented below:

Output 1: OLS, using observations 1–476
Dependent variable: $KIDS$

	Coefficient	Std. Error	t-ratio	p-value
const	−2,1966	5.0370	−0,4361	0.6630
AGE	0,4788	0.2178	2,1982	0.0284
AGE^2	−0,0054	0.0025	−2,1862	0.0293
$BLACK$	0,3640	0.2929	1,2429	0.2145
$EDUC$	−0,1381	0.0298	−4,6403	0.0000
$YEAR$	−0,0489	0.0152	−3,2135	0.0014

Mean dependent var	2.67	S.D. dependent var	1.67
Sum squared resid	1197.9	S.E. of regression	1.60
R^2	0.0993	Adjusted R^2	0.0897
$F(5, 470)$	10.36	P-value(F)	1.9e–09

Output 2: OLS, using observations 1–476
Dependent variable: $KIDS$

	Coefficient	Std. Error	t-ratio	p-value
const	−6,0500	4.8054	−1,2590	0.2087
AGE	0,4908	0.2179	2,2518	0.0248
AGE^2	−0,0055	0.0025	−2,2398	0.0256
$BLACK$	0,3814	0.2931	1,3014	0.1938
$EDUC$	−0,1374	0.0298	−4,6184	0.0000
$Y78$	−0,1001	0.1871	−0,5351	0.5929
$Y84$	−0,5794	0.1827	−3,1706	0.0016

Mean dependent var	2.67	S.D. dependent var	1.67
Sum squared resid	1194.3	S.E. of regression	1.60
R^2	0.1020	Adjusted R^2	0.0905
$F(6, 469)$	8.87	P-value(F)	3.5e–09

Coefficient covariance matrix (Output 2)

AGE	AGE^2	$BLACK$	$EDUC$	$Y78$	$Y84$	
?	−0,0005	0,0013	0,0007	0,0034	0,0036	AGE
	?	$-1{,}4 \times 10^{-5}$	$-7{,}4 \times 10^{-6}$	$-3{,}6 \times 10^{-5}$	$-3{,}6 \times 10^{-5}$	AGE^2
		?	0	0,0030	0,0012	$BLACK$
			?	−0,0003	−0,0008	$EDUC$
				?	0,0180	$Y78$
					?	$Y84$

Output 3: OLS, using observations 1–476
Dependent variable: $KIDS$

	Coefficient	Std. Error	t-ratio	p-value
const	−15,5816	7.4233	−2,0990	0.0363
AGE	0,4401	0.2172	2,0261	0.0433
AGE^2	−0,0050	0.0025	−2,0148	0.0445
$BLACK$	0,3984	0.2917	1,3660	0.1726
$EDUC$	0,9904	0.4627	2,1403	0.0328
$YEAR$	0,1321	0.0756	1,7473	0.0812
$(YEAR \times EDUC)$	−0,0143	0.0059	−2,4438	0.0149

Mean dependent var	2.67	S.D. dependent var	1.67
Sum squared resid	1182.8	S.E. of regression	1.588072
R^2	0.1106	Adjusted R^2	0.0992
$F(6, 469)$	9.7	P-value(F)	4.2e–10

Coefficient covariance matrix (Output 3)

AGE	AGE^2	$BLACK$	$EDUC$	$YEAR$	$(YEAR \times EDUC)$	
?	−0,0005	0,0009	−0,0066	−0,0009	$9{,}3 \times 10^{-5}$	AGE
	?	$-9{,}3 \times 10^{-6}$	$7{,}5 \times 10^{-5}$	1×10^{-5}	-1×10^{-6}	AGE^2
		?	0,0069884	0,0011365	$-8{,}3 \times 10^{-5}$	$BLACK$
			?	0,034144	−0,0027	$EDUC$
				?	−0,0004	$YEAR$
					?	$(YEAR \times EDUC)$

Output 4: TSLS, using observations 1–476
Dependent variable: $KIDS$
Instrumented: $EDUC$
Instruments: const AGE AGE^2 $BLACK$ $YEAR$ $RURAL$ $LPOP$ $(YEAR \times RURAL)$ $(YEAR \times LPOP)$

