This paper is not to be removed from the Examination Halls

UNIVERSITY OF LONDON **279 0020 ZA**

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

UL06/27

SECTION A

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

1. (a) Explain what is meant by consistency in a statistical estimator. Show that in a simple regression model of Y_t on X_t the ordinary least squares (OLS) estimate of the slope value is consistent.

(b) Explain what is meant by 'simultaneous equation bias' in the structural form of a simultaneous equation model? Why is simultaneous equation bias not a problem in the reduced form of the model? Explain carefully.

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

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

where Var is variance.

(d) Standard hypothesis testing on a simple linear model is achieved by using t tests on the parameter estimates and an F test defined as $F = \frac{R^2/(k-1)}{(1-R^2)/(T-k)}$ where R^2 is the goodness of fit statistic, T is the sample size and k is the number of coefficients including the constant term. What conclusions do you draw from

i. significant t tests but insignificant F test
ii. insignificant t tests but significant F test
iii. insignificant t tests and insignificant F test.

(e) An econometrician suggests an estimator for β given by $\tilde{\beta} = \frac{1}{N}\sum_{t=1}^{N}\frac{y_t}{x_t}$ where N is the sample size and the model is $y_t = \beta x_t + u_t$, $E(u_t) = 0$, $E(u_t^2) = \sigma^2$. Prove that $\tilde{\beta}$ is an unbiased estimator for β and that its variance is $\text{var}\{\tilde{\beta}\} = \frac{\sigma^2}{N^2}\left(\sum_{t=1}^{N}\frac{1}{x_t^2}\right)$ under certain assumptions. What extra assumptions have you used?

(f) Employing appropriate assumptions, show that least squares applied to a regression model with serially correlated error terms yields unbiased parameter estimators but incorrect standard errors.

(question continues on next page)

PLEASE TURN OVER

(g) 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.

(h) Show that the OLS estimate of β_0 in $Y_t = \beta_0 + \beta_1 X_t + u_t$ is given by $\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X}$. Under what conditions will $\hat{\beta}_0$ be zero?

SECTION B

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

2. (a) What is an adaptive expectations model? Explain how it can be used to analyse short and long term relationships between the dependent variable and an independent variable.

(b) Suppose business expenditure for new plant and equipment (Y_t) is explained by the relation

$$\ln(Y_t) = \alpha + \beta \ln(X_t^*) + u_t,$$

where u_t is a random variable, ln is the natural logarithm and X_t^* is the level of expected sales (which is unobserved) and is formed by

$$\ln(X_t^*) - \ln(X_{t-1}^*) = \gamma(\ln(X_{t-1}) - \ln(X_{t-1}^*)).$$

X_t is the level of actual sales.

i. Derive a linear relationship that can be used to estimate α, β and using observable variables Y_t and X_t.

ii. Given sufficient data on Y_t and X_t discuss the problems of estimating the parameters α, β and γ in the model you produced in part b(i). How would you estimate the parameters?

3. (a) Explain what is meant by (i) a non-stationary time series and (ii) a random walk with drift.

(b) Explain carefully how would you test a time series for non-stationarity?

(c) Non-stationary time series can give rise to a ‘spurious regression’. Explain what is meant by a spurious regression and how it arises. How would you decide whether your regression is spurious? Explain.

4. (a) Explain what you understand by the Durbin-Watson test. Why is it important in econometrics?

(b) The following regression was estimated by ordinary least squares on annual US data 1935-55.

$$\hat{C}_t = \underset{(152.0)}{249012} + \underset{(0.10)}{0.72}\ Y_t \qquad \text{(A)}$$

$R^2 = 0.73$, $F = 51.52$, $dw = 0.32$, where figures in brackets are standard errors, C_t is per capita personal consumption expenditure in 1958 prices, Y_t is per capita disposable income in 1958 prices and dw is the Durbin-Watson statistic.

i. What are the standard Gauss-Markov conditions which need to be satisfied if ordinary least squares is to give the best linear unbiased estimates? Assuming that the Gauss-Markov assumptions hold in the above estimated equation (A) test the hypothesis that the marginal propensity to consume is one.

ii. Explain carefully the Cochrane-Orcutt method of estimation. Under what conditions should it be used?

iii. The model (A) was re-estimated by the Cochrane-Orcutt method to give the following estimated equation,

$$\hat{C} = \underset{(228.0)}{745.92} + \underset{(0.14)}{0.40}\ Y_t;$$

$R^2 = 0.94$, $F = 273.7$, $dw = 1.21$, $\hat{\rho} = 0.94$.

where ρ is the first order serial correlation coefficient.

How would you account for the two different sets of results? Note that the estimated residuals from the regression (A) were:

1935	-12.11	1942	-195.43	1949	83.87
1936	-6.01	1943	-213.40	1950	81.32
1937	3.04	1944	-220.19	1951	62.37
1938	31.30	1945	-127.79	1952	63.19
1939	21.88	1946	29.23	1953	75.50
1940	19.00	1947	88.44	1954	87.18
1941	-40.41	1948	56.41	1955	112.64

PLEASE TURN OVER

5. 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.

6. (a) What do you understand by a logistic model and how would you estimate it? How does the logit model compare with a linear probability model? Explain.

(b) Data on 753 married women aged 30 -60 are used in a logistic regression with dependent variable equal to 1 if the wife works and 0 otherwise. The explanatory variables are:

K5 = number of children $\leq$ 5 years old,
K618 = number of children aged 6 -18,
Age = Wife's age in years,
Wc = 1 if wife attended college, 0 otherwise,
Hc = 1 if husband attended college, 0 otherwise,
Lwg = Logarithm of wife's estimated wage rate,
Inc = Family income excluding wife's earnings.

The estimated regression results are:

	Estimated Parameter	**Asymptotic t-Values**	**Mean of Variable**
Constant	3.182	4.94	
K5	-1.463	-7.43	0.24
K618	-0.065	-0.95	1.35
Age	-0.063	-4.92	42.54
Wc	0.807	3.51	0.28
Hc	0.112	0.54	0.39
Lwg	0.605	4.01	1.10
Inc	-0.034	-4.20	20.13

i. Interpret the estimated logistic regression results. Do the estimated coefficients have the expected sign?

ii. How would you measure the 'goodness of fit' of the above estimated equation?

7. 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.

PLEASE TURN OVER

8. Consider a two equation linear model

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

$t = 1,2,\cdots,T$; $E(u_t) = E(e_t) = 0$; $E(u_t^2) = \sigma_u^2$; $E(e_t^2) = \sigma_e^2$; $E(u_t e_t) = \sigma_{ue}$;
$E(u_s e_t) = 0$ 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) Discuss, without derivation, what is the indirect least squares estimator of α_1.

END OF PAPER