UNIVERSITY OF CAMBRIDGE

ECONOMICS TRIPOS PART IIA

Wednesday 5 June 2013 1:30-4:30

Paper 3

THEORY AND PRACTICE OF ECONOMETRICS I

The paper is divided into two Sections - A and B.

Answer **FOUR** questions from Section A and **TWO** questions from Section B.

Each Section carries equal weight.

Answers from each Section must be written in separte booklets with the letter of the Section written on each cover sheet.

Credit will be given for clear presentation of relevant statistics.

Write your **candidate number** not your name on the cover of each booklet.

This written exam carries 60% of the marks for Paper 3.

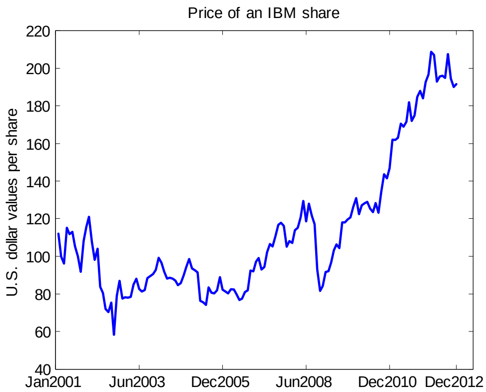
Write legibly

STATIONERY REQUIREMENTS	SPECIAL REQUIREMENTS
20 Page booklet x 2	Durbin Watson and Dicky Fuller Tables
Metric graph paper	New Cambridge Elementary Statistical Tables
Rough work pads	Approved calculators allowed
Tags	

You may not start to read the questions printed on the subsequent pages of this question paper until instructed that you may do so by the Invigilator

SECTION A

1 The following figure shows the weekly time series p_t for the price of the IBM share during 144 weeks from January 2001 to December 2012.



(a) How would you model this time series, and what is the economic justification for your model?

(b) A linear regression of p_t on a constant and p_{t-1} yields the following OLS results

$$\hat{p}_t = \underset{(1.958)}{0.900} + \underset{(0.017)}{0.997} p_{t-1},$$

where the standard errors of the coefficient estimates are given in parentheses. Test the hypothesis that p_t is integrated of order one.

(c) The sample standard deviation of $p_t - p_{t-1}$ is 7.15. Would you be surprised if the price of the IBM share in the last week of June of 2013 turns out to be 270 U.S. dollars? Explain your answer.

2 A production function for farms is

$$\log Y_{it} = \alpha + \beta \log X_{it} + \gamma \log S_{it} + u_{it}, \qquad (1)$$

where Y_{it} is the crop produced at farm i during period t, X_{it} is the amount of labour employed, S_{it} describes the quality of soil, and u_{it} consists of random factors such as weather. Suppose that you observe Y_{it} and X_{it} for $i = 1, ..., n$ and $t = 1, ..., T$, but that you do not observe S_{it}.

(a) Under what conditions is the OLS estimator of β, from the regression of Y_{it} on a constant and X_{it}, consistent for β? Are these conditions plausible from the microeconomic perspective? Explain.

(b) Describe the fixed effects estimator of β. Is it consistent? Explain.

(c) It was observed that producing an excessively large quantity of the crop during period t leads to lower productivity of plants during the next period. How would you modify equation (1) to model this phenomenon?

3 Person i, who is caught speeding, can choose to attend a day-long speed awareness course instead of taking three points on his or her license (12 or more penalty points disqualify you from driving). Let $Y_i = 1$ if person i is caught speeding again within the next 3 years, and let $Y_i = 0$ otherwise. Further, let $D_i = 1$ if individual i takes the course, and let $D_i = 0$ otherwise.

(a) Suppose that

$$Y_i = \begin{cases} Y_{0i} & \text{if} \quad D_i = 0 \\ Y_{1i} & \text{if} \quad D_i = 1 \end{cases}, \tag{2}$$

where Y_{1i} (Y_{0i}) denote the potential outcomes if individual i takes (does not take) the course. Assuming that the individual treatment effect of taking the course is constant and equal to β, show that equation (2) can be written as

$$Y_i = \alpha + \beta D_i + \eta_i, \text{ where } E(\eta_i) = 0.$$

(b) Let $E(Y_i|D_i) = \gamma + \delta D_i$. Explain why δ may not be equal to β.

