FINAL EXAM ECONOMETRICS

Answer each of the three parts on a different booklet. The duration of the exam is two hours and a half. The value of each individual question is one point out of ten.

1. Let (Y, X, W) be three random variables related to each other through the following linear regression model

$$Y = \beta_0 + \beta_1 X + \beta_2 W + u, \tag{1}$$

where we know that

$$\mathbb{E}\left(u \mid X, W\right) = \mathbb{E}\left(u \mid W\right) = \gamma_0 + \gamma_2 W. \tag{2}$$

   (a) Write equation (2) as a linear regression model explaining the variable u as a function of W and where you should show that the new error term v satisfies $\mathbb{E}\left(v \mid X, W\right) = 0$. Substitute this model for u in (1) and interpret the coefficients of the obtained model which relates Y to X, W and v.

   (b) We observe a simple random sample $\{Y_i, X_i, W_i\}_{i=1}^n$ of (Y, X, W) and consider the Ordinary Least Squares (OLS) estimator of the coefficients in model (1) where Y is the explained variable and (X, W) the explanatory variables. Justify whether this estimator is biased or unbiased for the parameters $(\beta_0, \beta_1, \beta_2)$ when (2) holds. What would happen if W was omitted from the regression?

   (c) A follow-up study confirms that condition (2) does not hold and that the following condition holds instead:

$$\mathbb{E}\left(u \mid X, W\right) = \alpha_0 + \alpha_1 X + \alpha_2 W, \quad \alpha_1 \neq 0. \tag{3}$$

   We have access to data on another variable Z which satisfies

$$\mathbb{E}\left(u \mid Z, W\right) = \mathbb{E}\left(u \mid W\right) = \delta_0 + \delta_2 W, \quad \delta_2 \neq 0. \tag{4}$$

   What would be your strategy to estimate model (1) under restrictions (3)-(4)? Which coefficients of this model could be consistently estimated? Justify your answer.

2. Using cross-sectional data for different states in the United States, we want to study the relationship between cannabis consumption and individual income. To this end we consider the following system of equations

$$\log(income) = \beta_0 + \beta_1 cannabis + u_1 \tag{5}$$

$$cannabis = \gamma_0 + \gamma_1 \log(income) + \gamma_2 fine + \gamma_3 prison + u_2, \tag{6}$$

where *cannabis* is monthly cannabis consumption, $fine$ is the standard fine imposed for cannabis possession in the state of residence and *prison* is a binary variable equal to 1 if in the state of residence one can be sent to prison for *cannabis* possession for personal consumption, 0 otherwise. We assume that $fine$ and *prison* vary across states.

   (a) Assuming that the variables $fine$ and *prison* are exogenous (are not correlated with any of the error terms), write down the reduced form regression for *cannabis* and show that *cannabis* is endogenous in equation (5).

   (b) What restrictions must hold on coefficients $\gamma's$ in equation (6) in order to be able to get an unbiased estimator of coefficients $\beta's$ in equation (5)? Explain if it would be possible to test these restrictions, directly or indirectly.

(c) Explain how you would estimate the $\beta' s$ coefficients in (5). How would you obtain standard errors robust to heteroskedasticity?

(d) Explain the steps you would follow to test the exogeneity (no correlation with the error term) of the instruments used to estimate the $\beta' s$ coefficients in equation (5). In particular, describe how you would obtain the relevant statistic needed to implement the test, as well as its critical values or p-values.

3. Suppose we want to investigate the demand function of subscriptions to scientific journals by university libraries. For this purpose, we have available data about 180 journals (within economics) for the year 2000. Consider the following linear regression model:

$$\log(subs) = \beta_0 + \beta_1 \log(ppcite) + \beta_2 (\log(ppcite))^2 + \beta_3 \log(age) + \beta_4 \log(ppcite) * \log(age) + u, \quad (7)$$

where *subs* is the number of libraries that have subscribed to a given journal, *ppcite* is the price per citation, i.e. the journal subscription price divided by the total number of article citations in the journal, and *age* is journal's seniority in years. $\log(x)$ is the logarithm x. The error term u satisfies the common OLS assumptions, so that all estimates are unbiased.

(a) Provide an expression, using the coefficients of model (7), that describes the expected effect of a 1% increase in the price per citation on the demand of subscriptions, for a journal with 20 years of seniority and a $ppcite = 10$. Explain how you would test the hypothesis that such effect is equal or larger than -1 , and under what conditions the test is to be considered a valid one.

(b) Using the estimates of model (7) in Table 1 below, test whether the demand elasticity of subscriptions to price depends also on the seniority of the journal. Under what conditions is the proposed test a valid one?

(c) Alternatively, we decide to estimate the following model, where $age40$ is a binary variable that takes value 1 if the journal's seniority is equal or below 40 years, and 0 otherwise:

$$\begin{aligned} \log(subs) &= \beta_0 + \beta_1 \log(ppcite) + \beta_2 (\log(ppcite))^2 + \beta_3 age40 \\ &\quad + \beta_4 \log(ppcite) * age40 + \beta_5 (\log(ppcite))^2 * age40 + u. \end{aligned} \quad (8)$$

Test the hipothesis that the demand elasticity of subscriptions to prices is constant in prices. Use the estimates of model (8) reported in Table 1 below. Under what conditions is the proposed test a valid one?

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.

Table 1: Regression table

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable:	log(*subs*)	log(*subs*)	log(*subs*)	log(*subs*)	log(*subs*)	log(*subs*)
log(*ppcite*)	-0.561	-0.460	-1.468	-0.382	-0.455	-0.351
	(0.0449)	(0.0446)	(0.385)	(0.0369)	(0.0474)	(0.0597)
$\log(ppcite)^2$	-0.0291		0.00721		-0.0142	0.0164
	(0.0177)		(0.0219)		(0.0170)	(0.0197)
log(*age*)		0.531	0.556			
		(0.222)	(0.247)			
$\log(ppcite) * \log(age)$			0.257			
			(0.106)			
*age*40				-0.534	-0.624	-0.491
				(0.117)	(0.126)	(0.146)
$\log(ppcite) * age40$				-0.205		-0.228
				(0.0805)		(0.0854)
$\log(ppcite)^2 * age40$						-0.0272
						(0.0415)
Constant	4.839	2.724	2.699	5.054	5.050	5.028
	(0.0680)	(0.841)	(0.934)	(0.0766)	(0.0827)	(0.0822)
Observations	180	180	180	180	180	180
R^2	0.564	0.584	0.607	0.632	0.620	0.633

Standard errors in parentheses