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

DURATION: 2 HOURS and 30 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:

$$[0{,}24 \times (\#\text{ correct answers}) - 0{,}08 \times (\#\text{ incorrect answers})] + 0{,}16$$

YOUR ANSWERS

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

Problem 1:

Consider the mass **conditional** probability distributions of Y given X, $P(Y|X = X_j)$, $X_j = 0, 5, 10$.

$P(Y\|X = X_j)$		X		
		0	**5**	**10**
Y	**5**	1/3	1/2	1/3
	10	1/3	1/2	1/3
	15	1/3	0	1/3

and the mass marginal probability distribution of X, $P(X)$

$P(X)$	X		
	0	**5**	**10**
	3/10	4/10	3/10

We want to characterize $E(Y|X)$, $E(X|Y)$, $L(Y|X)$, $L(X|Y)$, and use them to make predictions.

Problem 2:

We have data from a year auction of antique clocks organized by the German company Triberg Clock. The following model is considered:

$$P = \beta_0 + \beta_1 A + \beta_2 C + \beta_3 A^2 + \beta_4 C^2 + \beta_5 (A \times C) + \varepsilon,$$

where P is the price, in hundred euros, of the winning bid, A is the age of the clock (in years) and C is the number of bidders. Besides, the error term verifies, for any age and any number of bidders, $E(\varepsilon | A, C) = 0$ and $V(\varepsilon | A, C) = \sigma^2$.

The descriptive statistics of the variables are:

Output 0: Summary Statistics, using the obs 1–32

Variable	Mean	Std. Dev.	Minimum	Maximum
P	1328.1	393.6	729.0	2131.0
A	144.9	27.4	108.0	194.0
C	9.5	2.8	5.0	15.0

We will assume that the corresponding population characteristics of such variables coincide with their sample counterparts. Notice that some questions can require the knowledge of such population characteristics, particularly the range of values taken by the variables of interest.

We have obtained the following estimates:

Output 1: OLS, using observations 1–32
Dependent variable: P

	Coefficient	Std. Error	t-ratio	p-value
const	−399.31	777.48	−0.51	0.61
A	4.07			0.66
C	16.57	63.23	0.26	0.79
A^2	−0.005	0.027	−0.20	0.84
C^2	−4.23	2.18	−1.94	0.06
$(A \times C)$	1.11	0.23	4.74	0.00

Sum squared resid	199132.1		
R^2	0.9585	Adjusted R^2	0.9506

Output 2: OLS, using observations 1–32
Dependent variable: P

	Coefficient	Std. Error	t-ratio	p-value
const	−292.73	636.67	−0.45	0.64
$4A + C$	0.89	2.18	0.41	0.68
A^2	−0.005	0.027	−0.19	0.84
C^2	−3.75	1.02	−3.68	0.001
$(A \times C)$	1.16	0.15	7.50	0.00

Sum squared resid	199603.5		
R^2	0.9584	Adjusted R^2	0.9523
$F(4, 27)$	155.6989	P-value(F)	3.13e–18

Output 3: OLS, using observations 1–32
Dependent variable: P

	Coefficient	Std. Error	t-ratio	p-value
const	−399.31	777.48	−0.51	0.61
A	20.64	66.11	0.31	0.76
$C - A$	16.57	63.23	0.26	0.79
A^2	−0.005	0.027	−0.20	0.84
C^2	−4.23	2.18	−1.94	0.06
$(A \times C)$	1.11	0.23	4.74	0.00

Sum squared resid	199132.1		
R^2	0.9585	Adjusted R^2	0.9506

Problem 3:

A sport articles company wants to elucidate the impact of individual income and other individual characteristics on sales. For such purpose, it undertakes a survey for a sample of individuals who buy sport articles, considering the following specification:

$$\begin{aligned} SPORT &= \beta_0 + \beta_1 INC + \beta_2 AGE + \beta_3 AGE^2 + \beta_4 FEM + \beta_5 SING + \beta_6 SOUTH \\ &+ \beta_7 WGHT + \beta_8 WGHT \times FEM + \beta_9 SOUTH \times FEM + \varepsilon \end{aligned} \tag{S}$$

where, for each individual,

- $SPORT$ is annual expenditure in sport articles (in thousand euros);
- INC is annual income (in thousand euros);
- AGE is age (in years);
- FEM is a binary variable that equals 1 if the individual is a woman and 0 if the individual is a man;
- $SING$ is a binary variable that equals 1 if the individual is not married and 0 otherwise;
- $SOUTH$ is a binary variable that equals 1 if the individual lives in the South of the country and 0 otherwise;
- $WGHT$ is weight (in kilograms).