(c) Suppose that the course can only be taken by individuals who live in randomly chosen areas within the UK. Suppose that you know post codes of these areas. How would you use data on the post codes of the speed limit violators to consistently estimate β?

4 A researcher studies the effect of water pollution on the mortality rate in a fish population. He runs extensive laboratory experiments and determines that the 0.1% and 0.05% concentrations of a pollutant in water lead to 50% and 23% mortality rate, respectively.

(a) Assuming that the probability of death is described by a probit model, determine the mortality rate of fish in absolutely clean water.

(b) Let $\Phi(x)$ be the cumulative distribution function of $N(0,1)$. The values of the standard normal density at x_1 and x_2 such that $\Phi(x_1) = 0.5$ and $\Phi(x_2) = 0.23$ are 0.399 and 0.303, respectively. Using this information, compare the marginal effects of reducing the concentration of the pollutant from the initial values of 0.1% and 0.05%.

(c) In the context of the problem above, discuss potential pitfalls of the linear probability model.

5 An investigator analysing the relationship between wages and unemployment starts by estimating the following static linear regression model using aggregate time series data

$$w_t = \beta_0 + \beta_1 p_t + \beta_2 U_t + \varepsilon_t,$$

where w_t is the logarithm of the nominal wage, p_t the logarithm of the price level, and U_t the unemployment rate (measured as a percentage).

(a) A colleague criticises the investigator's estimates of this regression on the grounds that, since wages and prices are trended, the resulting relationship is 'spurious'. Explain what is meant by a spurious regression, and how you might detect it.

(b) The investigator now reformulates his regression to allow for dynamic effects, so that it takes the form

$$w_t = \theta_0 + \theta_1 p_t + \theta_2 p_{t-1} + \theta_3 U_t + \theta_4 U_{t-1} + \theta_5 w_{t-1} + \varepsilon_t$$

Explain how you could use estimates of the coefficients of this regression to compute the long run effect of a higher price level on nominal wages.

(c) A second colleague now argues that the regression should be reformulated so that nominal wages are allowed to respond to movements in the expected price level p_t^e, so that it becomes

$$w_t = \theta_0 + \theta_1 p_t^e + \theta_2 p_{t-1} + \theta_3 U_t + \theta_4 U_{t-1} + \theta_5 w_{t-1} + \varepsilon_t$$

Suggest two strategies which you might employ to estimate such a regression.

6 Consider a simple linear regression framework

$$y_t = \beta_0 + \beta_1 x_t + u_t \text{ with } t = 1, \ldots, T,$$

where $E\left(u_t | x_1, \ldots, x_t\right) = 0$ for all t. Suppose the error term follows a stationary AR(2) process, say $u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \varepsilon_t$, where $\{\varepsilon_t, t = 1, ..., T\}$ are i.i.d. with distribution $N\left(0, \sigma^2\right)$.

(a) In the above regression, what are the advantages of GLS relative to OLS?

(b) Describe in detail how you would compute the feasible GLS estimators for β_0 and β_1.

(c) In general, would you always prefer feasible GLS to OLS? Discuss.

7 Let $hy6_t$ be the three-month holding yield from buying a six-month U.S. Treasury bill at time $t-1$ and selling it at time t as a three-month Treasury bill. Further, let $hy3_t$ be the three-month holding yield from buying a three-month Treasury bill at time $t-1$. The expectation hypothesis implies that the slope coefficient in the regression of $hy6_t$ on $hy3_t$ should not be statistically different from one.

(a) Assuming that there is evidence of a unit root in $hy3_t$, discuss problems with standard OLS analysis of the above regression.

(b) Rewrite the following model in the error correction form

$$hy6_t = \theta_0 + \theta_1 hy6_{t-1} + \theta_2 hy3_t + \theta_3 hy3_{t-1} + u_t. \tag{3}$$

What constraint on the parameters $\theta_1, ..., \theta_3$ is implied by the expectation hypothesis?

(c) Assuming that the constraint holds, explain how you would estimate the error correction form of equation (3). How would you test for serial correlation in u_t?

SECTION B

8 A study of wage determination based on a random sample of 1472 individuals from the working population of Belgium in 1994 contains 893 males and 579 females. Table 1 reports OLS results from four different specifications of linear regression of $\log(wage_i)$ (in Belgian francs) on the explanatory variables defined as follows:

- $male_i$ is 1 if male, 0 if female;
- $educ_i$ is 1 if primary school, 2 if lower vocational training, 3 if intermediate level, 4 if higher vocational training, 5 if university level;
- $edu(j)_i$ is 1 if $educ_i = j$, 0 otherwise;
- $exper_i$ is experience in years.

