This paper is not to be removed from the Examination Halls

UNIVERSITY OF LONDON **279 0020 ZA**

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

Elements of Econometrics and Economic Statistics

Tuesday, 11 May 2004 : 10.00am to 1.00pm

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 securely fastened inside the answer book.

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

A hand held non-programmable calculator may be used when answering questions on this paper. The make and type of machine must be stated clearly on the front of the answer book.

PLEASE TURN OVER

UL04/433

Section A

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

1. (a) Explain how dummy variables can be used to correct for seasonal variation in economic data.

(b) Suppose the regression model $Y_t = \beta_0 + \beta_1 X_t + u_t$, where $E(u_t) = 0$, $E(u_t^2) = \sigma^2$ where σ^2 is known and $E(u_t u_s) = 0$ $t \neq s$ is estimated by ordinary least squares (OLS) using T_1 observations to produce estimates of β_0, β_1 and their standard errors. Suppose now that extra data is measured such that there are now T_2 observations ($T_2 > T_1$) how would the estimates of the parameters and their standard errors change? Explain.

(c) Suppose that the model $Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t$ where $E(u_t) = 0$, $E(u_t^2) = \sigma^2$ and $E(u_t u_s) = 0$ $t \neq s$ is estimated by ordinary least squares (OLS) using T observations and the null hypotheses H_0: $\beta_i = 0$, $i = 1,2,3$ are not rejected but the null hypothesis H_0: $\beta_1 = \beta_2 = \beta_3 = 0$ is rejected what conclusions might you draw? Explain.

(d) Show that the ordinary least squares (OLS) estimate of β in the model $y_t = \beta x_t + u_t$ where $u_t = \rho u_{t-1} + v_t$ is unbiased under certain assumptions. What are these assumptions?

(e) What do you understand by an instrumental variable? How would you estimate the model $y_t = \beta x_t + u_t$ by the instrumental variable method?

(f) In a regression model $Y_t = \beta_0 + \beta_1 X_t + u_t$ where $u_t = \rho u_{t-1} + v_t$ and $E(v_t) = 0$, $E(v_t^2) = \sigma^2$, $E(v_t v_s) = 0$ for $t \neq s$ the Durbin-Watson statistic is

computed as $dw = \dfrac{\sum_{t=2}^{n} (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^{n} \hat{u}_t^2}$ where $\hat{u}_t$ is the estimated least squares residual. It can be shown that dw is approximately equal to $2(1 - \rho)$ in large samples. What is the implication of this result for the distribution of the Durbin-Watson statistic and the operation of the Durbin-Watson test?

(g) Let X and Y be two random variables. Show that

$$var(aX + bY) = a^2 var(X) + b^2 var(Y) + 2ab cov(X,Y)$$

where a and b are known constants.
Explain how this result changes if X and Y are independent.

(question continues on next page)

PLEASE TURN OVER

(h) Let $\hat{u}_t$ be the residuals in the least squares fit of y_t against x_t and a constant term for t=1,2,3.....T. Derive the following results:-

$$\sum_{t=1}^{T} \hat{u}_t = 0 \qquad \text{and} \qquad \sum_{t=1}^{T} x_t \hat{u}_t = 0$$

Section B

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

2. Economists have tried to examine the catch-up hypothesis where it is predicted that poorer countries will grow faster than rich countries to 'catch-up' the rich countries. To assess this hypothesis an economist regresses the growth rate of GDP (gross domestic product) on GDP per capita and then on GDP per capita and other variables which might influence growth rates. The following regression results come from regressions using data from 96 countries

$$gr_i = \underset{(0.023)}{-0.058} - \underset{(0.002)}{0.009}\ln(gdp_i) + e_{1i} \qquad \text{(A)}$$

$n = 96$, $R^2 = 0.122$ and $s = 0.025$

where standard errors are given in brackets, gr_i is the growth rate of the ith country over the period 1985-1990, $\ln(gdp_i)$ is the logarithm of GDP for country i, e_{1i} is the estimated residual and s is the standard error of the residuals.

A second regression yields

$$gr_i = -0.019 + \underset{(0.023)}{0.001}\ln(gdp_i) + \underset{(0.002)}{0.12}\, inv_i + e_{2i} \qquad \text{(B)}$$

$n = 96$, $R^2 = 0.196$ and $s = 0.024$

where e_{2i} is the estimated residual and inv_i is the ratio of investment to GDP in 1985.

(a) Test the hypothesis that the coefficient on $\ln(gdp_i)$ is zero for both regressions. What does the result of your tests say about the 'catch-up' hypothesis?

(b) The R^2 measure is higher for equation (B) than for (A). Is this what you would expect? Explain.

(question continues on the next page)

(c) The regression results may be affected by heteroskedasticity. Explain what you understand about heteroskedasticity, why the estimation results might be affected by heteroskedasticity and what the effects are if heteroskedasticity is present.

3. A model where the dependent variable is the log of expenditure on beer at 1995 prices ($beer_t$) is regressed on the log of total household expenditure at 1995 prices (exp_t), the log of the price of beer relative to all consumer prices (pb_t) and the log of the price of alcoholic drinks excluding beer relative to all consumer prices (pa_t) gives the following results:-

$$\underset{}{beer_t} = \underset{(1.387)}{-5.272} + \underset{(0.114)}{1.266}\, exp_t \; \underset{(0.096)}{-0.989}\, pb_t \; \underset{(0.134)}{-0.412}\, pa_t + e_t$$

e_t is the estimated residual, standard errors are in brackets, the sample size is 45 and $R^2 = 0.906$.

(a) Interpret the estimated equation.

(b) Test the null hypothesis that the coefficient on exp_t is unity. What is the significance of this hypothesis?

(c) Construct a 95% confidence interval on the coefficient of pb_t. Explain why the confidence interval is a useful construct.

(d) If the variable pb_t was dropped from the regression explain what would happen to the parameter estimate of pa_t and why.

4. (a) Explain what is meant by a best linear unbiased estimator.

(b) In the model $y_t = \alpha x_t + u_t$ the ordinary least squares (OLS) estimator of α is $\hat{\alpha} = \dfrac{\sum_{t=1}^{n} x_t y_t}{\sum_{t=1}^{n} x_t^2}$. Show how this result is derived detailing all the assumptions you use.

(c) Explain why $\hat{\alpha}$ is a linear estimator.

(d) Show that $\hat{\alpha}$ is an unbiased estimator detailing all the assumptions you use.

PLEASE TURN OVER

5. The following set of regression results were obtained from annual UK data 1963-1989:-

$$\hat{S}_t = \underset{(2719)}{18455} + \underset{(0.071)}{0.230}\, Y_t - \underset{(220)}{450}\, INF_t + \underset{(340)}{1220}\, R_t - \underset{(270)}{750}\, DEM_t$$

$R^2 = 0.795$, Durbin-Watson = 1.40.

The variables are:-

S_t Aggregate Personal Savings (£'000)
Y_t Personal Disposable Income (£'000)
INF_t Rate of price inflation (%)
R_t Rate of interest (%)
DEM_t Percentage of the population in the 18-35 age group.
The numbers in brackets are estimated standard errors.

(a) Test the individual significance of the three variables INF, R and DEM. Do the coefficients have the expected sign?

(b) It is often asserted that regressions using time series data suffer from serial correlation. What is serial correlation? Why do time series regressions suffer from serial correlation? What is the effect of serially correlated errors on OLS results?

(c) Test the above estimated equation for the presence of first order serial correlation.

(d) Do the results of your test in (c) alter the conclusions you drew in (a)? Explain.

6. A simple model of supply and demand for a consumption good is:-

$$D_t = \alpha_0 + \alpha_1 P_t + \alpha_2 Y_t + \alpha_3 \overline{P}_t + u_{1t}$$
$$S_t = \beta_0 + \beta_1 P_t + \beta_2 P_{t-1} + u_{2t}$$
$$D_t = S_t$$

where
D_t is the demand for the good
S_t is the supply of the good
P_t is the price of the good
$\overline{P}_t$ is an index of retail prices
Y_t is consumers' income
u_{1t} and u_{2t} are random disturbances.

(a) What do you understand by exogenous and endogenous variables? Which variables in the above model would you consider to be endogenous and which exogenous? Explain.

(question continues on the next page)

(b) Examine the identification of each equation in the model.

(c) Explain why indirect least squares is an inappropriate estimation method for the supply equation. When can indirect least squares be used?

(d) Suppose the equations were estimated using two-stage least squares. Explain this method of estimation and describe the conditions under which this method can be used.

7. (a) Explain what you understand by a 'dummy variable'.

(b) Explain how dummy variables can be used to test for a 'change in structure'.

(c) Give details of an alternative test for a 'change in structure' and compare this technique with that of dummy variables.

(d) What other uses, apart from a test of a change in structure or seasonal correction, might an econometrician employ dummy variables? Explain carefully the details of this other use.

8. Let the regression equation be:

$$Y_t = \beta_1 + \beta_2 X_t + u_t \quad ; \quad t = 1,2,\ldots,T$$

where

$$E(u_t) = 0$$
$$E(u_s u_t) = \sigma^2 \quad \text{if } s = t$$
$$= 0 \quad \text{if } s \neq t$$

(a) Obtain the ordinary least squares (OLS) estimators of β_1 and β_2.

(b) Show that the OLS estimator $\hat{\beta}_2$ of β_2 is unbiased.

(c) Is $\hat{\beta}_2$ consistent? Explain in detail.

END OF PAPER

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 and Economic Statistics

Wednesday, 25 May 2005 : 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 securely fastened inside the answer book.

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

A hand held non-programmable calculator may be used when answering questions on this paper. The make and type of machine must be stated clearly on the front cover of the answer book.

PLEASE TURN OVER

SECTION A

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

1. (a) Explain the importance of R^2 (coefficient of determination) and $\overline{R}^2$ (R^2 adjusted for degrees of freedom). What advantages does $\overline{R}^2$ have over R^2?

(b) Explain what you understand by autocorrelation of the disturbance term in a regression model? What are the causes of autocorrelation?

(c) Explain the concepts of Type I error and Type II error in hypothesis testing. What do you understand by the power of a test?

(d) A simple random sample of size n is taken from a normal population with unknown mean μ but known standard deviation σ. The sample mean is $\overline{X}$.

Explain how would you test the null hypothesis $H_0 : \mu = \mu_0$ against the alternative

(i) $H_1 : \mu \neq \mu_0$

(ii) $H_1 : \mu > \mu_0$.

(e) Explain what you understand by omitted variable bias?

(f) Let a regression equation be:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t \quad ; \quad t = 1,2,\ldots,T.$$

Outline briefly, how would you test

(i) $\beta_2 = 1$

(ii) jointly β_2 and β_3 are significantly different from zero.

(g) Let $Y = \alpha + \beta X$, where Y and X are random variables and α and β are constants. Show that $\mathrm{Cov}(X, Y) = \beta^2 \mathrm{Var}(X)$.

(h) Explain how dummy variables can be used to verify structural change?

SECTION B

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

2. Let a regression equation be:

$$Y_t = \beta_1 + \beta_2 X_t + u_t \quad ; \quad t = 1,2,\ldots,T.$$

(a) Obtain the ordinary least squares (OLS) estimator of β_2.

(b) Stating all your assumptions carefully, show that the OLS estimator of β_2 is best linear unbiased.

3. In order to model the demand for motor vehicles, an econometrician proposes the general linear regression model

$$Y_t = \beta_0 + \beta_P P_t + \beta_E E_t + \beta_B B_t + u_t \quad ; t = 1965, 1966, \cdots, 1986$$

where
Y is the logarithm of an index of consumer expenditure on motor vehicles, spares and accessories at constant prices,
P is the logarithm of a relative price index of motor vehicles,
E is the logarithm of real total household expenditure,
B is the logarithm of a relative price index of public road transport,
u is the error term.

This model and a restricted version of the model were fitted using ordinary least squares (OLS) to annual data covering the period 1965-1986 and the following results were obtained

$$\underset{}{\hat{Y}_t} = \underset{(0.56)}{6.27} - \underset{(.067)}{0.705 P_t} \qquad \text{(A)}$$

$R^2 = 0.0738$, RSS=0.636;

$$\hat{Y}_t = -\underset{(3.03)}{2.05} - \underset{(0.347)}{0.926 P_t} + \underset{(0.644)}{1.78 E_t} + \underset{(0.310)}{0.0608 B_t} \qquad \text{(B)}$$

R^2=0.720, RSS=0.192,

where R^2 is the coefficient of determination, RSS denotes residual sum of squares and estimated standard errors are given in parentheses.

(question continues on the next page)

(a) Test the hypothesis $H_0 : \beta_P = 0$ in both fitted models (A) and (B). Comment on your results.

(b) Test the individual hypotheses $H_0 : \beta_E = 0$ and $H_0 : \beta_B = 0$ and the joint hypothesis $H_0 : \beta_E = \beta_B = 0$.

(c) Which of the fitted models (A) and (B) is preferable? Explain your answer.

(d) Discuss the economic implications of the fitted model (B). Is there any evidence that public road transport acts as a substitute for private motor travel?

(e) How might a plot of the OLS residuals from fitted model (B) against time assist you in ascertaining whether or not the model for the demand for motor vehicles is misspecified? Explain.

4. Let a model be:

$$Y_t = \beta_0 + \beta_1 X_{t1} + \beta_2 X_{t2} + u_t \quad ; t = 1,2,\cdots,T$$

$E(u_t) = 0$ for all t. A researcher suspects that the variance of the disturbance term is $V(u_t) = \sigma^2 X_{t1}$.

(a) Explain how the researcher should proceed to test the null hypothesis $H_0 : V(u_t) = \sigma^2$ against the alternative hypothesis $H_1 : V(u_t) = \sigma^2 X_{t1}$, for all $t = 1,2,\cdots,T$.

(b) If the researcher's suspicion is correct then how will it affect the properties of the ordinary least squares estimators?

(c) Suggest in detail an estimation procedure, which will give best linear unbiased estimates of the parameters.

PLEASE TURN OVER

5. The following set of models for the demand for holidays abroad by UK residents produced the estimates:

	I	II	III
EXCH	0.226 (0.19)	0.338 (0.40)	1.127 (0.95)
RDISPY	0.240 (8.65)	0.232 (10.46)	0.147 (2.60)
HOL_{-1}	-	-	0.86 (3.20)
TEMP	-	-0.792 (1.61)	-0.592 (1.11)
RAIN	-	0.0045 (2.74)	0.0042 (2.37)
Constant	19.436 (2.11)	26.43 (2.74)	-461.91 (0.91)
R^2	0.60	0.67	0.68
Adjusted R^2	0.57	0.65	0.65
dw	0.71	1.71	2.25
N	20	19	19

(figures in brackets are t values) and where

HOL is the number of overseas holidays taken by UK residents in the current year.
EXCH is the average exchange rate ($/£)
RDISPY is the real disposable income at 1980 prices
HOL_{-1} is the number of holidays taken by UK residents in the previous year
TEMP is the mean daily temperature in the UK in the previous year
RAIN is the annual level of rainfall in the UK in the previous year
N is the sample size, dw is the Durbin-Watson statistic.

(a) Interpret the estimated equations. Why is the sample size different for model I?

(b) The coefficient on RAIN is very small. Does this suggest that RAIN is not important? Explain your answer.

(c) Which of the 3 models do you prefer as an explanation of the demand for holidays abroad and why?

6. Consider the simple linear regression model

$$Y_t = \beta_0 + \beta_1 X_t + u_t \quad ; t = 1,2,\cdots,T$$

where β_0 and β_1 are an unknown coefficients. t=1,...,T , are T observations on the random variables Y and X. A "goodness of fit" measure for ordinary least squares (OLS) estimation of the above model is defined by

(question continues on the next page)

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

(a) Prove that

R^2=1-RSS/TSS,

where the residual sum of squares $RSS \equiv \sum_{t=1}^{n} \hat{u}_t^2$, $\hat{u}_t = Y_t - \hat{\beta}_0 - \hat{\beta}_1 X_t$,

for all t.

(b) Hence, or otherwise, show that

$0 \leq R^2 \leq 1.$

(c) Briefly detail what happens to R^2 if an extra explanatory variable is added to the regression model.

(d) Has R^2 any drawback? Explain.

7. Consider a two equation model

$$y_t = \alpha z_t + u_{1t}$$
$$y_t = \beta_1 z_t + \beta_2 x_t + u_{2t} \quad ; \quad t = 1, 2, \ldots, T$$

where y_t and z_t are endogenous variables, x_t is an exogenous variable, u_{1t} and u_{2t} are serially uncorrelated disturbances with zero means, variances σ_1^2 and σ_2^2 and covariance σ_{12} for all t.

(a) Write down the reduced form corresponding to y and z. Suggest a method of estimation of the reduced form parameters, which will give unbiased and consistent estimators.

(b) Comment on the identifiability of the above equations. Is there any point in estimating an underidentified equation? Explain.

(c) Obtain the indirect least squares estimator of α. Is this estimator consistent? Explain.

PLEASE TURN OVER

8. Consider a two variable regression model

$$Y_t = \beta_1 + \beta_2 X_t + u_t \quad ; \quad t = 1,2,\ldots,T$$

where

$$u_t = \rho u_{t-1} + \varepsilon_t \quad \text{for all } t$$

$$\begin{aligned} E(\varepsilon_t) &= 0 \\ E(\varepsilon_s \varepsilon_t) &= \sigma_\varepsilon^2 \quad \text{if } s = t \\ &= 0 \quad \text{if } s \neq t \end{aligned}$$

(a) Derive:-

(i) the variance of u_t and

(ii) $E(u_t u_{t-1})$.

(b) Explain in detail how you would test the null hypothesis $\rho = 0$? Specify all the assumptions needed for this test.

(c) Discuss in detail a method of estimation which gives best linear unbiased estimates of β_1 and β_2.

END OF PAPER

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 and Economic Statistics

Wednesday, 25 May 2005 : 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 securely fastened inside the answer book.

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

A hand held non-programmable calculator may be used when answering questions on this paper. The make and type of machine must be stated clearly on the front cover of the answer book.

PLEASE TURN OVER

UL05/34

SECTION A

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

1. (a) Show that $\overline{Y}/\overline{X}$ is an unbiased estimator of α in $y_t = \alpha x_t + u_t$. What assumptions are necessary for this result? Is $\overline{Y}/\overline{X}$ a good estimator for β_2 in $Y_t = \beta_0 + \beta_2 X_t + u_t$? Explain.

(b) Show that the infinite distributed lag model

$$Y_t = \alpha + \beta \sum_{j=0}^{\infty} \lambda^j X_{t-j} + \varepsilon_t$$

can be written in terms of X_t and a single lag Y_{t-1}. What estimation problems does this model have?

(c) How can you test for the presence of heteroskedasticity in a regression model?

(d) There are 3 standard ways of measuring the explanatory power of a regression equation: R^2 (adjusted and unadjusted) and the standard F statistic. Explain how these are related to each other and how they measure the explanatory power of the equation.

(e) Assume that consumption is a function of income and suppose you believe that the oil crisis of 1973/74 reduced consumption expenditure in 1974 but then returned to the original level in 1975. How would you test this hypothesis using dummy variables?

(f) Why is the assumption of normality of the disturbance term an important one for econometricians?

(g) What is meant by indirect least squares? Explain.

(h) If Y is a random variable formed by $\alpha+\beta X$, where X is a random variable. α and β are constants, show that $\text{cov}(X,Y) = \beta^2 \text{ var}(X)$. What is the correlation between X and Y?

PLEASE TURN OVER

SECTION B

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

2. A study of applications for home mortgages used the linear probability model

$$MORT_i = \beta_0 + \beta_1 INC_i + \beta_2 AGE_i + \beta_3 PROP_i + u_i$$

where

$MORT_i$ = 1 if a mortgage is granted: 0 otherwise
INC_i = income of applicant in £000
AGE_i = age of applicant in years
$PROP_i$ = age of the property for which the mortgage is being applied.

(a) The estimated coefficient for INC_i was 1.48 with standard error 0.51. What is the interpretation of this coefficient?

(b) What are the problems associated with the linear probability model?

(c) What is a logit model? Does it overcome the problems you listed in (b)?

3. A regression of consumption (C) on income (Y) and unemployment (U) using annual data 1961-82 for the UK produced the following results:

$$\hat{C}_t = \underset{(2817.0)}{17880} + \underset{(0.026)}{0.7527}Y_t + \underset{(0.798)}{0.930}U_t \qquad R^2 = 0.992$$

(figures in brackets are standard errors) with a table of correlation coefficients between variables of:

	C	Y	U
C	1.00	0.996	0.783
Y	0.996	1.00	0.771
U	0.783	0.771	1.00

(a) Describe carefully the process of hypothesis testing for ordinary least squares parameter estimates of a standard linear regression using t and F distributions. What assumptions have you used?

(question continues on next page)

(b) Interpret the estimated equation and the table of correlation coefficients. From the evidence given do you think that multicollinearity is a problem? Explain.

(c) If consumption was to be regressed on income alone what do you think would happen to the coefficient on income? Explain.

4. The following set of models for the demand for holidays abroad by UK residents produced the estimates:

	I	II	III
EXCH	0.226 (0.19)	0.338 (0.40)	1.127 (0.95)
RDISPY	0.240 (8.65)	0.232 (10.46)	0.147 (2.60)
HOL_{-1}	-	-	0.86 (3.20)
TEMP	-	-0.792 (1.61)	-0.592 (1.11)
RAIN	-	0.0045 (2.74)	0.0042 (2.37)
Constant	19.436 (2.11)	26.43 (2.74)	-461.91 (0.91)
R^2	0.60	0.67	0.68
Adjusted R^2	0.57	0.65	0.65
dw	0.71	1.71	2.25
N	20	19	19

(figures in brackets are t values) and where
HOL is the number of overseas holidays taken by UK residents in the current year.
EXCH is the average exchange rate (\$/£)
RDISPY is the real disposable income at 1980 prices
HOL_{-1} is the number of holidays taken by UK residents in the previous year
TEMP is the mean daily temperature in the UK in the previous year
RAIN is the annual level of rainfall in the UK in the previous year
N is the sample size dw is the Durbin-Watson statistic.

(a) Explain the use of the Durbin-Watson statistic.

(b) Interpret the estimated equations. Why is the sample size different for model I?

(c) The coefficient on RAIN is very small. Does this suggest that RAIN is not important? Explain your answer.

(d) Which of the 3 models do you prefer as an explanation of the demand for holidays abroad and why?

PLEASE TURN OVER

5. Consider the following model of the banking sector where the variable of interest is the number of bank loans made to businesses.

Demand $Q_t = \alpha_0 + \alpha_1 R_t + \alpha_2 RD_t + \alpha_3 X_t + u_{1t}$

Supply $Q_t = \beta_0 + \beta_1 R_t + \beta_2 RI_t + \beta_3 Y_t + u_{2t}$

where

Q_t = total number of business loans
R_t = interest rate on business loans
RI_t = interest rate charged to non-commercial customers
RD_t = interest rate charged to commercial customers
X_t = business expectations
Y_t = total bank deposits

and Q_t, R_t are endogenous.

(a) Why would ordinary least squares (OLS) applied to these equations produce inconsistent parameter estimates?

(b) Explain carefully the concept and application of two-stage least squares (2SLS) and illustrate your answer using the above model. Would the resulting estimates have a smaller variance than the instrumental variable estimator using one of the exogenous variables?

6. (a) Explain what is meant by instrumental variable estimation. Show that instrumental variable estimates are consistent.

(b) Give examples of two areas where instrumental variable estimation is useful in econometrics.

7. Let a model be:

$$Y_t = \beta_0 + \beta_1 X_{t1} + \beta_2 X_{t2} + u_t \quad ; t = 1,2,\cdots,T$$

$E(u_t) = 0$ for all t. A researcher suspects that the variance of the disturbance term is $V(u_t) = \sigma^2 X_{t1}$.

(question continues on the next page)

(a) Explain how the researcher should proceed to test the null hypothesis $H_0 : V(u_t) = \sigma^2$ against the alternative hypothesis $H_1 : V(u_t) = \sigma^2 X_{t1}$, for all $t = 1, 2, \cdots, T$.

(b) If the researcher's suspicion is correct then how will it affect the properties of the ordinary least squares estimators?

(c) Suggest, in detail, an estimation procedure which will give best linear unbiased estimates of the parameters.

8. In order to investigate the effects of the 1970 equal pay act which came into force at the end of 1975, two models were estimated (i) the dependent variable FEMALE (female hourly earnings) was regressed against male hourly earnings (MALE) and other variables using data collected from 477 firms who employed both male and female workers and (ii) the dependent variable FMRATIO (the ratio of female to male hourly earnings) was regressed against other variables defined below.

Dependent variable	FEMALE	FMRATIO
Constant	22.84 (7.06)	0.714 (15.08)
AFTER	39.56 (9.84)	0.104 (5.60)
MALE	0.39 (15.19)	-
GNP80	-	-0.335 (1.45)
R^2	0.94	0.64
N	477	477

Figures in brackets are t values. AFTER is a dummy variable which takes the value 1 after 1975, 0 before. GNP80 is the level of GNP in 1980 prices.

(a) How would you test the hypothesis that males and females had equal earnings? From these results does it look like as if the equal pay act was successful?

(b) What is the interpretation of the coefficient of GNP80?

(c) What econometric problems might a researcher encounter with cross-section data of this form and how should the researcher proceed?

END OF PAPER

Examiner's report 2006

Zone A

General remarks

The paper is divided into two sections. Section A is compulsory and contains eight questions which are intended to examine across the whole syllabus. Section B is designed to examine a selection of topics in greater depth.

Candidates need to know the basic theory of econometrics, the technique of ordinary least squares (OLS), and the properties of the OLS estimators together with a good grasp of the effects of the standard OLS assumptions not being satisfied. In addition, candidates will need to be able to interpret the results of OLS estimation.

Just as last year, it appears that some candidates are confused about the difference between sample and population values when discussing covariance values – candidates would be well advised to make sure that they understand this distinction, perhaps even using a notation which distinguishes between the two cases, for example, using 'côv' to denote sample covariance values and 'cov' to denote population values. Also, just as last year, many candidates did not appear to read questions carefully enough and often omitted to give answers to parts of questions which ask for details, such as, the assumptions necessary for a particular result to be true. This loses marks.

Specific comments on questions

Section A
Question 1

a. An estimator of β is said to be consistent if $p\lim(\hat{\beta}) = \beta$. Operationally if the estimator is unbiased (or asymptotically unbiased) and the variance tends to zero as the sample size increases then the estimator is consistent.

$$\hat{\beta}_1 = \frac{\operatorname{cov}(x, y)}{\operatorname{var}(x)} = \frac{\operatorname{cov}(x, \beta x + u)}{\operatorname{var}(x)} = \beta + \frac{\operatorname{cov}(x, u)}{\operatorname{var}(x)}$$

to determine consistency take plim to give

$$p\lim(\hat{\beta}_1) = p\lim\left(\frac{\operatorname{cov}(x, u)}{\operatorname{var}(x)}\right) = \frac{p\lim(\operatorname{cov}(x, u))}{p\lim(\operatorname{var}(x))} = \frac{0}{\sigma_x^2}$$

assuming Gauss-Markov assumptions hold. Hence $\hat{\beta}_1$ is a consistent estimator of β_1.

b. Simultaneous equation bias occurs because the RHS endogenous variables are correlated with the error term. This correlation causes OLS parameter estimates to be biased. The reduced form does not suffer from this problem because the reduced form is created from the structural form by substituting out all RHS endogenous variables with exogenous (or lagged endogenous).