	Coefficient	Std. Error	z-stat	p-value
const	−43,3158	34.9236	−1,2403	0.2149
AGE	0,6207	0.2728	2,2749	0.0229
AGE^2	−0,0069	0.0031	−2,2416	0.0250
$BLACK$	0,6224	0.3591	1,7331	0.0831
$EDUC$	3,0089	2.8246	1,0652	0.2868
$YEAR$	0,3817	0.4369	0,8736	0.3824
$(YEAR \times EDUC)$	−0,0360	0.0350	−1,0299	0.3031

Mean dependent var	2.67	S.D. dependent var	1.67
Sum squared resid	1497.2	S.E. of regression	1.79
R^2	0.0061	Adjusted R^2	−0.0066
$F(6, 469)$	4.28	P-value(F)	0.00033

Coefficient covariance matrix (Output 4)

AGE	AGE^2	$BLACK$	$EDUC$	$YEAR$	$(YEAR \times EDUC)$	
?	−0,0008	0,0058	−0,1016	−0,0195	0,0015	AGE
	?	$-5{,}7 \times 10^{-5}$	0,0012	0,0002	$-1{,}8 \times 10^{-5}$	AGE^2
		?	0,3550	0,0521	−0,0042	$BLACK$
			?	1,2291	−0,0986	$EDUC$
				?	−0,0153	$YEAR$
					?	$(YEAR \times EDUC)$

Output 5A: OLS, using observations 1–476
Dependent variable: $EDUC$

	Coefficient	Std. Error	t-ratio	p-value
const	15,7871	8.2430	1,9152	0.0561
AGE	−0,6551	0.3326	−1,9693	0.0495
AGE^2	0,0071	0.0038	1,8905	0.0593
$BLACK$	−0,3120	0.4519	−0,6904	0.4903
$YEAR$	0,1492	0.0358	4,1732	0.0000
$RURAL$	11,1516	4.1572	2,6825	0.0076
$LPOP$	5,0678	3.6998	1,3698	0.1714
$(YEAR \times RURAL)$	−0,1509	0.0530	−2,8455	0.0046
$(YEAR \times LPOP)$	−0,0586	0.0473	−1,2389	0.2160

Mean dependent var	12.71	S.D. dependent var	2.53
Sum squared resid	2743.1	S.E. of regression	2.42
R^2	0.0974	Adjusted R^2	0.0819
$F(8, 467)$	6.3	P-value(F)	9.2e–08

Output 5B: OLS, using observations 1–476
Dependent variable: $EDUC$

	Coefficient	Std. Error	t-ratio	p-value
const	24,0167	7.7205	3,1108	0.0020
AGE	−0,7663	0.3354	−2,2849	0.0228
AGE^2	0,0083	0.0038	2,1778	0.0299
$BLACK$	−0,5512	0.4528	−1,2174	0.2241
$YEAR$	0,0779	0.0233	3,3433	0.0009

Mean dependent var	12.71	S.D. dependent var	2.53
Sum squared resid	2878.1	S.E. of regression	2.47
R^2	0.0530	Adjusted R^2	0.0449
$F(4, 471)$	6.6	P-value(F)	0.00004

Output 5C: OLS, using observations 1–476
Dependent variable: $(YEAR \times EDUC)$

	Coefficient	Std. Error	t-ratio	p-value
const	322,8543	649.5532	0,4970	0.6194
AGE	−54,4209	26.2127	−2,0761	0.0384
AGE^2	0,5941	0.2974	1,9977	0.0463
$BLACK$	−22,7724	35.6073	−0,6395	0.5228
$YEAR$	24,2667	2.8177	8,6124	0.0000
$RURAL$	907,9992	327.5882	2,7718	0.0058
$LPOP$	355,6616	291.5432	1,2199	0.2231
$(YEAR \times RURAL)$	−12,3139	4.1784	−2,9471	0.0034
$(YEAR \times LPOP)$	−4,0834	3.7263	−1,0958	0.2737

Mean dependent var	997.41	S.D. dependent var	220.29
Sum squared resid	17033368	S.E. of regression	190.98
R^2	0.2610	Adjusted R^2	0.2484
$F(8, 467)$	20.6	P-value(F)	8.4e–27

Output 5D: OLS, using observations 1–476
Dependent variable: $(YEAR \times EDUC)$