(a) Suppose that a woman is expected to earn five million Belgian francs. According to the OLS results for specification (1), how much would a man with the same amount of education and experience be expected to earn?

(b) From an economic point of view, explain why you may prefer specification (2) to specification (1). Test specification (1) against specification (2) at the 1% significance level.

(c) Use results from Table 1 to perform the Chow test (at the 1% significance level) of a hypothesis that wage determination is the same for men and women.

(d) The results in Table 1 do not take into account a potential heteroskedasticity problem. What effects would heteroskedasticity have on the results reported? In particular, comment on the validity of the estimates, their standard errors, R^2, adjusted R^2 and the F-statistic.

(e) Consider specification (2). You suspect that the source of heteroskedasticity is related to gender alone. Describe in detail how you would test for this. Propose an alternative estimator that may be preferred to the OLS estimator in case your suspicion is correct.

Table 1

OLS results for four specifications of wage regression. Standard errors are in parentheses. F-statistic corresponds to the test of the regression's significance.

Regressor	(1)	(2)	(3)	(4)
$constant$	1.145 (0.041)	1.272 (0.045)	1.372 (0.040)	1.216 (0.078)
$male_i$	0.118 (0.016)	0.128 (0.015)		0.154 (0.095)
$\log(educ_i)$	0.437 (0.018)			
$\log(exper_i)$	0.231 (0.011)	0.230 (0.011)	0.310 (0.021)	0.207 (0.017)
$edu(2)_i$		0.144 (0.033	0.113 (0.021)	0.224 (0.068)
$edu(3)_i$		0.305 (0.032)	0.343 (0.024)	0.433 (0.063)
$edu(4)_i$		0.474 (0.033)	0.374 (0.025)	0.602 (0.063)
$edu(5)_i$		0.639 (0.033)	0.469 (0.026)	0.755 (0.065)
$male_i \times edu(2)_i$				−0.097 (0.078)
$male_i \times edu(3)_i$				−0.167 (0.073)
$male_i \times edu(4)_i$				−0.172 (0.074)
$male_i \times edu(5)_i$				−0.146 (0.076)
$male_i \times \log(exper_i)$				0.041 (0.021)
Summary Statistics				
R^2	0.376	0.398	0.388	0.403
adjusted R^2	0.375	0.395	0.385	0.399
F-statistic	294.96	161.14	166.24	89.69

(TURN OVER)

9 A researcher is investigating the relationship between oil price 'shocks' and the business cycle in the United States over the period 1988Q1-2011Q4. She estimates the following pair of regressions

$$\widehat{\Delta p_t} = \underset{(0.1408)}{0.3343} + \underset{(0.0008)}{0.0020}t - \underset{(0.0471)}{0.1175}p_{t-1} \text{ with } AIC = -0.9944,$$

and

$$\begin{aligned}\widehat{\Delta p_t} &= \underset{(0.1470)}{0.2844} + \underset{(0.0008)}{0.0017}t - \underset{(0.0491)}{0.1002}p_{t-1} + \underset{(0.1026)}{0.3330}\Delta p_{t-1} - \underset{(0.1090)}{0.3662}\Delta p_{t-2} \\ &\quad + \underset{(0.1052)}{0.2164}\Delta p_{t-3} - \underset{(0.1038)}{0.1844}\Delta p_{t-4} \text{ with } AIC = -1.1040,\end{aligned}$$

where p_t is the log of the real oil price at date t, Δ is the first difference operator such that $\Delta p_t = p_t - p_{t-1}$, AIC is the Akaike Information Criterion, and standard errors are in parentheses.

(a) Using these equations, conduct a test of the null hypothesis that the log of the real oil price is a random walk with drift. Explain the reasoning behind your choice of test equation.

(b) The researcher now estimates a Vector Autoregression (VAR) of the form

$$\begin{aligned}\Delta p_t &= \alpha_0 + \sum_{j=1}^{4} \beta_j \Delta gdp_{t-j} + \sum_{k=1}^{4} \gamma_k \Delta p_{t-k} + \varepsilon_{1t} \\ \Delta gdp_t &= \psi_0 + \sum_{j=1}^{4} \theta_j \Delta gdp_{t-j} + \sum_{k=1}^{4} \phi_k \Delta p_{t-k} + \varepsilon_{2t}\end{aligned}$$

where gdp_t is the log of real US GDP at date t. The disturbance terms $\{\varepsilon_{1t}, \varepsilon_{2t}\}$ are jointly normally distributed with mean zero and variance-covariance matrix Ω, and are serially uncorrelated.