To answer the question the best route to take is to specify simple simultaneous equation model and show that the RHS endogenous variable is correlated with the error term.

c. $var(X+Y) = E(X+Y-E(X+Y))^2 = E((X-E(X)) + (Y-E(Y))^2 = E(X-E(X))^2 + E(Y-E(Y))^2 + 2E(X-E(X))(Y-E(Y)) = var(X) + var(Y) + 2cov(X,Y)$

similarly $var(X-Y) = var(X) + var(Y) - 2cov(X,Y)$.

Note that the question is phrased in terms of expectations so that answers which use sample values are wrong.

d. i. significant t tests but insignificant F test is not possible. In the 2 variable case the F statistic is the square of the t statistic – both statistics will tell the same story.

ii. Insignificant t tests but significant F test would suggest multicollinearity – correlation between variables leads to larger s.e's and hence low t values. The F test is not affected.

iii. insignificant t tests and insignificant F statistic tell the same story – the model is not correct.

e. $$\tilde{\beta} = \frac{1}{N}\sum_{t=1}^{N}\frac{y_t}{x_t}; \; y_t = \beta x_t + u_t \qquad ; \; E(u_t) = 0 \quad ; E(u_t^2) = \sigma^2$$

$$E\tilde{\beta} = \frac{1}{N}\sum_{t=1}^{N}\frac{E(y_t)}{x_t} = \frac{1}{N}\sum_{t=1}^{N}\frac{\beta x_t}{x_t} = \frac{N\beta}{\beta} = \beta \Rightarrow \text{unbiased}$$

$$V(\tilde{\beta}) = V\left[\frac{1}{N}\sum_{t=1}^{N}\frac{y_t}{x_t}\right] = \frac{1}{N^2}\sum_{t=1}^{N}\frac{V(y_t)}{x_t^2} \quad ; \text{Assuming } E(u_s u_t) = 0 \text{ if } s \neq t$$

$$= \frac{1}{N^2}\sum_{t=1}^{N}\frac{V(u_t)}{x_t^2} = \frac{\sigma^2}{N^2}\sum_{t=1}^{N}\frac{1}{x_t^2}$$

Extra assumptions used:

(i) $E(u_s u_t) = 0$ if $s \neq t$

(ii) x's are non-stochastic.

f. $Yt = \beta_0 + \beta_1 X_t + u_t; \; u_t = \rho u_{t-1} + v_t$ is a model with first order serial correlation. The OLS estimate of β_1 is

$$\frac{cov(X,Y)}{var(X)} = \frac{cov(X, \beta_0 + \beta_1 X_t + u_t)}{var(X_t)} = \beta_1 + \frac{cov(X,u)}{var(X)}$$

hence $E(\hat{\beta}_1) = \beta_1 + E\left(\frac{cov(X,u)}{var(X)}\right) = \beta_1$ since $E(u) = 0$ and X is non-stochastic.

$var(\hat{\beta}_1) \neq \dfrac{\sigma^2}{\sum x_t^2}$ (which is the OLS estimate) hence standard errors and t values are not correct.

g. The likelihood function is defined by

$$L = \prod_{i=1}^{n}\frac{1}{\sigma\sqrt{2\pi}}e^{-(x_i-\mu)^2/2\sigma^2} = \left(2\pi\sigma^2\right)^{-n/2}e^{-\sum(x_i-\mu)^2/2\sigma^2}$$

hence, taking logs, we get $\ln L = \dfrac{n}{2}\ln\left(2\pi\sigma^2\right) - \dfrac{1}{2\sigma^2}\sum_{i=1}^{n}\left(x_i - \mu\right)^2$

To maximise differentiate w.r.t μ and set the derivative = 0

$$\frac{\partial \ln L}{\partial \mu} = -\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu) = 0$$

gives $\tilde{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$

h. The OLS estimate of β_0 is obtained by differentiating the residual sum of squares by β_0 and putting the derivative =0.

$$\text{RSS} = \sum_{i=1}^{n}(Y_t - \hat{\beta}_0 - \hat{\beta}_1 X_t)^2$$

Hence $\frac{\partial \text{RSS}}{\partial \beta_0} = -2\sum_{i=1}^{n}(Y_t - \hat{\beta}_0 - \hat{\beta}_1 X_t)$

and $\frac{\partial \text{RSS}}{\partial \beta_0} = 0$ gives $\sum_{t=1}^{n} Y_t = \hat{\beta}_0 + \hat{\beta}_1 \sum_{t=1}^{n} X_t$ or, dividing through by n

gives $\overline{Y} = \hat{\beta}_0 + \hat{\beta}_1 \overline{X}$.

If $\beta_0 = 0$ then $\overline{Y} = \hat{\beta}_1 \overline{X}$ and the line passes through the origin.

Question 2

a. The adaptive expectations model is covered in Dougherty section 12.3. Short term relationships are described by parameter values on current variables, that is, the coefficient of X_t where Y_t is a function of X_t and lagged values. The long term relationship is measured by putting in equilibrium values for all variables to find the relationship between the equilibrium values of X and Y. The procedure is described in Dougherty, p32.

b. Expectations are formed by $\ln(X_t^*) - (1-\gamma)\ln(X_{t-1}^*) = \gamma\ln(X_{t-1})$

To get an estimable equation we need to eliminate the expected values from the equation. Multiply through by (1-γ) and lag to give

$$(1-\gamma)\ln(Y_{t-1}) = (1-\gamma)\alpha + (1-\gamma)\beta\ln(X_{t-1}^*) + (1-\gamma)u_{t-1}$$

Now subtract from the original equation

$$\ln(Y_t) - (1-\gamma)\ln Y_{t-1} = \alpha\gamma + \beta(\ln(X_t^*) - (1-\gamma)\ln(X_{t-1}^*)) + (u_t - (1-\gamma)u_{t-1})$$

$$\ln(Y_t) - (1-\gamma)\ln Y_{t-1} = \alpha\gamma + \beta(\gamma\ln(X_{t-1})) + (u_t - (1-\gamma)u_{t-1})$$

or

$$\ln(Y_t) = \alpha\gamma + \beta\gamma\ln(X_{t-1}) + (1-\gamma)\ln Y_{t-1} + (u_t - (1-\gamma)u_{t-1})$$

Parameters are estimated by non-linear techniques. If these procedures are not available then a grid search can be used where g is given values between 0 and 1 in steps of 0.1 and the remaining parameters estimated by OLS (see Dougherty p319 for details).

Question 3

a. A (weakly) stationary time series has population mean and population variance independent of time. A non-stationary time series, therefore, is a time series where either the mean and/or variance does depend on time. An example of a non-stationary time series is a random walk $y_t = y_{t-1} + u_t$, where u_t is a standard (white noise) error term. The random walk can be written as $y_t = y_0 + u_t + u_{t-1} + u_{t-2} + \ldots + u_1$ which has mean 0 and variance $t\sigma^2$ where σ^2 is the variance of u_t. As $t \to \infty$ the variance also $\to \infty$, that is, it depends on time.

A random walk with drift is defined by $y_t = \beta_1 + y_{t-1} + u_t$. If the series starts at y_0 then $y_t = y_0 + \beta_1 t + u_t + u_{t-1} + \ldots + u_0$ so the mean is now also a function of time ($\beta_1 t$).

b. The standard test for a unit root is due to Dickey and Fuller and is based on the model $y_t = \beta_1 + \beta_2 y_{t-1} + \gamma t + u_t$ which can be re-written as $\Delta y_t = \beta_1 + (1-\beta_2)y_{t-1} + \gamma t + u_t$ where $\Delta y_t = y_t - y_{t-1}$. The null hypothesis for stationarity is H_0: $1-\beta_2 = 0$, H_A: $1-\beta_2 \neq 0$. We cannot use the standard t-test procedure in this case because the distribution of the t-statistic is not a t-distribution so critical values have been computed by Dickey and Fuller using Monte Carlo techniques.

   The test is sensitive to the presence of serial correlation in the error term so we need to take steps to remove the effects of this serial correlation – this is done by including lagged values of y_t in the regression, that is, $y_t = \beta_1 + \beta_2 y_{t-1} + \beta_3 y_{t-2} + \gamma t + u_t$ for an AR(1) serial correlation. This is more easily tested by using the model $\Delta y_t = \beta_1 + (1-\beta_2-\beta_3)y_{t-1} - \beta_3 \Delta y_{t-2} + \gamma t + u_t$ with null hypothesis H_0**:** $1-\beta_1-\beta_2 = 0$. Once again using Dickey–Fuller tables.

c. Spurious regression was first demonstrated by Granger and Newbold who showed, using Monte Carlo techniques, that a regression involving two non-stationary series could give rise to spurious results in that the t-statistics over-rejected the null hypothesis of a zero coefficient for two independent random walk series. Since the spurious regression result applies if a simple regression of y_t on xt has non-stationary residuals when y_t and x_t are themselves non-stationary, the Dickey–Fuller test is of high importance. We first test to see if the series are non-stationary (using Dickey–Fuller) then we test the residuals from this regression to see if that was non-stationary. If the residuals are non-stationary then the relationship is not co-integrating and the results are spurious.

Question 4

a. The Durbin–Watson test is a test for the breakdown of the Gauss–Markov assumption that $E(u_t u_s) = 0,\ t \neq s$

   This condition is termed serial correlation or autocorrelation – this often arises from the omission of relevant variables.

   Correlation between u_t and u_{t-k} is called autocorrelation of order k.

   If we have n observations there are n-1 possible autocorrelations – this is too complicated to model – we try to simplify. The usual model involves just first order autocorrlation, that is,

   $u_t = \rho\, u_{t-1} + v_t$; where $E(v_t)=0$ and $E(v_t^2) = \sigma_v^2$, $E(v_t v_s) = 0$ for $t \neq s$.

   The simplest test is one where errors u_t and u_{t-1} have correlation ρ.

   H_0: $\rho = 0$

   H_A: $\rho \neq 0$

   Durbin Watson Test

$$dw = \frac{\sum_{t=2}^{n}(\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^{n}\hat{u}_t^2} = \frac{\sum_{t=2}^{n}\hat{u}_t^2 + \sum_{t=2}^{n}\hat{u}_{t-1}^2 - 2\sum_{t=2}^{n}\hat{u}_t\hat{u}_{t-1}}{\sum_{t=1}^{n}\hat{u}_t^2} \cong 2(1-\hat{\rho})$$

$\rho = +1 \quad dw = 0$

$\rho = -1 \quad dw = 4$

$\rho = 0 \quad dw = 2$

the distribution of dw depends on the values of explanatory variables hence exact limits are not calculable so tables give two limiting values d_u and d_l.

Problems with DW test.

1. test is only for 1st order serial correlation
2. test is inconclusive if dw lies between dl and du
3. test cannot be applied with lagged endogenous variables
4. test cannot be applied if equation has no constant term.

b. i. Standard t test t = (0.72 – 1)/0.1 = -2.8

with 21-2 df we have t* = 2.093 hence we reject H_0: β = 0.

If you wanted to be correct you should really do a one-tail test since MPC>1 is not a realistic option in which case t* = 1.729 and we reject H_0 again.

ii. The Cochrane–Orcutt procedure is described in Dougherty, p346.

iii. The important thing to notice here is the distribution of the errors from regression (A) which has a classic serial correlation pattern – in general positive values are followed by positive values, negative values by negative values. The errors are serially correlated hence Cochrane–Orcutt should give improved estimates by removing the autocorrelation. We can see that this was substantial since ρ = 0.94 and the dw statistic is now large enough to accept H_0: ρ = 0. The critical values are d_l = 0.97, d_u = 1.16 although this step is not necessary (or valid).

Note, however, that the equation is probably mis-specified. The years in which the US was involved in the Second World War (1941–1945) are characterised by large negative errors hence there was a change in structure for these years. This could have been modelled by including a dummy variable with a value of 1 for the years 1941–1945, 0 otherwise. Using Cochrane–Orcutt instead of the dummy variable will not completely solve the serial correlation problem because the technique only deals with first order serial correlation and the missing dummy variable will be more complex than this.

Question 5

a. The variable age has a +ve sign as expected – income increases with age and since the income/age relationship is not linear a quadratic term is also expected (income eventually drops after retirement). The signs on S and S^2 are not perhaps as expected: S can be thought of as measuring work experience so a –ve sign is counter-intuitive but might be due to the fact that individuals with higher qualifications are likely to be earning more and have less years in the job market so a –ve sign is appropriate (the S^2 sign is not predictable).

b. R^2 are low (usual in cross-section models) but t and F tests for significance of parameter estimates all show significant coefficients hence the hypothesis of no relationship is rejected so there is no cause for concern.

c. Omitted variable bias. Model (ii) is an improvement on (i) hence you would expect omitted variable bias and since age and age^2 are positively related and the coefficient on age^2 is expected to be –ve the bias will be –ve. A good answer would show the derivation of the omitted variable bias.

d. $F=\frac{\left(R_U^2-R_R^2\right)/m}{\left(1-R_U^2\right)/(N-k)}=\frac{(0.11-0.05)/2}{(1-0.11)/(3866-5)}=130.15$ and $F^*_{2,3861}=19.5$

Hence we reject H_0: $\beta_3 = \beta_4 = 0$.

e. The Goldfeldt–Quandt test is described on pages 227–228 of Dougherty. A good answer would note that the data should be ordered by a heteroskedastic variable (age in this case). The Goldfeldt–Quandt test was suggested because cross-section data on variables like income (which has wide spread) is characterised by heteroskedasticity. The consequence of heteroskedasticity is that t tests and F tests are invalid under OLS.

Question 6

a. The logistic model and the linear probability model are fully described in sections 11.1 and 11.2 of Dougherty. The important point is that the predicted value from a linear probability model can be interpreted as a probability but this can lie outside the interval [0,1]. The logit model cannot have this problem.

b. i. Asymptotic t values show that the coefficients on K5, Age, Wc, Lwg and Inc are all significantly different from 0. One would expect K5, K618, Inc and, probably, Age to have a negative effect on the probability that the wife works. One would also expect that Wc and Lwg would be positive. The sign of Hc is unpredictable. The significant coefficients all follow the expected signs.

ii. The 'goodness of fit' can be assessed by pseudo-R^2, the number of outcomes correctly predicted, the sum of squared residuals and the correlation between the outcomes and predicted values. Descriptions of these measures can be found on pages 287–288 of Dougherty. Since all these measures have some disadvantages, particularly when dealing with events with low probability, it is sensible to use more than one measure.

Question 7

a. The OLS estimator is derived by minimising the sum of squares of errors, that is, $I=\sum_{t=1}^{T}\hat{u}_t^2=\sum_{t=1}^{T}\left(y_t-\hat{\beta}x_t\right)^2$. To achieve the minimum differentiate w.r.t $\hat{\beta}$ and set the derivative equal to 0 and solve, that is,

$$\frac{\partial I}{\partial\hat{\beta}}=-2\sum\left(y_t-\hat{\beta}x_t\right)x_t=0 \text{ which has the solution } \hat{\beta}=\frac{\sum_{t=1}^{T}x_t y_t}{\sum_{t=1}^{T}x_t^2}$$

b. Unbiased estimator of σ^2 is $\frac{\sum_{t=1}^{T}\hat{u}_t^2}{T-k}$ (in this case k = 1).

Where $\hat{u}_t = y_t-\hat{\beta}x_t$ where $\hat{\beta}$ is the OLS estimator of β and $\sum_{t=1}^{T}\hat{u}_t^2$ = sum of squares of residuals.

Candidates are required to explain these. Note that the divisor is T-k rather than T (see Dougherty, p82 for an explanation).

c. OLS results follow $E(\hat{\beta})=\beta$ and $V(\hat{\beta})=\frac{\sigma^2}{\sum x_t^2}$

It follows that $\lim_{T\to\infty}V(\hat{\beta})\to 0$ since as the sample size increases $\sum x_t^2$ must increase.

Hence a sufficient condition for consistency holds $\Rightarrow \hat{\beta}$ is a consistent estimator of β.

Question 8

a. Only Order condition of identification is required.

Order Condition: $R \geq G - 1$, where R is the number of restrictions imposed on the equation under consideration and G is the number of jointly dependent variables or total number of equations in the complete model.

1st equation:

$R = 0$ and $G - 1 = 2 - 1 = 1 \Rightarrow R < G - 1$. Equation is under-identified.

2nd equation:

$R = 1$ and $G - 1 = 2 - 1 = 1 \Rightarrow R - G - 1$. Equation is exactly identified.

b. Our model is:

$Q_t = \beta_0 + \beta_1 P_t + \beta_2 Y_t + u_t$ (i) Demand equation

$Q_t = \alpha_0 + \alpha_1 P_t + e_t$ (ii) Supply equation

$t = 1,2,\cdots,T$. T is the sample size. u and e = disturbance terms

$E(u_t) = E(e_t) = 0$; $E(u_t^2) = \sigma_u^2$; $E(e_t^2) = \sigma_e^2$;

$E(u_t e_t) = \sigma_{ue}$; $E(u_t e_s) = 0$; $s, t = 1,2\cdots,T$

Q and P are endogenous variables.

In 2SLS we manipulate the equation in such a way that we write in the RHS, instead of the endogenous variable the estimated value of it (in this case). After that we apply OLS to the equation. Taking deviation from the mean we can write equations (i) and (ii) as

$q_t = \beta_1 p_t + \beta_2 y_t + u_t$ (iii)

$q_t = \alpha_1 p_t + e_t$ (iv)

We want to estimate α_2. Estimation involves two stages.

First Stage:

Write down the reduced form (RF) corresponding to the RHS endogenous variable, in this case p. Let it be

$P_t = \Pi_1 y_t + v_t$ (v)

As RF parameters can always be consistently estimated by the OLS, apply OLS to (v) to get

$$\hat{p}_t = \hat{\Pi}_1 y_t \quad ; \hat{\Pi}_1 = \frac{\sum_{t=1}^{T} p_t y_t}{\sum_{t=1}^{T} y_t^2} \qquad \text{(vi)}$$

We can also write pt as

$p_t = \hat{p}_t + \hat{v}_t$; where $\hat{v}_t$ is the OLS estimate of v_t (vii)

Second Stage:

In the second stage instead of P_t write $\hat{p}_t + \hat{v}_t$ in (iv) to get

$q_t = \alpha_1 \hat{p}_t + \alpha_2 \hat{v}_t + e_t$ (viii)

Apply OLS to (viii) to get 2SLS estimator of α_2 as

$$\hat{\alpha}_{1,2SLS} = \frac{\sum_{t=1}^{T} \hat{p}_t q_t}{\sum_{t=1}^{T} \hat{p}_t^2}$$

$$= \frac{\hat{\Pi}_1 \sum_{t=1}^{T} q_t y_t}{\hat{\Pi}_1^2 \sum_{t=1}^{T} y_t^2} \quad ; \hat{p}_t = \hat{\Pi}_1 y_t \text{ from (vi)}$$

$$= \frac{\sum_{t=1}^{T} q_t y_t}{\sum_{t=1}^{T} y_t^2} \frac{\sum_{t=1}^{T} y_t^2}{\sum_{t=1}^{T} p_t y_t} \quad ; \hat{\Pi}_1 = \frac{\sum_{t=1}^{T} p_t y_t}{\sum_{t=1}^{T} y_t^2} \quad \text{from (vi)}$$

$$= \frac{\sum_{t=1}^{T} q_t y_t}{\sum_{t=1}^{T} p_t y_t} \quad \text{2SLS estimator of } \alpha_2 \qquad \text{(ix)}$$

Consistency of the 2SLS estimator:

$$\hat{\alpha}_{1,2SLS} = \frac{\sum_{t=1}^{T} q_t y_t}{\sum_{t=1}^{T} p_t y_t}$$

$$= \frac{\sum_{t=1}^{T} y_t (\alpha_2 p_t + e_t)}{\sum_{t=1}^{T} p_t y_t}$$

$$= \alpha_1 \frac{\sum_{t=1}^{T} p_t y_t}{\sum_{t=1}^{T} p_t y_t} + \frac{\sum_{t=1}^{T} y_t e_t}{\sum_{t=1}^{T} p_t y_t}$$

$$= \alpha_2 + \frac{\sum_{t=1}^{T} y_t e_t}{\sum_{t=1}^{T} p_t y_t}$$

$$\text{plim}(\hat{\alpha}_{1,2SLS}) = \alpha_1 + \frac{\text{plim} \frac{1}{T} \sum_{t=1}^{T} y_t e_t}{\text{plim} \frac{1}{T} \sum_{t=1}^{T} p_t y_t}$$

$$= \alpha_1 + \frac{\text{Cov}(y_t, e_t)}{\text{Cov}(p_t, y_t)}$$

$= \alpha_1$; $\text{Cov}(y_t, e_t) = 0$ by definition, as y_t is exogenous.

c. ILS and 2SLS estimators are the same as the equation under consideration is exactly identified.

Examination paper for 2007

There will be no change to the format, style or number of questions in the examination paper for 2007.

Examiners' commentary 2008

20 Elements of econometrics

Specific comments on questions Zone A

Section A

Answer all **eight** parts of Question 1 (5 marks each, 40 marks in total).

Question 1

a)

[Note: Concept of **sufficient condition for consistency** has to be used **(Ref: Dougherty, C. *Introduction to Econometrics* (third edition) Ch. R.8)** combined with simple calculation of expectation and variance.]

$$\mathrm{E}[\hat{\mu}_1] = \mathrm{E}\left[\frac{\sum_{t=1}^{T} X_t}{T}\right] = \frac{T\mu}{T} = \mu \Rightarrow \text{unbiased.}$$

$$\mathrm{V}(\hat{\mu}_1) = \frac{\sum_{t=1}^{T} \mathrm{V}(X_t)}{T^2} = \frac{T\sigma^2}{T^2} = \frac{\sigma^2}{T},$$

hence $\lim_{T\to\infty} \mathrm{V}(\hat{\mu}_1) \to 0$. Sufficient condition for consistency holds. $\hat{\mu}_1$ is consistent.

$$\mathrm{E}[\hat{\mu}_2] = \mathrm{E}\left[\frac{\sum_{t=1}^{T-1} X_t}{T}\right] = \frac{(T-1)\mu}{T} = \mu - \frac{\mu}{T} \Rightarrow \text{biased,}$$

but $\lim_{T\to\infty} \mathrm{E}[\hat{\mu}_2] \to 0 \Rightarrow$ asymptotically unbiased.

$$\mathrm{V}(\hat{\mu}_2) = \frac{\sum_{t=1}^{T-1} \mathrm{V}(X_t)}{T^2} = \frac{(T-1)\sigma^2}{T^2} = \frac{\sigma^2}{T} - \frac{\sigma^2}{T^2},$$

hence $\lim_{T\to\infty} \mathrm{V}(\hat{\mu}_2) \to 0$. Sufficient condition for consistency holds. $\hat{\mu}_2$ is consistent.

b)

$$\hat{\alpha}_{\text{OLS}} = \frac{\sum x_t y_t}{\sum x_t^2} = \frac{\sum x_t(\alpha x_t + \beta z_t + \epsilon_t)}{\sum x_t^2} = \alpha + \frac{\beta \sum x_t z_t}{\sum x_t^2} + \frac{\sum x_t \epsilon_t}{\sum x_t^2}.$$

$$\mathrm{E}[\hat{\alpha}_{\text{OLS}}] = \alpha + \beta \mathrm{E}\left[\frac{\sum x_t z_t}{\sum x_t^2}\right].$$

If x_t and z_t are **not orthogonal** to each other then $\hat{\alpha}_{\text{OLS}}$ is a biased estimator of α. Bias is $\beta \sum x_t z_t / \sum x_t^2$. Direction of the bias will depend upon the signs of β and $\sum x_t z_t$. Bias will disappear if x_t and z_t are **orthogonal** to each other.

(Ref: Dougherty, C. *Introduction to Econometrics* (third edition) Ch. 6.2]

c)

β_1 is the intercept. If X is zero, then Y will take value β_1. β_2 is the slope. If there is a one unit increase in X, Y will increase by β_2 units.

Assumptions:

i. $\mathrm{E}[u_t] = 0$. Required for unbiasedness.
ii. $\mathrm{E}[u_t^2] = \sigma^2 \Rightarrow$ Homoscedasticity.
iii. $\mathrm{E}[u_t u_s] = 0 \; \forall \; s \neq t \Rightarrow$ No autocorrelation.
iv. X's are fixed in repeated samples.
v. Number of observations is greater than or equal to the number of parameters being estimated.

Assumptions i.–iv. are required to show that the OLS estimator is BLUE. Assumption v. is required to obtain the estimates. If the number of observations is less than the number of parameters being estimated then estimates cannot be obtained.

(Ref: Dougherty, C. *Introduction to Econometrics* (third edition) Ch. 2.2]

d)

Disturbance term is heteroscedastic, hence OLS estimators will be inefficient. Use Weighted Least Squares (WLS). Transform the model by dividing by X_t.

$$\begin{aligned} \frac{Y_t}{X_t} &= \frac{\alpha}{X_t} + \beta + \frac{u_t}{X_t}, \\ \mathrm{E}\left[\frac{u_t}{X_t}\right] &= 0, \\ \mathrm{V}\left(\frac{u_t}{X_t}\right) &= \frac{\sigma^2 X_t^2}{X_t^2} = \sigma^2, \\ \mathrm{E}\left[\frac{u_s u_t}{X_s X_t}\right] &= 0. \end{aligned}$$

The transformed disturbance term in (.17) has all the assumptions required for OLS estimators to be BLUE. Hence applying OLS to (.17), $\hat{\alpha}$ and $\hat{\beta}$ can be obtained which will be BLUE. These are WLS estimators.

(Ref: Dougherty, C. *Introduction to Econometrics* (third edition) Ch. 7.3]

e)

If after removing the trend from a non-stationary series the resulting variable becomes stationary, then the variable is called **trend-stationary**. Let

$$Z_t = X_t - \alpha_1 t = \alpha_0 + u_t,$$

where $\mathrm{E}[u_t] = 0$, $\mathrm{Var}(u_t) = \sigma^2$ and $\mathrm{E}[u_t u_{t-s}] = 0 \; \forall \; s \neq t$. Then,

$$\begin{aligned} \mathrm{E}[Z_t] &= \mathrm{E}[\alpha_0 + u_t] = \alpha_0 \\ \mathrm{Var}(Z_t) &= \mathrm{Var}(\alpha_0 + u_t) = \sigma^2 \\ \mathrm{Cov}(Z_t, Z_{t-s}) &= \mathrm{E}[Z_t - \mathrm{E}[Z_t]][Z_{t-s} - \mathrm{E}[Z_{t-s}]] \\ &= \mathrm{E}[u_t u_{t-s}] = 0. \end{aligned}$$

This means that Z_t has constant mean and variance for all t, and covariance is zero for $s \neq 0$. It implies that the series is trend-stationary. If a non-stationary process can be transformed into a stationary process by differencing, then the series is said to be **difference-stationary**. Let X_t be a random walk with a drift,

$$X_t = \beta_0 + X_{t-1} + \epsilon_t$$

where $\mathrm{E}[\epsilon_t] = 0$, $\mathrm{Var}(\epsilon_t) = \sigma^2$ and $\mathrm{E}[\epsilon_t \epsilon_s] = 0 \; \forall \; s \neq t$. Subtract X_{t-1} from both sides of (.26) to get

$$\Delta X_t = X_t - X_{t-1} = \beta_0 + \epsilon_t.$$

It can easily be checked that $\mathrm{E}[\Delta X_t] = \beta_0$, $\mathrm{Var}(\Delta X_t) = \sigma_\epsilon^2 = \sigma^2$ and $\mathrm{Cov}(\Delta X_t, \Delta X_{t-1}) = 0 \; \forall \; s \neq t$. This means that ΔX_t is stationary. This implies that X_t is difference-stationary.