Besides, income can be correlated with unobserved characteristics that might also affect expenditure in sport articles. Hence, it can be that $C(INC, \varepsilon) \neq 0$. The remaining explanatory variables in the model above are not correlated with the error term.

In addition to the explanatory variables in the model above, there is also information about the individual's years of education ($EDUC$) and the years of education of the individual's father ($FEDUC$). These variables are uncorrelated with any unobserved characteristic that might also affect expenditure in sport articles.

We have obtained the following estimates:

Output 1: OLS estimates using the 935 observations 1–935
Dependent variable: $SPORT$

	Coefficient	Std. Error	t-ratio	p-value
const	37,7486	31.5318	1,1972	0.2316
INC	0,6671	0.6138	1,0869	0.2774
AGE	0,3401	1.9090	0,1782	0.8586
AGE^2	−0,0036	0.0014		
FEM	−2,2428	0.6031	−3,7190	0.001
$SING$	0,7775			0.002
$SOUTH$	−0,1803	0.5536	−0,3257	0.7447
$WGHT$	−0,1025	0.0512	-2.0023	0.0455
$WGHT \times FEM$	−0,0323	0.1506	−0,2144	0.8303
$SOUTH \times FEM$	−0,0692	1.4793	−0,0468	0.9627

Mean dependent var	0.043929	S.D. dependent var	0.007224
Sum squared resid	0.047795	S.E. of regression	0.007188
R^2	0.019495	Adjusted R^2	0.009955

NOTE to Output 1: The R^2 of the OLS estimation of a similar specification that omits FEM, $WGHT \times FEM$ and $SOUTH \times FEM$ is 0,009495

Coefficient covariance matrix (Output 1)

INC	AGE	AGE^2	FEM	$SING$	$SOUTH$	$WGHT$	$(WGHT \times FEM)$	$(SOUTH \times FEM)$	
?	0.05	−0.10	−0.08	0.05	0	0	0	0	INC
	?	−0.05	0	0.02	0	−0.10	0.02	−0.05	AGE
		$0{,}029^2$	0	0	0	0	−0.01	0	AGE^2
			?	0.1	0.10	0.02	−0.20	−1.30	FEM
				0,063	0.01	−0.40	−0.10	−0.10	$SING$
					?	0	−0.002	−0.30	$SOUTH$
						?	−0.10	0	$WGHT$
							?	0.01	$(WGHT \times FEM)$
								2.188	$(SOUTH \times FEM)$

Output 2: OLS estimates using the 935 observations 1–935
Dependent variable: INC

	Coefficient	Std. Error	t-ratio	p-value
const	1,2033	1.8148	0,6631	0.5075
AGE	−0,0577	0.1103	−0,5233	0.6010
AGE^2	0,0006	0.0017	0,3597	0.7192
FEM	0,1523	0.0995	1,5305	0.1263
$SING$	−0,1739	0.0436	−3,9930	0.0001
$SOUTH$	0,0589	0.0315	1,8688	0.0620
$WGHT$	−0,0035	0.0029	−1,2100	0.2267
$WGHT \times FEM$	−0,0165	0.0102	−1,6209	0.1055
$SOUTH \times FEM$	0,0977	0.0966	1,0111	0.3123
$FEDUC$	0,0138	0.0047	2,9464	0.0033
$EDUC$	0,0470	0.0068	6,9315	0.0000

Mean dependent var	0.975920	S.D. dependent var	0.405896
Sum squared resid	99.17078	S.E. of regression	0.368579
R^2	0.186567	Adjusted R^2	0.175424

NOTE to Output 2: The R^2 of the OLS estimation of a similar specification that omits $FEDUC$ and $EDUC$ is 0,1717

Output 3: OLS estimates using the 935 observations 1–935
Dependent variable: $SPORT$

	Coefficient	Std. Error	t-ratio	p-value
const	48,0705	35.4399	1,3564	0.1754
INC	0,4594	0.2069	2,2199	0.0267
AGE	−0,4097	2.1563	−0,1900	0.8494
AGE^2	0,0055	0.0324	0,1696	0.0899
FEM	−0,5713	1.9915	−0,2869	0.7743
$SING$	−0,2548	0.9057	−0,2813	0.7786
$SOUTH$	0,1037	0.6348	0,1634	0.8703
$WGHT$	−0,1139	0.0568	−2,0047	0.0454
$WGHT \times FEM$	−0,2221	0.2004	−1,1082	0.2681
$SOUTH \times FEM$	1,6185	1.8847	0,8587	0.3908
RES_INC	6,3914	2.1915	2,9164	0.0036

Mean dependent var	0.043929	S.D. dependent var	0.007224
Sum squared resid	0.037615	S.E. of regression	0.007178
R^2	0.028753	Adjusted R^2	0.015448

NOTE to Output 3: The variable RES_INC denotes the residuals of Output 2.