	Coefficient	Std. Error	t-ratio	p-value
const	957,0200	609.1059	1,5712	0.1168
AGE	−63,0297	26.4602	−2,3821	0.0176
AGE^2	0,6831	0.3003	2,2748	0.0234
$BLACK$	−40,9939	35.7203	−1,1476	0.2517
$YEAR$	18,7665	1.8383	10,2087	0.0000

Mean dependent var	997.41	S.D. dependent var	220.29
Sum squared resid	17914806	S.E. of regression	195.03
R^2	0.2228	Adjusted R^2	0.2162
$F(4, 471)$	33.7	P-value(F)	8.8e–25

Output 6: OLS, using observations 1–476
Dependent variable: $KIDS$

	Coefficient	Std. Error	t-ratio	p-value
const	−43,3158	30.9547	−1,3993	0.1624
AGE	0,6207	0.2418	2,5665	0.0106
AGE^2	−0,0069	0.0027	−2,5291	0.0118
$BLACK$	0,6224	0.3183	1,9553	0.0511
$EDUC$	3,0089	2.5036	1,2018	0.2300
$YEAR$	0,3817	0.3873	0,9856	0.3249
$(YEAR \times EDUC)$	−0,0360	0.0310	−1,1619	0.2459
$RES5A$	−2,0124	2.5476	−0,7899	0.4300
$RES5C$	0,0214	0.0316	0,6790	0.4975

NOTA: $RES5A$ y $RES5C$ son los respectivos residuos de las Output 5A y 5C.

Mean dependent var	2.67	S.D. dependent var	1.67
R^2	0.1250	Adjusted R^2	

Output 7: OLS, using observations 1–476
Dependent variable: $RES4$

	Coefficient	Std. Error	t-ratio	p-value
const	−0,9642	6.0759	−0,1587	0.8740
AGE	0,0135	0.2452	0,0551	0.9561
AGE^2	−0,0001	0.0028	−0,0536	0.9572
$BLACK$	−0,0349	0.3331	−0,1047	0.9166
$YEAR$	0,0098	0.0264	0,3716	0.7104
$RURAL$	−0,8046	3.0643	−0,2626	0.7930
$LPOP$	2,3062	2.7271	0,8456	0.3982
$(YEAR \times RURAL)$	0,0082	0.0391	0,2105	0.8334
$(YEAR \times LPOP)$	−0,0315	0.0349	−0,9027	0.3671

NOTA: $RES4$ son los residuos de la Salida 4.

Mean dependent var	0.0000	S.D. dependent var	1.77
Sum squared resid	1490.4	S.E. of regression	1.79
R^2	0.0046	Adjusted R^2	−0.0125
$F(8, 467)$	0.2689	P-value(F)	0.9757

Statistical Tables:

Critical Values $N(0,1)$	
	Acumulated Probability
99,5 %	2,576
99 %	2,326
97,5 %	1,960
95 %	1,645
90 %	1,282

Critical Values χ^2_m				
		Acumulated Probability		
	m	90 %	95 %	99 %
	1	2,7	3,8	6,6
	2	4,6	6,0	9,2
	3	6,2	7,8	11,3
	4	7,8	9,5	13,3
	5	9,2	11,1	15,1

1. Assume that model (I) verifies the assumptions of the classical regression model. An appropriate estimate of $V\left(KIDS|\, AGE, BLACK, EDUC, YEAR\right)$ (rounded to 1 decimal), is:

   a) 2,8.

   b) 2,6.

   c) 1,6.

2. Assume that model (I) verifies the assumptions of the classical regression model. If the variable $EDUC$ was measured with error (and such error had strictly positive variance), estimates in Output 1 will be:

   a) Always inconsistent.

   b) Inconsistent, only if the measurement error is correlated with the error term of the model.

   c) Consistent, but less efficient than if the variable were measured without error.

3. If a relevant variable were omitted in model (I), estimates in Output 1 will be:

   a) Always inconsistent.

   b) Consistent, provided that the omitted variable is uncorrelated with the remaining explanatory variables in the model.

   c) None of the other statements is true.

4. Assume that model (I) verifies the assumptions of the classical regression model. Consider two white women in the same year and with the same age, but the first one having 7 years of education less than the second one. The first woman, on average, will have approximately (rounding to the closest integer number):

   a) 1 child less than the second one.

   b) The same number of children than the second one.

   c) 1 child more than the second one.

