EXTRAORDINARY EXAM OF ECONOMETRICS

You have two hours. Answer each question on a different booklet.

1. Consider the model that relates variables Y, X_1, X_2 and Z as follows

$$\log Y = \begin{cases} \beta_0 + \beta_1 X_1 + \beta_2 \log X_2 + U \text{ if } Z \geq 0 \\ \gamma_0 + \gamma_1 X_1 + \gamma_2 \log X_2 + U \text{ if } Z < 0, \end{cases}$$

with U an error term whose expectation conditional on X_1, X_2 and Z is zero. Z is unobserved. Instead of Z, we observe the variable W

$$W = \begin{cases} 0 \text{ if } Z \geq 0 \\ 1 \text{ if } Z < 0. \end{cases}$$

   (a) Express the previous model in a single equation as a function of the observed variables and relate the parameters of the new model to $(\beta_0, \beta_1, \beta_2)$ and $(\gamma_0, \gamma_1, \gamma_2)$.

   (b) Provide expressions, as a function of the observed variables and of the coefficients, when X_1 and X_2 are respectively equal to values x_1^0 and x_2^0, to approximate the expected elasticity of Y with respect to X_1 and the expected elasticity of Y with respect to X_2.

   (c) Provide an expression that approximates the expected percentage variation of Y when Z switches from negative to positive.

2. Suppose that we want to study whether there exists racial discrimination regarding the extension of mortgages in the US. To this end we rely on a random sample of mortgage applications in the country, from which we extract information on the binary variable HIP, equal to 1 if an application is accepted and to zero otherwise, variable $BLACK$, equal to 1 if the applicant is a person of color and 0 otherwise and variable H/I, the ratio of the value of the requested mortgage over the annual income of the applicant. Consider the regression model

$$HIP = \beta_0 + \beta_1 \cdot BLACK + \beta_2 \cdot H/I + \beta_3 \cdot (BLACK \cdot H/I) + U,$$

where U has a zero mean conditional on all the explanatory variables.

   (a) What is the effect of being black on the acceptance of mortgage applications in this model?

   (b) Suppose that we want to test the global significance of the explanatory variables $BLACK$ and H/I. Could you do this test using the R^2 of the model estimated by ordinary least squares? Justify your answer.

   (c) Write down the expression for the conditional probability of obtaining a mortgage, for a white individual. Keeping in mind the expected sign of β_2, would the predicted values obtained from this model make sense for every possible value of the variable H/I? Justify your answer.

3. In order to study the effect of fertility on female labour supply, suppose we have access to a sample (coming from the 1980 US census) of married women aged between 21 and 35 with 2 or more kids. For this purpose, the following model is estimated using OLS:

$$weeks = \beta_0 + \beta_1 morekids + \beta_2 age + \beta_3 white + U$$

where *weeks* indicate the number of weeks of work each mother has supplied in 1979 and *morekids* is a dummy variable taking value 1 if the mother has more than 2 kids, and 0 otherwise. The variables *age* (mother's age in years) and *white* (a dummy variable equal to 1 if the woman is white) can be considered exogenous.

   (a) We suspect that the regression above might not deliver an unbiased estimator of the causal effect of fertility (*morekids*) on labour supply (*weeks*). For such reason, we consider the variable *samesex*, a dummy taking value 1 if the two first kids are of the same sex (male-male or female-female) and 0 otherwise, as an instrumental variable. Carefully explain the conditions under which *samesex* can be considered a valid instrument for *morekids*.

(b) Using the information provided in Table 1, investigate the relevance as well as the exogeneity of the variable *samesex*.

(c) The interaction between *samesex* and the exogenous variable *white* is proposed as an additional instrumental variable: $sswhite = samesex * white$. Using the information provided in Table 1, investigate the relevance and the exogeneity of the two instruments $(samesex, sswhite)$.

(d) Would it be correct to use the standard errors reported in the column (4) of Table 1 to test the hypothesis that parameter β_1 is not negative? Justify your answer.

Tabla 1 note: All regressions presented are estimated using OLS.
The variable *morekidshat* represents the predicted values obtained from the regression in column (3).
The variable *uhat* is obtained as:

$$uhat = weeks - \hat{\beta}_0 - \hat{\beta}_1 morekids - \hat{\beta}_2 age - \hat{\beta}_3 white$$

where the estimated coefficients correspond to those obtained in column (4), in the same order.

Table 1: Regression table

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable:	morekids	morekids	morekids	weeks	uhat	uhat
Constant	0.493	0.426	0.421	12.97	0.000000116	1.145
	(0.200)	(0.202)	(0.210)	(10.93)	(8.441)	(8.889)
morekidshat				-7.411		
				(14.39)		
samesex		0.136	0.146			-2.134
		(0.0424)	(0.127)			(5.515)
sswhite			-0.0106			2.455
			(0.135)			(5.897)
age	0.00109	0.000924	0.000928	0.453	-2.18e-10	-0.000862
	(0.00652)	(0.00650)	(0.00651)	(0.281)	(0.278)	(0.279)
white	-0.198	-0.190	-0.185	-4.152	-0.000000122	-1.268
	(0.0686)	(0.0682)	(0.0979)	(4.153)	(2.973)	(4.319)
Observations	500	500	500	500	500	500
R^2	0.01820	0.03850	0.03851	0.00610	0.00001	0.00030

Robust standard errors in parentheses

RELEVANT CRITICAL VALUES: $Z_{0.90} = 1.282$, $Z_{0.95} = 1.645$, $Z_{0.975} = 1.96$, $\chi^2_{2,95} = 5.99$, $\chi^2_{2,975} = 7.378$, $\chi^2_{3,95} = 7.81$, $\chi^2_{3,975} = 9.3484$, $\chi^2_{4,95} = 9.49$, $\chi^2_{4,975} = 11.1433$, $\chi^2_{4,95} = 9.49$, $\chi^2_{4,975} = 11.1433$, $\chi^2_{5,95} = 11.07$, $\chi^2_{5,975} = 12.8325$, where $\mathbb{P}(Z \leq Z_\alpha) = \alpha$ and $\mathbb{P}\left(\chi^2_m \leq \chi^2_{m,\alpha}\right) = \alpha$, Z is distributed as a normal with mean 0 and variance equal 1, and χ^2_m as a chi-squared with m degrees of freedom.