Output 4: TSLS,using the 935 observations 1–935
Dependent variable: $SPORT$
Instrumented: INC
Instruments: const AGE AGE^2 FEM $SING$ $SOUTH$ $WGHT$
$WGHT \times FEM$ $SOUTH \times FEM$ $FEDUC$ $EDUC$

	Coefficient	Std. Error	z	p-value
const	48,0705	37.2741	1,2896	0.1972
INC	0,4594	0.2177	2,1107	0.0348
AGE	−0,4097	2.2679	−0,1806	0.8566
AGE^2	0,0055	0.0341	0,1620	0.8713
FEM	−0,5713	2.0946	−0,2727	0.7850
$SING$	−0,2548	0.9526	−0,2675	0.7891
$SOUTH$	0,1037	0.6676	0,1553	0.8766
$WGHT$	−0,1139	0.0598	−1,9060	0.0566
$WGHT \times FEM$	−0,2221	0.2107	−1,0537	0.2920
$SOUTH \times FEM$	1,6185	1.9822	0,8165	0.4142

Mean dependent var	0.043929	S.D. dependent var	0.007224
Sum squared resid	0.041666	S.E. of regression	0.007549
R^2	0.000006	Adjusted R^2	-0.012306

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. (Problem 1) Consider the following statements:
   I. X and Y are uncorrelated.
   II. The higher the X is, the higher Y is.
   III. X and Y are independent.

   a) None of the three statements is true.
   b) Only II. is true.
   c) Only I. and III. are true.
   d) Only I. is true.

2. (Problem 1) Consider the following statements:
   I. Y and X are uncorrelated.
   II. The higher the Y is, the higher X is.
   III. X and Y are independent.

   a) None of the three statements is true.
   b) Only II. is true.
   c) Only I. and III. are true.
   d) Only I. is true.

3. (Problem 1) Consider the following statements:
   I. $L(Y|X)$ is constant for all X.
   II. $E(X|Y)$ is constant for all Y.
   III. The conditional expectation of Y given X is linear in X.

   a) Only I. is true.
   b) Only I. and II. are true.
   c) The three statements are true.
   d) None of the three statements is true.

4. (Problem 1) Consider the following statements:
I. $L(X|Y) = E(X|Y)$.
II. $L(X|Y)$ has zero slope.
III. $L(Y|X)$ has zero slope.

   a) The three statements are true.
   b) Only II. and III. are true.
   c) Only II. is true.
   d) Only III. is true.

5. (Problem 1) The linear projection of Y given X is approximately:

   a) $9 + 0X$.
   b) $1{,}68 + 0{,}05X$.
   c) None of the other answers is true.
   d) We cannot answer with the information available.

6. (Problem 1) If we consider the best prediction of Y for a randomly chosen observation:
I. It is approximately 9, if we ignore the value of X for such observation.
II. It is approximately 9, if we know that $X = 5$ for such observation.
III. It is approximately 9, if we know that $X = 10$ for such observation.

   a) Only I. and III. are true.
   b) Only I. and II. are true.
   c) The three statements are true.
   d) Only I. is true.

7. (Problem 1) If we consider the best prediction of X for a randomly chosen observation:
I. It is approximately 5, if we ignore the value of Y for such observation.
II. It is approximately 5, if we know that $Y = 5$ for such observation.
III. It is approximately 5, if we know that $Y = 15$ for such observation.

   a) Only I. and III. are true.
   b) Only I. and II. are true.
   c) The three statements are true.
   d) Only I. is true.

8. (Problem 1) Given the information available, conditioning on $X = 0$, the best prediction of Y would approximately be:

   a) 9.
   b) 10.
   c) None of the other answers is true.
   d) 1,7.

9. (Problem 1) If X changes from 5 to 10, the causal effect on Y is approximately equal to (rounding to one decimal digit):

   a) 0.

b) $-2{,}5$.
c) 2,5.
d) None of the other answers is true.

10. (Problem 1) The conditional expectation of X given Y is:

    a) Constant for all Y.
    b) Strictly increasing with Y.
    c) None of the other answers is true.
    d) We cannot answer with the information available.

