ECONOMETRICS EXAM

UNIVERSIDAD CARLOS III DE MADRID

2015-16

Answer all questions in 2 hours and a half. Critical values at the end of the exam.

1 From a random sample of sale of 258 houses in Spain in a given period it has been studied the relationship between housing prices in euros (Y) and its size in square meters (X_1), number of rooms (X_2), if the home is new construction or second hand (qualitative information collected by the dummy X_3 that equals 0 if it is second-hand and equal to 1 if new), if the house has a garage (qualitative information gathered by the dummy X_4 that takes the value 0 if garage and 1 if not), if the district to which it belongs is qualified as a tourist place (qualitative information collected by the dummy X_5, which takes the value 0 if tourism and the value 1 if it is not). The following model is considered

$$\log(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_1 X_2 + \beta_7 X_4 X_5 + U,$$

where $\beta' s$ are unknown parameters and U is an error term. Suppose all classical assumptions (conditions of Gauss-Markov theorem) hold. Here are the model estimates as well as two other restricted models (standard errors in parentheses).

Dependent Variable: $\log(Y)$			
Explicative V.	Model 1	Model 2	Model 3
Constant	0.005 (0.01)	0.004 (0.01)	0.004 (0.01)
X_1	0.014 (0.0002)	0.015 (0.0002)	0.015 (0.0002)
X_2	0.043 (0.001)	0.046 (0.001)	0.046 (0.001)
X_3	0.026 (0.002)	0.025 (0.002)	0.021 (0.002)
X_4	0.063 (0.003)	0.062 (0.002)	0.059 (0.002)
X_5	-0.032 (0.004)	-0.031 (0.004)	
$X_1 X_2$	0.022 (0.01)	0.019 (0.01)	0.019 (0.01)
$X_4 X_5$	-0.053 (0.03)		
R^2	0.7601	0.7571	0.7533

All confidence intervals are 95% and tests at the 5% significance level.

a. 0.5 To quantify the effect on the price of an increase in the size of 10 m^2 in a house of 100 m^2 in terms of the number of rooms. In what units it is measured this effect? Is this effect of equal magnitude for new housing and for second-hand?

b. 0.5 Test if the variation in the price caused by building an additional room is independent of house size.

c. 0.75 Does it has any effect on the price that a property is in a tourist area? Interpret the estimated value of this effect. Would you answer the same with a significance level of 1%?

d. 0.75 Suppose that the estimated covariance between the estimated coefficient of X_1 and $X_2 X_1$ is 0.00051, provide a confidence interval for the causal effect of an increase in size of $10m^2$ in a one bedroom house.

2. The following model is considered in a nationwide study on the relationship between average rental housing (RENT) (in euros) and the price (PRICE) (thousands of euros), controlling for the fact that housing is in an urban center,

$$RENT = \beta_0 + \beta_1 PRICE + \beta_2 URBAN + U,$$

where $URBAN$ takes the value 1 if the property is in an urban area and 0 otherwise, $\beta's$ are parameters and U is an error term. Based on a random sample of size $n = 50$ of the country's districts, the following models have been estimated,

	(1)	(2)	(3)	(4)
Dependent V.	RENT	PRICE	PRICE	RENT
Constant	125.9 (14.19)	−18.67 (12.00)	7.225 (8.936)	120.7 (15.71)
$URBAN$	0.525 (0.249)	0.182 (0.115)	0.616 (0.131)	0.0815 (0.305)
$PRICE$	1.5121 (0.228)			2.240 (0.339)
$INCOME$		2.731 (0.682)		
$REG2$		−5.095 (4.122)		
$REG3$		−1.778 (4.073)		
$REG4$		13.41 (4.048)		
R^2	0.669	0.691	0.317	0.599
SCR	20259.6	3767.6	8322.2	24565.7

a. 0.50 Why we might think that $PRICE$ is an endogenous variable? What would be the consequences for the OLS estimators of the model?

b. 0.50 Consider 5 instruments: median family income (income in thousands of euros) and four binary variables to describe the region of the country ($REG1, REG2, REG3$ and $REG4$): 5 instruments are considered. We know that these instruments are independent of the error term. Test the relevance of all instruments and establish whether or not the instruments are weak.

c. 0.75 Which of the price predictors, based on (2) or (3), is used in the second stage of the two stages least squares (2SLS)? Column (4) provides the 2SLS estimate. Standard errors, R^2 and SCR are provided by the package GRETL. That is, the β coefficients are estimated by OLS using as dependent variable $RENT$ and as explanatory variables $URBAN$ and $PRICE$ predicted in the regression of the first stage. Is it possible to make a joint significance of the variable $PRICE$ and $URBAN$ using the available information? Justify your answers.

d. 0.75 Explain how you would test that all instruments used are exogenous. Suggest a test statistic and set the criteria for rejecting the null hypothesis of exogeneity of all instruments.