It is important to know whether a variable is difference- or trend-stationary because for difference-stationary variables shocks have a permanent effect whereas for trend-stationary variables shocks are transitory.

(Ref: Dougherty, C. *Introduction to Econometrics* (third edition) Ch. 13.1]

f)

Consider the model

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

If X_t is not independently distributed of u_t, then the OLS estimator of β will be inconsistent. Consider a variable Z that is correlated with X but not correlated with u. Z can be considered as an instrumental variable. An estimator of β based on Z is known as an instrumental variable (IV) estimator. It is defined as

$$\hat{\beta}_{\text{IV}} = \frac{\sum Z_t Y_t}{\sum Z_t X_t}.$$

It can be shown that $\hat{\beta}_{\text{IV}}$ is a consistent estimator of β.

$$\hat{\beta}_{\text{IV}} = \frac{\sum Z_t Y_t}{\sum Z_t X_t} = \frac{\sum Z_t(\beta X_t + u_t)}{\sum Z_t X_t} = \beta + \frac{\sum Z_t u_t}{\sum Z_t X_t}.$$

$$\text{plim}(\hat{\beta}_{\text{IV}}) = \beta + \frac{\text{plim}\left(\sum Z_t u_t/T\right)}{\text{plim}\left(\sum Z_t X_t/T\right)} = \beta \Rightarrow \hat{\beta}_{\text{IV}}$$

is a consistent estimator of β.

[Note: plim $(\sum Z_t u_t/T) = 0$ and plim $(\sum Z_t X_t/T) \neq 0$.]

(Ref: Dougherty, C. *Introduction to Econometrics* (third edition) Ch. 8.6]

g)

We transform the equation into an estimable form by taking logs to give:

$$\log\left(\frac{M_t}{Y_t}\right) = \log(\alpha) + \beta\log(r_t),$$

i.e. a log-log model. Now we can estimate the parameters by using OLS — you would be awarded extra marks if the specification of the error term was discussed at this point.

The estimate of α is given by exp(constant) and the slope parameter is the estimate of β.

OR

Use a non-linear estimation process.

(Ref: Dougherty, C. *Introduction to Econometrics* (third edition) Chs. 4.2, 4.3 and 4.4]

h)

$$\hat{\alpha} = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{y_t}{x_t}\right) = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{\alpha x_t + u_t}{x_t}\right) = \alpha + \frac{1}{T}\sum_{t=1}^{T}\left(\frac{u_t}{x_t}\right),$$

$$\text{hence } \mathrm{E}[\hat{\alpha}] = \alpha + \frac{1}{T}\sum_{t=1}^{T}\frac{\mathrm{E}[u_t]}{x_t} = \alpha \Rightarrow \text{unbiased}.$$

$$\mathrm{Var}(\hat{\alpha}) = \mathrm{E}[\hat{\alpha} - \alpha]^2 = \mathrm{E}\left(\frac{1}{T}\sum_{t=1}^{T}\left(\frac{u_t}{x_t}\right)\right)^2 = \sigma^2\frac{1}{T^2}\sum_{t=1}^{T}\left(\frac{1}{x_t^2}\right)$$

which will tend to zero as $T \to \infty$.

As the estimator is unbiased and also the variance of the estimator tends to zero as $T \to \infty$, the sufficient condition of consistency holds, hence the estimator is consistent. The given estimator is not efficient because it is not the ordinary least squares estimator.

(Ref: Dougherty, C. *Introduction to Econometrics* (third edition) Chs. R.6 and R.7]

Section B

Section B requires candidates to answer three questions in this section (20 marks each, 60 marks in total).

Question 2

a)

This should be answered explaining the Dickey-Fuller (DF) test. Candidates also need to explain that as this test is sensitive to the presence of serial correlation in the error term, we need to take steps to remove the effects of this serial correlation. This is done by including lagged values of the dependent variable in the regression. This development is known as the augmented Dickey-Fuller (ADF) test. Standard t-test procedure in this case is not valid because the distribution of the t-statistic is not a t-distribution, so critical values have been computed by Dickey and Fuller using Monte Carlo techniques. **To answer this question technical details should be given.**

(Ref: Dougherty, C. *Introduction to Econometrics* (third edition) Ch. 13.4]

b)

In an adaptive expectations model, the actual value of the variable is compared with the expected value. If the actual value is greater, the expected value is adjusted upwards for the next period. **To answer this question technical details should be given. Process of estimation should be given.**

(Ref: Dougherty, C. *Introduction to Econometrics* (third edition) Ch. 11.3]

c)

To answer this question the definition of dummy variables should be given and also how dummy variables can be used to test for a change in the intercept or a change in the slope. How change in the intercept or slope and also the joint test for change in the intercept and slope can be conducted should be explained. Dummy variable trap should be explained. **Technical details should be given.**

(Ref: Dougherty, C. *Introduction to Econometrics* (third edition) Ch. 5]

d)

Discuss the motivation for using the Chow test of predictive failure. Explain the technique itself, namely estimate the regression over a long sub-period (incorporating a large amount of the data) and then using the estimated coefficients to predict y values in the other period. Then proceed to compare the predictions with the actual values. **Technical details should be given.**

(Ref: Dougherty, C. *Introduction to Econometrics* (third edition) Ch. 11.6]

Question 3

a)

In general a linear combination of two time series will be non-stationary if at least one of them is non-stationary. The degree of integration of the combination will be equal to that of the most highly integrated individual series.

For example a combination of I(1) and I(0) series will be I(1) and a combination of I(1) and I(1) series will be I(1).

If a long-run relationship exists between the time series then the result may be different. Suppose Y_t and X_t are both I(1). A linear combination of Y_t and X_t may be written as $u_t = Y_t - \beta X_t$. If the linear combination u_t is I(0), then Y_t and X_t are said to be cointegrated.

If Y_t and X_t are cointegrated then it implies that a long-run relationship exists between Y_t and X_t. This concept can be generalised. Consider a general linear model

$$Y_t = \beta_1 + \beta_2 X_{2t} + \cdots + \beta_K X_{Kt} + u_t.$$

Then the disturbance term, u_t, can be thought of as measuring the deviation between components of the model. In the short-run the divergence between the components will fluctuate, but if the model is correctly specified there will be a limit to the divergence. Hence though Y's and X's are non-stationary, 'u' will be stationary. If there are K variables in the model, the maximum number of cointegrating relationships will be $K-1$.

Cointegration is an overriding requirement for any economic model using non-stationary time series data. If the variables do not cointegrate then we have a problem of spurious regression and econometric work becomes almost meaningless.

If a cointegrating relationship exists then OLS estimators can be shown to be superconsistent.

(Ref: Dougherty, C. *Introduction to Econometrics* (third edition) Ch. 13.4]

b)

$$\begin{aligned}
\mathrm{E}[y_t] &= \mathrm{E}[u_t + \theta u_{t-1}] = 0. \\
\mathrm{E}[y_t^2] &= \mathrm{E}[u_t^2] + \theta^2 \mathrm{E}[u_{t-1}^2] + 2\theta \mathrm{E}[u_t u_{t-1}] \\
&= (1+\theta^2)\sigma^2 \text{ since } \mathrm{E}[u_t u_{t-1}] = 0. \\
\mathrm{E}[y_t y_{t-1}] &= \mathrm{E}[(u_t + \theta u_{t-1})(u_{t-1} + \theta u_{t-2})] \\
&= \mathrm{E}[u_t u_{t-1}] + \theta \mathrm{E}[u_t u_{t-2}] \\
&\quad + \theta \mathrm{E}[u_{t-1}^2] + \theta^2 \mathrm{E}[u_{t-1} u_{t-2}] \\
&= \theta\sigma^2 \text{ since all terms in} \\
&\qquad \mathrm{E}[u_t u_{t-s}] = 0, s > 0. \\
\mathrm{E}[y_t y_{t-2}] &= \mathrm{E}[(u_t + \theta u_{t-1})(u_{t-2} + \theta u_{t-3})] \\
&= \mathrm{E}[u_t u_{t-2}] + \theta \mathrm{E}[u_t u_{t-3}] \\
&\quad + \theta \mathrm{E}[u_{t-1} u_{t-2}] + \theta^2 \mathrm{E}[u_{t-1} u_{t-3}] \\
&= 0.
\end{aligned}$$

Thus both first and second moments are independent of t and so y_t must be (weakly) stationary.

(Ref: Dougherty, C. *Introduction to Econometrics* (third edition) Ch. 13 and Dougherty, C. *Introduction to Econometrics* (2004), Ch. 14]

c)

Regression of this type is known as *spurious regression*. If we regress Y_t on X_t, i.e.

$$Y_t = \pi_0 + \pi_1 X_t + v_t,$$

Granger and Newbold have shown that although there is no relationship between Y and X, the regression will produce a t-ratio which **will reject the null hypothesis** $\mathrm{H}_0 : \pi_1 = 0$.

The reason for this result is that if $\mathrm{H}_0 : \pi_1 = 0$, then

$$Y_t = \pi_0 + v_t$$

and since Y_t is I(1) and π_0 is constant, it follows that v_t must be I(1). This violates the standard distributional theory based on the assumption that v_t is stationary, i.e. v_t is I(0). Hence the misleading result.

(Ref: Dougherty, C. *Introduction to Econometrics* (third edition) Ch. 13.2]

Question 4

a)

i. Likelihood function is

$$L = (2\pi\sigma^2)^{-T/2} \exp\left[-\frac{\sum(y_t - \beta x_t)^2}{2\sigma^2}\right].$$

(Ref: Dougherty, C. *Introduction to Econometrics* (third edition) Ch. 10.6]

ii.

$$\begin{aligned}
\log L &= -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log(\sigma^2) - \frac{\sum(y_t - \beta x_t)^2}{2\sigma^2}. \\
\frac{\partial \log L}{\partial \beta} &= -\frac{2\sum(y_t - \beta x_t)(-x_t)}{2\sigma^2} = 0. \\
\hat{\beta}_{\mathrm{MLE}} &= \frac{\sum x_t y_t}{T}. \\
\frac{\partial \log L}{\partial \sigma^2} &= -\frac{T}{2\sigma^2} + \frac{\sum(y_t - \beta x_t)^2}{2\sigma^4} = 0. \\
\hat{\sigma}^2_{\mathrm{MLE}} &= \frac{\sum(y_t - \hat{\beta}_{\mathrm{MLE}} x_t)^2}{T}.
\end{aligned}$$

(Ref: Dougherty, C. *Introduction to Econometrics* (third edition) Ch. 10.6]

b)

As the model has been estimated by maximum likelihood, F-test cannot be used. All the slope coefficients equal to zero can be tested using the likelihood ratio statistic $2(\log L - \log L_0)$. This is asymptotically distributed as a chi-square with $K - 1$ degrees of freedom. $K - 1$ is the number of explanatory variables in the model. In this case $K - 1 = 2$.

$\log L$ is the log of the unrestricted likelihood and $\log L_0$ is the log of the restricted likelihood. $\log L_0$ has been obtained with only the intercept in the regression.

(Ref: Dougherty, C. *Introduction to Econometrics* (third edition) Ch. 10.6]

Question 5

a)

i. Constant term is the same in both regions.

ii. The effect of income is the same in both regions.

iii. The effect of size is the same in both regions.

iv. The same model fits both regions, i.e. there is no difference between the North and the South.

(Ref: Dougherty, C. *Introduction to Econometrics* (third edition) Ch. 5]

b)

Testing for each of the hypotheses involves using the dummy variable REG and testing the parameter on the dummy variable for significance.

i. $EXP = \alpha + \gamma_1 INC + \gamma_2 INC \times REG + \delta_1 SIZE + \delta_2 SIZE \times REG + \epsilon$

ii. $EXP = (\alpha_1 + \alpha_2 REG) + \gamma_1 INC + \delta_1 SIZE + \delta_2 SIZE \times REG + \epsilon$

iii. $EXP = (\alpha_1 + \alpha_2 REG) + \gamma_1 INC + \gamma_2 INC \times REG + \delta_1 SIZE + \epsilon$

iv. $EXP = (\alpha_1 + \alpha_2 REG) + \gamma_1 INC + \gamma_2 INC \times REG + \delta_1 SIZE + \delta_2 SIZE \times REG + \epsilon$

(Ref: Dougherty, C. *Introduction to Econometrics* (third edition) Ch. 5]

c)

The F-test can be used to test restrictions. In iv. it can be done by estimating the overall regression of EXP on INC and $SIZE$ and comparing the resulting SSE with the sum of those obtained by running regressions on data from the North and data from the South.

The other hypotheses would be tested by using the dummy variable regressions in i., ii. and iii. given above. Estimate the SSE from the regressions with the dummy variable and compare these between models i., ii., iii. and iv. using the F-test on restricted and unrestricted SSEs.

(Ref: Dougherty, C. *Introduction to Econometrics* (third edition) Ch. 5]

Question 6

a)

Order condition: R $\geq$ G$-$1 where R is the number of restrictions imposed on the equation. G is the number of endogenous variables in the complete model which is also equal to the number of equations.

- R $>$ G$-$1 $\Rightarrow$ overidentified
- R $=$ G$-$1 $\Rightarrow$ exactly identified
- R $<$ G$-$1 $\Rightarrow$ underidentified.

(Ref: Dougherty, C. *Introduction to Econometrics* (third edition) Ch. 9]

b)

i. First equation R $= 1$, G$-1 = 2 - 1 = 1 \Rightarrow$ exactly identified.
Second equation R $= 0$, G$-1 = 2 - 1 = 1$. As R $<$ G-1, equation is underidentified.

ii.

$$\begin{aligned} y_{1t} &= \alpha y_{2t} + u_{1t} \\ y_{2t} &= \beta_1 x_t + \beta_2 y_{1t} + u_{2t}. \end{aligned}$$

Substituting (.58) into (.57), we get

$$y_{1t} = \alpha(\beta_1 x_t + \beta_2 y_{1t} + u_{2t}) + u_{1t} = \alpha\beta_1 x_t + \alpha\beta_2 y_{1t} + \alpha u_{2t} + u_{1t}$$

$$\text{or, } (1 - \alpha\beta_2) y_{1t} = \alpha\beta_1 x_t + \alpha u_{2t} + u_{1t}$$

$$\Rightarrow y_{1t} = \frac{\alpha\beta_1 x_t}{1-\alpha\beta_2} + \frac{\alpha u_{2t} + u_{1t}}{1-\alpha\beta_2}.$$

$$\text{Similarly, } y_{2t} = \frac{\beta_1 x_t}{1-\alpha\beta_2} + \frac{\beta_2 u_{1t} + u_{2t}}{1-\alpha\beta_2}.$$

(.61) and (.62) are reduced forms corresponding to y_{1t} and y_{2t} respectively. These will exist only when $\alpha\beta_2 \neq 1$. OLS estimator of α is

$$\hat{\alpha} = \frac{\sum y_{1t}y_{2t}}{\sum y_{2t}^2} = \frac{\sum y_{2t}(\alpha y_{2t} + u_{1t})}{\sum y_{2t}^2} = \alpha + \frac{\sum y_{2t}u_{1t}}{\sum y_{2t}^2}$$

$$\text{plim } \hat{\alpha} = \alpha + \frac{\text{plim}\frac{1}{T}\sum y_{2t}u_{1t}}{\text{plim}\frac{1}{T}\sum y_{2t}^2} = \alpha + \frac{\text{Cov}(y_{2t}, u_{1t})}{\text{Var}(y_{2t})}$$

$$\begin{aligned}
\text{Cov}(y_{2t}, u_{1t}) &= \text{E}[y_{2t} - \text{E}[y_{2t}]]u_{1t} \\
&= \text{E}\left[\frac{\beta_2 u_{1t} + u_{2t}}{1-\alpha\beta_2}\right] u_{1t} \text{ from (.62)} \\
&= \frac{\beta_2\sigma_1^2 + \sigma_{12}}{1-\alpha\beta_2} \neq 0.
\end{aligned}$$

$$\text{Hence, plim } \hat{\alpha} = \alpha + \left(\frac{\beta_2\sigma_1^2 + \sigma_{12}}{1-\alpha\beta_2}\right)/\text{Var}(y_{2t}) \neq \alpha \Rightarrow \text{ inconsistency.}$$

(Ref: Dougherty, C. *Introduction to Econometrics* (third edition) Ch. 9]

iii. Equation (.57) is exactly identified hence ILS or 2SLS can be used. Brief description of any method is required.

(Ref: Dougherty, C. *Introduction to Econometrics* (third edition) Ch. 9]

Question 7

a)

Model is:

$$Y_i = \beta_0 + \beta_1 X_i + u_i, \qquad i = 1, 2, \ldots, n$$

$$\begin{aligned}
Y_i &= 1 \text{ if the event takes place} \\
&= 0 \text{ otherwise}
\end{aligned}$$

We note that

$$\text{E}[Y_i|X_i] = \beta_0 + \beta_1 X_i$$

$$\text{Also E}[Y_i|X_i] = 1 \cdot \text{P}(Y_i = 1) + 0 \cdot \text{P}(Y_i = 0) = \text{P}(Y_i = 1) = P_i$$

From (.72) and (.73), $\text{E}[Y_i|X_i] = \beta_0 + \beta_1 X_i = P_i$, hence we can interpret $\text{E}[Y_i|X_i] = \beta_0 + \beta_1 X_i$ as the probability that the event will occur, given X_i. As Y_i takes only two values 1 or 0, therefore u_i can take only two values: $1 - \beta_0 - \beta_1 X_i$ when $Y_i = 1$ and $-\beta_0 - \beta_1 X_i$ when $Y_i = 0$. Based on this we can write the probability distribution of u_i as

Y_i	u_i	$f(u_i)$
1	$1 - \beta_0 - \beta_1 X_i$	$\beta_0 + \beta_1 X_i$
0	$-\beta_0 - \beta_1 X_i$	$1 - \beta_0 - \beta_1 X_i$

We can write $\mathrm{V}(u_i)$ as

$$\begin{aligned}
\mathrm{V}(u_i) &= \mathrm{E}[u_i^2] \\
&= (1-\beta_0-\beta_1 X_i)^2(\beta_0+\beta_1 X_i) \\
&\quad +(-\beta_0-\beta_1 X_i)^2(1-\beta_0-\beta_1 X_i) \\
&= (1-\beta_0-\beta_1 X_i)(\beta_0+\beta_1 X_i) \\
&\quad \times[(1-\beta_0-\beta_1 X_i)+(\beta_0+\beta_1 X_i)] \\
&= (\beta_0+\beta_1 X_i)(1-\beta_0-\beta_1 X_i) \\
&= \mathrm{E}[Y_i](1-\mathrm{E}[Y_i]) \\
&= P_i(1-P_i) \quad \forall\ i=1,2,\ldots,n.
\end{aligned}$$

Hence the disturbance term is heteroscedastic. This will make OLS estimators inefficient. Weighted least squares can be used to obtain efficient estimators of β_0 and β_1.

<u>Problem:</u>

i. As the distribution of the disturbance term only takes two values, it is not continuous. This implies that usual test statistics are invalid.

ii. In many cases the estimated probability $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$ will be negative or greater than 1.

(Ref: Dougherty, C. *Introduction to Econometrics* (third edition) Ch. 10.1]

b)

i.

$$\hat{Y}_i = -83.94 + 158.39(0.9) - 76(0.9)^2 + 1.16(2) + 0.93(1) = 0.301.$$

$$\hat{P}_i = \mathrm{P}(Y_i = 1) = \frac{\exp(\hat{Y}_i)}{1+\exp(\hat{Y}_i)} = \frac{\exp(0.301)}{1+\exp(0.301)} = 0.5746.$$

(Ref: Dougherty, C. *Introduction to Econometrics* (third edition) Ch. 10.2]

ii. As the model has been estimated by maximum likelihood, the F-test cannot be used. In these situations the likelihood ratio test is used. Likelihood ratio statistic is

$$LR = 2(\ln L - \ln L_0) = 2(-321.25 - (-416.01)) = 189.52.$$

Critical value of χ^2_6 at 5% level of significance is 12.592. Critical value of χ^2_6 at 1% level of significance is 16.812. Therefore reject H_0.

(Ref: Dougherty, C. *Introduction to Econometrics* (third edition) Ch. 10.6]

Question 8

a)

If $\hat{\theta}$, based on a sample of size T, is a consistent estimator of θ then

$$\text{Prob}\left(|\hat{\theta}-\theta| > \epsilon\right) \to 0$$

as $T \to \infty$ for every $\epsilon > 0$. Another way of expressing consistency is that $\hat{\theta}$ converges in probability to θ. In short we can write the above statement as plim $\hat{\theta} = \theta$; where plim stands for the probability limit. Hence if plim $\hat{\theta} = \theta$ then $\hat{\theta}$ is a consistent estimator of θ.

Sufficient conditions for consistency are:

$$\mathrm{E}[\hat{\theta}] = \theta \text{ or } \lim_{T\to\infty} \mathrm{E}[\hat{\theta}] = \theta \text{ and } \lim_{T\to\infty} \mathrm{V}(\hat{\theta}) \to 0.$$

The concept is useful if small sample properties (e.g. unbiasedness) do not hold. Consistency is often easier to prove than unbiasedness.

(Ref: Dougherty, C. *Introduction to Econometrics* (third edition) Ch. R.8]

b)

i. $P_t^* - (1-\lambda)P_{t-1}^* = \lambda p_{t-1}$, hence

$$\begin{aligned} Q_t - (1-\lambda)Q_{t-1} &= [\alpha - (1-\lambda)\alpha] \\ &\quad +\beta(P_t^* - (1-\lambda)P_{t-1}^*) \\ &\quad +\gamma(Z_t - (1-\lambda)Z_{t-1}) \\ &\quad +u_t - (1-\lambda)u_{t-1} \\ &= \alpha\lambda + \beta\lambda P_{t-1} + \\ &\quad \gamma Z_t - (1-\lambda)Z_{t-1} \\ &\quad +u_t - (1-\lambda)u_{t-1}. \end{aligned}$$

$$Q_t = \alpha\lambda + (1-\lambda)Q_{t-1} + \beta\lambda P_{t-1} + \gamma Z_t - (1-\lambda)Z_{t-1} + u_t - (1-\lambda)u_{t-1}.$$

Note that the error term is serially correlated.

ii. If serial correlation is present the OLS parameter estimates would be unbiased but inefficient. Their standard errors as estimated by OLS would be incorrect hence the t-test is invalid.

(Ref: Dougherty, C. *Introduction to Econometrics* (third edition) Ch.12.3]

iii. The equation in i. can be rewritten in terms of past P_t and past Z_t by back substitution to eliminate the lagged dependent variable Q_{t-1}. Now use non-linear least squares or a grid search to estimate the parameters.

(Ref: Dougherty, C. *Introduction to Econometrics* (third edition) Ch.11.3]

Examiners' commentaries 2009

20 Elements of econometrics

Specific comments on questions – Zone B

Section A

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

Question 1

(a) Possible uses of the F test are (i) a joint test of two (or more) parameter values in a regression equation, (ii) a Chow test of coefficient stability, (iii) goodness of fit of regression equation and (iv) Goldfeldt-Quandt test of homoskedasticity.

A good answer would include the model, null hypothesis, test statistic, degrees of freedom and distribution of the test statistic.

[See Dougherty, 2007, Chapters 3.5, 6.5, 7.2, 11.6.]

(b) i. Minimising the sum of squares of errors under the assumptions of zero mean for the error term and non-stochastic independent variable which yields an unbiased parameter estimate. (Note: other Gauss Markov assumptions are not needed.)

ii. Since $\hat{\beta}$ is a random variable and β is a fixed parameter value equality occurs when $\sum u_t X_t = 0$ which occurs with probability zero. (Note: this is not the same as $\sum \hat{u}_t X_t = 0$ which must be true from the Normal equations.)

[See Dougherty, 2007, Chapters 2.2, 2.5.]

(c) Taking logs gives $\log Y = \log \beta_1 + \beta_2 \log X$. Differentiating with respect to X produces:

$$\frac{1}{Y}\frac{dY}{dX} = 0 + \beta_2 \frac{1}{X}.$$

Elasticity is given by $\dfrac{x}{y}\dfrac{dy}{dx} = \beta_2$, which is constant. The elasticity is estimated by the slope parameter of a regression of $\log y$ on $\log X$.

A good answer would note that the error term should be additive in this log-log model.

[See Dougherty, 2007, Chapter 4.2.]

(d) Consider the model:

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

If X_t is not independently distributed of u_t then the OLS estimator of β will be inconsistent. Consider a variable Z that is correlated with X but not correlated with u and is not an explanatory variable in the equation in its own right. Z can be considered as an instrumental variable. An estimator of β based on Z is known an instrumental variable (IV) estimator. It is defined as:

$$\hat{\beta}_{IV} = \frac{\sum Z_t Y_t}{\sum Z_t X_t}.$$

It can be shown that $\hat{\beta}_{IV}$ is a consistent estimator of β.

$$\hat{\beta}_{IV} = \frac{\sum Z_t Y_t}{\sum Z_t X_t} = \frac{\sum Z_t(\beta X_t + u_t)}{\sum Z_t X_t} = \beta + \frac{\sum Z_t u_t}{\sum Z_t X_t},$$

$$p\lim(\hat{\beta}_{IV}) = \beta + \frac{\text{plim}\left(\sum Z_t u_t / T\right)}{\text{plim}\left(\sum Z_t X_t / T\right)} = \beta.$$

Thus $\hat{\beta}$ is a consistent estimator of β.

Note: plim $(\sum Z_t u_t/T) = 0$ and plim $(\sum Z_t X_t/T) \neq 0$.

[See Dougherty, 2007, Chapter 8.6.]

(e) Let emp = 1 if employed, emp = 0 if not employed. This is a binary dependent model where we would regress emp on the other variables by OLS (linear probability model) or (better) use probit or logit estimation.

Brief description of any of these estimation methods is required.

[See Dougherty, 2007, Chapters 1.1, 10.2, 10.3.]

(f) A time series is stationary (weakly) if the mean, variance and autocovariance of the series are independent of time. Autocovariance may depend on length of the lag.

In this example $X_t = \theta X_{t-1} + u_t$; $t = 1, 2, \ldots, T$, where $E(u_t) = 0$; $\text{Var}\,(u_t) = \sigma^2$ and $E(u_s u_t) = 0$ for all s and t, $s \neq t$. Assume X_0 is fixed. We can write:

$$\begin{aligned}
t = 1: &\quad X_1 = \theta X_0 + u_1, \\
t = 2: &\quad X_2 = \theta X_1 + u_2 = \theta(\theta X_0 + u_1) + u_2 = \theta^2 X_0 + \theta u_1 + u_2, \\
t = 3: &\quad X_3 = \theta X_2 + u_3 = \theta^3 X_0 + \theta^2 u_1 + \theta u_2 + u_3, \\
&\vdots
\end{aligned}$$

Doing these recursive substitutions, we can write:

$$X_t = \theta X_{t-1} + u_t = \theta^t X_0 + u_t + \theta u_{t-1} + \cdots + \theta^{t-1} u_1.$$

Therefore:

$$E(X_t) = E(\theta^t X_0 + u_t + \theta u_{t-1} + \cdots + \theta^{t-1} u_1) = \theta^t X_0,$$

$$\begin{aligned}
\text{Var}\,(X_t) &= \text{Var}\left(\theta^t X_0 + u_t + \theta u_{t-1} + \cdots + \theta^{t-1} u_1\right) \\
&= \sigma_\varepsilon^2 (1 + \theta^2 + \theta^4 + \cdots + \theta^{2(t-1)}) \\
&= \sigma^2 \sum_{s=0}^{t-1} \theta^{2s}.
\end{aligned}$$

If $\theta \geq 1$ then for large 't', it is easy to see that $E(X_t) \to \infty$ and $\text{Var}\,(X_t) \to \infty$. So the variable X_t is non-stationary and standard analysis test statistics are not valid as these processes assume stationarity and finite mean and variance for the random variable. If this is the case then the variable is growing at the exponential rate which is rare for economic variables.