11. (Problem 2) Given Output 1:

    a) The explanatory variables account for more than 95 % of price variation.
    b) The estimated model is inappropriate, since it predicts negative prices for certain values of the explanatory variables (for instance, when $A = 1$ and $C = 1$).
    c) The model does not hold the assumptions of the classical regression model, since the sum of squared residuals is greater than zero.
    d) None of the other answers is true.

12. (Problem 2) Given Output 1, to test the null hypothesis that neither age nor the number of bidders affect price:

    a) We reject the null hypothesis at the 0.1 % significance level.
    b) None of the other answers is true.
    c) We do not reject the null hypothesis at the 1 % significance level.
    d) We cannot answer with the information available.

13. (Problem 2) Given Output 1, to test that $\beta_1 = 0$:

    a) We do not reject the null hypothesis at the usual significance levels.
    b) We reject the null hypothesis at the 5 % significance level.
    c) We cannot answer with the information available.
    d) None of the other answers is true.

14. (Problem 2) Given Output 1, the *ceteris paribus* effect of age on the clock's price is:

    a) Positive and marginally increasing with age.
    b) Positive and marginally decreasing with age.
    c) Positive and independent of clock's age.
    d) None of the other answers is true.

15. (Problem 2) Given Output 1, the *ceteris paribus* effect of age on the clock's price is:

    a) Positive and marginally increasing with the number of bidders.
    b) Positive and independent of the number of bidders.
    c) None of the other answers is true.
    d) Positive and marginally decreasing with the number of bidders.

16. (Problem 2) Given Output 1, for an auction with 10 bidders of a 120 years old clock, increasing the number of bidders yields an estimated marginal change in price of approximately:

    a) 65,17 euros.
    b) 6517 euros.
    c) 16,57 euros.
    d) 1657 euros.

17. (Problem 2) Given Output 1, for an auction with 5 bidders of a 120 years old clock, increasing the number of bidders yields an estimated marginal change in price of approximately:

    a) 10747 euros.
    b) 107,47 euros.
    c) 16,57 euros.
    d) 1657 euros.

18. (Problem 2) Consider the following null hypothesis: $H_0 : \beta_1 = 4\beta_2$. Then:

    a) The appropriate test statistic is: $\frac{0{,}9585 - 0{,}9584}{1 - 0{,}9585} \times 32$.
    b) The appropriate test statistic is: $\frac{0{,}9585 - 0{,}9523}{1 - 0{,}9585}$.
    c) The appropriate test statistic is: $\frac{199603 - 199132}{1 - 199603}$.
    d) We cannot answer with the information available.

19. (Problem 2) Consider the following null hypothesis: $H_0 : \beta_1 = 4\beta_2$. Then:

    a) We cannot reject the null hypothesis at usual significance levels.
    b) We reject the null hypothesis at the 1 % significance level.
    c) We reject the null hypothesis at the 5 % significance level.
    d) We cannot answer with the information available.

20. (Problem 2) Consider the following null hypothesis: $H_0 : \beta_1 + \beta_2 = 20$. Then:

    a) The appropriate test statistic is: $\frac{20{,}64 - 20}{66{,}11}$.
    b) The appropriate test statistic is: $\frac{20{,}64}{66{,}11}$.
    c) The appropriate test statistic is: $\frac{16{,}57 - 20}{63{,}23}$.
    d) We cannot answer with the information available.

21. (Problem 2) Consider the following null hypothesis: $H_0 : \beta_1 + \beta_2 = 20$. Then:

    a) We cannot reject the null hypothesis at usual significance levels.
    b) We reject the null hypothesis at the 1 % significance level.
    c) We reject the null hypothesis at the 5 % significance level.
    d) We cannot answer with the information available.

22. (Problem 2) Regarding Outputs 1 to 3, consider the following statements:
I. Outputs 1 and 2 correspond to alternative representations of the same model.
II. Outputs 1 and 3 correspond to alternative representations of the same model.
III. Output 3 corresponds to a restricted version of the model for Output 1.

   a) Only I. is true.
   b) Only III. is true.
   c) Only II. is true.
   d) None of the three statements is true.

23. (Problem 2) Regarding Outputs 1 to 3, consider the following statements:
I. Output 2 corresponds to a restricted version of the model for Output 1.
II. Outputs 1 and 3 correspond to alternative representations of the same model.
III. Output 3 corresponds to a restricted version of the model for Output 2.

   a) Only I. and II. are true.
   b) Only I. and III. are true.
   c) Only I. is true.
   d) None of the three statements is true.