5. Assume that model (I) verifies the assumptions of the classical regression model. Under the evidence of Output 1, the effect of education on the number of children is:

   a) Negative (on average) for all the women in the sample.

   b) Positive (on average) for all the black women in the sample, as the coefficient of $BLACK$ is higher in absolute value than the coefficient of $EDUC$.

   c) None of the other statements is true.

6. Assume that model (I) verifies the assumptions of the classical regression model. Under the evidence of Output 1, we can assert that the mean fertility rates:

   a) Have remained constant over time.

   b) Have decreased over time.

   c) We do not have conclusive evidence.

7. Assume that model (I) verifies the assumptions of the classical regression model. Under the evidence of Output 1, for a given age, race and education level (rounding to 1 decimal):

   a) A woman in 1984 had on average 0,6 children less than a woman in 1972.

b) A woman in 1978 had on average 0,3 children more than a woman in 1972.

c) A woman in 1978 had on average 0,3 children less than a woman in 1984.

8. Comparing models (I) and (II):

   a) Model (I) is more restrictive, as it imposes that the effect of education on the number of children in 1972 is null.

   b) Model (II) is less restrictive, as it allows, for a given race, age, and education level, the fertility rate to change differently over time.

   c) Models (I) and (II) are not comparable, since model (I) includes variables that model (II) does not, and vice versa.

9. Comparing models (I) and (II):

   a) Models (I) and (II) are different models, since none of them is a particular case of the other one.

   b) Model (I) imposes the constraint that the coefficients of $Y78$ and $Y84$ were equals.

   c) Modelo (I) imposes the constraint that the coefficient of $Y78$ is exactly half of the coefficient of $Y84$.

10. Using $KIDS$ as dependent variable, consider models that include a constant, AGE, AGE^2, $BLACK$ and $EDUC$. Then:

    a) If we also included $YEAR$ and $Y78$ as explanatory variables and estimate by OLS, the R^2 would be higher than the one in Output 2.

    b) If we also included $YEAR$ and $Y84$ as explanatory variables and estimate by OLS, the estimated coefficients of AGE, AGE^2, $BLACK$ and $EDUC$ would be the same than those in Output 2.

    c) If we also included $YEAR$, $Y78$ and $Y84$ as explanatory variables, such model would be more general than model (I) or model (II).

11. Assume that the error of model (II) satisfies $E\left(\varepsilon_2 | AGE, BLACK, EDUC, Y78, Y84\right) \neq 0$ for any combination of values of the conditioning variables, while the homoskedasticity does not hold. Additionally, we have four additional variables Z_1, Z_2, Z_3, Z_4, non included in the model and uncorrelated with ε_2. Then, in any case:

    a) If we estimated the model (II) by 2SLS using Z_1, Z_2, Z_3, Z_4 as instruments, the estimators obtained for the coefficients δ_1, δ_2, δ_3, δ_4, would be consistent.

    b) If we estimated the model (II) by OLS, the estimators obtained for the coefficients δ_1, δ_2, δ_3, δ_4, would be inconsistent.

    c) If we estimated the model (II) by OLS including Z_1, Z_2, Z_3, Z_4 as additional variables, the estimators obtained for the coefficients δ_1, δ_2, δ_3, δ_4, would be consistent.

12. Assume that model (I) verifies the assumptions of the classical regression model. Given the estimated coefficients, for a given age, educational level and race:

    a) For every 50 women, there are approximately 29.3 more children in 1972 than in 1984.

    b) A woman in 1978 has approximately 29.3 % less children than a woman in 1972.

    c) For every 50 women, there are approximately 2.4 more children in 1972 than in 1984.

13. Assume that model (III) verifies the assumptions of the classical regression model. Given the estimated coefficients, in 1972, the mean difference in the number of children between a black and a white woman with the same age but the second one with 5 years of education less is (rounding to 1 decimal):

    a) 0,2 more children.
    b) 5,3 more children.
    c) 4,5 less children.

14. Assume that model (III) verifies the assumptions of the classical regression model. Given the estimated coefficients, for a 20 year old black woman with 5 years of education the mean number of children has decreased between 1978 and 1984 (rounding to 1 decimal) by:

    a) $-0{,}7$.
    b) $-0{,}4$.
    c) $-0{,}8$.