If $|\theta| < 1$, then for large 't':

$$E(X_t) = \theta^t X_0 = 0,$$

and:

$$\text{Var}\,(X_t) = \sigma^2 \sum_{s=0}^{t-1} \theta^{2s} = \sigma^2 (1 + \theta^2 + \theta^4 + \cdots) = \frac{\sigma^2}{1 - \theta^2},$$

which is a constant.

For large t, covariance is given by:

$$\begin{aligned}
\text{Cov}\,(X_t, X_{t-s}) &= E[X_t - E(X_t)][X_{t-s} - E(X_{t-s})] = E[X_t X_{t-s}] \\
&= E[u_t + \theta u_{t-1} + \theta^2 u_{t-2} \cdots + \theta^{t-1} u_1][u_{t-s} + \theta u_{t-s-1} + \cdots + \theta^{t-1} u_{-s}] \\
&= \theta^s (1 + \theta^2 + \theta^4 + \cdots) \sigma^2 \\
&= \frac{\theta^s \sigma^2}{1 - \theta^2},
\end{aligned}$$

which depends only upon the value of 's'. Therefore if $-1 < \theta < 1$. Hence the variable X_t is stationary if $|\theta| < 1$.

[See Dougherty, 2007, Chapter 13.1.]

(g) RF corresponding to z_t and y_t are:

$$z_t = \frac{\beta_2}{\alpha - \beta_1} x_t + \frac{u_{2t} - u_{1t}}{\alpha - \beta_1} \quad \cdots, \tag{1.1}$$

$$\begin{aligned} y_t &= \frac{\alpha\beta_2}{\alpha - \beta_1} x_t + \frac{\alpha(u_{2t} - u_{1t})}{\alpha - \beta_1} + u_{1t} \\ &= \frac{\alpha\beta_2}{\alpha - \beta_1} x_t + \frac{\alpha u_{2t} - \beta u_{1t}}{\alpha - \beta_1} \cdots \end{aligned} \tag{1.2}$$

(1.1) and (1.2) will exist only if $\alpha - \beta_1 \neq 0$.

In the reduced form, the explanatory variable and the disturbance term are independent, hence reduced form parameters can be consistently estimated by OLS.

[See Dougherty, 2007, Chapter 9.1.]

(h) Given the equations:

$$Y_t = \beta_1 + \beta_2 X_t + u_t,$$
$$u_t = \rho u_{t-1} + v_t,$$

where v has zero mean, constant variance and zero autocovariance, combine the two equations to give:

$$Y_t = \beta_1(1-\rho) + \rho Y_{t-1} + \beta_2 X_t - \beta_2 \rho X_{t-1} + v_t,$$

which is the restricted of the general form (an (ADL(1, 1)) model):

$$Y_t = \lambda_1 + \lambda_2 Y_{t-1} + \lambda_3 X_t + \lambda_4 X_{t-1} + v_t,$$

and is subject to the restriction $\lambda_4 = -\lambda_2\lambda_3$. The test of this restriction is the common factor test.

Note that the usual F test of the restriction is not appropriate because the restriction is non-linear so we have to use the test statistic:

$$N \log\left(\frac{RSS_R}{RSS_U}\right) \sim \chi_1^2,$$

where RSS_R and RSS_U are residual sum of squares from the restricted and unrestricted models respectively. N is the sample size and the test statistic is asymptotically chi-square with 1 degree of freedom.

[See Dougherty, 2007, Chapter 12.6.]

Section B

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

Question 2

(a) The quarterly dummies are attached to the level of sales implying that the sales effect on profits varies over quarters. In this case quarter 4 is the base quarter (hence Q_4 is not needed in the equation) and the sales effect on profits is increased in quarters 1 and 3 but reduced in quarter 2. If Q_4 is included in the equation then we will face a situation of perfect muticollinearity (dummy variable trap).

(b) With sample size 24 the critical t value is 2.093 under a 2 tail test. The coefficients are all significantly different from 0 with the exception of quarter 2. The slope increases over the quarter 4 figure in quarters 1 and 3.

(c) The effect of using the quarterly dummies by themselves is to allow for intercept adjustments between quarters. This is often a sensible procedure when variables, such as profits or sales, vary seasonally. The quarterly dummies are one way of effectively seasonally adjusting the data.

(d) With both slope and intercept variation the effect is to effectively produce a different model for each quarter.

[See Dougherty, 2007, Chapters 5.1, 5.2 and 5.3.]

Question 3

(a) To answer this question you have to define R^2.

Show the decomposition $TSS = ESS + RSS$.

Show that $0 \leq R^2 \leq 1$. Mention drawback of R^2.

Technical details should be given.

[See Dougherty, 2007, Chapter 1.7.]

(b) Omitted variable bias occurs when a valid variable is omitted from the estimated model. Suppose the 'true' model is $y_t = \beta_1 x_{1t} + \beta_2 x_{2t} + u_t$ but the estimated model is $y_t = \beta_1 x_{1t} + v_t$. The OLS estimator of β_1 from the estimated model is:

$$\begin{aligned}\hat{\beta}_1 &= \frac{\sum x_{1t} y_t}{\sum x_{1t}^2} \\ &= \frac{\sum x_{1t} \left(\beta_1 x_{1t} + \beta_2 x_{2t} + u_t\right)}{\sum x_{1t}^2} \\ &= \beta_1 + \beta_2 \frac{\sum x_{1t} x_{2t}}{\sum x_{1t}^2} + \frac{\sum x_{1t} u_t}{\sum x_{1t}^2},\end{aligned}$$

with:

$$E\left(\hat{\beta}_1\right) = \beta_1 + \beta_2 \frac{\sum x_{1t} x_{2t}}{\sum x_{1t}^2},$$

since $E(u_t) = 0$.

The bias is $\beta_2 \frac{\sum x_{1t} x_{2t}}{\sum x_{1t}^2}$ which depends on the value of β_2 and the covariance between x_{1t} and x_{2t}.

A good answer will explore this further.

[See Dougherty, 2007, Chapter 6.2.]

(c) To answer this question it has to be shown that measurement error in independent variables induces OLS estimator to be inconsistent. Technical details are required.

[See Dougherty, 2007, Chapter 8.4.]

(d) Koyck lag transformation allows us to explore the dynamics of an adaptive expectations model This transformation expresses the dependant variable in terms of the current values of the explanatory variables and lagged dependant variable.

Technical details are required.

[See Dougherty, 2007, Chapter 11.3.]

(e) [See Dougherty, 2007, Chapter 12.5.]

Question 4

(a) t values for equation (1) are 4.20, 4.70, 5.25, 4.57, 1.10 and 0.72, with critical value 2.032; and for equation (2), 4.24, 4.66, 5.25 and 4.73, with critical value 2.028. The coefficients are of expected sign – since we have a log-log model the coefficients are elasticities. The second equation is better because there are no non-significant variables but need to check for multicollinearity (and heteroscedasticity).

[See Dougherty, 2007, Chapter 2.8.]

(b) Definition of heteroskedasticity should be given.

You might expect heteroscedasticity because the data set is cross section and cities are probably very different in size (POP and DEN).

(c) Use Goldfeldt-Quandt test.

POP and DEN as possible variables causing heteroskedasticity.

Please note that in Goldfeldt-Quandt test it is important to specify which variable is being used for ordering the sample.

(d) OLS estimators are unbiased, consistent but inefficient. Standard errors of OLS estimators are biased hence t and F tests are invalid.

[See Dougherty, 2007, Chapter 7.1.]

Question 5

(a) In a multiple regression model if two explanatory variables are correlated then we face the situation of multicollinearity. Two situations may arise:

- There is an exact relationship between two explanatory variables (perfect multicollinearity). In this case, the OLS estimator would not exist.
- Explanatory variables are correlated but not exactly. Suppose in the equation there are only two explanatory variables X_{1t} and X_{2t} and they are correlated but not exactly. In this case OLS estimates of the parameters can be obtained. The regression coefficients will be unbiased and the standard errors remain valid. The variances of the regression coefficients are:

$$\mathrm{Var}\left(\hat{\beta}_1\right) = \frac{\sigma^2}{(1-r^2)\sum x_{1t}^2}, \qquad \mathrm{Var}\left(\hat{\beta}_2\right) = \frac{\sigma^2}{(1-r^2)\sum x_{2t}^2},$$

  where r is the correlation coefficient between X_{1t} and X_{2t} and:

$$x_{1t} = X_{1t} - \overline{X}_1, \qquad x_{2t} = X_{2t} - \overline{X}_2.$$

  Hence high correlation leads to large coefficient variances and erratic estimates of the coefficients. The effect on variances may not be detrimental since there are other factors which may mitigate this result. A large sample or large variation in the X values will reduce the multicollinearity effect.

[See Dougherty, 2007, Chapter 3.4.]

(b)

i. When there is a fixed relationship between the independent variables we have perfect multicollinearity and an OLS estimator cannot be obtained.

ii. A relationship between the independent variables will lead to multicollinearity with the result covered in (a). The larger the variance of the error term the smaller the correlation between the X variables, i.e. the smaller will be r.

iii. The effect of including a third dependent variable in general is to make the issue of multicollinearity more difficult to pin down since the collinearity might be between linear combinations of variables. In this case, with high pairwise collinearity, the inclusion of a third variable will not reduce the multicollinearity and hence the problems mentioned above will still hold.

[See Dougherty, 2007, Chapter 3.4.]

Question 6

(a) If after removing the trend from a non-stationary series the resulting variables becomes stationary, then the variable is called **trend stationary**. Let:

$$Z_t = X_t - \alpha_1 t = \alpha_0 + u_t,$$

where $E(u_t) = 0$, $\operatorname{Var}(u_t) = \sigma^2$ and $E(u_t u_{t-s}) = 0$ for all s and t. Then:

$$\begin{aligned} E(Z_t) &= E(\alpha_0 + u_t) = \alpha_0, \\ \operatorname{Var}(Z_t) &= \operatorname{Var}(\alpha_0 + u_t) = \sigma^2, \\ \operatorname{Cov}(Z_t, Z_{t-s}) &= E[Z_t - E(Z_t)][Z_{t-s} - E(Z_{t-s})] = E(u_t u_{t-s}) = 0. \end{aligned}$$

This means that Z_t has constant mean and variance for all t, and covariance is zero for all s. It implies that the series is trend-stationary.

If a non-stationary process can be transformed into a stationary process by differencing then the series is said to be **difference stationary**.

Let X_t be a random walk with a drift:

$$X_t = \beta_0 + X_{t-1} + \varepsilon_t, \tag{6.1}$$

where $E(\varepsilon_t) = 0$, $\operatorname{Var}(\varepsilon_t) = \sigma^2$ and $E(\varepsilon_t \varepsilon_s) = 0$ for all s and t, $s \neq t$.

Subtract X_{t-1} from both sides of (6.1) to get:

$$\Delta X_t = X_t - X_{t-1} = \beta_0 + \varepsilon_t.$$

It can be easily checked that $E(\Delta X_t) = \beta_0$, $\operatorname{Var}(\Delta X_t) = \sigma_\varepsilon^2$ and $\operatorname{Cov}(\Delta X_t, \Delta X_{t-s}) = 0$ for all s and t. This means that ΔX_t is stationary. This implies that X_t is difference stationary.

It is important to know whether a variable is difference or trend stationary because for difference stationary variables shocks have a permanent effect whereas for trend stationary variables shocks are transitory.

[See Dougherty, 2007, Chapter 13.1.]

(b) Description of Dickey-Fuller test and ADF test is required.

[See Dougherty, 2007, Chapter 13.3.]

(c) Consider a simple ADL(1,1) [this is also known as ARDL(1, 1)] model:

$$Y_t = \alpha_1 + \alpha_2 Y_{t-1} + \alpha_3 X_t + \alpha_4 X_{t-1} + u_t. \tag{6.2}$$

Rewrite (6.2) as:

$$\begin{aligned} \Delta Y_t &= Y_t - Y_{t-1} \\ &= \alpha_1 + \alpha_2 Y_{t-1} - Y_{t-1} + \alpha_3 X_t - \alpha_3 X_{t-1} + \alpha_3 X_{t-1} + \alpha_4 X_{t-1} + u_t \\ &= \alpha_1 - (1-\alpha_2) Y_{t-1} + \alpha_3 \Delta X_t + (\alpha_3 + \alpha_4) X_{t-1} + u_t \\ &= \alpha_3 \Delta X_t - (1-\alpha_2) \left[Y_{t-1} - \frac{\alpha_1}{(1-\alpha_2)} - \frac{(\alpha_3+\alpha_4)}{(1-\alpha_2)} X_{t-1} \right] + u_t \\ &= \alpha_3 \Delta X_t - (1-\alpha_2) \left[Y_{t-1} - \beta_1 - \beta_2 X_{t-1} \right] + u_t, \end{aligned}$$

or:

$$\Delta Y_t = \alpha_3 \Delta X_t - \pi \left[Y_{t-1} - \beta_1 - \beta_2 X_{t-1} \right] + u_t, \tag{6.3}$$

where:

$$\pi = (1-\alpha_2), \quad \beta_1 = \frac{\alpha_1}{(1-\alpha_2)}, \quad \beta_2 = \frac{(\alpha_3+\alpha_4)}{(1-\alpha_2)}.$$

Equation (6.3) is the ECM.

When the two variables Y and X are cointegrated the ECM incorporates not only the short-run but also log-run effects. The long run equilibrium $Y_{t-1} - \beta_1 - \beta_2 X_{t-1}$ is included in the model together with the short-run effect captured by the differenced term.

6

All the terms in the ECM, given by (6.3), are stationary. As Y and X are $I(1)$, then ΔX and ΔY are $I(0)$. As Y and X are cointegrated their linear combination:

$$u_{t-1} = Y_{t-1} - \beta_1 - \beta_2 X_{t-1} \sim I(0).$$

The coefficient π provides us with the information about the speed of adjustment in cases of disequilibrium:

i. If $\pi = 1$ then 100% of the adjustment takes place within the period. In other words adjustment is instantaneous and full.

ii. If $\pi = 0.5$ then 50% adjustment takes place each period.

iii. If $\pi = 0$ then there is no adjustment.

[See Dougherty, 2007, Chapter 13.5.]

Question 7

(a) $H_0^{d3} : \beta_{d3} = 0$ vs $H_1^{d3} : \beta_{d3} < 0$

(A) $t_0^{\beta_{d3}} = \dfrac{-0.0636}{0.0899} = -0.707$. Not reject H_0^{d3} at 0.05 level.

(B) $t_0^{\beta_{d3}} = \dfrac{-0.0390}{0.0901} = -0.433$. Not reject H_0^{d3} at 0.05 level

$t_{361-7}(0.05) \approx 1.645 \approx t_{361-9}$.

Secondary school children have little or no effect on female labour supply.

(b) $H_0^{\text{ed}} : \beta_{\text{ed}} = 0 \quad$ vs $\quad H_1^{\text{ed}} : \beta_{\text{ed}} > 0$

$t_0^{\beta_{\text{ed}}} = \dfrac{0.00282}{0.0125} = 0.226$. Not reject at 0.005 level.

$H_0^{\text{ed2}} : \beta_{\text{ed2}} = 0 \quad$ vs $\quad H_1^{\text{ed2}} : \beta_{\text{ed2}} \neq 0$

$t_0^{\beta_{\text{ed2}}} = \dfrac{0.00256}{0.00117} = 2.19$. Reject at 0.05 level.

$t_{361-9}(0.025) = 1.96$.

$H_0^{\text{ed,ed2}} : \beta_{\text{ed}} = 0, \beta_{\text{ed2}} = 0 \quad$ vs $\quad H_1^{\text{ed,ed2}}$: either $\beta_{\text{ed}} \neq 0$ or $\beta_{\text{ed2}} \neq 0$.

$$F = \frac{(106.502 - 104.818)/2}{104.818/(361-9)} = 2.83$$

$F_{2,\infty}(0.05) = 3.00$. Not reject at 0.05 level.

The F test shows that the coefficients on ed and ed2 are not jointly significantly different from 0 but the t-test shows that the coefficient of ed2 is hence there is a conflict. Note, however, that the t test is just significant and the F test is just below the significance level. Testing at a lower than 0.05 level would produce agreement. Model B has a higher R^2 value but this is to be expected because the number of independent variables is greater but they do not contribute a significant effect hence model A is probably preferable.

(c) The signs of the coefficients are $\widehat{\beta}_{\text{oinc}} < 0$, $\widehat{\beta}_{\text{age}} < 0$ and $\widehat{\beta}_{\text{age2}} < 0$ (but insignificant) which indicates the depletion of skills as age increases. On the other hand note that $\widehat{\beta}_{\text{ed}} > 0$ (but insignificant), $\widehat{\beta}_{\text{ed2}} > 0$ which shows hours worked increasing and accelerating with respect to education. Lastly $\widehat{\beta}_{d2} < 0$ and $\widehat{\beta}_{d1} > 0$ (but both insignificant) and there is a negative effect of education for primary school children but positive for pre-school children. Insignificant variables are, however, not distinguishable from zero effect hence they actually play no part in the explanation of hours worked.

(d) Since the parameter is insignificant you should probably remove the variable associated with years of education, i.e. ed together with age2 and d3. Note that the coefficient on d1 is insignificant in model B but strongly significant in model A which might suggest some omitted variable bias in model A and is a reason to tread carefully with deletion.

[See Dougherty, 2007, Chapters 2.8 and 3.5.]

Question 8

(a) Serial correlation of the error term is the violation of the hypothesis $E(u_s u_t) = 0$ which has implications for OLS estimates of the parameters – they will be unbiased but not efficient and the standard errors are not computed correctly leading to invalid t and F tests. It may be due to misspecification since an omitted variable can be reflected in the estimated residual. If the omitted variable is serially correlated (very likely for economic variables) then the residual, displaying the behavior of the omitted variable, will also be serially correlated.

To answer this question technical details should be given.

[See Dougherty, 2007, Chapter 12.3.]

(b)
i. t values are 5.40, 4.23 and 3.85, hence parameters are significantly different from 0 and D&P are supported.

ii. $N = 58$, $k' = 2$ hence $d_l = 1.514$ and $d_u = 1.652$. DW lies in inconclusive region so we cannot say.

iii. DW rejects the null hypothesis of $\rho = 0$ hence serial correlation is present and the parameter estimates are unbiased, not efficient and the estimate of the standard errors are incorrect hence t testing is invalid.

iv. If the serial correlation is due to misspecification then the estimated disturbance will display the behavior of the missing variable hence an approach would be to attempt to match up the estimated error with potential omitted variables.

[See Dougherty, 2007, Chapter 12.3.]

Changes in the format of the examination

This examination is three hours long. The paper is divided into two sections. The format will change in 2010. In 2009, Section A, which was compulsory, contained eight questions which were intended to examine the whole syllabus. Section B was designed to examine a selection of topics in greater depth. Section B had seven questions, of which students were required to answer three. From 2010 the new format will be: In Section A, Question 1 will have 5 parts (instead of 8) worth 5 marks each. Thus Question 1 will be worth 25 marks (instead of 40). Question 1 will remain compulsory. Section B will consist of 5 questions (instead of 7), of which students have to answer three. Each question in Section B is worth 25 marks.

Examiners' commentaries 2010

20 Elements of econometrics

Format of the examination

This commentary reflects the examination and assessment arrangements for this unit in the academic year 2009–10. In 2011 the format of the examination will change to:

> Candidates should answer FOUR of the following SIX questions: QUESTION 1 of Section A (25 marks in total) and THREE questions from Section B (25 marks each). Candidates are strongly advised to divide their time accordingly.
> Candidates should note that Section B will now consist of five questions rather than seven questions as previously.

The format and structure of the examination may change again in future years, and any such changes will be publicised on the virtual learning environment (VLE).

Specific comments on questions – Zone A

Candidates should answer **FOUR** of the following **EIGHT** questions: **QUESTION 1** of Section A (25 marks in total) and **THREE** questions from Section B (25 marks each). **Candidates are strongly advised to divide their time accordingly.**

Extracts from statistical tables are given after the final question on this paper

Graph paper is provided at the end of this question paper. If used, it must be detached and fastened securely inside the answer book.

A calculator may be used when answering questions on this paper and it must comply in all respects with the specification given with your Admission Notice. The make and type of machine must be clearly stated on the front cover of the answer book.

Section A

Answer **all** parts of question 1 (25 marks each).

Question 1

(a) **If $Y_t = \widehat{\alpha} + \widehat{\beta} X_t + \widehat{u}_t$ is the result of fitting a linear relationship by ordinary least squares show that:**

$$\sum_{t=1}^{T} \left(Y_t - \bar{Y}\right)^2 = \sum_{t=1}^{T} \left[\widehat{\beta}\left(X_t - \bar{X}\right) + \widehat{u}_t\right]^2,$$

and hence show that:

Total sum of squares = explained sum of squares + residual sum of squares.

Explain the significance of this result.

(b) **Consider a model:**

$$y_t = \alpha x_t + u_t, \qquad t = 1, 2, \ldots, T,$$

where $E(u_t) = 0$, $E(u_t^2) = \sigma^2 x_t^2$ and $E(u_s u_t) = 0$ if $s \neq t$ for all $s, t = 1, 2, \ldots, T$. x's are fixed.

i. **Derive the weighted least squares (WLS) estimator, $\widehat{\alpha}$, of α and also derive the variance of $\widehat{\alpha}$.**

ii. **Is the WLS estimator of α consistent? Explain.**

(c) **In the model:**

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

x_t is measured with error. Data is only available on $x_t^\star$, where

$$x_t^\star = x_t + v_t, \qquad t = 1, 2, \ldots, T,$$

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 $x_t^\star$ have zero means.

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

(d) **Let the regression equation be:**

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t, \qquad t = 1, 2, \ldots, T.$$

Outline briefly, how you would test:

i. **$\beta_2 = 1$,**

ii. **jointly β_2 and β_3 are zero.**

Specify the assumptions required for these tests.

(e) **If a random variable X has a distribution with probability density function:**

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right),$$

show that the maximum likelihood estimator of the mean (μ) of the random variable X is the sample mean.

Reading for this question

C. Dougherty, '**Introduction to Econometrics**' (third edition) Chapters 1.7, 2.8, 7.3, 8.4 and 10.6.

Approaching the question

(a)

$$\begin{aligned}
TSS &= \sum \left[Y_t - \bar{Y}\right]^2 \\
&= \sum \left[\widehat{\alpha} + \widehat{\beta} X_t + \widehat{u}_t - (\widehat{\alpha} + \widehat{\beta}\bar{X})\right]^2 \\
&= \sum \left[\widehat{\beta}(X_t - \bar{X}) + \widehat{u}_t\right]^2 \\
&= \sum \left[(\widehat{Y}_t - \bar{Y}) + \widehat{u}_t\right]^2 \\
&= \sum \left(\widehat{Y}_t - \bar{Y}\right)^2 + \sum \widehat{u}_t^2 + 2\sum \widehat{u}_t (Y_t - \bar{Y})
\end{aligned}$$

Since $\sum \widehat{u}_t \left(\widehat{Y}_t - \bar{Y}\right) = \sum \widehat{u}_t \widehat{\beta}\left(X_t - \bar{X}\right) = 0$, this becomes:

$$= \sum \left(\widehat{Y}_t - \bar{Y}\right)^2 + \sum \widehat{u}_t^2.$$

Thus $TSS = ESS + RSS$, and:

$$R^2 = \frac{ESS}{TSS} = \frac{TSS - RSS}{TSS} = 1 - \frac{RSS}{TSS},$$

where RSS is the residual sum of squares. We therefore know RSS must lie between 0 and TSS and TSS must be positive, hence:

$$0 \leq R^2 \leq 1.$$

(b) i. Weighted least squares estimator is derived by estimating the following equation by OLS:

$$\frac{y_t}{x_t} = \alpha + \frac{u_t}{x_t}, \qquad E\left(\frac{u_t}{x_t}\right) = 0, \qquad \text{Var}\left(\frac{u_t}{x_t}\right) = \sigma^2, \qquad E\left(\frac{u_s}{x_s}\frac{u_t}{x_t}\right) = 0,$$

if $s \neq t$. Applying OLS we get the weighted least squares estimator as:

$$\widehat{\alpha} = \frac{1}{T}\sum \frac{y_t}{x_t} = \frac{1}{T}\sum \frac{\alpha x_t + u_t}{x_t} = \alpha + \frac{1}{T}\sum \frac{u_t}{x_t}.$$

$E(\widehat{\alpha}) = \alpha$ and the variance of $\widehat{\alpha}$ is:

$$\begin{aligned} E\left(\widehat{\alpha} - \alpha\right)^2 &= E\left(\frac{1}{T}\sum \frac{u_t}{x_t}\right)^2 \\ &= \frac{1}{T^2}E\left(\sum \left(\frac{u_t}{x_t}\right)^2\right) + \frac{1}{T^2}E\left(\sum_{s\neq t}^{T} \frac{u_s u_t}{x_s x_t}\right) \\ &= \frac{1}{T^2}\sum \left(\frac{\sigma^2 x_t^2}{x_t^2}\right) \\ &= \frac{\sigma^2}{T}. \end{aligned}$$

ii. Use sufficient condition for consistency. $E(\alpha) = \alpha$ and $\lim \text{Var}\left(\widehat{\alpha}\right) \mapsto 0$ as $T \mapsto \infty$. This implies $\widehat{\alpha}$ is a consistent estimator of α.

(c) $y_t = \beta x_t + u_t$ where $x_t^\star = x_t + v_t$.

$$\widehat{\beta} = \frac{\sum x_t^\star y_t}{\sum (x_t^\star)^2} = \frac{\sum (x_t + v_t)(\beta x_t + u_t)}{(x_t + v_t)^2} = \frac{\beta \sum x_t^2 + \sum x_t u_t + \beta \sum x_t v_t + \sum v_t u_t}{\sum x_t^2 + \sum v_t^2 + 2\sum x_t v_t}.$$

$$\begin{aligned} \text{plim}\left(\widehat{\beta}\right) &= \frac{\frac{1}{T}\text{plim}\left(\beta \sum x_t^2 + \sum x_t u_t + \beta \sum x_t v_t + \sum v_t u_t\right)}{\frac{1}{T}\text{plim}\left(\sum x_t^2 + \sum v_t^2 + 2\sum x_t v_t\right)}, \\ &= \frac{\beta \sigma_x^2}{\sigma_x^2 + \sigma_v^2} \\ &\neq \beta. \end{aligned}$$

Thus $\widehat{\beta}$ is inconsistent.