24. (Problem 2) Suppose we want to test $H_0 : \beta_1 = \beta_2$ vs. $H_1 : \beta_1 \neq \beta_2$. Consider the following statements:
I. Such test could be implemented through a $t-$ statistic based on $\widehat{\beta}_1$ y $\widehat{\beta}_2$, from Output 1, provided that we also have their corresponding estimated variances and the estimated covariance between them.
II. Such test could be implemented through a $t-$ statistic for the hypothesis $H_0 : \gamma_2 = 0$ vs. $H_0 : \gamma_2 \neq 0$, if we implemented the OLS estimation of the model $P = \gamma_0 + \gamma_1 (A + C) + \gamma_2 C + \gamma_3 A^2 + \gamma_4 C^2 + \gamma_5 (A \times C) + v$.
III. Such test could be implemented through a $t-$ statistic for the hypothesis $H_0 : \delta_2 = 0$ vs. $H_0 : \delta_2 \neq 0$, if we implemented the OLS estimation of the model $P = \delta_0 + \delta_1 (A + C) + \delta_2 A + \delta_3 A^2 + \delta_4 C^2 + \delta_5 (A \times C) + u$.

   a) Only I. is true.
   b) Only III. is true.
   c) Only II. is true.
   d) The three statements are true.

25. (Problem 2) Given Output 1, for the mean values of the variables, we can state that the elasticity of the expected price of a clock with respect to its age is, approximately:

   a) 0,44.
   b) 13,17.
   c) We cannot answer with the information available.
   d) None of the other answers is true.

26. (Problem 2) Given Output 1, we can state that the elasticity of the expected price of a clock with respect to its age is:

   a) Constant, for any age and any number of bidders.

b) Approximately 1,44, for the mean values of the variables.
c) Approximately 13,17, for the mean values of the variables.
d) None of the other answers is true.

27. (Problem 2) Given Output 1, we can state that a consistent estimate of $V(P|\, A, C)$ is approximately:

a) 199132,1.
b) $(393{,}6)^2$.
c) We cannot answer with the information available.
d) None of the other answers is true.

28. (Problem 2) Given Output 1, we can state that a consistent estimate of $V(P)$ is approximately:

a) 6222,9.
b) $(393{,}6)^2$.
c) We cannot answer with the information available.
d) None of the other answers is true.

29. (Problem 2) Given Output 1, we can state that a consistent estimate of $V(\varepsilon|\, A, C)$ is approximately:

a) 199132,1.
b) $(393{,}6)^2$.
c) We cannot answer with the information available.
d) None of the other answers is true.

30. (Problem 3) Asume that model (S) holds the assumptions of the classical regression model. Consider the following statements:
I. The linear projection of $SPORT$ given INC, AGE, FEM, $SING$, $SOUTH$, $WGHT$ coincides with the conditional expectation of $SPORT$ given INC, AGE, FEM, $SING$, $SOUTH$, $WGHT$.
II. The conditional expectation of $SPORT$ given the variables INC, AGE, FEM, $SING$, $SOUTH$, $WGHT$ is a linear function of such variables.
III. The linear projection of $SPORT$ given INC, AGE, FEM, $SING$, $SOUTH$, $WGHT$ is the best predictor of $SPORT$.

a) The three statements are true.
b) None of the three statements is true.
c) Only I. and III. are true.
d) Only II. and III. are true.

31. (Problem 3) Let $E(\varepsilon|\, INC, AGE, FEM, SING; SOUTH, WGHT) = 0$ for any combination of the values of $INC, AGE, FEM, SING; SOUTH, WGHT$. Given Output 1, the causal effect of age on the expenditure on sport products is:

a) Constant.
b) Decreasing with age.
c) Increasing with age.

d) Under such conditions, it is not ensured that Output 1 provides such causal effect.

32. (Problem 3) Asume that model (S) holds the assumptions of the classical regression model. We want to elucidate whether the effect of age is linear. Consider the following statements:
I. The null hypothesis is $H_0 : \beta_2 - 2\beta_3 = 0$.
II. We can conclude that AGE^2 must be dropped from the model.
III. The appropriate test statistic is approximately $-2{,}6$, and it is approximately distributed as a standard normal.

   a) The three statements are true.
   b) None of the three statements is true.
   c) Only II. is true.
   d) Only III. is true.

