This paper is not to be removed from the Examination Halls

UNIVERSITY OF LONDON **279 0020 ZB**

BSc degrees and Diplomas for Graduates in Economics, Management, Finance and the Social Sciences, the Diploma in Economics and Access Route for Students in the External Programme

Elements of Econometrics

Tuesday, 23 May 2006 : 2.30pm to 5.30pm

Candidates should answer **FOUR** of the following **EIGHT** questions: **QUESTION 1** of Section A (40 marks) and **THREE** questions from Section B (20 marks each).
Candidates are strongly advised to divide their time accordingly.

Graph paper is provided. If used, it must be fastened securely inside the answer book.

New Cambridge Statistical Tables (second edition) and Durbin Watson d-Statistical Tables are provided.

A hand held calculator may be used when answering questions on this paper but it must not be pre-programmed or able to display graphics, text or algebraic equations. The make and type of machine must be stated clearly on the front cover of the answer book.

PLEASE TURN OVER

UL06/28

SECTION A

Answer all **eight** parts of question 1 (5 marks each)

1. (a) Explain what do you understand by the Durbin-Watson (DW) test. State the assumptions required for using the DW test.

(b) A simple random sample of size n , $X_1, X_2, \cdots, X_n$ is drawn from a population with mean μ and variance σ^2. Consider the following estimators of μ

$$\hat{\mu}_1 = X_1 + \frac{\sum_{i=2}^{n} X_i}{n} \qquad \text{and} \qquad \hat{\mu}_2 = \frac{\sum_{i=1}^{n-1} X_i}{n}$$

Explain which estimator is consistent.

(c) Explain why a disturbance term in a regression equation may be heteroskedastic. Explain how the properties of ordinary least squares estimators are affected if the disturbance term is heteroskedastic.

(d) Let X and Y be two random variables. Derive, using expectation

Var(X+Y) and Var(X-Y)

where Var is variance.

(e) If a random variable X has a distribution with density function $\frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$ show that the maximum likelihood estimator of the mean (μ) of the random variable X is the sample mean.

(f) In the linear regression model

$$Y_t = \alpha + \beta X_t + u_t \quad ; \quad t = 1,2,\cdots,T$$

prove that $Cov(X_t, \hat{u}_t) = 0$ where $\hat{u}_t = Y_t - \hat{\alpha} - \hat{\beta} X_t$ and Cov is the sample covariance. $\hat{\alpha}$ and $\hat{\beta}$ are ordinary least squares estimators of α and β.

(question continues on next page)

PLEASE TURN OVER

(g) Consider the model:

$$Y_t = \alpha + \beta X_t + u_t \quad ; \qquad t = 1, \cdots, 6$$

where $E(u_t) = 0$; $E(u_t^2) = \sigma^2$ and $E(u_s u_t) = 0$ if $s \neq t$.

The observations on X_t's are

X_1	X_2	X_3	X_4	X_5	X_6
1	2	3	4	5	6

Consider an estimator of β as $\tilde{\beta} = \frac{1}{8}\left[Y_6 + Y_5 - Y_2 - Y_1\right]$.
Derive the sampling variance of $\tilde{\beta}$.

(h) R^2, the coefficient of determination is defined as

$$R^2 = \frac{\text{Explained Sum of Squares}}{\text{Total Sum of Squares}}$$

Show that R^2 lies between 0 and 1 when there is an intercept term in the model.

SECTION B

Answer **three** questions from this section (20 marks each)

2. Let the regression equation be

$$y_t = \beta x_t + u_t \quad ; \quad t = 1, 2, \ldots, T$$

where $E(u_t) = 0$; $E(u_t^2) = \sigma^2$ and $E(u_s u_t) = 0$ if $s \neq t$.

(a) Obtain the ordinary least squares (OLS) estimator of β.

(b) Explain in detail how would you obtain an unbiased estimator of σ^2.

(c) Is the OLS estimator of β consistent? Explain in detail.

3. Consider a two equation linear model

$$Q_t = \beta_0 + \beta_1 P_t + \beta_2 Y_t + u_t \qquad \text{(i)} \qquad \text{demand equation}$$
$$Q_t = \alpha_0 + \alpha_1 P_t + e_t \qquad \text{(ii)} \qquad \text{supply equation}$$

$t = 1,2,\cdots,T$; $E(u_t) = E(e_t) = 0$; $E(u_t^2) = \sigma_u^2$; $E(e_t^2) = \sigma_e^2$; $E(u_t e_t) = \sigma_{ue}$;
$E(u_s e_t) = 0$ if $s \neq t$, for all $s,t = 1,2,\cdots,T$ and variables are defined as:
Q_t = demand for good
P_t = price for good
Y_t = personal disposable income
quantity demanded is equal to the quantity supplied
u_t and e_t are disturbance terms.

(a) Examine the identifiability of the above two equations.

(b) Derive the two-stage least squares estimator of α_1 and also examine its consistency.

(c) Derive the indirect least squares estimator of α_1. Comment on the result obtained.

4. In the model $y_t = \beta x_t + u_t$; $t = 1,2,\ldots\ldots\ldots,T$

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

$$x_t^* = x_t + v_t \quad ; \qquad t = 1,2,\ldots\ldots,T$$

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.