(d) i. This is a standard two-tail t test of the form:

$$t = \frac{\left(\widehat{\beta}_2 - 1\right)}{\text{se}_{\widehat{\beta}_2}}.$$

The degrees of freedom for the t test are $T - 3$.

ii. The test of the null hypothesis $H_0 : \beta_2 = \beta_3 = 0$ against the alternative hypothesis $H_A : \beta_2 \neq 0$ and/or $\beta_3 \neq 0$ is achieved by applying the F test where:

$$F = \frac{R^2}{k-1} \bigg/ \frac{1-R^2}{T-k}.$$

T is the sample size and k is the number of parameters in the regression, i.e. 3. R^2 is the coefficient of determination. The F test has $(2, T-3)$ degrees of freedom. Note that it is important to tailor your answer to the specification of the question, putting in actual values where appropriate rather than quoting the formulae in general terms.

Assumptions required are that disturbance term should not be autocorrelated and should be homoscedastic otherwise these tests won't be valid. Disturbance term should be normally distributed.

(e) The likelihood function is defined by:

$$L = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(X_i-\mu)^2}{2\sigma^2}\right) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{\sum (X_i-\mu)^2}{2\sigma^2}\right).$$

Hence, taking logs, we get:

$$\ln L = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2.$$

To maximise, differentiate w.r.t. μ and set the derivative equal to 0:

$$\frac{\partial \ln L}{\partial \mu} = -\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu) = 0,$$

which gives:

$$\widehat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}.$$

Section B

Answer **three** questions from this section (25 marks each).

Question 2

Consider a two equation model

$$\boldsymbol{q_t = \beta_1 p_t + \beta_2 y_t + u_t,}$$

$$\boldsymbol{q_t = \alpha p_t + e_t,}$$

$t = 1, 2, \ldots T$, where:

- **T is the sample size,**
- **q_t and p_t are endogenous variables,**
- **y_t is an exogenous variable, and**
- **u_t and e_t are serially uncorrelated disturbances with zero means, respective variances σ_1^2 and σ_2^2 and covariance σ_{12} for all t.**

(a) **Examine the identifiability of the above given equations.**

(b) **Examine the consistency of the ordinary least squares estimator of α.**

(c) Derive the two stage least squares estimator of α.

(d) Without derivation, explain what would be the indirect least squares estimator of α.

Reading for this question

C. Dougherty, '**Introduction to Econometrics**' (third edition) Chapters 9.2 and 9.3.

Approaching the question

(a) Only Order condition of identification is required.

Order Condition: $R \geq G - 1$, where R is the number of restrictions imposed on the equation under consideration and G is the number of jointly dependent variables or total number of equations in the complete model.

1st equation

$$R = 0, \quad G - 1 = 2 - 1 = 1.$$

As $R < G - 1$, this equation is underidentified.

2nd equation

$$R = 1, \quad G - 1 = 2 - 1 = 1.$$

This equation is exactly identified.

(b) The model is:

$$q_t = \beta_1 p_t + \beta_2 y_t + u_t, \tag{2.1}$$

$$q_t = \alpha p_t + e_t. \tag{2.2}$$

Applying OLS to equation (2.2), we get:

$$\widehat{\alpha}_{OLS} = \frac{\sum_{t=1}^{T} p_t q_t}{\sum_{t=1}^{T} p_t^2}. \tag{2.3}$$

Inconsistency of the OLS estimator

$$\widehat{\alpha}_{OLS} = \frac{\sum_{t=1}^{T} p_t q_t}{\sum_{t=1}^{T} p_t^2} = \frac{\sum_{t=1}^{T} p_t(\alpha p_t + e_t)}{\sum_{t=1}^{T} p_t^2} = \alpha \frac{\sum_{t=1}^{T} p_t^2}{\sum_{t=1}^{T} p_t^2} + \frac{\sum_{t=1}^{T} p_t e_t}{\sum_{t=1}^{T} p_t^2} = \alpha + \frac{\sum_{t=1}^{T} p_t e_t}{\sum_{t=1}^{T} p_t^2}.$$

To check consistency of $\widehat{\alpha}$ take plim of $\widehat{\alpha}$:

$$\text{plim}\,(\widehat{\alpha}) = \alpha + \frac{\text{plim}\left(\frac{1}{T}\sum_{t=1}^{T} p_t e_t\right)}{\text{plim}\left(\frac{1}{T}\sum_{t=1}^{T} p_t^2\right)} = \alpha + \frac{\text{Cov}\,(p_t, e_t)}{\text{Var}\,(p_t)},$$

where $\text{Var}\,(p_t) = \sigma_p^2$.

To derive $\text{Cov}\,(p_t, e_t)$, first we have to obtain the RF corresponding to p. To obtain the RF substitute (2.2) in (2.1) to get:

$$\alpha p_t + e_t = \beta_1 p_t + \beta_2 y_t + u_t, \qquad \Rightarrow \qquad (\alpha - \beta_1) p_t = \beta_2 y_t + u_t - e_t.$$

Thus the RF corresponding to p is:

$$p_t = \frac{\beta_2}{\alpha - \beta_1} y_t + \frac{u_t - e_t}{\alpha - \beta_1}. \tag{2.4}$$

$E(u_t) = E(e_t) = 0$ implies both:

$$E(p_t) = \frac{\beta_2}{\alpha - \beta_1} y_t,$$

and:

$$\begin{aligned}
\text{Cov}\,(p_t, e_t) &= E\,[p_t - Ep_t]\,e_t \\
&= E\left[\frac{u_t - e_t}{\alpha - \beta_1}\right] \\
&= \frac{E(u_t e_t)}{\alpha - \beta_1} - \frac{E(e_t^2)}{\alpha - \beta_1} \\
&= \frac{\sigma_{12}}{\alpha - \beta_1} - \frac{\sigma_2^2}{\alpha - \beta_1} \\
&= \frac{\sigma_{12} - \sigma_2^2}{\alpha - \beta_1} \\
&\neq 0.
\end{aligned}$$

Thus $\widehat{\alpha}_{OLS}$ is an inconsistent estimator of α.

We can easily check that:

$$\text{plim}\,(\widehat{\alpha}) = \alpha + \frac{\sigma_{12} - \sigma_2^2}{(\alpha - \beta_1)\sigma_p^2} \neq \alpha.$$

(c) Estimation involves two stages.

First Stage

Write down the RF corresponding to the RHS endogenous variable, in this case p. Let it be:

$$p_t = \Pi_1 y_t + v_t. \tag{2.5}$$

As RF parameters can always be consistently estimated by the OLS, apply OLS to (2.5) to get:

$$\widehat{p}_t = \widehat{\Pi}_1 y_t, \quad \widehat{\Pi}_1 = \frac{\sum_{t=1}^{T} p_t y_t}{\sum_{t=1}^{T} y_t^2}. \tag{2.6}$$

We can also write p_t as:

$$p_t = \widehat{p}_t + \widehat{v}_t, \tag{2.7}$$

where $\widehat{v}_t$ is the OLS estimate of v_t.

Second Stage

In the second stage instead of p_t, write $\widehat{p}_t + \widehat{v}_t$ in the original equation to get:

$$q_t = \alpha\widehat{p}_t + \alpha\widehat{v}_t + e_t. \tag{2.8}$$

Apply OLS to (2.8) to get 2SLS estimator of α as:

$$\widehat{\alpha}_{2,2SLS} = \frac{\sum_{t=1}^{T} \widehat{p}_t q_t}{\sum_{t=1}^{T} \widehat{p}_t^2} = \frac{\widehat{\Pi}_1 \sum_{t=1}^{T} q_t y_t}{\widehat{\Pi}_1^2 \sum_{t=1}^{T} y_t^2},$$

where we have used $\widehat{p}_t = \widehat{\Pi}_1 y_t$. By noting that $\widehat{\Pi}_1 = \dfrac{\sum_{t=1}^{T} p_t y_t}{\sum_{t=1}^{T} y_t^2}$, we then have:

$$= \frac{\sum_{t=1}^{T} q_t y_t}{\sum_{t=1}^{T} y_t^2} \frac{\sum_{t=1}^{T} y_t^2}{\sum_{t=1}^{T} p_t y_t},$$

which, upon cancellation, yields the 2SLS estimator of α:

$$= \frac{\sum_{t=1}^{T} q_t y_t}{\sum_{t=1}^{T} p_t y_t}.$$

(d) Equation (2.6) is exactly identified, hence 2SLS and ILS estimators will be same.

Question 3

Write brief notes on the following:

(a) Durbin-Watson *d* statistic.

(b) The Koyck distributed lag.

(c) Type I, type II errors and power of a test.

Reading for this question

C. Dougherty, '**Introduction to Econometrics**' (third edition) Chapters 12.3, 11.2, 11.3 and 2.

Approaching the question

(a) Durbin-Watson (DW) Statistic

The DW statistic is defined as:

$$dw = \frac{\sum_{t=2}^{T} (\widehat{u}_t - \widehat{u}_{t-1})^2}{\sum_{t=1}^{T} \widehat{u}_t^2}.$$

The DW statistic can be applied only if:

i. the disturbance term follows AR(1) process,

ii. the model has an intercept term, and

iii. there is no lagged dependent variable as an explanatory variable.

In the Durbin-Watson table (which can be found in the Appendix of any Econometrics book) there are two values:

d_L**:** lower limit,

d_U**:** upper limit,

where:

- if $dw < d_L$ then there is positive autocorrelation;
- if $dw > 4 - d_L$ then there is negative autocorrelation;
- if $d_L \leq dw \leq d_U$ then no conclusion can be drawn;
- if $4 - d_U \leq dw \leq 4 - d_L$ then no conclusion can be drawn.

Candidates should also discuss that in large samples $dw \simeq 2(1 - \rho)$.

(b) Write down Koyck (infinite) lag form and show how it can be transformed into an estimable equation with a lagged dependent variable. The resulting equation has a serially correlated error term and hence cannot be estimated by OLS so use a non-linear estimation process or a grid search. **Technical details should be given.**

(c) Type I error (usually denoted by α) is the error of rejecting the null hypothesis when it is true, type II error (denoted by β) is the error of not rejecting the null hypothesis when the alternative hypothesis is true. The power of a test is $1 - \beta$. A good answer here will give a diagrammatic representation of the relationship between α, β and the power.

Question 4

(a) Consider a model:

$$\boldsymbol{Y_i = \beta_1 + \beta_2 X_i + u_i, \qquad i = 1, 2, \ldots, n,}$$

where:

$$\boldsymbol{Y_i = \begin{cases} 1, & \text{if the event takes place,} \\ 0, & \text{otherwise.,} \end{cases}}$$

and $E(u_i) = 0$, $i = 1, 2, \ldots, n$.

i. Explain fully the problems which arise 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:

$$\boldsymbol{Y_i = \begin{cases} 1, & \text{if the } i^{\text{th}} \text{ household purchases a newspaper,} \\ 0, & \text{otherwise.} \end{cases}}$$

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 in full time education,

$u_i =$ unobserved disturbance term.

The model was estimated by logit with the following results:

	Estimated Coefficients	Asymptotic Standard Errors
S	0.521	0.10
E	0.067	0.012
Constant	-2.56	1.57

$$\log L_U = -321.25 \qquad \log L_R = -416.01.$$

$\log L_U$ is the log likelihood from the unrestricted model and $\log L_R$ is the log likelihood of the model where all the slope coefficients are restricted to zero.

i. Explain how the coefficients were estimated.

ii. Test the null hypothesis that all the slope coefficients are jointly equal to zero.

Reading for this question

C. Dougherty, '**Introduction to Econometrics**' (third edition) Chapters 10.1, 10.2 and 10.6.

Approaching the question

(a) i. LPM is used to denote a model in which the dependent variable is binary whereby it takes value 1 if the event occurs and 0 if it does not. It is estimated by the ordinary least squares (OLS).

Let the model be:

$$Y_i = \beta_0 + \beta_1 X_i + u_i, \qquad i = 1, 2, \ldots, n. \tag{4.1}$$

where:

$$Y_i = \begin{cases} 1, & \text{if event occurs,} \\ 0, & \text{if not.} \end{cases}$$

Assume $E(u_i) = 0$, then:

$$E\left[Y_i | X_i\right] = \beta_0 + \beta_1 X_i. \tag{4.2}$$

Also:

$$E\left[Y_i | X_i\right] = 1 \times P\left[Y_i = 1\right] + 0 \times P\left[Y_i = 0\right] = P_i. \tag{4.3}$$

From (4.2) and (4.3):

$$E\left[Y_i | X_i\right] = \beta_0 + \beta_1 X_i = P_i,$$

hence, we can interpret $E\left[Y_i | X_i\right] = \beta_0 + \beta_1 X_i$ as the probability that the event will occur given X_i.

If we denote $\widehat{\beta}_0$ and $\widehat{\beta}_1$ as estimates of β_0 and β_1 respectively, then we can write:

$$\widehat{Y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 X_i = \widehat{P}_i \tag{4.4}$$

as the estimated probability that the event will occur.

As Y_i takes only two values, 1 or 0, therefore u_i can take only two values:

$$u_i = \begin{cases} \beta_0 - \beta_1 X_i, & Y_i = 0, \\ 1 - \beta_0 - \beta_1 X_i, & Y_i = 1. \end{cases}$$

Based on this we can write the probability distribution of u_i as:

$$f(u_i) = \begin{cases} 1 - \beta_0 - \beta_1 X_i, & Y_i = 0, \\ \beta_0 + \beta_1 X_i, & Y_i = 1. \end{cases}$$

This probability distribution also satisfies the assumption that:

$$E(u_i) = (1 - \beta_0 - \beta_1 X_i)\left(\beta_0 + \beta_1 X_i\right) + \left(-\beta_0 - \beta_1 X_i\right)\left(1 - \beta_0 - \beta_1 X_i\right) = 0.$$

We can write $V(u_i)$ as:

$$
\begin{aligned}
V(u_i) &= E\left(u_i^2\right) \\
&= (1-\beta_0-\beta_1X_i)^2\,(\beta_0+\beta_1X_i) + (+\beta_0-\beta_1X_i)^2\,(1-\beta_0-\beta_1X_i) \\
&= (1-\beta_0-\beta_1X_i)\,[(1-\beta_0-\beta_1X_i)+(\beta_0+\beta_1X_i)]\,(\beta_0+\beta_1X_i) \\
&= (\beta_0+\beta_1X_i)\,(1-\beta_0-\beta_1X-i) \\
&= E(Y_i)\,[1-E(Y_i)] \\
&= P_i(1-P_i),
\end{aligned}
$$

for all $i = 1, 2, \ldots, n$. Hence the disturbance term is heteroscedastic. This will make OLS estimators inefficient.
In many cases the estimated probability $\widehat{Y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 X_i$ will be negative or greater than 1.

ii. Weighted least squares

We can see from (4.4), that an estimator of P_i is $\widehat{P}_i = \widehat{Y}_i$, therefore $\widehat{Y}_i\left(1-\widehat{Y}_i\right)$ can be used as an estimator of:

$$V(u_i) = E(Y_i)\,[1-E(Y_i)] = P_i(1-P_i).$$

Weights can be obtained as:

$$W_i = \left[\widehat{Y}_i\left(1-\widehat{Y}_i\right)\right]^{\frac{1}{2}}.$$

Divide (4.1) by W_i to get:

$$\frac{Y_i}{W_i} = \frac{\beta_0}{W_i} + \beta_1\frac{X_i}{W_i} + \frac{u_i}{W_i}, \quad i = 1, 2, \ldots, n,$$

and apply OLS to obtain the WLS estimator of β_0 and β_1. This will give an efficient estimator.

Problem:

In practice the estimated variance of u_i, $\widehat{Y}_i\left(1-\widehat{Y}_i\right)$, may be negative as again the estimated probability $\widehat{Y}_i$ may be negative or greater than 1.
The obvious correction of the estimated negative probability is to constrain estimated probabilities within $[0, 1]$ interval. If we do this we might predict an occurrence with probability 1 when it is possible that it might not occur, or we might predict an occurrence with probability 0 when it might actually occur. The estimation process may give unbiased estimates but predictions obtained from it will be biased.

(b) i. See Chapter 10.2 of Dougherty.

ii. $H_0 =$ all slope coefficients are equal to zero.

$$-2\,[\log L_R - \log L_u] \sim \chi^2_q,$$

where q is the number of restrictions imposed by H_0. This gives:

$$-2\,[-416.01 - (-321.25)] = 189.52 \sim \chi^2_2.$$

Critical value of chi-square with 2 df at 5% level of significance is 5.991.
We therefore reject H_0.

Question 5

The following 3 equations were estimated by the ordinary least squares 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** (0.34)	**3.66** (0.13)	**2.57** (0.21)
age	**-0.00** (0.007)	**0.075** (0.006)	**0.14** (0.01)
age^2	-	**-0.0008** (0.0001)	**-0.001** (0.0001)
S	-	-	**-0.05** (0.005)
S^2	-	-	**0.0004** (0.0001)
R^2	**0.0007**	**0.05**	**0.11**

where age = age in years, age^2 = age $\times$ age, S = age the individual completed full time education, and $S^2 = S \times S$. Figures in brackets are standard errors.

(a) Using the model (i), briefly explain the method of ordinary least squares estimation.

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

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

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

(e) 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.

(f) 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.

Reading for this question

C. Dougherty, '**Introduction to Econometrics**' (third edition) Chapters 1.7, 3.1, 3.5, 6.2 and 7.2.

Approaching the question

(a) The model is:

$$\text{earning}_t = \beta_0 + \beta_1 \text{age}_t + u_t, \qquad t = 1, 2, \ldots, T.$$

OLS estimators minimises the residual sum of squares (RSS). Let:

$$\widehat{u}_t = \widehat{\beta}_0 - \widehat{\beta}_1 \text{age}_t$$

be the residual, then the RSS is:

$$\sum \widehat{u}_t = \sum \left(\text{earning}_t - \widehat{\beta}_0 - \widehat{\beta}_1 \text{age}_t\right)^2.$$

Minimising this with respect to $\widehat{\beta}_0$ and $\widehat{\beta}_1$, we get the first order conditions:

$$\frac{\partial \sum \widehat{u} + t^2}{\partial \widehat{\beta}_0} = 0, \qquad \frac{\partial \sum \widehat{u} + t^2}{\partial \widehat{\beta}_1} = 0,$$

Solving these first order conditions we obtain the OLS estimators $\widehat{\beta}_0$ and $\widehat{\beta}_1$.

(b) The variable age has a positive sign as expected – income increases with age and since the income/age relationship is not linear a quadratic term is also expected (income eventually drops after retirement). The signs on S and S^2 are not perhaps as expected – S can be thought of as measuring work experience so a negative sign is counter-intuitive but might be due to the fact that individuals with higher qualifications are likely to be earning more and have less years in the job market so a negative sign is appropriate (the S^2 sign is not predictable).

(c) R^2 are low (usual in cross-section models) but t and F tests for significance of parameter estimates all show significant coefficients, hence the hypothesis of no relationship is rejected so there is no cause for concern.

(d) Omitted variable bias. Model (ii) is an improvement on (i) hence you would expect omitted variable bias and since age and age^2 are positively related and the coefficient on age is expected to be negative the bias will be negative. A good answer would derive the omitted variable bias.

(e)
$$F = \frac{\left(R_U^2 - R_R^2\right)/m}{\left(1 - R_U^2\right)/(N-k)} = \frac{(0.11 - 0.04)/2}{(1 - 0.11)/(3866 - 5)} = 130.15,$$

and $F^{\star}_{2,3861} = 19.5$. Hence we reject $H_0 : \beta_3 = \beta_4 = 0$.

(f) The Goldfeldt-Quandt test is described on pages 227–228 of Dougherty. A good answer would note that the data should be ordered by a heteroskedastic variable (age in this case). It was suggested because cross-section data on variables like income (which has wide spread) is characterised by heteroskedasticity, and also because t tests and F tests are invalid.

Question 6

(a) Consider a model:

$$\boldsymbol{y_0 = 0, \qquad y_t = \theta y_{t-1} + u_t, \quad t = 1, 2, \ldots T,}$$

where $\boldsymbol{E(u_t) = 0}$, $\boldsymbol{E(u_t^2) = \sigma^2}$ and $\boldsymbol{E(u_s u_t) = 0}$ when $\boldsymbol{s \neq t}$, for all $\boldsymbol{s, t = 1, 2, \ldots, T}$.

Derive the mean and variance of $\boldsymbol{y_t}$ when $\boldsymbol{|\theta| = 1}$ and comment on the result.

(b) Describe in detail the Dickey-Fuller and the augmented Dickey-Fuller (ADF) procedure for testing for the order of integration of a time series variable.

(c) Consider a model:

$$\boldsymbol{Y_t = \alpha_1 + \alpha_2 Y_{t-1} + u_t, \qquad t = 1, 2, \ldots, T,}$$
$$\boldsymbol{u_t = \rho u_{t-1} + \varepsilon_t,}$$

where $\boldsymbol{E(\varepsilon_t) = 0}$, $\boldsymbol{E(\varepsilon_t^2) = \sigma^2}$ and $\boldsymbol{E(\varepsilon_s \varepsilon_t) = 0}$ for all $\boldsymbol{s, t = 1, 2, \ldots, T}$.

Derive the specification for the ADF test.

(d) Consider a model:

$$\boldsymbol{Y_t = \beta X_t + u_t, \qquad t = 1, 2, \ldots, T,}$$
$$\boldsymbol{u_t = \theta e_{t-1} + e_t,}$$

where $\boldsymbol{E(e_t) = 0}$, $\boldsymbol{E(e_t^2) = \sigma^2}$ and $\boldsymbol{E(e_s e_t) = 0}$ if $\boldsymbol{s \neq t}$ for all $\boldsymbol{s, t = 1, 2, \ldots, T}$.

Are $\boldsymbol{Y_t}$ and $\boldsymbol{X_t}$ cointegrated? Explain.

Reading for this question

C. Dougherty, '**Introduction to Econometrics**' (third edition) Chapters 13.1, 13.3 and 13.4.

C. Dougherty, '**Subject Guide**' (2008) pp.223–224.

Approaching the question

(a) The model is:

$$Y_t = Y_{t-1} + u_t, \qquad t = 1, 2, \ldots T. \tag{6.1}$$

We can write:

$$\begin{aligned} t = 1, &\qquad Y_1 = u_1, \\ t = 2, &\qquad Y_2 = Y_1 + u_2 = u_1 + u_2, \\ t = 3, &\qquad Y_3 = Y_2 + u_3 = u_1 + u_2 + u_3 \\ \vdots &\qquad \quad \vdots \end{aligned}$$

Doing these recursive substitutions, we can write:

$$Y_t = Y_{t-1} + u_t = u_t + u_{t-1} + \ldots + u_1 = \sum_{s=1}^{t} u_s,$$

where $E(Y_t) = 0$ and:

$$\text{Var}\,(Y_t) = \text{Var}\left(\sum_{s=1}^{t} u_s\right) = t\sigma^2.$$

Thus Y_t is non-stationary as the variance of Y_t is dependent on time. Y_t is a random walk.

(b) The standard test for a unit root is due to Dickey and Fuller and is based on the model:

$$y_t = \beta_1 + \beta_2 y_{t-1} + \gamma_t + u_t,$$

which can be re-written as:

$$\Delta y_t = \beta_1 + (1 - \beta_2) y_{t-1} + \gamma_t + u_t,$$

where $\Delta y_t = y_t - y_{t-1}$.

The null hypothesis for stationarity is:

$$H_0 : 1 - \beta_2 = 0, \qquad H_A : 1 - \beta_2 \neq 0.$$

We cannot use the standard t-test procedure in this case because the distribution of the t-statistic is not a t-distribution, so critical values have been computed by Dickey and Fuller using Monte-Carlo techniques.

The test is sensitive to the presence of serial correlation in the error term so we need to take steps to remove the effects of this serial correlation – this is done by including lagged values of y_t in the regression, i.e.:

$$y_t = \beta_1 + \beta_2 y_{t-1} + \beta_3 y_{t-2} + \gamma_t + u_t$$

for an AR(1) serial correlation. This is more easily tested by using the model:

$$\Delta y_t = \beta_1 + (1 - \beta_2 - \beta_3) y_{t-1} - \beta_3 \Delta y_{t-2} + \gamma_t + u_t,$$

with null hypothesis $H_0 : 1 - \beta_1 - \beta_2 = 0$. Once again, the Dickey-Fuller tables should be used.

(c) The model is:

$$Y_t = \alpha_1 + \alpha_2 Y_{t-1} + u_t, \qquad u_t = \rho u_{t-1} + \varepsilon_t, \tag{6.2}$$

where ε_2 is $I(0)$. Lag (6.2) by one period and multiply it by ρ to get:

$$\rho Y_{t-1} = \alpha_1 \rho + \alpha_2 \rho Y_{t-2} + u_{t-1}. \tag{6.3}$$

Subtract (6.3) from (6.2) and rearrange to get:

$$\begin{aligned} Y_t &= \alpha_1(1 - \rho) + (\alpha_2 + \rho) Y_{t-1} - \alpha_2 \rho Y_{t-2} + (u_t - \rho u_{t-1}) \\ &= \alpha_1(1 - \rho) + (\alpha_2 + \rho) Y_{t-1} - \alpha_2 \rho Y_{t-2} + \varepsilon_t. \end{aligned}$$

This can be written as:

$$Y_t = \beta_1 + \beta_2 Y_{t-1} + \beta_3 Y_{t-2} + \varepsilon_t.$$

If we subtract Y_{t-1} from both sides and introduce $\beta_3 Y_{t-1}$ into the right hand side we get:

$$Y_t - Y_{t-1} = \beta_1 + \beta_2 Y_{t-1} - Y_{t-1} + \beta_3 Y_{t-1} - \beta_3 Y_{t-1} + \beta_3 Y_{t-2} + \varepsilon_t,$$

or:

$$\Delta Y_t = \beta_1 + (\beta_2 + \beta_3 - 1) Y_{t-1} - \beta_3 \Delta Y_{t-1} + \varepsilon_t.$$

To test for unit root we test the coefficient of Y_{t-1}, i.e.:

$$H_0 : \beta_2 + \beta_3 - 1 = 0, \qquad H_1 : \beta_2 + \beta_3 - 1 < 0.$$

(d) Y_t and X_t are cointegrated if a linear combination of Y_t and X_t is I(0). $u_t = Y_t - \beta X_t$ is a linear combination of Y_t and X_t. Hence, we have to examine the stationarity of u_t:

$$E(u_t) = E(e_t + \theta e_{t-1}) = 0.$$

$$E\left(u_t^2\right) = E\left(e_t^2\right) + \theta^2 E\left(e_{t-1}^2\right) + E(e_t e_{t-1}) = \left(1 + \theta^2\right)\sigma^2,$$

since $E(e_t e_{t-1}) = 0$.

$$\begin{aligned} E(u_t u_{t-1}) &= E(e_t + \theta e_{t-1})(e_{t-1} + \theta e_{t-2}) \\ &= E(e_t e_{t-1}) + \theta E(e_t e_{t-2}) + \theta E\left(e_{t-2}^2\right) + \theta^2 E(e_{t-1} e_{t-2}) \\ &= \theta\sigma^2, \end{aligned}$$

since $E(e_t e_{t-s}) = 0$, for all $s > 0$.

$$\begin{aligned} E(u_t u_{t-2}) &= E(e_t + \theta e_{t-1})(e_{t-2} + \theta e_{t-3}) \\ &= E(e_t e_{t-2}) + \theta E(e_t e_{t-3}) + \theta E(e_{t-1} e_{t-2}) + \theta^2 E(e_{t-2} e_{t-3}) \\ &= 0. \end{aligned}$$

Thus both first and second moments are independent of t and u_t must be (weakly) stationary. This implies that Y_t and X_t are cointegrated.