15. Assume that model (III) verifies the assumptions of the classical regression model. Given the estimated coefficients (and rounding to two decimals), for a 20 year old white woman with 10 years of education, in 1972, the mean number of children is approximately:

    a) 0,34.
    b) 1,12.
    c) 0,98.

16. Assume that model (I) verifies the assumptions of the classical regression model. Given the estimated coefficients, the mean difference in the number of children between two women with the same characteristics, but one in 1972 and the other in 1978, is:

    a) Significantly different from zero.
    b) Statistically equal to zero.
    c) The question cannot be answered with the provided information.

17. Assume that model (III) verifies the assumptions of the classical regression model. If we wanted to test whether the mean number of children in 1972 is the same for a black woman with 10 years of education than the one for a black woman with the same age but with 12 years of education, the null hypothesis would be:

    a) $H_0 : \gamma_4 + 72\gamma_6 = 0$.
    b) $H_0 : 2\gamma_4 + 72\gamma_6 = 0$.
    c) $H_0 : \gamma_4 = -144\gamma_6$.

18. Assume that model (II) verifies the assumptions of the classical regression model. If we wanted to test that the causal effect of age on the number of children is constant, the null hypothesis would be:

    a) $H_0 : \delta_1 = \delta_2 = 0$.
    b) $H_0 : \delta_2 = 0$.
    c) $H_0 : \delta_1 - \delta_2 = 0$.

19. Assume that model (III) verifies the assumptions of the classical regression model. Under the evidence of Output 3, and taking into account only women younger than 40 years of age:

    a) More educated women have, on average, more children.
    b) Older women have, on average, more children.
    c) The causal effect of education is the same for all women considered.

20. Assume that model (III) verifies the assumptions of the classical regression model. Under the evidence of Output 3:

    a) The causal effect of education is positive.
    b) The causal effect of education is more negative in 1978 than in 1972.
    c) The causal effect of education is more negative in 1978 than in 1984.

21. Assume that model (III) verifies the assumptions of the classical regression model. If we wanted to test that education does not affect fertility, the null hypothesis is:

    a) $H_0 : \begin{cases} \gamma_4 - \gamma_6 = 0 \\ \gamma_6 = 0 \end{cases}$.
    b) $H_0 : \gamma_4 = \gamma_6$.
    c) $H_0 : \gamma_4 = 0$.

22. Assume that model (III) verifies the assumptions of the classical regression model. If we wanted to test that the causal effect of education on the number of children does not depend on the year, the null hypothesis is:

    a) $H_0 : \gamma_6 = 0$.
    b) $H_0 : \gamma_4 = \gamma_6 = 0$.
    c) $H_0 : \gamma_5 = \gamma_6 = 0$.

23. Assume that model (III) verifies the assumptions of the classical regression model. If we wanted to test that the mean number of children does not depend on the year, the null hypothesis is:

    a) $H_0 : \gamma_5 = \gamma_6 = 0$.
    b) $H_0 : \gamma_5 = 0$.
    c) $H_0 : \gamma_5 = \gamma_6$.

24. Assume that model (III) verifies the assumptions of the classical regression model. If we wanted to test that the mean number of children does not depend on the year:

    a) We do not reject it, because the p-value of the corresponding statistic is equal to 0.
    b) We reject it, given the the magnitude of the corresponding statistic obtained when comparing the unrestricted model and the model that imposes such restriction.
    c) The question cannot be answered with the provided information.

25. Comparing models (I), (II) and (III):

    a) Model (I) is the most restrictive.
    b) Model (III) is the least restrictive.

c) Models (I) and (II) are not comparable, because none of them is a particular case of the other.

26. Assuming that models (I) and (II) verify the assumptions of the classical regression model, if race ($BLACK$) was an irrelevant variable:

    a) Output 1 will provide inconsistent estimates for model (I) parameters.

    b) Output 2 will provide consistent estimates for model (II) parameters only if we dropped the variable $BLACK$ from the explanatory variables.

    c) Output 1 and Output 2 will provide consistent estimates for the parameters of models (I) and (II), respectively.

27. Assume that model (II) verifies the assumptions of the classical regression model. If the race ($BLACK$) was an irrelevant variable, the variance of the estimators of the coefficients of the relevant variables will be higher:

    a) The higher the correlation of $BLACK$ with the relevant variables.

    b) The lower the correlation of $BLACK$ with the relevant variables.

    c) The higher the proportion of black women in the sample.