33. (Problem 3) We want to elucidate whether the effect of age is linear. Consider the following statements:
I. The null hypothesis is $H_0 : \beta_2 - 2\beta_3 = 0$.
II. We can conclude that AGE^2 must be dropped from the model.
III. The appropriate test statistic is aproximately $-2{,}6$, and it is approximately distributed as a standard normal.

   a) The three statements are true.
   b) None of the three statements is true.
   c) Only II. is true.
   d) Only III. is true.

34. (Problem 3) Asume that model (S) holds the assumptions of the classical regression model. We want to elucidate whether the expenditure in sport products is independent of gender. Consider the following statements:
I. The null hypothesis is $H_0 : \beta_4 = \beta_8 = \beta_9$.
II. The appropriate test statistic is $W^0 = 935 \times \dfrac{(0{,}019495 - 0{,}009495)}{1 - 0{,}019495}$.
III. We can build an appropriate test statistic, which will be approximately distributed as a χ^2_3.

   a) The three statements are true.
   b) None of the three statements is true.
   c) Only II. and III. are true.
   d) Only I. and II. are true.

35. (Problem 3) Assume that model (S) holds the assumptions of the classical regression model. Comparing a single man living in the South and a married man living in the North, both with the same age, weight and income, consider the following statements:
I. We can statistically conclude that, on average, a single man living in the South spends more in sport articles than a married man living in the North.
II. If we want to test that there are no differences in the mean expenditure in sport articles between such individuals, the null hypothesis would be $H_0 : \beta_5 = -\beta_6$.
III. We can build a $t-$test to assess statistically whether there are differences in the average expenditure on sport articles between such individuals.

   a) The three statements are true.

b) Only II. and III. are true.

c) Only III. is true.

d) We cannot answer with the information available.

36. (Problem 3) Given the information available:
I. We reject, at the 1 % significance level, that INC is exogenous.
II. We can conclude that both $EDUC$ and $FEDUC$ are valid instruments.
III. The model estimated in Output 4 is exactly identified.

a) The three statements are true.

b) Only II. and III. are true.

c) None of the three statements is true.

d) Only I. and II. are true.

37. (Problem 3) Consider the following statements:
I. Given the information available, we cannot elucidate whether $EDUC$ and $FEDUC$ are valid instruments for income.
II. Provided that $EDUC$ and $FEDUC$ are valid instruments for income, inference about model (S) should be based on Output 4.
III. Provided that $EDUC$ and $FEDUC$ are valid instruments for income, we can conclude that INC is endogenous.

a) The three statements are true.

b) Only II. and III. are true.

c) None of the three statements is true.

d) Only I. and II. are true.

38. (Problem 3) Consider the following statements:
I. Inference about model (S) should be based on Output 1, since it yields a higher R^2 than Output 4.
II. The model estimated in Output 4 is exactly identified.
III. Output 1 yields consistent estimates of the parameters of model (S).

a) The three statements are true.

b) Only I. and III. are true.

c) None of the three statements is true.

d) Only II. is true.

39. (Problem 3) Assume that model (S) holds the assumptions of the classical regression model. Consider two single men living in the South, both with the same age, weight and income. Suppose that one moves to the North, while the other experiences an increase in annual income of 2000 euros. Our conjecture is that the mean change in expenditure on sport articles is similar for both individuals.
I. The null hypothesis is $H_0 : 2\beta_1 + \beta_6 = 0$.
II. We cannot reject such conjecture.
III. If we impose the constraint implied by such conjecture, the restricted model should drop the variables INC and $SOUTH$ and include the variable $(INC - 2 \times SOUTH)$ instead.

a) The three statements are true.

b) None of the three statements is true.

c) Only II. is true.

d) Only I. and III. are true.

40. (Problem 3) Using the appropriate estimates, the estimated causal effect on the expenditure on sport articles if income increases by 500 euros is:

a) None of the other answers is true.

b) Significant, and around 230 euros.

c) Significant, and around 334 euros.

d) We cannot ascertain the causal effect with the available information.

41. (Problem 3) In model (S), suppose that we drop $SING$, and add the variable $MARR$ (which takes on the value 1 if the individual is married and 0 if the individual is not married) instead. Given the estimation results, and using the appropriate estimates:
I. The estimated coefficient of $MARR$ would equal 0,2548.
II. The sum of squared residuals of the corresponding estimated model would equal 0,958334.
III. The estimated coefficients of the remaining variables would also change.

a) We cannot answer with the available information.

b) Only I. and II. are true.

c) Only I. and III. are true.

d) Only I. is true.

Answer Keys to exam type D

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