Question 7

(a) Consider a model:

$$\boldsymbol{y_t = \beta x_t + u_t,}$$

and $E(u_t) = 0$, $E\left(u_t^2\right) = \sigma^2$ and $E(u_s u_t) = 0$ if $s \neq t$ for all $s, t = 1, 2, \ldots, T$. Show that the ordinary least squares estimator of β is unbiased and consistent.

(b) Consider a model:

$$\boldsymbol{y_t = \alpha y_{t-1} + u_t,}$$

where:

$$\boldsymbol{u_t = \theta\varepsilon_{t-1} + \varepsilon_t, \qquad t = 1, 2, \ldots, T,}$$

and $E(\varepsilon_t) = 0$, $E\left(\varepsilon_t^2\right) = \sigma^2$ and $E(\varepsilon_s \varepsilon_t) = 0$ if $s \neq t$ for all $s, t = 1, 2, \ldots, T$.

i. Show that the OLS estimator of α is inconsistent.

ii. What properties does the OLS estimator of α have if $\theta = 0$?

Reading for this question

C. Dougherty, '**Introduction to Econometrics**' (third edition) Chapters 2 and 12.5.

Approaching the question

(a)

$$\widehat{\beta} = \frac{\sum x_t y_t}{\sum x_t^2} = \beta + \frac{\sum x_t u_t}{\sum x_t^2},$$

$E\left(\widehat{\beta}\right) = \beta$ as $E(u_t) = 0$, and:

$$V\left(\widehat{\beta}\right) = \frac{\sum x_t^2 V(u_t)}{\left(\sum x_t^2\right)^2} = \frac{\sigma^2 \sum x_t^2}{\left(\sum x_t^2\right)^2} = \frac{\sigma^2}{\sum x_t^2}.$$

Use sufficient condition for consistency: $E\left(\widehat{\beta}\right) = \beta$ and $V\left(\widehat{\beta}\right) \to 0$ as $T \to \infty$. So the sufficient condition of consistency holds, therefore $\widehat{\beta}$ is a consistent estimator of β.

(b) i. The OLS estimator of α is:

$$\widehat{\alpha} = \frac{\sum y_t y_{t-1}}{\sum y_{t-1}^2} = \alpha + \frac{\sum u_t y_{t-1}}{\sum y_{t-1}^2} = \frac{\sum \varepsilon_t y_{t-1}}{\sum y_{t-1}^2} + \frac{\theta \sum \varepsilon_{t-1} y_{t-1}}{\sum y_{t-1}^2},$$

where y_t has been substituted as $y_t = \alpha y_{t-1} + \varepsilon_{t-1} + \theta \varepsilon_{t-1}$.

Since $y_{t-1} = \alpha y_{t-2} + \varepsilon_{t-2} + \theta \varepsilon_{t-2}$ hence $\text{plim}\left(\frac{1}{T}\sum \varepsilon_t y_{t-1}\right) = 0$. It is easy to check that $\text{plim}\left(\frac{1}{T}\sum \varepsilon_{t-1} y_{t-1}\right) \neq 0$. It is also assumed $\text{plim}\left(\frac{1}{T}\sum y_{t-1}^2\right)$ is greater than zero and finite. Thus $\text{plim}(\widehat{\alpha}) \neq \alpha$ and therefore $\widehat{\alpha}$ is inconsistent.

ii. Since $\theta = 0$, $u_t = \varepsilon_t$, and:

$$\widehat{\alpha} = \frac{\sum y_t y_{t-1}}{\sum y_{t-1}^2} = \alpha + \frac{\sum u_t y_{t-1}}{\sum y_{t-1}^2} = \alpha + \frac{\sum \varepsilon_t y_{t-1}}{\sum y_{t-1}^2},$$

As $\text{plim}\left(\frac{1}{T}\sum \varepsilon_t y_{t-1}\right) = 0$, $\text{plim}(\widehat{\alpha}) = \alpha$, so $\widehat{\alpha}$ is consistent.

However, the expectation of $\sum \varepsilon_t y_{t-1} \big/ \sum y_{t-1}^2$ is not zero. To see this expand, to get:

$$\frac{\sum \varepsilon_t y_{t-1}}{\sum y_{t-1}^2} = \frac{(\varepsilon_2 y_1 + \varepsilon_3 y_2 + \varepsilon_4 y_3 + \ldots + \varepsilon_T y_{T-1})}{(y_1^2 + y_2^2 + y_3^2 + \ldots + y_{T-1}^2)}. \tag{7.1}$$

Since the model says $y_t = \alpha y_{t-1} + \varepsilon_t$ the same values of ε_t appear in both the numerator and denominator of (7.1). Thus the expectation of (7.1) is not zero and $\widehat{\alpha}$ is biased.

Question 8

In order to model the demand for motor vehicles, an econometrician proposes the general linear regression model:

$$\boldsymbol{Y_t = \beta_0 + \beta_P P_t + \beta_E E_t + \beta_B B_t + u_t, \qquad t = 1965, 1966, \ldots, 1986,}$$

where:

- $\boldsymbol{Y}$ **is the logarithm of an index of consumer expenditure on motor vehicles, spares and accessories at constant prices,**
- $\boldsymbol{P}$ **is the logarithm of a relative price index of motor vehicles,**
- $\boldsymbol{E}$ **is the logarithm of real total household expenditure,**

- **B is the logarithm of a relative price index of public road transport,**
- **u is the error term.**

This model was fitted using ordinary least squares (OLS) to annual data covering the period 1965–1986 and the following results were obtained:

$$\hat{Y}_t = \underset{(0.56)}{6.27} - \underset{(0.067)}{0.705}P_t, \tag{8.1}$$

$$R^2 = 0.0738, \qquad RSS = 0.636,$$

$$\hat{Y}_t = \underset{(3.03)}{-2.05} - \underset{(0.347)}{0.926}P_t + \underset{(0.644)}{1.78}E_t + \underset{(0.310)}{0.0608}B_t, \tag{8.2}$$

$$R^2 = 0.720, \qquad RSS = 0.192,$$

where R^2 is the coefficient of determination, RSS denotes residual sum of squares and estimated standard errors are given in parentheses.

(a) Test the hypothesis $H_0 : \beta_P = 0$ in both fitted models (8.1) and (8.2). Comment on your results.

(b) Test the individual hypotheses $H_0 : \beta_E = 0$ and $H_0 : \beta_B = 0$ and the joint hypothesis $H_0 = \beta_E = \beta_B = 0$.

(c) Which of the fitted models (8.1) and (8.2) is preferable? Explain your answer.

(d) Discuss the economic implications of the fitted model (8.2). Is there any evidence that public road transport acts as a substitute for private motor travel?

(e) How might a plot of the OLS residuals from fitted model (8.2) against time assist you in ascertaining whether or not the model for the demand for motor vehicles is misspecified? Explain.

Reading for this question

C. Dougherty, '**Introduction to Econometrics**' (third edition) Chapters 2 and 6.2.

Approaching the question

(a) $H_0^P : \beta_P = 0$ vs $H_1^P : \beta_P < 0$. (Note the one-tail alternative hypothesis).

For model (8.1), $T = 22$ and $k = 2$, and our test statistics are:

$$t_{\text{calc}} = -0.705/0.0673 = -10.48, \qquad t_{20}(0.05) = -1.725.$$

We therefore reject H_0^P for model (8.1).

For model (8.2), $T = 22$ and $k = 2$, and our test statistics are:

$$t_{\text{calc}} = -0.926/0.346 = -2.67, \qquad t_{18}(0.05) = -1.734.$$

We therefore reject H_0^P for model (8.2).

(b) $H_0^E : \beta_E = 0$ vs $H_1^E : \beta_E > 0$, (once again note the one-tail alternative)

$$t_{\text{calc}} = \frac{1.78}{0.644} = 2.76.$$

Reject.

$H_0^B : \beta_B = 0$ vs $H_1^B : \beta_B > 0$,

$$t_{\text{calc}} = \frac{0.0608}{0.310} = 0.196.$$

Do not reject.

$H_0^{BE}: \beta_B = \beta_E = 0$ vs $H_1^{BE}: \beta_B \neq 0$ and/or $\beta_E \neq 0$.

$$F_{\text{calc}} = \frac{(0.636 - 0.192)/2}{0.192/18} = 20.81$$

$$F_{2,18}(0.05) = 3.49.$$

Reject.

(c) There are important variables omitted in model (8.1). Model (8.2) is preferable.

(d) Very slight; see (b). Note that since the model is a log-log model the estimated coefficients are estimates of elasticities. The estimated coefficients have the correct signs according to standard economic relationships.

(e) Patterns and range: do the plotted residuals violate independence and homoscedasticity assumptions?

Examiners' commentaries 2010

20 Elements of econometrics

Format of the examination

This commentary reflects the examination and assessment arrangements for this unit in the academic year 2009–10. In 2011 the format of the examination will change to:

> Candidates should answer FOUR of the following SIX questions: QUESTION 1 of Section A (25 marks in total) and THREE questions from Section B (25 marks each). Candidates are strongly advised to divide their time accordingly.
> Candidates should note that Section B will now consist of five questions rather than seven questions as previously.

The format and structure of the examination may change again in future years, and any such changes will be publicised on the virtual learning environment (VLE).

Specific comments on questions – Zone B

Candidates should answer **FOUR** of the following **EIGHT** questions: **QUESTION 1** of Section A (25 marks in total) and **THREE** questions from Section B (25 marks each). **Candidates are strongly advised to divide their time accordingly.**

Extracts from statistical tables are given after the final question on this paper

Graph paper is provided at the end of this question paper. If used, it must be detached and fastened securely inside the answer book.

A calculator may be used when answering questions on this paper and it must comply in all respects with the specification given with your Admission Notice. The make and type of machine must be clearly stated on the front cover of the answer book.

Section A

Answer **all** parts of question 1 (25 marks each).

Question 1

(a) **In the linear regression model:**

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t, \qquad t = 1, 2, \ldots, T,$$

prove that:

i. $\sum_{t=1}^{T} \widehat{u}_t = 0$ **, where** $\widehat{u}_t$ **is the residual defined as:**

$$\widehat{u}_t = \widehat{Y}_t - \widehat{\beta}_1 - \widehat{\beta}_2 X_{2t} - \widehat{\beta}_3 X_{3t}.$$

$\widehat{\beta}_1$, $\widehat{\beta}_2$ and $\widehat{\beta}_3$ are the ordinary least squares (OLS) estimators of β_1, β_2 and β_3, respectively.

ii.

$$\sum_{t=1}^{T} X_{2t}\widehat{u}_t = 0 = \sum_{t=1}^{T} X_{3t}\widehat{u}_t.$$

(b) The logistic model is given by:

$$F(\theta) = \frac{e^{\theta}}{1+e^{\theta}}.$$

Sketch the shape of this function as θ varies from $-\infty$ to $+\infty$. Explain how this function is used to model the probability of an event occurring when θ is replaced with an index $X\beta$ and the use of this model in econometrics.

(c) Show that

$$\sum_{t=1}^{T} \left(Y_t - \bar{Y}\right)^2 = \sum_{t=1}^{T} \left[\widehat{\beta}\left(X_t - \bar{X}\right) + \widehat{u}_t\right]^2,$$

where $Y_t = \widehat{\alpha} + \widehat{\beta}X_t + \widehat{u}_t$ is the result of fitting a linear relationship by OLS and hence show that:

Total sum of squares = explained sum of squares + residual sum of squares.

Explain the significance of this result.

(d) Explain why a disturbance term in a regression equation may be serially correlated and describe how serial correlation affects the ordinary least squares results.

(e) If $\widehat{Y}_{T+p} = \widehat{\alpha} + \widehat{\beta}X_{T+p}$ is the prediction of Y_t in time period $T+p$ where the linear relationship has been estimated by OLS on observations $1, 2, \ldots, T$, show that the prediction error $\widehat{Y}_{T+p} - Y_{T+p}$ has zero mean and explain why the prediction error itself is unlikely to be zero. Give details of the assumptions you have used.

Reading for this question

C. Dougherty, '**Introduction to Econometrics**' (third edition) Chapters 1.7, 10.1, 10.2, 12.3 and11.5.

Approaching the question

(a) To apply OLS we minimise the sum of squares of errors:

$$I = \sum_{t=1}^{T} \widehat{u}^2 = \sum_{t=1}^{T} \left[Y_t - \widehat{\beta}_1 - \widehat{\beta}_2 X_{2t} - \widehat{\beta}_3 X_{3t}\right]^2,$$

by differentiating with respect to $\widehat{\beta}_1$, $\widehat{\beta}_2$ and $\widehat{\beta}_3$. The resulting equations are:

$$\frac{\partial I}{\partial \widehat{\beta}_1} = 2\sum_{t=1}^{T}\left(Y_t - \widehat{\beta}_1 - \widehat{\beta}_2 X_{2t} - \widehat{\beta}_3 X_{3t}\right)(-1) = -2\sum_{t=1}^{T}\widehat{u}_t = 0,$$

$$\frac{\partial I}{\partial \widehat{\beta}_1} = 2\sum_{t=1}^{T}\left(Y_t - \widehat{\beta}_1 - \widehat{\beta}_2 X_{2t} - \widehat{\beta}_3 X_{3t}\right)(-X_{2t}) = -2\sum_{t=1}^{T}\widehat{u}_t X_{2t} = 0,$$

$$\frac{\partial I}{\partial \widehat{\beta}_1} = 2\sum_{t=1}^{T}\left(Y_t - \widehat{\beta}_1 - \widehat{\beta}_2 X_{2t} - \widehat{\beta}_3 X_{3t}\right)(-X_{3t}) = -2\sum_{t=1}^{T}\widehat{u}_t X_{3t} = 0,$$

which gives the three results.

(b) The sketch should show the elongated S shape of the logistic function and leads to the logit function:

$$\frac{e^{X\beta}}{1+e^{X\beta}},$$

which can be written:

$$\frac{1}{1+e^{-X\beta}}.$$

The function takes the values 0 at $X\beta = -\infty$, 0.5 at 0 and 1 at $+\infty$. The parameters β are estimated by maximum likelihood and the object is to model a binary dependent variable using independent variables X.

(c)

$$\begin{aligned}
TSS &= \sum \left[Y_t - \bar{Y}\right]^2 \\
&= \sum \left[\widehat{\alpha} + \widehat{\beta}X_t + \widehat{u}_t - (\widehat{\alpha} + \widehat{\beta}\bar{X})\right]^2 \\
&= \sum \left[\widehat{\beta}(X_t - \bar{X}) + \widehat{u}_t\right]^2 \\
&= \sum \left[(\widehat{Y}_t - \bar{Y}) + \widehat{u}_t\right]^2 \\
&= \sum \left(\widehat{Y}_t - \bar{Y}\right)^2 + \sum \widehat{u}_t^2 + 2\sum \widehat{u}_t(Y_t - \bar{Y})
\end{aligned}$$

Since $\sum \widehat{u}_t \left(\widehat{Y}_t - \bar{Y}\right) = \sum \widehat{u}_t \widehat{\beta}\left(X_t - \bar{X}\right) = 0$, this becomes:

$$= \sum \left(\widehat{Y}_t - \bar{Y}\right)^2 + \sum \widehat{u}_t^2,$$

(see the answer to 1(a)). Thus $TSS = ESS + RSS$ and

$$R^2 = \frac{ESS}{TSS} = \frac{TSS - RSS}{TSS} = 1 - \frac{RSS}{TSS},$$

where RSS is the residual sum of squares. We therefore know RSS must lie between 0 and TSS and TSS must be positive, hence:

$$0 \leq R^2 \leq 1.$$

(d) Serial correlation is the breakdown of the Gauss-Markov assumption that the error terms are uncorrelated through time ($E(u_t u_s) = 0$ for $t \neq s$). Serial correlation is a common problem in econometric estimation when time series data is being analysed (note that serial correlation is not feasible with cross-section data). There are many reasons why relationships might display serial correlation of the errors but the main reasons are:

(i) Inertia, or 'persistence of effect'. Economic relationships are not usually instantaneous but effects are spread over several periods. Any excluded variables in a regression relation will produce an error term which will change slowly over time since most economic variables are generally not random in behaviour.

(ii) Specification bias. If the regression equation has omitted an important variable, or the equation is specified with the incorrect functional form, i.e. a variable is entered as a linear instead of a quadratic form then the effect of the 'omitted variable' will appear in the error term. Once again, since most economic variables are not random in behaviour but tend to change slowly over time, the error term will contain elements of the behaviour of the missing variable and hence will be serially correlated.

(iii) Cobweb phenomenon. Certain types of economic models tend to produce serially correlated behaviour. Under cobweb phenomena economic agents react to information from the previous time period and hence build in a relationship between successive time periods. The classic cobweb behaviour is in agriculture where farmers plant crops in response to prices obtained in the previous harvest. If prices were high then there will be a large increase in planting (leading to low prices in the current growing season) and if prices were low then the planting will decrease in the current season (leading to high prices in the current growing season). This type of behaviour leads to a relationship between successive observations and hence serial correlation effects.

(iv) Manipulation of data. Sometimes data is manipulated to remove seasonal variation. This involves averaging data over a number of time periods which will produce inertia in behaviour and hence serial correlation.

The effect of serially correlated errors is that OLS coefficients are unbiased but inefficient. Their standard errors are incorrectly estimated hence the t values are incorrect leading to incorrect hypothesis tests.

(e)

$$\begin{aligned}
E\left(Y_{t+P}-\widehat{Y}_{t+P}\right) &= E\left(\alpha+\beta X_{t+P}+u_{t+P}\right)-E\left(\widehat{\alpha}+\widehat{\beta} X_{t+P}\right)\\
&= \alpha+\beta X_{t+P}+E\left(u_{t+P}\right)-E\left(\widehat{\alpha}\right)-X_{t+P}E\left(\widehat{\beta}\right)\\
&= \alpha+\beta X_{t+P}-\alpha-\beta X_{t+P}\\
&= 0.
\end{aligned}$$

This result is on expectations. The actual values will give 0 with probability zero since the prediction error is a stochastic variable. We need the standard Gauss-Markov assumptions to hold. If the Gauss Markov conditions do not hold then we cannot assume that $E\left(\widehat{\alpha}\right)=\alpha$ and $E\left(\widehat{\beta}\right)=\beta$, hence the result no longer holds.

Section B

Answer **three** questions from this section (25 marks each).

Question 2

(a) Explain what is meant by a consistent estimator. Why is the concept useful in econometrics?

(b) An economist develops a theory that the demand for money, M_t, in period t is negatively related to the expected inflation rate, i^e_{t+1}, in period $t+1$:

$$M_t=\beta_0+\beta_1 i^e_{t+1}+u_t,$$

where u_t is a disturbance term with $E(u_t)=0$, $E\left(u_t^2\right)=\sigma^2$ and $E\left(u_t u_s\right)=0$ for $s\neq t$.

The expected interest rate is determined by an adaptive expectations process:

$$i^e_{t+1}-i^e_t=\lambda\left(i_t-i^e_t\right),\qquad 0\leq\lambda\leq 1,$$

where i_t is the actual inflation rate in period t. The economist uses the following model to fit the relationship

$$M_t=\alpha_0+\alpha_1 i_t+\alpha_2 M_{t-1}+v_t.$$

i. Show how the adaptive expectations process can be written as an infinite sum of all past observations on i_t and explain the implications of your formulation for the behaviour of M_t.

ii. Explain how estimates of α_0, α_1 and α_2 can be used to estimate the parameters β_0, β_1 and λ.

iii. **Explain why Ordinary Least Squares (OLS) produces inconsistent estimates of the parameters α_0, α_1 and α_2.**

iv. **How could you estimate the model to obtain consistent estimates of α_0, α_1 and α_2.**

Reading for this question

C. Dougherty, '**Introduction to Econometrics**' (third edition) Chapter 11.3.

Approaching the question

(a) An estimator $\widehat{\theta}$ is consistent if plim $\left(\widehat{\theta}\right) = \theta$. If an estimator is asymptotically unbiased and variance tends to 0 then it is consistent. We often use plim when expected values do not exist – we use this concept when dealing with instrumental variables for example.

(b) i.

$$\begin{aligned} i^e_{t+1} &= \lambda i_t + (1-\lambda) i^e_t \\ &= \lambda i_t + (1-\lambda)\left(\lambda i_{t-1} + (1-\lambda) i^e_{t-1}\right) \\ &= \lambda\left(i_t + (1-\lambda) i_{t-1} + (1-\lambda)^2 i_{t-2} + \ldots\right). \end{aligned}$$

The weights decline with $(1-\lambda)$.
Hence:

$$\begin{aligned} M_t &= \beta_0\lambda + \beta_1\lambda i_t + (1-\lambda)M_{t-1} + u_t - (1-\lambda)u_{t-1} \\ &= \alpha_0 + \alpha_1 i_t + \alpha M_{t-1} + v_t \end{aligned}$$

ii.

$$\lambda = 1-\alpha_2, \qquad \beta_1 = \frac{\alpha_1}{\lambda}, \qquad \beta_0 = \frac{\alpha_0}{\lambda}.$$

iii. OLS gives biased estimates because B_{t-1} and the error term are correlated – a proof involving OLS on a simple model such as $y_t = \beta x_t + u_t$, where x_t and u_t are correlated is expected here.

iv. IV using x_t and x_{t-1} as instruments. Alternatively use grid search on values of λ.

Question 3

(a) **Show that ordinary least squares (OLS) applied to the model:**

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

provides an unbiased estimator of β as long as certain assumptions are true. Explain carefully what these assumptions are.

(b) **What do you understand by instrumental variable (IV) estimation? Explain how you would estimate the model in (a) by instrumental variables and give details of the properties of your estimators.**

(c) **Give TWO examples of models where instrumental variable estimation is an improvement on OLS. Explain why IV estimation is superior to OLS in these cases.**

Reading for this question

C. Dougherty, '**Introduction to Econometrics**' (third edition) Chapters 2, 8.6 and 9.3.

Approaching the question

(a) Consider the simple model:

$$y_t = \beta x_t + u_t.$$

Then

$$\widehat{\beta} = \frac{\mathrm{Cov}\,(x,y)}{\mathrm{Var}\,(x)} = \beta + \frac{\mathrm{Cov}\,(x,u)}{\mathrm{Var}\,(x)} = \beta + \frac{\sum_{t=1}^{T} x_t u_t}{\sum_{t=1}^{T} x_t^2}.$$

Hence:

$$E\left(\widehat{\beta}\right) = \beta + \frac{\sum_{t=1}^{T} x_t E\,(u_t)}{\sum_{t=1}^{T} x_t^2} = \beta + 0,$$

since $E\,(u_t) = 0$ and x_t and u_t are independently distributed.

(b) IV estimator is:

$$\widehat{\beta}_{IV} = \frac{\sum y_t z_t}{\sum x_t z_t},$$

where z_t is the instrumental variables such that it is correlated with x_t but uncorrelated with the error term u_t, and not a variable in the equation.

$$\widehat{\beta}_{IV} = \frac{\sum y_t z_t}{\sum x_t z_t} = \frac{\sum z_t\,(\beta x_t + u_t)}{\sum x_t z_t} = \beta + \frac{\sum z_t u_t}{\sum x_t z_t}.$$

So:

$$\mathrm{plim}\left(\widehat{\beta}_{IV}\right) = \mathrm{plim}\,(\beta) + \mathrm{plim}\left(\frac{\sum z_t u_t}{\sum x_t z_t}\right) = \beta + \frac{\mathrm{plim}\left(\frac{1}{T}\sum x_t u_t\right)}{\mathrm{plim}\left(\frac{1}{T}\sum x_t z_t\right)}.$$

By assumption:

$$\mathrm{plim}\left(\frac{1}{T}\sum x_t u_t\right) = 0, \qquad \mathrm{plim}\left(\frac{1}{T}\sum x_t z_t\right) \neq 0,$$

hence $\widehat{\beta}_{IV}$ is consistent.

(c) There are several instances where instrumental variables are useful. Two examples are adaptive expectations where OLS will give biased parameter estimates and simultaneous equations where we use 2SLS (which is a form of IV) where again OLS will give biased results.

Question 4

You have been given the task of estimating an earnings function using data from a data set on 2868 individuals. You have observations on:

EARN: log of earnings measured in dollars per hour

S: years of schooling

ASVABC: the score on a test of cognitive ability

MALE: a dummy which = 1 if male, 0 if female

UNION: a dummy which =1 if the individual belonged to a union in 2004; 0 otherwise.

You estimate 4 regressions using Ordinary Least Squares (OLS): regression (1) and (4) use the whole sample, regression (2) uses only those individuals who belonged to a union in 2004 and regression (3) uses only those individuals who did not belong to a union in 2004.

Dependent variable: EARN

	Whole sample (1)	Union only (2)	Non-union only (3)	Whole sample (4)
S	0.066 (0.004)	0.028 (0.012)	0.070 (0.005)	0.066 (0.004)
ASVABC	0.013 (0.001)	0.011 (0.003)	0.013 (0.001)	0.013 (0.001)
MALE	0.214 (0.017)	0.286 (0.049)	0.199 (0.018)	0.209 (0.017)
UNION	-	-	-	0.189 (0.028)
Constant	0.819 (0.055)	1.545 (0.164)	0.750 (0.058)	0.803 (0.055)
R^2	0.249	0.195	0.260	0.261
RSS	588.3	43.7	522.5	579.7
Sample size	2868	286	2582	2868

(Standard errors in parentheses)

(a) **What, precisely, do the coefficients on MALE and UNION tell us?**

(b) **Using the above results analyse whether there is a difference between earnings for union and non-union workers. Use whatever method you like but explain the method carefully, state all the assumptions necessary and discuss the limitations of your approach in answering the question of whether there is a difference.**

(c) **Another researcher points out that earnings are dependent on age and the length of time the individual has worked and suggests you include these two variables. You tell him you are concerned about multicollinearity if both these variables are included in this regression. What is the problem with multicollinearity and is it likely to occur in the revised equation? Explain.**

Reading for this question

C. Dougherty, '**Introduction to Econometrics**' (third edition) Chapters 2.8, 3.1, 3.2, 3.4 and 3.5.

Approaching the question

(a) Coefficients on UNION give the monetary increase in wages through being a member of a Union – the individual should, of course, be working in a unionised firm. MALE is equivalent. The candidate should do a t test to establish that the parameters are significantly different from 0: $t_{\text{MALE}} = 12.29$ and $t_{\text{UNION}} = 6.75$ with critical $t_{563} = 1.96$.

(b) Two possible approaches:

i. use t test on Union dummy – this will only give information on intercept shift.

ii. use F test:

$$F = \frac{(588.3 - 579.7)/1}{(579.7/(2868 - 4))} = 42.5.$$

Critical $F_{4,562} = 2.21$, hence we reject H_0 and conclude there is a difference.

(c) Multicollinearity causes increased coefficient standard errors which leads to insignificant t-values. Work experience = age - schooling assuming no periods of unemployment. Hence work experience and age could well be correlated and we would expect to see insignificant parameters.

Question 5

(a) Consider a model:

$$\boldsymbol{y_0 = 0, \qquad y_t = \theta y_{t-1} + u_t, \quad t = 1, 2, \ldots T,}$$

where $\boldsymbol{E(u_t) = 0}$, $\boldsymbol{E(u_t^2) = \sigma^2}$ and $\boldsymbol{E(u_s u_t) = 0}$ when $\boldsymbol{s \neq t}$, for all $\boldsymbol{s, t = 1, 2, \ldots, T}$.