3. Consider the following Working-Leser Engel curve for food expenditure,

$$Y = \beta_0 + \beta_1 D + \beta_2 \ln X + \beta_3 M + \beta_4 D \ln X + \beta_5 DM + U, \quad (1)$$

where Y is the percentage of food expenditure with respect to total expenditure of a household in thousands of euros a year, X is total expenditure in the year, M is the number of household members, D is a variable that takes the value 1 if the family lives in a town of over 10,000 inhabitants and 0 otherwise, $\beta's$ are parameters and U is an error term. On the other hand, it is considered the next alternative model

$$\ln Y = \gamma_0 + \gamma_1 D + \gamma_2 \ln X + \gamma_3 M + \gamma_4 D \ln X + \gamma_5 DM + V,$$

where $\gamma's$ are parameters and V an error term.

a. 0.50 In which of the two models the elasticity of Y versus X for a family living in a city of 20,000 inhabitants is constant, that is, the elasticity does not depend on any variable in the model? Justify your answer.

b. 0.50 For which values of the parameters γ and β the elasticity of Y versus X is identical in the two models for a family living in a village of 100 inhabitants which spent on food 10% of its budget?

c. 0.5 Can you compare the fit of the two models using the coefficient of determination R^2? Justify your answer.

d. 1.5 Tastes are very different for families living in small and big cities. This implies that the variance of Y is different depending on the value taken by D, with X and M constant, which can be expressed using mathematical notation as $Var(Y|\, M, X, D) = \sigma_1^2 D + \sigma_2^2(1-D)$. Explain the consequences of this heterogeneity on the inferences made by OLS using the usual output from GRETL for model (1). Are inferences on β_2 and β_3 based on standard GRETL output valid using only the subpopulation of families living in towns of less than 10,000 inhabitants? Explain how would you test $H_0 : \beta_2 = \beta_3$ vs $H_1 : \beta_2 \neq \beta_3$ using this subpopulation.

4. We want to estimate the causal relationship between performing a work in public administration and the fact of being male or female and years of education. To do this we consider a random sample of 700 employees from a district, employees were asked if they were employed by the government ($GOV =$ 1 if working in the administration and zero otherwise), gender ($MALE$= 1 if man and 0 if female) and years of education ($EDUC$). With these data we estimate the following linear probability model (standard errors in parentheses)

$$\widehat{GOV} = \underset{(0.027)}{0.152} + \underset{(0.003)}{0.035}\, EDUC - \underset{(0.025)}{0.050}\, MALE$$

a. 0.5 Provide an expression for the mean and conditional variance of GOV given $EDUC$ and $MALE$ in terms of the true unknown parameters under the linear probability model specification.

b. 0.5 Interpret the result of the fit for a man with 16 years of education.

c. 0.75 How much more likely is the fact of finding a woman working for the administration with respect to finding a man with the same educational level?

d. 0.75 Explain the limitations of linear probability model and how these difficulties are overcome by the probit and logit models.

Critical values:

$\Pr(F_{1,\infty} > 3.84) = 0.050$	$\Pr(\chi_1 > 3.84) = 0.050$	$\Pr(N(0,1) > 1.645) = 0.050$
$\Pr(F_{1,\infty} > 5.02) = 0.025$	$\Pr(\chi_1 > 5.02) = 0.025$	$\Pr(N(0,1) > 1.960) = 0.025$
$\Pr(F_{1,\infty} > 6.63) = 0.010$	$\Pr(\chi_1 > 6.63) = 0.010$	$\Pr(N(0,1) > 2.326) = 0.010$
$\Pr(F_{1,\infty} > 7.88) = 0.005$	$\Pr(\chi_1 > 7.88) = 0.005$	$\Pr(N(0,1) > 2.576) = 0.005$
$\Pr(F_{2,\infty} > 3.00) = 0.050$	$\Pr(\chi_2 > 5.99) = 0.050$	$\Pr(t_2 > 6.31) = 0.050$
$\Pr(F_{2,\infty} > 3.69) = 0.025$	$\Pr(\chi_2 > 7.38) = 0.025$	$\Pr(t_2 > 2.92) = 0.025$
$\Pr(F_{2,\infty} > 4.61) = 0.010$	$\Pr(\chi_2 > 9.21) = 0.010$	$\Pr(t_2 > 6.96) = 0.010$
$\Pr(F_{2,\infty} > 5.30) = 0.005$	$\Pr(\chi_2 > 10.6) = 0.005$	$\Pr(t_2 > 9.92) = 0.005$
$\Pr(F_{4,\infty} > 2.37) = 0.050$	$\Pr(\chi_4 > 9.49) = 0.050$	$\Pr(t_4 > 2.13) = 0.050$
$\Pr(F_{4,\infty} > 2.79) = 0.025$	$\Pr(\chi_4 > 11.1) = 0.025$	$\Pr(t_4 > 2.78) = 0.025$
$\Pr(F_{4,\infty} > 3.32) = 0.010$	$\Pr(\chi_4 > 13.3) = 0.010$	$\Pr(t_4 > 3.74) = 0.010$
$\Pr(F_{4,\infty} > 3.71) = 0.005$	$\Pr(\chi_4 > 14.8) = 0.005$	$\Pr(t_4 > 4.60) = 0.005$
$\Pr(F_{7,\infty} > 2.01) = 0.050$	$\Pr(\chi_7 > 14.1) = 0.050$	$\Pr(t_7 > 1.89) = 0.050$
$\Pr(F_{7,\infty} > 2.29) = 0.025$	$\Pr(\chi_7 > 16.0) = 0.025$	$\Pr(t_7 > 2.36) = 0.025$
$\Pr(F_{7,\infty} > 2.51) = 0.010$	$\Pr(\chi_7 > 17.6) = 0.010$	$\Pr(t_7 > 3.00) = 0.010$
$\Pr(F_{7,\infty} > 2.90) = 0.005$	$\Pr(\chi_7 > 20.3) = 0.005$	$\Pr(t_7 > 3.50) = 0.005$