UNIVERSITY OF LONDON
External System

Model exam 3.

SECTION A

Answer **ALL** questions from this section (questions **1-5**). Each question in this section bears 8 marks

Question 1. In the model

$$y = \beta x_t + u_t\,,\; t = 1, 2, ..., T$$

x_t is measured with error. Data is only available on x_t^*, where

$$x_t^* = x_t + v_t\,;\; t = 1, 2, ..., T$$

and $Eu_t = Ev_t = 0$, $E(u_t v_t) = E(x_t u_t) = E(x_t v_t) = 0$. y_t, x_t and x_t^* have zero means.

If $\hat{\beta}$ is the ordinary least squares estimator of β from regressing y_t on x_t^*, show that $\hat{\beta}$ is inconsistent.

Question 2 Briefly explain what is instrumental variable estimation. Consider the model

$$Y_t = \beta X_t + u_t\,;\; t = 1, 2, ..., T\,.$$

What is instrumental variable estimator for this model? Is it consistent? Give an example of a model where instrumental variable estimation is an improvement on ordinary least squares (OLS). Explain why IV estimation is superior to OLS in this case.

Give an example of a model where instrumental variable estimation is an improvement on ordinary least squares (OLS).

Explain why IV estimation is superior to OLS in this case.

Question 3 Write short explanations on the Durbin-Wu-Hausman test for measurement error.

Question 4 Explain the concept of the likelihood function and the maximum likelihood estimator, and briefly describe its properties.
Derive the expression for the loglikelihood function for the logit model.
(It is not expected here the students to solve maximization problem)

Question 5 Let the probability density function of a random variable X be $f(x, \theta)$.
Explain the procedure to use the Wald test for testing the null $H_0 : \theta = \theta_0$.

SECTION B

Question 6. Consider the simple linear regression model:

$$Y_i = \alpha + \beta X_i + u_i,\ i = 1, ..., n$$

in the presence of correlation between the errors and regressors. The regressors exhibit variability in the sample, i.e. $\sum_{i=1}^{n}(X_i - \overline{X})^2 \neq 0$. Under assumptions of homoskedasticity and the absence of autocorrelation, the IV estimator for β that uses the instrument Z has the following (asymptotic) variance (no need to prove this statement):

$$\text{Var}(\hat{\beta}_{IV}) = \frac{\sigma_u^2}{\sum_{i=1}^{n}(X_i - \overline{X})^2} \times \frac{1}{r_{XZ}^2},$$

where $r_{XZ} \neq 0$ is the sample correlation between X and Z and σ_u^2 is the variance of the disturbance term.

(a) Give the formula for $\hat{\beta}_{IV}$ (you are not asked to derive it).

(b) Provide at least three factors that will help obtain more <u>precise</u> IV parameter estimates for β. In your answer explain why the precision of parameter estimates is important.

(c) Discuss the following statement: 'If X is not correlated with u, the best choice of instrument is using the regressor itself.'

Question 7. We are interested in explaining the willingness of households to buy ecologically produced apples. We use data where each family was presented with a description of ecologically friendly apples, along with prices (in $) of regular apples (*regprc*) and prices of the hypothetical ecolabeled apple (*ecoprc*). The variable we want to explain is the dummy variable, *ecobuy* which equals 1 if the household wants to buy ecologically friendly apples and 0 otherwise. Additional household variables we have are family income in $1000s, *faminc*, household size, *hhsize*, years of schooling, *educ*, and *age*.
Using a sample of 660 households, the following results were obtained:

	OLS A	Probit A	Probit B
constant	0.890	1.088	−0.244
	[.068]	(.206)	(.474)
Regprc	0.735	2.029	2.030
	[.132]	(.378)	(.381)
Ecoprc	−0.845	−2.344	−2.267
	[.106]	(.318)	(.321)
Faminc			0.0014
			(.0015)
Hhsize			0.069
			(.037)
Educ			0.071
			(.024)
Age			−0.001
			(.004)
R^2	.086		
log *L*		−407.60	−399.04

The heteroskedasticity robust standard errors are reported in squared brackets and the (asymptotic) standard errors are reported in parentheses.

(a) Carefully interpret the estimated coefficient on the price of ecologically friendly apples reported in 'OLS A' and discuss whether the effect is statistically significant. In your answer explain why it is important to use robust standard errors.

(b) It is argued that using the Probit model is better than using the linear probability model when explaining the binary variable *ecobuy*. Discuss the benefits/drawbacks of using the Probit model when trying to explain a binary variable. In your answer explain what the linear probability model refers to.

(c) The Probit model B postulates that:

$$\Pr(ecobuy = 1 | x) =$$
$$= \Phi(\beta_0 + \beta_1 regprc + \beta_2 ecoprc + \beta_3 faminc + \beta_4 hhsize + \beta_5 educ + \beta_6 age),$$

where $\Phi(z)$ is the standard normal cumulative distribution function.

Use the likelihood ratio test, to test the joint significance of the nonprice variables. Clearly indicate the null and alternative hypothesis, the test statistic and the rejection rule.

(d) In this question we are interested in the marginal effect of the price of ecologically friendly apples using Probit model A holding constant the price of regular apples.

(i) Indicate how you can obtain the marginal effect of *ecoprc* using the Probit model.

(ii) Unlike in the LPM this marginal effect will not be constant. Discuss what computations you would carry out to obtain the marginal effect of a \$0.10 reduction in *ecoprc* when evaluated at the mean of our explanatory variables (*regprc* mean equals \$0.884 and *ecoprc* mean equals \$1.082). You are not expected to use your calculator. Clarity of the computations required is enough.

Question 8. It is postulated that a reasonable demand-supply model for the wine industry in Australia would be given by:

$$Q_t = \alpha_0 + \alpha_1 P_t^w + \alpha_2 P_t^b + \alpha_3 Y_t + \alpha_4 A_t + u_t \quad \text{demand}$$
$$Q_t = \beta_0 + \beta_1 P_t^w + \beta_3 S_t + v_t \quad \text{supply,}$$

where Q_t = real per capita consumption of wine, P_t^w = price of wine relative to CPI, P_t^b = price of beer relative to CPI, Y_i = real per capita disposable income, A_t = real per capital advertising expenditure, and S_t = storage cost. CPI is the Consumer Price Index.

The endogenous variables in this model are *Q* and P^w, and the exogenous variables are P^b, *Y*, *A* and *S*.

The variance of u_t and v_t are, respectively, σ_u^2 and σ_v^2, and $\text{Cov}(u_t, v_t) = \sigma_{uv} \neq 0$. The errors do not exhibit any correlation over time.

(a) Provide the reduced form for P_t^w.

(b) The OLS estimation of the demand function, based on annual data from 1955-1975 ($T = 20$), gave the following results (all variables are in logs and figures in parentheses are *t*-ratios).

$$\hat{Q}_t = \underset{(-6.04)}{-23.651} + \underset{(4.0)}{1.158} P_t^w - \underset{(-0.45)}{0.275} P_t^b + \underset{(4.5)}{3.212} Y_t - \underset{(-1.3)}{0.603} A_t .$$

All the coefficients except that of *Y* have the wrong signs. The coefficient of P^w (price elasticity of demand, α_1) not only has the wrong sign but also appears significant.

Explain why the OLS parameter estimator may give rise to these counter-intuitive results. You are expected to use your results in part (a) to support your answer.

(c) The supply equation is overidentified. Clearly explain this terminology. What distinguishes overidentification from exact identification and underidentification? Provide one set of assumptions that would render the supply equation exactly identified.

(d) Discuss how you should estimate the supply equation in light of the overidentification.