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

(b) In the above given model, suppose x_t was measured without error, y_t was measured with error and data was only available on y_t^* where $y_t^* = y_t + w_t$ and $E(w_t) = 0$; $E(u_t w_t) = E(x_t w_t) = 0$ and $E(v_t w_t) = 0$ Let $\hat{\beta}$ be the OLS estimator of β from regressing y_t^* on x_t. Is $\hat{\beta}$ consistent? Explain in detail.

(c) Suppose in the above given model, both y_t and x_t are measured with errors and data is available only on y_t^* and x_t^* where x_t^* and y_t^* are defined above, respectively. Discuss (without derivation) whether the OLS estimator of β, from regressing y_t^* on x_t^* will be consistent or inconsistent.

PLEASE TURN OVER

5. (a) Consider a model:

$$y_i = \beta_1 + \beta_2 X_i + u_i \quad ; \quad i = 1,2,\ldots,n$$

where $y_i = 1$ if the event takes place
$= 0$ otherwise.

i. Explain fully the problem which arises if the above model is estimated by ordinary least squares.

ii. How would you estimate the model by weighted least squares where weights are standard errors of the disturbance term. Discuss advantages and disadvantages of this procedure.

(b) A researcher wants to examine the newspaper reading habits of households. For this she collects data on fifty households and defines

$y_i = 1$ if the i-th household purchases a newspaper
$= 0$ otherwise.

She estimates the model defining $y_i = f(S_i, E_i, u_i)$ where

S_i = years spent by the head of the i-th household in full time education
E_i = average earnings of the head of the i-th household
u_i = unobserved disturbance term.

The model was estimated by ordinary least squares (OLS) and logit with the following results:

	OLS	logit
S	0.099 (4.07)	0.521 (3.10)
E	0.012 (2.29)	0.067 (1.84)
Constant	0.015 (0.16)	-2.56 (-3.57)

the figures in parentheses are the t values.

i. Interpret the results of these estimated equations.

ii. Obtain the predicted probability for the i-th household if $S_i = 7$ and $E_i = 40$ from both sets of estimates.

6. Given cross-section data for 41 countries the following regression results were obtained:

$$\ln(\text{revenue}_i) = 4.05 + 0.896\ln(\text{trade}_i) - 0.726\ln(\text{gnp}_i) + \hat{u}_i$$
$$(0.868) \quad (0.200) \qquad (0.105)$$

where estimated standard errors are given in (.), $\hat{u}_i$ is the least squares residual for country i and ln is the natural logarithm. Variables are defined as follows:

revenue: percentage government revenue from tariffs,
trade: total foreign trade (exports and imports) as a percentage of GNP,
gnp: GNP per capita.

(a) Interpret the above results.

(b) Construct a 90% confidence interval on the elasticity of the share of tariffs in total government revenue with respect to share of foreign trade in GNP. What does your estimated confidence interval tell you statistically and economically?

(c) As the data are cross-section, a concern is about the possible presence of heteroskedasticity. Consider the following two regressions

$$\ln(\hat{u}_i) = -4.958 + 0.857\ln(\text{trade}_i) + \hat{\varepsilon}_{1i}$$
$$(2.17) \quad (0.623)$$

$$\ln(\hat{u}_i) = -3.601 + 0.274\ln(\text{gnp}_i) + \hat{\varepsilon}_{2i}$$
$$(1.96) \quad (0.131)$$

where estimated standard errors are given in (.) and $\hat{\varepsilon}_{1i}$ and $\hat{\varepsilon}_{2i}$ are the respective least squares residuals for country i. Explain why these regressions might be an appropriate way to investigate the possible presence of heteroskedasticity. Do they indicate any potential problems with heteroskedasticity?

(d) Explain, if heteroskedasticity is found, how it can be remedied. What effect does finding heteroskedasticity have on your answer in parts (a) and (b).

7. (a) Explain why distributed lag models are of importance in specifying economic relationships.

(b) Explain what is meant by a 'partial adjustment model'. Describe carefully how would you estimate the parameters of this model.

PLEASE TURN OVER

8. The following 3 equations were estimated using 3,866 observations from the 1985 Family Expenditure Survey. The dependent variable is the log of male gross earnings.

	(i)	**(ii)**	**(iii)**
constant	5.20	3.66	2.57
	(0.34)	(0.13)	(0.21)
age	-0.00	0.075	0.14
	(0.007)	(0.006)	(0.01)
age^2	-	-0.0008	-0.001
		(0.0001)	(0.0001)
S	-	-	-0.05
			(0.005)
S^2	-	-	0.0004
			(0.0001)
R^2	0.0007	0.05	0.11

where S = age - age left full time education and the figures in brackets are standard errors.

(a) Are the signs of the coefficients as you would expect? Explain.

(b) The R^2 statistics are very low in absolute terms. Is this a cause for concern? Explain.

(c) Why, in your opinion, has the coefficient of the age variable changed in the way it has between equations (i) and (ii)? Explain fully.

(d) Test the joint significance of the S and S^2 variables. On what assumptions does this test rely? Are they likely to be true in this case? Explain.

(e) It is suggested by a colleague that a Goldfeldt-Quandt test statistic should have been calculated for the models (i), (ii) and (iii). What exactly is a Goldfeldt-Quandt test, what do you infer from it and why was it suggested? Explain in detail.

END OF PAPER