Derive the mean and variance of $\boldsymbol{y_t}$ when $\boldsymbol{|\theta| = 1}$ and comment on the result.

(b) Describe in detail the Dickey-Fuller and the augmented Dickey-Fuller (ADF) procedure for testing for the order of integration of a time series variable.

(c) Consider a model:

$$\boldsymbol{Y_t = \alpha_1 + \alpha_2 Y_{t-1} + u_t, \qquad t = 1, 2, \ldots, T,}$$

$$\boldsymbol{u_t = \rho u_{t-1} + \varepsilon_t,}$$

where $\boldsymbol{E(\varepsilon_t) = 0}$, $\boldsymbol{E(\varepsilon_t^2) = \sigma^2}$ and $\boldsymbol{E(\varepsilon_s \varepsilon_t) = 0}$ for all $\boldsymbol{s, t = 1, 2, \ldots, T}$.

Derive the specification for the ADF test.

(d) Consider a model:

$$\boldsymbol{Y_t = \beta X_t + u_t, \qquad t = 1, 2, \ldots, T,}$$

$$\boldsymbol{u_t = \theta e_{t-1} + e_t,}$$

where $\boldsymbol{E(e_t) = 0}$, $\boldsymbol{E(e_t^2) = \sigma^2}$ and $\boldsymbol{E(e_s e_t) = 0}$ if $\boldsymbol{s \neq t}$ for all $\boldsymbol{s, t = 1, 2, \ldots, T}$.

Are $\boldsymbol{Y_t}$ and $\boldsymbol{X_t}$ cointegrated? Explain.

Reading for this question

C. Dougherty, '**Introduction to Econometrics**' (third edition) Chapters 13.1, 13.3 and 13.4.

C. Dougherty, '**Subject Guide**' (2008) pp.223–224.

Approaching the question

(a) The model is:

$$Y_t = Y_{t-1} + u_t, \qquad t = 1, 2, \ldots T. \tag{5.1}$$

We can write:

$$\begin{aligned}
t &= 1, & Y_1 &= u_1, \\
t &= 2, & Y_2 &= Y_1 + u_2 = u_1 + u_2, \\
t &= 3, & Y_3 &= Y_2 + u_3 = u_1 + u_2 + u_3 \\
&\vdots & &\vdots
\end{aligned}$$

Doing these recursive substitutions, we can write:

$$Y_t = Y_{t-1} + u_t = u_t + u_{t-1} + \ldots + u_1 = \sum_{s=1}^{t} u_s,$$

where $E(Y_t) = 0$ and:

$$\text{Var}\,(Y_t) = \text{Var}\left(\sum_{s=1}^{t} u_s\right) = t\sigma^2.$$

Thus Y_t is non-stationary as the variance of Y_t is dependent on time. Y_t is a random walk.

(b) The standard test for a unit root is due to Dickey and Fuller and is based on the model:

$$y_t = \beta_1 + \beta_2 y_{t-1} + \gamma_t + u_t,$$

which can be re-written as:

$$\Delta y_t = \beta_1 + (1 - \beta_2) y_{t-1} + \gamma_t + u_t,$$

where $\Delta y_t = y_t - y_{t-1}$.

The null hypothesis for stationarity is:

$$H_0 : 1 - \beta_2 = 0, \qquad H_A : 1 - \beta_2 \neq 0.$$

We cannot use the standard t-test procedure in this case because the distribution of the t-statistic is not a t-distribution, so critical values have been computed by Dickey and Fuller using Monte-Carlo techniques.

The test is sensitive to the presence of serial correlation in the error term so we need to take steps to remove the effects of this serial correlation – this is done by including lagged values of y_t in the regression, i.e.:

$$y_t = \beta_1 + \beta_2 y_{t-1} + \beta_3 y_{t-2} + \gamma_t + u_t$$

for an AR(1) serial correlation. This is more easily tested by using the model:

$$\Delta y_t = \beta_1 + (1 - \beta_2 - \beta_3) y_{t-1} - \beta_3 \Delta y_{t-2} + \gamma_t + u_t,$$

with null hypothesis $H_0 : 1 - \beta_1 - \beta_2 = 0$. Once again, the Dickey-Fuller tables should be used.

(c) The model is:

$$Y_t = \alpha_1 + \alpha_2 Y_{t-1} + u_t, \qquad u_t = \rho u_{t-1} + \varepsilon_t, \tag{5.2}$$

where ε_2 is $I(0)$. Lag (5.2) by one period and multiply it by ρ to get:

$$\rho Y_{t-1} = \alpha_1 \rho + \alpha_2 \rho Y_{t-2} + u_{t-1}. \tag{5.3}$$

Subtract (5.3) from (5.2) and rearrange to get:

$$\begin{aligned} Y_t &= \alpha_1(1-\rho) + (\alpha_2 + \rho) Y_{t-1} - \alpha_2 \rho Y_{t-2} + (u_t - \rho u_{t-1}) \\ &= \alpha_1(1-\rho) + (\alpha_2 + \rho) Y_{t-1} - \alpha_2 \rho Y_{t-2} + \varepsilon_t. \end{aligned}$$

This can be written as:

$$Y_t = \beta_1 + \beta_2 Y_{t-1} + \beta_3 Y_{t-2} + \varepsilon_t.$$

If we subtract Y_{t-1} from both sides and introduce $\beta_3 Y_{t-1}$ into the right hand side we get:

$$Y_t - Y_{t-1} = \beta_1 + \beta_2 Y_{t-1} - Y_{t-1} + \beta_3 Y_{t-1} - \beta_3 Y_{t-1} + \beta_3 Y_{t-2} + \varepsilon_t,$$

or:

$$\Delta Y_t = \beta_1 + (\beta_2 + \beta_3 - 1) Y_{t-1} - \beta_3 \Delta Y_{t-1} + \varepsilon_t.$$

To test for unit root we test the coefficient of Y_{t-1}, i.e.:

$$H_0 : \beta_2 + \beta_3 - 1 = 0, \qquad H_1 : \beta_2 + \beta_3 - 1 < 0.$$

(d) Y_t and X_t are cointegrated if a linear combination of Y_t and X_t is I(0). $u_t = Y_t - \beta X_t$ is a linear combination of Y_t and X_t. Hence, we have to examine the stationarity of u_t:

$$E(u_t) = E(e_t + \theta e_{t-1}) = 0.$$

$$E\left(u_t^2\right) = E\left(e_t^2\right) + \theta^2 E\left(e_{t-1}^2\right) + E(e_t e_{t-1}) = \left(1+\theta^2\right)\sigma^2,$$

since $E(e_t e_{t-1}) = 0$.

$$\begin{aligned} E(u_t u_{t-1}) &= E(e_t + \theta e_{t-1})(e_{t-1} + \theta e_{t-2}) \\ &= E(e_t e_{t-1}) + \theta E(e_t e_{t-2}) + \theta E\left(e_{t-2}^2\right) + \theta^2 E(e_{t-1} e_{t-2}) \\ &= \theta \sigma^2, \end{aligned}$$

since $E(e_t e_{t-s}) = 0$, for all $s > 0$.

$$\begin{aligned} E(u_t u_{t-2}) &= E(e_t + \theta e_{t-1})(e_{t-2} + \theta e_{t-3}) \\ &= E(e_t e_{t-2}) + \theta E(e_t e_{t-3}) + \theta E(e_{t-1} e_{t-2}) + \theta^2 E(e_{t-2} e_{t-3}) \\ &= 0. \end{aligned}$$

Thus both first and second moments are independent of t and u_t must be (weakly) stationary. This implies that Y_t and X_t are cointegrated.

Question 6

Write short explanations on the following:

(a) The Durbin-Wu-Hausman test for measurement error.

(b) The Box-Cox test for functional form.

(c) The Common Factor test.

Reading for this question

C. Dougherty, '**Introduction to Econometrics**' (third edition) Chapters 4.5, 8.6, 9.1 and 12,6.

Approaching the question

(a) Under measurement error OLS will be inconsistent and IV is to be preferred. If there is no measurement error both OLS and IV will be consistent and OLS will be preferred because it is more efficient. Under H_0: no measurement error, OLS and IV coefficients will not be systematically different and the test statistic is based on the differences between them.

(b) When we want to compare a linear function with a semi-log function (i.e. the dependent variable Y is replaced by $\log Y$) we use the Box-Cox test which has test statistic:

$$\frac{n}{2} \log \frac{RSS_L}{RSS_S},$$

where RSS_L and RSS_S are the larger and smaller RSS respectively and n is the number of observations. The test statistic has a chi-square distribution with 1 df under the null hypothesis of no difference in fit.

(c) The disturbance term being subject to AR(1) autocorrelation is a restricted version of the more general ADL(1,1) model:

$$Y_t = \lambda_1 + \lambda_2 Y_{t-1} + \lambda_3 X_t + \lambda_4 X_{t-1} + \varepsilon_t,$$

with the restriction $\lambda_4 = -\lambda_2\lambda_3$.

The usual F test of a restriction is not appropriate since the restriction is non-linear, hence we use the large sample test:

$$-n \log (RSS_R/RSS_U),$$

where n is the number of observations and RSS_L, RSS_U are the residual sum of squares from the restricted and unrestricted sum of squares of errors respectively.

Question 7

Consider the following model:

$$\boldsymbol{R_t = \alpha_0 + \alpha_1 M_t + \alpha_2 Y_t + u_{1t},} \tag{7.1}$$

$$\boldsymbol{Y_t = \beta_0 + \beta_1 R_t + \beta_2 I_t + u_{2t},} \tag{7.2}$$

where M = Money supply, Y = gross domestic product (GDP), R = the rate of interest and I = capital investment.

(a) Specify which variables you consider are endogenous and which are exogenous. Give reasons for your answers.

(b) Examine the identification of both equations. How would you estimate each of the equations? Describe your method fully.

(c) What difference to your answer in (b) would it make if the GDP equation was to include the variable Y_{t-1}?

Reading for this question

C. Dougherty, '**Introduction to Econometrics**' (third edition) Chapters 9.1 and 9.3.

Approaching the question

(a) Endogenous variables are R_t and Y_t.

Exogenous variables are M_t and I_t.

A brief definition is that endogenous variables are determined within the model, but exogenous variables are determined outside the model. Note that lagged endogenous variables do not contribute towards simultaneous equation bias and hence are lumped together with exogenous to form the 'predetermined' variables in a simultaneous equation model.

(b) The order condition for identification is that the number of restrictions (k) should be greater than the number of included endogenous variables on the RHS.

Denoting the number of endogenous variables (or the number of equations) by G, then:

for equation (7.1), the number of restrictions = 1, $G - 1 = 1$, so this equation is exactly identified;

for equation (7.2), the number of restrictions = 1, $G - 1 = 1$, so this equation is exactly identified.

(c) Since both equations are exactly identified they can both be estimated by Indirect Least Squares (ILS) where the reduced form is estimated by OLS and then the structural form parameters are recovered from these estimates.

(d) If Y_{t-1} is included in the Y_t equation there are now 3 pre-determined variables in the model (2 exogenous and 1 lagged endogenous). The R_t equation now has 3 restrictions so is overidentified and we have to use two stage least squares (2SLS). The Y_t equation has only one restriction, hence is still exactly identified.

Question 8

A regression of operating profits (PROFIT) for 50 UK companies on sales (SALES) and number of employees (NEMP) yields:

$$\mathbf{PROFIT}_i = \underset{(-0.08)}{-1143.9} + \underset{(11.88)}{0.439}\mathbf{SALES}_i - \underset{(-8.46)}{34.44}\mathbf{NEMP}_i + e_i, \qquad R^2 = 0.87,$$

where figures in brackets are t values, i is the company and e_i is the estimated residual.

(a) **Interpret the results of the reported regression.**

(b) **The accountant, who supplied the data, warns that heteroskedasticity is likely to be a problem. Explain why this is so and how the interpretation of the above results are affected if heteroskedasticity is present.**

(c) **If the residuals (e_i) from the above regression are computed, squared and logged and then regressed on the log of sales the estimated equation is:**

$$\log e_i^2 = \underset{(0.60)}{1.032} + \underset{(10.37)}{1.536} \log(\mathbf{SALES}_i) + v_i, \qquad R^2 = 0.72$$

where the figures in brackets are t values and v_i is the estimated residual. What conclusions can you draw from the results of this regression? Explain fully.

Reading for this question

C. Dougherty, '**Introduction to Econometrics**' (third edition) Chapters 3.1, 3.2, 3.8 and 7.1.

Approaching the question

(a) The sample size is 50 and the equation has 3 parameters. The critical t value for 47 df is 2.021 (although you could make a case for a one tail test), hence the coefficients on SALES and NEMP both reject the null that the slope is zero. The coefficients show that a unit increase in sales will increase profits by 0.439 and a unit increase in employees will decrease profits by 34.44 (because costs have increased more than revenue).

R^2 is high indicating that the equation fits well (the equation explains 87% of the variation in profits).

(b) Heteroskedasticity is likely to be a problem because the data is cross-section where firms are likely to be of very different sizes – small firms (in terms of sales and employees) will not see much variation in profits whilst larger firms (in sales and employees) can experience large variations in profits. Heteroskedasticity will give unbiased but inefficient parameter estimates, but the standard errors will be incorrect leading to invalid t and F tests.

(c) The regression shows that heteroskedasticity is present and is caused by the sales variable. The regression is an estimate of the equation $\sigma^2 = A \times \text{sales}^{\beta}$, i.e. the variance of the error term is proportional to sales raised to the power β. With the information that the error variance is proportional to the level of sales^2 we can apply weighted least squares by dividing through by sales.

Examiners' commentaries 2011

20 Elements of econometrics

Important note

This commentary reflects the examination and assessment arrangements for this course in the academic year 2010–11. The format and structure of the examination may change in future years, and any such changes will be publicised on the virtual learning environment (VLE).

Specific comments on questions – Zone A

SECTION A

Answer **all** parts of Question 1 (25 marks in total).

Question 1

(a) **Consider a model:**

$$Y_i = \alpha + \beta X_i + u_i; \quad i = 1, \ldots, 6$$

where $E(u_i) = 0$, $E(u_i^2) = \sigma^2$ and $E(u_i u_j) = 0$ if $i \neq j$.

The observations on X_i's are

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

The OLS estimator of β is $\hat{\beta}$ and

$$\textbf{Var}(\hat{\beta}) = \frac{\sigma^2}{17.5}.$$

An alternative estimator of β is

$$\tilde{\beta} = \frac{1}{8}\left[Y_6 + Y_5 - Y_2 - Y_1\right].$$

Compare the sampling variance of $\tilde{\beta}$ with that of $\hat{\beta}$.

(5 marks)

(b) **Show that the infinite distributed lag model $Y_t = \alpha + \beta \sum_{j=0}^{\infty} \lambda^j X_{t-j} + \epsilon_t$ can be written in terms of X_t and a single lag Y_{t-1}. What estimation problems does this model have?**

(5 marks)

(c) **Explain what is meant when variables are cointegrated. Why is this considered to be important?**

(5 marks)

(d) **Let the regression equation be:**

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

where $E(u_t) = 0$; $E(u_t^2) = \sigma^2$ and $E(u_s u_t) = 0$ if $s \neq t$. X's are fixed in repeated samples.

Obtain the ordinary least squares estimator of β. Show that the OLS estimator of β is linear and unbiased.

(5 marks)

(e) **Explain what you understand by the Durbin-Watson (DW) test. State the assumptions required for performing the DW test.**

(5 marks)

Reading for this question

C. Dougherty, **Introduction to Econometrics**. (third edition) Chapters R.5, R.6, 1.4, 2.5, 12.3 and 12.4.

D. N. Gujarati, **Basic Econometrics**. (fourth edition) Chapters 3A.2, 3.1, 12.6 17.4 and 21.11.

(a) **Approaching the question**
Replace the values of X_is in the equation given by $\tilde{\beta}$. The variance of $\tilde{\beta}$ has to be calculated and compared with the variance of $\hat{\beta}$ (variance of $\hat{\beta}$ is given). While deriving the variance be careful to take into account that $E(u_i u_j) = 0$ if $i \neq j$. The solution follows.
Replace the values of X_is

$$\begin{aligned}\tilde{\beta} &= \frac{1}{8}(Y_6 + Y_5 - Y_2 - Y_1) \\ &= \frac{1}{8}[(\alpha + 6\beta + u_6) + (\alpha + 5\beta + u_5) - (\alpha + 2\beta + u_2) - (\alpha + \beta + u_1)] \\ &= \frac{1}{8}(8\beta + u_6 + u_5 - u_2 - u_1). \qquad (.1)\end{aligned}$$

From (.1), it is easy to see that

$$\begin{aligned}\mathrm{Var}(\tilde{\beta}) &= \mathrm{Var}\left[\frac{1}{8}(u_6 + u_5 - u_2 - u_1)\right] \quad \text{since } \mathrm{Var}(8\beta) = 0 \\ &= \frac{1}{8^2}[\mathrm{Var}(u_6) + \mathrm{Var}(u_5) + \mathrm{Var}(u_2) + \mathrm{Var}(u_1)] \quad \text{since } E(u_i u_j) = 0 \text{ if } i \neq j \\ &= \frac{4\sigma^2}{64} \quad \text{since } \mathrm{Var}(u_i) = \sigma^2 \text{ for all } i \\ &= \frac{\sigma^2}{16}.\end{aligned}$$

(b) **Approaching the question**
Subtract λY_{t-1} from Y_t. The corresponding equation has Y_{t-1} (lagged dependent variable) as an explanatory variable, also the disturbance term becomes autocorrelated. Hence Y_{t-1} and the disturbance term are correlated and this makes the OLS estimator inconsistent. The solution follows.

$$Y_t = \alpha + \beta X_t + \beta\lambda X_{t-1} + \beta\lambda^2 X_{t-2} + \beta\lambda^3 X_{t-3} + \ldots + \epsilon_t. \qquad (.2)$$

If we multiply this equation through by λ and lag we get

$$\lambda Y_{t-1} = \alpha\lambda + \beta\lambda X_{t-1} + \beta\lambda^2 X_{t-2} + \beta\lambda^3 X_{t-3} + \beta\lambda^4 X_{t-4} + \ldots + \lambda\epsilon_{t-1}. \qquad (.3)$$

Now subtract (.3) from (.2) to give

$$\begin{aligned}Y_t - \lambda Y_{t-1} &= \alpha(1-\lambda) + \beta X_t + \epsilon_t - \lambda\epsilon_{t-1} \\ Y_t &= \alpha(1-\lambda) + \beta X_t + \lambda Y_{t-1} + (\epsilon_t - \lambda\epsilon_{t-1})\end{aligned}$$

which we could estimate by OLS except that Y_{t-1} and ϵ_{t-1} are correlated hence a RHS variable is correlated with the error term so OLS produces inconsistent parameter estimates.

(c) **Approaching the question**
Definition of cointegration is required and the relationship between the long-run relationship and cointegration should be explored. The solution follows.
In general, a linear combination of two time series will be non-stationary if one or more of them is non-stationary. The degree of integration of the combination will be equal to that of most highly integrated individual series. For example, a combination of I(1) and I(0) series will be I(1) and a combination of I(1) and I(1) series will be I(1).
If a long-run relationship exists between the time series then the result may be different. Suppose Y_t and X_t are both I(1). A linear combination of Y_t and X_t may be written as $u_t = Y_t - \beta X_t$. If the linear combination u_t is I(0), then Y_t and X_t are said to be cointegrated.
If Y_t and X_t are cointegrated then it implies that a long-run relationship exists between Y_t and X_t. This concept can be generalised. Consider a general linear model

$$Y_t = \beta_1 + \beta_2 X_{2t} + \ldots + \beta_K X_{Kt} + u_t.$$

Then the disturbance term u_t can be thought of as measuring the deviation between components of the model. In the short run the divergence between the components will fluctuate, but if the model is correctly specified there will be a limit to the divergence. Hence, though Ys and Xs are non-stationary, u will be stationary. If there are K variables in the model the maximum number of cointegrating relationships will be $K-1$.
Cointegration is an overriding requirement for any economic model using non-stationary time series data. If the variables do not cointegrate then we have a problem of spurious regression and econometric work becomes almost meaningless. If a cointegrating relationship exists then OLS estimators can be shown to be superconsistent.

(d) **Approaching the question**
It is required to obtain the residual sum of squares (RSS) and minimise the RSS with respect to $\hat{\beta}$ to obtain the OLS estimator. To show unbiasedness it has to be shown that $E[\hat{\beta}] = \beta$. The solution follows.

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

The residual sum of squares (RSS) is

$$\sum \hat{u}_t^2 = \sum (y_t - \hat{\beta} x_t)^2$$

and minimising RSS with respect to $\hat{\beta}$ we get

$$\frac{\partial \sum \hat{u}_t^2}{\partial \hat{\beta}} = 2 \sum (y_t - \hat{\beta} x_t)(-x_t) = 0.$$

Solving we get

$$\hat{\beta}_{OLS} = \frac{\sum x_t y_t}{\sum x_t^2}.$$

<u>Linear:</u>
Define w_t as

$$w_t = \frac{x_t}{\sum x_t^2}$$

then

$$\hat{\beta} = \frac{\sum x_t y_t}{\sum x_t^2} = \sum w_t y_t,$$

this shows that $\hat{\beta}$ is linear in Y.
<u>Unbiasedness:</u>

$$\hat{\beta} = \sum w_t y_t = \sum w_t (\beta x_t + u_t) = \beta \sum w_t x_t + \sum w_t u_t = \beta + \sum w_t u_t$$

since $\sum w_t x_t = 1$. It follows that

$$E(\hat{\beta}) = \beta + \sum w_t E(u_t) = \beta \quad \Rightarrow \quad \text{unbiased.}$$

(e) **Approaching the question**
It is required to define the DW test statistic. Null and alternative hypotheses should be clearly stated. DW test in large samples should be explored. The solution follows.
The Durbin-Watson test is a test for the breakdown of the Gauss-Markov assumption that $E(u_t u_s) = 0$, $t \neq s$. Consider a model

$$\begin{aligned} Y_t &= \beta_0 + \beta_1 X_t + u_t, \quad t = 1, 2, \ldots, T \\ u_t &= \rho u_{t-1} + v_t, \end{aligned}$$

where $E(v_t) = 0$, $E(v_t^2) = \sigma_v^2$ and $E(v_t v_s) = 0$ for $t \neq s$.
The hypotheses are $\mathrm{H}_0 : \rho = 0$ vs. $\mathrm{H}_1 : \rho \neq 0$.
The Durbin-Watson test is

$$dw = \frac{\sum_{t=2}^{T} (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^{T} \hat{u}_t^2} = \frac{\sum_{t=2}^{T} \hat{u}_t^2 + \sum_{t=2}^{T} \hat{u}_{t-1}^2 - 2 \sum_{t=2}^{T} \hat{u}_t \hat{u}_{t-1}}{\sum_{t=1}^{T} \hat{u}_t^2} \cong 2(1 - \hat{\rho})$$

$\rho = +1 \Rightarrow dw = 0$, $\rho = -1 \Rightarrow dw = 4$, $\rho = 0 \Rightarrow dw = 2$.
The distribution of dw depends on the values of explanatory variables hence exact limits are not calculable so tables give two limiting values d_U and d_L. The test is inconclusive if dw lies between d_L and d_U.
Assumptions:

i. Test is only for first order serial correlation.
ii. There is an intercept term in the model.
iii. Test cannot be applied with lagged endogenous variables.

SECTION B

Answer **three** questions from this section (25 marks each).

Question 2

Let the model be:

$$Y_t = \beta_0 + \beta_1 X_{t1} + \beta_2 X_{t2} + u_t; \quad t = 1, 2, \ldots, T.$$

$E(u_t) = 0$ **for all** t**. A researcher suspects that the variance of the disturbance term is Var**$(u_t) = \sigma^2 X_{t1}$**.**

(a) **Explain how the researcher should proceed to test the null hypothesis** $\mathbf{H}_0 : \mathbf{Var}(u_t) = \sigma^2$ **against the alternative hypothesis** $\mathbf{H}_1 : \mathbf{Var}(u_t) = \sigma^2 X_{t1}$**.**

(7 marks)

(b) **If the researcher's suspicion is correct then how will it affect the properties of the ordinary least squares estimators?**

(3 marks)

(c) **Suggest in detail an estimation procedure, which will give best linear unbiased estimates of the parameters when Var**$(u_t) = \sigma^2 X_{t1}$**.**

(5 marks)

(d) **Consider the model**

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

where $E(u_t) = 0$**;** $E(u_t^2) = \sigma^2 x_t^2$ **and** $E(u_s u_t) = 0$ **if** $s \neq t$**, for all** s **and** t**.** x_t **is an observed non-random variable.**

The density function of u_t **is**

$$f(u_t) = \left(2\pi\sigma^2 x_t^2\right)^{-1/2} \exp\left[-\frac{1}{2\sigma^2}\left(\frac{u_t}{x_t}\right)^2\right].$$

Derive the maximum likelihood estimators of α **and** σ^2**.**

(10 marks)

Reading for this question

C. Dougherty, **Introduction to Econometrics**. (third edition) Chapters 7 and 12.6.

C. Dougherty, **Subject Guide** (2011) Chapter 7.

D. N. Gujarati, **Basic Econometrics**. (fourth edition) Chapters 4A, 11.2, 11.4 and 11.5

(a) **Approaching this question**

In Dougherty, both the Goldfeldt-Quandt test and White test are given. Candidates are required to describe either one of them. Both tests are explained below:

Goldfeld-Quandt Test:

This test assumes that $\text{Var}(u_t)$ is proportional to the size of one of the RHS variables (say X_1). The observations are ranked by X_1. Run a separate regression for the first n_1 $(< n/2)$ and the last n_1 observations – the middle $(n - 2n_1)$ observations are not used. If heteroskedasticity is present the RSS from the two regressions will differ. Form the test RSS_2/RSS_1 where RSS_1 is the residual sum of squares from the first n_1 observations and RSS_2 is the residual sum of squares from the last n_1 observations. The test statistic will have an F-distribution with $(n_1 - k, n_1 - k)$ degrees of freedom where k is the number of parameters in the equation. The null hypothesis is that the variances are homoskedastic.

White Test:

The White test looks for more general evidence of association between the variance of the error term and the regressors. Regress the squared residuals from the original regression on the regressors from that model, together with the squares and the cross-products of those variables. The test statistic is nR^2 where n is the sample size and R^2 is the R^2 from the White regression. This test statistic has a chi-square distribution with degrees of freedom equal to the number of regressors. The test assumes the sample size is large.

(b) **Approaching this question**

i. OLS estimator is unbiased and consistent, but it is no longer efficient, the standard errors are wrong.

ii. t- and F-tests are invalid.

(c) **Approaching this question**

Weighted least squares should be used. Model has to be transformed in such a way that the transformed disturbance term is homoskedastic. OLS is used to estimate the parameters of the transformed model. The resulting estimator is known as the weighted least squares estimator. The solution follows.