28. If education was an endogenous variable:

    a) The estimated coefficients in Output 1 would be inconsistent for model (I), but those of Output 2 would be consistent for model (II).

    b) The estimated coefficients in Output 2 would be inconsistent for model (II), but those of Output 3 would not be inconsistent for model (III).

    c) None of the other statements is true.

29. We want to obtain the causal effect of education on the number of children for model (III). If education was an endogenous variable, given the available information:

    a) We would need at least one variable, not included in the model and uncorrelated with ε_3, to get consistent estimates of the parameters of interest by 2SLS.

    b) We would need at least two different variables, not included in the model and uncorrelated with ε_3, to get consistent estimates of the parameters of interest by 2SLS, since model (III) includes the interaction of education with the variable $YEAR$.

    c) None of the other statements are correct.

30. If education was an endogenous variable, in order to test whether both $RURAL$ and $LPOP$ are valid instruments, we would have to:

    a) In a regression of $EDUC$ on the exogenous variables of the model and both instruments and on their corresponding interactions with the variable $YEAR$, test whether such instruments and their corresponding interactions are jointly significant.

    b) Test the hypothesis that the residual of the reduced form (linear projection of $EDUC$ on the exogenous variables of the model and both instruments) has a significant effect on education.

    c) In a regression of $EDUC$ on the exogenous variables of the model and both instruments, test whether such instruments are individually significant.

31. The 2SLS estimated coefficients in Output 4 could have been obtained in an equivalent way:

a) Estimating by OLS a model with $KIDS$ as the dependent variable and including as regressors all the exogenous explanatory variables and the corresponding predictions of education and its interaction with $YEAR$, by using Outputs 5B and 5D, respectively.

b) Estimating by OLS a model with $KIDS$ as the dependent variable and including as regressors all the exogenous explanatory variables, as well as the variables $RURAL$ and $LPOP$ and their corresponding interactions with $YEAR$.

c) Estimating by OLS a model with $KIDS$ as the dependent variable and including as regressors all the exogenous explanatory variables and the corresponding predictions of education and its interaction with $YEAR$, by using Outputs 5A and 5C, respectively.

32. The 2SLS estimated coefficients in Output 4 could have been obtained in an equivalent way:

a) Estimating by 2SLS a model with $KIDS$ as the dependent variable and including as regressors all the exogenous explanatory variables, the education and its interaction with $YEAR$, using as instruments the predictions based on Outputs 5B and 5D.

b) Estimating by 2SLS a model with $KIDS$ as the dependent variable and including as regressors all the exogenous explanatory variables, the education and its interaction with $YEAR$, using as instruments the predictions based on Outputs 5A and 5C.

c) None of the other statements is true.

33. If $RURAL$ and $LPOP$ were uncorrelated with ε_3, and we wanted to test that the variables $RURAL$ and $LPOP$ are valid instruments for $EDUC$, the test statistic would approximately be:

a) 51,4.

b) 23,4.

c) 7,8.

34. Suppose that we can ensure that AGE, $BLACK$ and, of course, $YEAR$, are uncorrelated with ε_2. Also, assume that $RURAL$ and $LPOP$ are uncorrelated with ε_2. If we had estimated model (II) by 2SLS but using $RURAL$ as the only instrument for $EDUC$, the estimators obtained for model (II) parameters:

a) Would be inconsistent.

b) Would be less efficient than the ones by 2SLS estimator but using both $RURAL$ and $LPOP$ as instruments.

c) The Gretl program would indicate us that we do not have enough instruments.

35. Suppose that we are interested in model (III). The information provided by Output 6 allows to assess whether:

a) We can reject the null hypothesis about the exogeneity of education.

b) We can reject the null hypothesis about instruments'validity.

c) $RURAL$ is a better instrument than $LPOP$.

36. Suppose that we are interested in model (III). Given the results:

a) We reject that $EDUC$ is exogenous.

b) We do not reject that the correlation of the instruments with $EDUC$ is equal to zero.

c) We do not reject that the correlation of the instruments with the error in model (III) is equal to zero.