The results of these regressions and their constrained versions are given in Table 2 below. Using these results, and a 5% significance level, test the hypothesis that movements in real oil prices do not Granger-cause movements in real US GDP.

(c) The researcher now re-estimates her equations over two selected subperiods: 1988Q1-2007Q4 (the 'Great Moderation') and 2008Q1-2011Q4 (the 'Great Recession'). The residual sums of squares from these subperiod regressions are shown in the lower part of Table 2. Using these results, and a 5% significance level, test the hypothesis that the coefficients of the equation relating growth in real GDP to the lagged growth in GDP and real oil prices are structurally stable.

(d) Comment on the implications of the additional results from part (c) for your conclusions in part (b) of this question.

Table 2
OLS estimates of the VAR from question 9. Standard errors are in parentheses

	A	*B*	*C*	*D*
Dependent	$\Delta oilp_t$	Δgdp_t	$\Delta oilp_t$	Δgdp_t
$\Delta oilp_{t-1}$	0.2455 (0.1075)	−0.0073 (0.0043)	0.3001 (0.1018)	
$\Delta oilp_{t-2}$	−0.4175 (0.1112)	−0.0027 (0.0044)	−0.4256 (0.1053)	
$\Delta oilp_{t-3}$	0.2088 (0.1093)	−0.0079 (0.0044)	0.1965 (0.1057)	
$\Delta oilp_{t-4}$	−0.2143 (0.1088)	−0.0060 (0.0043)	−0.2308 (0.1020)	
Δgdp_{t-1}	3.6963 (2.6955)	0.3692 (0.1072)		0.3845 (0.1035)
Δgdp_{t-2}	2.8905 (2.8177)	0.2951 (0.1121)		−0.4256 (0.1099)
Δgdp_{t-3}	−4.0373 (2.8008)	−0.1018 (0.1114)		−0.1627 (0.1099)
Δgdp_{t-4}	1.1741 (2.6931)	0.1186 (0.1071)		0.0772 (0.1038)
Intercept	−0.0094 (0.0236)	0.0021 (0.0009)	0.0136 (0.0141)	0.0026 (0.0009)
R^2	0.2332	0.3471	0.1870	0.2490
SSR $(88Q1-11Q4)$	1.60851	0.00254	1.70532	0.00293
Results for subsamples				
SSR $(88Q1-07Q4)$	0.95080	0.00148	1.09563	0.00161
SSR $(08Q1-11Q4)$	0.33948	0.00043	0.48177	0.00072

(TURN OVER)

10 An analyst wishes to examine the effect of family size on the employment status of mothers. Data have been taken from the 1980 census on married women aged 21-35 with at least two children. Columns of Table 3 marked as (2), (4) and (6) report OLS and IV (Wald) estimates of the effects of a third birth on labour supply using Twins and Sex Composition as instruments. Employment is a binary indicator equal to one if the woman is in the labour force and zero otherwise. Twins is a binary indicator equal to one if the first two children are twins, and Sex Composition is a binary indicator equal to one if the first two children are of the same sex.

(a) In what sense might an observed association between family size and employment be difficult to interpret?

(b) Write down the regression equation for the structural, first-stage and reduced form equation. Discuss how each of these equations can be used to estimate the impact of family size on labour supply.

(c) Explain the basis for the use of Sex Composition as an instrument for labour supply. Interpret the first stage estimates for this instrument.

(d) Compare the results for the OLS and the Wald estimates based upon the use of Twins and Sex Composition as instruments.

(e) In what sense are the individual Wald estimates inefficient?

Table 3

Estimates of the effects of family size on labour supply. All models include controls for mother's age, age at first birth, dummies for the sex of first and second births, and dummies for race.

			IV Estimates Using			
			Twins		Sex Composition	
Dependent Variable	Mean (1)	OLS (2)	First Stage (3)	Wald Estimates (4)	First Stage (5)	Wald Estimates (6)
Employment	0.528	−0.167 (0.002)	0.625 (0.011)	−0.083 (0.017)	0.067 (0.002)	−0.135 (0.029)

END OF PAPER