37. Suppose that we are interested in model (II). Consider the following statement: "In 1972, the fertility rate of a 30 year old black woman with 10 years of education is equal to the fertility rate of black woman with the same educational level but 1 year younger". Given the results (rounding to 2 decimals):

    a) The test statistic is approximately $t = 0{,}06$.
    b) The test statistic is approximately $t = 0{,}63$.
    c) The test statistic is approximately $t = 1{,}64$.

38. Suppose that we are interested in model (II). Consider the following statement: "In 1972, the fertility rate of a 30 year old black woman with 10 years of education is equal to the fertility rate of black woman with the same educational level but 1 year younger". Given the results:

    a) At the 1 % significance level, we can reject such assertion.
    b) At the 5 % significance level, we can reject such assertion.
    c) We cannot reject such assertion at the 5 % significance level.

39. Focusing on models (I) and (II):

    a) Model (II) is misspecified, since it omits the variable $Y72$.
    b) Model (I) is a particular case of model (II).
    c) None of the other statements are correct.

40. Suppose that we are interested in model (II) and its relation with model (I). Consider the following conjecture: "For a given age, race, and educational level, the decrease in the fertility rate is constant over time". If such conjecture was true, it must occur that:

    a) The constant terms of both models are equal, $\beta_0 = \delta_0$.
    b) $6\beta_5 = \delta_6 - \delta_5$.
    c) $\delta_5 = \delta_6$.

41. Comparing models (I) and (II), model (II) can be expressed as model (I) with the following restriction:

    a) $\delta_6 = \delta_5$.
    b) $\delta_6 = 2\delta_5$.
    c) $\delta_6 = 6\delta_5$.

42. Assume that model (III) verifies the assumptions of the classical regression model. If we wanted to test whether the causal effect of education for women observed in 1978 is the same as on women observed in 1984, the null hypothesis would be:

    a) $\gamma_6 = 0$.
    b) $\gamma_4 = \gamma_5 = \gamma_6$.
    c) $\gamma_4 = \gamma_6 = 0$.

43. Assume that model (III) verifies the assumptions of the classical regression model. If we wanted to test whether the causal effect of education for women observed in 1978 is the same as on women observed in 1984, the test statistic, in absolute value (rounding to 1 decimal) would be:

   a) 1,7.
   b) 2,4.
   c) 14,7.

44. Assume that model (III) verifies the assumptions of the classical regression model. If we wanted to test whether the causal effect of education for women observed in 1978 is the same as on women observed in 1984, we can conclude that:

   a) We do not reject the null hypothesis at the 5 % significance level.
   b) We reject the null hypothesis at the 5 %, but not at the 1 % significance level.
   c) We reject the null hypothesis at the 1 % significance level.

45. In model (III), suppose that education is endogenous. Results in Output 7 allow us to test whether:

   a) None of the instruments used in Output 4 is correlated with ε_3.
   b) Education is not correlated with ε_3.
   c) Education is not correlated with the instruments used in Output 4.

46. In model (III), suppose that education is endogenous. Given the instruments used in Output 4, we can assert that the model is:

   a) Exactly identified.
   b) Over-identified, the number of over-identifying restrictions being equal to 1.
   c) Over-identified, the number of over-identifying restrictions being equal to 2.

47. In model (III), suppose that education is endogenous. Given the instruments used in Output 4, if we wanted to perform a test of over-identifying restrictions, the appropriate test statistic would be:

   a) The R^2 from Output 7 multiplied by the number of observations.
   b) The R^2 from Output 5A multiplied by the number of observations.
   c) The R^2 from Output 4 multiplied by the number of observations.

48. In model (III), suppose that education is endogenous. Given the instruments used in Output 4, and given the results provided:

   a) We do not reject the instruments'validity.
   b) We reject the instruments'validity.
   c) There is not information to assess whether the instruments are valid or not.

Answer Keys to exam type B

1	b
2	a
3	b
4	c
5	a
6	b
7	a
8	b
9	c
10	b
11	b
12	a
13	a
14	b
15	a
16	a
17	a
18	b
19	b
20	b
21	a
22	a
23	a
24	c
25	a
26	c
27	a
28	c
29	a
30	a
31	c
32	b
33	b
34	b
35	a
36	a
37	c
38	c
39	b
40	b
41	b
42	a
43	b
44	b
45	a
46	c
47	a
48	a