<u>Weighted least squares:</u>

The model is

$$Y_t = \beta_0 + \beta_1 X_{t1} + \beta_2 X_{t2} + u_t, \quad t = 1, 2, \ldots, T. \tag{.4}$$

Divide (.4) by $\sqrt{X_{t1}}$ to get

$$\frac{Y_t}{\sqrt{X_{t1}}} = \frac{\beta_0}{\sqrt{X_{t1}}} + \beta_1\sqrt{X_{t1}} + \frac{\beta_2 X_{t2}}{\sqrt{X_{t1}}} + \frac{u_t}{\sqrt{X_{t1}}}, \quad t = 1, 2, \ldots, T. \tag{.5}$$

Equation (.5) is the transformed model. Variance of the transformed disturbance term is

$$\text{Var}\left(\frac{u_t}{\sqrt{X_{t1}}}\right) = \frac{\text{Var}(u_t)}{X_{t1}} = \frac{\sigma^2 X_{t1}}{X_{t1}} = \sigma^2.$$

We see that the variance of the transformed disturbance term is constant. Hence, OLS can be applied to (.5) to obtain efficient estimators of β_0, β_1 and β_2.

(d) **Approaching the question**

To obtain the MLE, the likelihood function (L) has to be obtained. The estimator which maximises L is the MLE. As L in general is highly non-linear in parameters, instead of L always $\ln L$ is maximised. Both give the same maximum. Detailed working follows.

The log likelihood function is

$$\ln L = -\frac{T}{2}\ln 2\pi - \frac{T}{2}\ln\sigma^2 - \frac{1}{2}\sum \ln x_t^2 - \frac{1}{2\sigma^2}\sum\left(\frac{y_t}{x_t}-\alpha\right)^2.$$

The first order conditions are:

$$\frac{\partial \ln L}{\partial \alpha} = \frac{1}{\sigma^2}\sum\left(\frac{y_t}{x_t}-\alpha\right) = 0 \qquad (.6)$$

$$\frac{\partial \ln L}{\partial \sigma^2} = -\frac{T}{2\sigma^2} + \frac{1}{2\sigma^4}\sum\left(\frac{y_t}{x_t}-\alpha\right)^2 = 0 \qquad (.7)$$

Solving (.6) and (.7) ML estimators of α and σ^2 are obtained as:

$$\hat{\alpha}_{MLE} = \frac{1}{T}\sum\left(\frac{y_t}{x_t}\right)$$

$$\hat{\sigma}^2_{MLE} = \frac{1}{T}\sum\left(\frac{y_t}{x_t}-\hat{\alpha}_{MLE}\right)^2.$$

Question 3

Explain and discuss the following:

(a) Difference stationary and trend stationary processes.

(9 marks)

(b) Effects on the properties of ordinary least squares estimator when relevant variables are excluded and irrelevant variables are included in the equation.

(8 marks)

(c) Dummy variables.

(8 marks)

Reading for this question

C. Dougherty, **Introduction to Econometrics**. (third edition) Chapters 5, 6.2, 6.3 and 13.1.

D. N. Gujarati, **Basic Econometrics**. (fourth edition) Chapters 9.1, 9.2, 9.5, 9.6, 9.7 13.3 and 21.10.

(a) **Approaching the question**

Definitions of difference stationary series and trend stationary series are required. These should also be illustrated with an example. Behaviour of time series in both cases should be explained. The solution follows.

If after removing the trend from a nonstationary series the resulting variable becomes stationary, then the variable is called *trend stationary.* Let

$$Z_t = X_t - \alpha_1 t = \alpha_0 + u_t, \quad t = 1, 2, \ldots, T,$$

where $E(u_t) = 0$, $\text{Var}(u_t) = \sigma^2$ and $E(u_t u_{t-s}) = 0$ for all s and t. Then

$$\begin{aligned} E(Z_t) &= E(\alpha_0 + u_t) = \alpha_0 \\ \text{Var}(Z_t) &= \text{Var}(\alpha_0 + u_t) = \sigma^2 \\ \text{Cov}(Z_t, Z_{t-s}) &= E\left[(Z_t - E(Z_t))(Z_{t-s} - E(Z_{t-s}))\right] = E(u_t u_{t-s}) = 0. \end{aligned}$$

This means that Z_t has constant mean and variance for all t, and covariance is zero for all s. It implies that the series X_t is trend-stationary.

If a nonstationary process can be transformed into a stationary process by differencing then the series is said to be difference-stationary.

Let X_t be a random walk with a drift

$$X_t = \beta_0 + X_{t-1} + \epsilon_t, \tag{.8}$$

where $E(\epsilon_t) = 0$, $\mathrm{Var}(\epsilon_t) = \sigma^2$ and $E(\epsilon_t \epsilon_s) = 0$ for all s and t, $s \neq t$.

Subtract X_{t-1} from both sides of (.8) to get

$$\Delta X_t = X_t - X_{t-1} = \beta_0 + \epsilon_t.$$

It can be easily checked that $E[\Delta X_t] = \beta_0$, $\mathrm{Var}(\Delta X_t) = \sigma_\epsilon^2$ and $\mathrm{Cov}(\Delta X_t, \Delta X_{t-s}) = 0$ for all s and t. This means that ΔX_t is stationary. This implies that X_t is difference-stationary.

It is important to know whether a variable is difference- or trend-stationary because for difference-stationary variables shocks have a permanent effect whereas for trend-stationary variables shocks are transitory.

(b) **Approaching the question**

Omitted variable bias should be explained and derived with the help of a model. The solution follows.

Omitted variable bias occurs when a relevant variable is omitted from the estimated model. Suppose the 'true' model is

$$y_t = \beta_1 x_{1t} + \beta_2 x_{2t} + u_t$$

but the estimated model is

$$y_t = \beta_1 x_{1t} + v_t.$$

The OLS estimator of β_1 is

$$\begin{aligned}
\hat{\beta}_1 &= \frac{\sum x_{1t} y_t}{\sum x_{1t}^2} = \frac{\sum x_{1t}(\beta_1 x_{1t} + \beta_2 x_{2t} + u_t)}{\sum x_{1t}^2} = \beta_1 + \beta_2 \frac{\sum x_{1t} x_{2t}}{\sum x_{1t}^2} + \frac{\sum x_{1t} u_t}{\sum x_{1t}^2} \\
E(\hat{\beta}_1) &= \beta_1 + \beta_2 \frac{\sum x_{1t} x_{2t}}{\sum x_{1t}^2}
\end{aligned}$$

since $E(u_t) = 0$. The bias is

$$\beta_2 \frac{\sum x_{1t} x_{2t}}{\sum x_{1t}^2}$$

which depends on the value of β_2 and the covariance between x_{1t} and x_{2t}. A good answer would explore this result further.

Suppose the true model is

$$y_t = \beta_1 x_{1t} + v_t$$

but the estimated model is

$$y_t = \beta_1 x_{1t} + \beta_2 x_{2t} + u_t$$

then $\hat{\beta}_1$ will be unbiased, but inefficient.

(c) **Approaching the question**

To answer this question the definition of a dummy variable should be given and also how dummy variables can be used to test for a change in intercept or a change in slope. How a change in slope or intercept and also joint test in change in intercept and slope can be conducted should be explained. The dummy variable trap should be explained. Technical details should be given.

Question 4

The following estimates were calculated from a sample of 7,634 women respondents from the General Household Survey 1995. The dependent variable takes the value 1 if the woman was in paid employment, and 0 otherwise.

	OLS	**Logit**	**Probit**
high	**0.093**	**0.423**	**0.259**
	(0.015)	**(0.071)**	**(0.043)**
noqual	**−0.210**	**−0.898**	**−0.554**
	(0.013)	**(0.056)**	**(0.035)**
age	**0.038**	**0.173**	**0.107**
	(0.003)	**(0.124)**	**(0.008)**
age2	**−0.051**	**−0.230**	**−0.142**
	(0.003)	**(0.069)**	**(0.009)**
mar	**0.024**	**0.103**	**0.063**
	(0.009)	**(0.057)**	**(0.035)**
Constant	**−0.068**	**−2.587**	**−1.593**
	(0.049)	**(0.225)**	**(0.137)**

Where high is 1 if the respondent has a higher educational qualification, 0 otherwise; noqual is 1 if the respondent has no qualifications, 0 otherwise; age is age in years; age2 is (age $\times$ age)/100; mar is 1 if married, 0 otherwise. Conventionally calculated standard errors are in brackets for the ordinary least squares (OLS) results, asymptotic standard errors are in brackets elsewhere.

(a) Explain how Probit estimates are calculated when the model has no intercept and only one explanatory variable.

(7 marks)

(b) Using all three sets of estimates, test the null hypothesis that the coefficient of mar is zero. Which test statistics would you consider more reliable? Explain.

(8 marks)

(c) Using OLS and Probit estimates, calculate the estimated probabilities of being in employment for a married woman aged 40 with a higher educational qualification. Comment on your results.

(6 marks)

(d) Test the null hypothesis that all the slope coefficients of the probit **model are jointly equal to zero. It is given that**

$$\ln L_R = -416.01$$
$$\ln L_U = -321.25$$

where $\ln L_R$ **and** $\ln L_U$ **are the log of the likelihood from the restricted and the unrestricted** probit **models respectively.**

(4 marks)

Reading for this question

C. Dougherty, **Introduction to Econometrics**. (third edition) Chapters 10.3 and 10.6.

C. Dougherty, **Subject Guide** (2011) Chapter 10.

D. N. Gujarati, **Basic Econometrics**. (fourth edition) Chapters 8A, and 15.9

(a) **Approaching the question**

The probit model uses the cumulative standardised normal distribution. The maximum likelihood technique is used to obtain the estimates of the parameters. Estimates have the standard maximum likelihood properties, i.e. the estimators are consistent, asymptotically efficient and asymptotically normally distributed. [For technical details see Dougherty (third edition) Ch. 10.3].

- **Approaching the question**

  To test the null hypothesis a t-test should be used. t tables are attached with the examination paper. Candidates should know how to look at the critical values. Null and alternative hypotheses should be clearly stated. The solution follows.

$$\begin{aligned} \mathrm{H}_0 : \text{coefficient of mar } &= 0 \\ \mathrm{H}_1 : \text{coefficient of mar } &\neq 0. \end{aligned}$$

t-test statistics are 2.67 (OLS), 1.81 (logit), 1.8 (probit).

The test based on the OLS estimates gives outright rejection. However, the standard errors for this test are wrongly calculated because of the heteroskedasticity of the error term. Thus we prefer probit or logit estimates.

(c) **Approaching the question**

A large number of students wrongly thought that to obtain the probability in the case of Probit a sophisticated calculator is needed. The probability in the case of Probit can be read directly from the attached normal tables. The solution follows:

$$\begin{aligned} 0.093 + 0.038 \times 40 - 0.051 \times 16 + 0.024 - 0.068 &= 0.753 \quad \text{(OLS)} \\ 0.259 + 0.107 \times 40 - 0.142 \times 16 + 0.063 - 1.593 &= 0.737 \quad \text{(Probit)} \end{aligned}$$

gives a probability of 0.74 (approximately).

They are all fairly close. The OLS estimates do not fall outside the probability bounds.

(d) **Approaching the question**

To test the null hypothesis a large sample likelihood ratio test should be used. In large samples $-2[\ln L_R - \ln L_U]$ is distributed as a chi-square with degrees of freedom equal to the number of restrictions imposed by the null hypothesis. Null and alternative hypotheses should be clearly stated. The solution follows.

$$\begin{aligned} \mathrm{H}_0 : \text{All the slope coefficients are } &= 0 \\ \mathrm{H}_1 : \text{All the slope coefficients are } &\neq 0 \\ -2[\ln L_R - \ln L_U] &\sim \chi^2_5 \\ -2[-416.01 - (-321.25)] &= 189.92 \end{aligned}$$

Critical value of χ^2_5 at 5% level of significance is 11.07, hence reject H_0.

Question 5

(a) **Explain what you understand by autocorrelation of the disturbance term in a regression model? What are the causes of autocorrelation?**

(5 marks)

(b) **The following equation was estimated by Ordinary Least Squares using 37 annual observations of UK aggregate data. The dependent variable (cloth$_t$) is the log of expenditure on clothing at 1995 prices, yd$_t$ is the log of aggregate disposable income at 1995 prices, pc$_t$ is the log of the price of clothing relative to all consumer prices, ps$_t$ is the log of the price of shoes relative to all consumer prices.**

$$\mathbf{cloth}_t = \underset{(1.531)}{-3.256} + \underset{(0.118)}{1.021}\mathbf{yd}_t - \underset{(0.132)}{0.240}\mathbf{pc}_t - \underset{(0.185)}{0.429}\mathbf{ps}_t + e_t$$

standard errors in brackets, e_t is an OLS residual, $n = 37$, $R^2 = 0.992$, $F = 1,364.0$, $s = 0.041$, **DW = 0.94. DW is the Durbin-Watson statistic.**

i. Test the hypothesis that the coefficient of yd_t is one.

(3 marks)

ii. Construct a 95% confidence interval for the coefficient of pc_t.

(3 marks)

iii. Give any assumptions which your results in i. and ii. above require.

(3 marks)

iv. Using the statistics given above, would you conclude that any of the assumptions you have given in iii. above are not valid here? Give reasons.

(5 marks)

v. What information do these estimates provide about the demand for clothing in the UK?

(6 marks)

Reading for this question

C. Dougherty, **Introduction to Econometrics**. (third edition) Chapters 2.8, 12.1, 12.2 and 12.3.

C. Dougherty, **Subject Guide** (2011) Chapter 12.

D. N. Gujarati, **Basic Econometrics**. (fourth edition) Chapters 5.6, 5.7, 12.1, 12.2 and 12.5.

(a) **Approaching the question**

Autocorrelation should be defined and the consequences of autocorrelation on OLS estimates should be discussed. What causes autocorrelation should also be discussed. The solution follows.

If the error term u_t in the model $Y_t = \beta_0 + \beta_1 X_t + u_t$ is such that $E(u_t u_s) \neq 0$ then the error term is said to be serially correlated. This a very general condition and it is generally necessary to assume the more restrictive condition that $u_t = \rho u_{t-1} + v_t$ where $|\rho| < 1$. The consequence is that ordinary least squares (OLS) parameter estimates are unbiased but inefficient and that their standard errors, and hence t-values, are incorrect.

The most common cause of serially correlated errors is that the model is mis-specified by omitting a variable. If this omitted variable is itself serially correlated then the error term will mirror this behaviour. Other causes of serial correlation are 'cobweb' type behaviour and mis-specified dynamic behaviour.

(b) i. **Approaching the question**

To test the null hypothesis, t-statistics should be used. Null and alternative hypotheses should be clearly stated. The solution follows.

Suppose the population model is

$$\text{cloth}_t = \beta_0 + \beta_1 \text{yd}_t + \beta_2 \text{pc}_t + \beta_3 \text{ps}_t + u_t$$

We test $\text{H}_0 : \beta_1 = 1$ vs. $\text{H}_1 : \beta_1 \neq 1$.

Then $(1.021 - 1)/0.118 = 0.1779$; this is t with 33 degrees of freedom.

The 5% critical value is 2.035 (approximately), thus do not reject the null hypothesis.

ii. **Approaching the question**

A 95% confidence interval for the coefficient of pc_t is

$$\beta_1 = -0.240 \pm 2.035 \times 0.132 = -0.240 \pm 0.2686,$$

that is $-0.509 < \beta_1 < 0.0286$.

iii. **Approaching the question**

Assumptions: the model is correct, $E(u_t^2) = \sigma^2$, $E(u_t u_s) = 0$ if $t \neq s$ and $u_t \sim N(0, \sigma^2)$.

iv. **Approaching the question**

To test autocorrelation the DW test statistic should be used. The DW table is attached with the examination paper. Null and alternative hypotheses should be clearly stated. The solution follows.

We test H_0 : No serial correlation vs. H_1 : Serial correlation.

From the DW table for $n = 37$, $k' = 3$ we see that $d_L = 1.31$, $d_U = 1.66$ thus reject the null hypothesis of no serial correlation. As a result estimates remain unbiased, but are no longer efficient and the standard errors are wrong. Thus both inferences made in (i) and (ii) above are based on incorrect standard errors. t- and F-tests are invalid.

v. **Approaching the question** As it is a double log model, coefficients are elasticities. Candidates are required to discuss income elasticity, own-price elasticity and cross-price elasticity. The solution follows.

Income elasticity close to one (reasonable for clothes as they are a mixture of luxuries and necessities). Own-price elasticity is negative, as expected, and the point estimate in absolute terms is comparatively small.

The cross-price elasticity with shoes is estimated to be -0.429. This suggests that shoes and clothes are complements not substitutes. The cross-price effect appears to be stronger than the own-price effect but because the standard errors are wrongly calculated not much can be concluded from this.

Question 6

In the model

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

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

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

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

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

(10 marks)

(b) **Obtain an expression for plim**$(\hat{\beta} - \beta)$**. Comment on the sign of this expression.**

(3 marks)

(c) **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** $E(w_t) = 0$**,** $E(u_t w_t) = E(x_t w_t) = E(y_t w_t) = 0$**. Let** $\hat{\beta}$ **be the OLS estimator of** β **from regressing** y_t^* **on** x_t**. Is** $\hat{\beta}$ **consistent? Explain in detail.**

(7 marks)

(d) **Suppose in the given model, both** y_t **and** x_t **are measured with errors and data is available only on** y_t^* **and** x_t^* **where** y_t^* **and** x_t^* **are defined above, respectively. Discuss whether the OLS estimator of** β**, from regressing** y_t^* **on** x_t^* **will be consistent or inconsistent.**

(5 marks)

Reading for this question

C. Dougherty, **Introduction to Econometrics**. (third edition) Chapter 8.4.

C. Dougherty, **Subject Guide** (2011) Chapter 8.

D. N. Gujarati, **Basic Econometrics**. (fourth edition) Chapter 13.5.

(a) **Approaching the question**

plim should be used to answer this question. A detailed derivation is required. The solution follows.

$$y_t = \beta x_t + u_t \quad \text{where} \quad x_t^* = x_t + v_t$$

$$\begin{aligned}
\hat{\beta} = \frac{\sum x_t^* y_t}{\sum x_t^{*2}} &= \frac{\sum (x_t + v_t)(\beta x_t + u_t)}{\sum (x_t + v_t)^2} \\
&= \frac{\beta \sum x_t^2 + \sum x_t u_t + \beta \sum x_t v_t + \sum v_t u_t}{\sum x_t^2 + \sum v_t^2 + 2 \sum x_t v_t} \\
\text{plim } \hat{\beta} &= \frac{\text{plim}[\beta \sum x_t^2 + \sum x_t u_t + \beta \sum x_t v_t + \sum v_t u_t]/T}{\text{plim}[\sum x_t^2 + \sum v_t^2 + 2 \sum x_t v_t]/T} \\
&= \frac{\beta \sigma_x^2}{\sigma_x^2 + \sigma_v^2} \neq \beta \quad \Rightarrow \quad \text{inconsistent.}
\end{aligned}$$

(b) **Approaching the question**

$$\text{plim } (\hat{\beta} - \beta) = \text{plim } \hat{\beta} - \beta = \frac{\beta \sigma_x^2}{\sigma_x^2 + \sigma_v^2} - \beta = -\frac{\beta \sigma_v^2}{\sigma_x^2 + \sigma_v^2}.$$

If $\beta > 0$, then $\text{plim}(\hat{\beta} - \beta) < 0$.

(c) **Approaching the question**

plim should be used to answer this question. A detailed derivation is required. The solution follows.

$$\begin{aligned}
\hat{\beta} = \frac{\sum x_t y_t^*}{\sum x_t^2} &= \frac{\sum x_t (y_t + w_t)}{\sum x_t^2} \\
&= \frac{\sum x_t (\beta x_t + u_t + w_t)}{\sum x_t^2} \\
&= \frac{\beta \sum x_t^2 + \sum x_t u_t + \sum x_t w_t}{\sum x_t^2}.
\end{aligned}$$

Then

$$\text{plim } \hat{\beta} = \frac{\text{plim}[\beta \sum x_t^2 + \sum x_t u_t + \sum x_t w_t]/T}{\text{plim}[\sum x_t^2]/T} = \frac{\beta \sigma_x^2}{\sigma_x^2} = \beta \quad \Rightarrow \quad \text{consistent.}$$

(d) **Approaching the question**

As the question does not ask for a detailed derivation, only an intuitive discussion is required.

This will give the same result as part (a).

Examiners' commentaries 2011

20 Elements of econometrics

Important note

This commentary reflects the examination and assessment arrangements for this course in the academic year 2010–11. The format and structure of the examination may change in future years, and any such changes will be publicised on the virtual learning environment (VLE).

Specific comments on questions – Zone B

SECTION A

Answer **all** parts of Question 1 (25 marks in total).

Question 1

(a) **Show that least squares applied to a linear model with serially correlated errors can yield unbiased parameter estimates. Discuss the consequences for conventionally calculated least squares estimates of parameter standard errors.**

(5 marks)

(b) **Unbiasedness is not the only criterion by which to judge a statistical estimator. What other criteria are used and why are they used? Explain fully.**

(5 marks)

(c) **Show that** $\mathrm{var}(Y_t) = \mathrm{var}(\hat{Y}_t) + \mathrm{var}(\hat{u}_t)$ **for the simple linear model** $Y_t = \beta_0 + \beta_1 X_t + u_t$ **where** b_0 **and** b_1 **are ordinary least squares estimates of** β_0 **and** β_1 **and** $\hat{u}_t = Y_t - b_0 - b_1 X_t$**. Explain how this expression is related to the properties of** R^2**, the coefficient of determination.**

(5 marks)

(d) **What is meant by the term 'spurious regression'? Explain how you would determine whether or not an estimated equation is a 'spurious regression'?**

(5 marks)

(e) **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.**

(5 marks)

Reading for this question

C. Dougherty, **Introduction to Econometrics**. (third edition) Chapters R.6, 1.7, 8.6, 9.3, 12.3 and 13.2.

C. Dougherty, **Subject Guide** (2011) Chapters 8 and 9.

D. N. Gujarati, **Basic Econometrics**. (fourth edition) A.7, 3.5, 12.2, 13.5, 20.4, 21.7.

(a) **Approaching the question**

Candidates should explain their answer using a simple two-variable linear model. A detailed derivation of results is required. The solution follows.

Let the model be

$$y_t = \beta x_t + u_t$$

where $u_t = \rho u_{t-1} + \epsilon_t$, $E(\epsilon_t) = 0$, $\mathrm{Var}(\epsilon_t) = \sigma_\epsilon^2$ and $E(\epsilon_s \epsilon_t) = 0$ if $s \neq t$. Then

$$E(b) = \beta + E\left(\frac{\sum x_t u_t}{\sum x_t^2}\right) = \beta + \frac{1}{\sum x_t^2} E\left(\sum x_t u_t\right) = \beta + \frac{\sum x_t E(u_t)}{\sum x_t^2} = \beta$$

as $E(u_t) = 0$. Then

$$\begin{aligned}
\mathrm{Var}(b) = E(b-\beta)^2 &= E\left(\frac{\sum x_t u_t}{\sum x_t^2}\right)^2 \\
&= \frac{1}{\left(\sum x_t^2\right)^2} \mathrm{Var}\left(\sum x_t u_t\right) \\
&= \frac{\sigma^2}{\left(\sum x_t^2\right)^2}\left(\sum x_t^2 + 2\rho \sum x_t x_{t-1} + 2\rho^2 \sum x_t x_{t-2} + \ldots\right) \quad \text{since } \mathrm{Var}(u_t) = \sigma^2 \\
&= \frac{\sigma^2}{\sum x_t^2}\left(1 + 2\rho \frac{\sum x_t x_{t-1}}{\sum x_t^2} + 2\rho^2 \frac{\sum x_t x_{t-2}}{\sum x_t^2} + \ldots\right)
\end{aligned}$$

which is $\neq \sigma^2 / \sum x_t^2$ (the OLS estimator of $\mathrm{Var}(b)$).

Hence the OLS standard error is incorrect and the resulting t-tests and F-tests are invalid.

(b) **Approaching the question**

The obvious candidates for alternative criteria are small variance (efficiency), MSE and consistency. A reasonable discussion of consistency is expected.

(c) **Approaching the question**

It has to be shown that $\mathrm{Cov}(\hat{Y}_t, \hat{u}_t) = 0$. To show this it should be explained that $\mathrm{Cov}(X_t, \hat{u}_t) = 0$. R^2 should be defined and it should be discussed that Total sum of squares = Explained sum of squares + Residual sum of squares. The solution follows.

$$Y_t = \hat{Y}_t + \hat{u}_t \quad \text{hence} \quad \mathrm{Var}(Y_t) = \mathrm{Var}(\hat{Y}_t) + \mathrm{Var}(\hat{u}_t) + 2\mathrm{Cov}(\hat{Y}_t, \hat{u}_t).$$

Now,

$$\mathrm{Cov}(\hat{Y}_t, \hat{u}_t) = \mathrm{Cov}(b_0 + b_1 X_t, \hat{u}_t) = \mathrm{Cov}(b_0, \hat{u}_t) + b_1 \mathrm{Cov}(X_t, \hat{u}_t),$$

which is 0 since $\mathrm{Cov}(b_0, \hat{u}_t) = 0$ since b_0 is a constant and $\mathrm{Cov}(X_t, \hat{u}_t) = 0$ by the normal equations. Hence $\mathrm{Var}(Y_t) = \mathrm{Var}(\hat{Y}_t) + \mathrm{Var}(\hat{u}_t)$ and multiplying by T gives TSS = ESS + RSS. Now $R^2 = 1 - \mathrm{RSS}/\mathrm{TSS} = \mathrm{ESS}/\mathrm{TSS}$ as $R^2 \geq 0$ since ESS and TSS must be positive and TSS > 0. Also $R^2 \leq 1$ since ESS $\leq$ TSS as RSS ≥ 0. [Note: Variances and covariances are sample variances and sample covariances.]

(d) **Approaching the question**

Meaning of spurious regression should be discussed. With the help of simple models it should be demonstrated that, though there may be no relationship between the dependent variable and explanatory variables, spurious regression may result in showing that explanatory variables significantly explain variations in the dependent variable. A detailed discussion follows.

The assumption that X_t and Y_t are stationary is crucial for the standard properties of OLS. When the series are non-stationary but independent, however, it is often the case that both series show an apparent trend which gives rise to a regression with high R^2 and a significant t-value for the slope parameter leading to the conclusion that there is a significant relationship between Y_t and X_t. This is 'spurious regression'.

Suppose

$$X_t = \beta_0 + X_{t-1} + u_t$$

and

$$Y_t = \beta_0 + Y_{t-1} + \epsilon_t.$$

X_t and Y_t are both non-stationary (suppose they are both I(1)). If we regress Y_t on X_t, i.e.

$$Y_t = \pi_0 + \pi_1 X_t + v_t$$

Granger and Newbold ['Spurious Regressions in Econometrics', *Journal of Econometrics*, 2, pp. 111–120] have shown that although there is no relationship between Y and X, the regression will produce a t-ratio which *will reject the null hypothesis* $\text{H}_0 : \pi_1 = 0$.

The reason for this result is that if $\text{H}_0 : \pi_1 = 0$, then

$$Y_t = \pi_0 + v_t$$

and since Y_t is I(1) and π_0 is constant it follows that v_t must be I(1). This violates the standard distributional theory based on the assumption that v_t is stationary, i.e. v_t is I(0). Hence the misleading result.

According to Granger-Newbold, $R^2 >$ Durbin-Watson statistic is a good rule-of-thumb to suspect that the estimated regression is spurious.

(e) **Approaching the question**

Ordinary least squares (OLS) gives inconsistent parameter estimates when the independent variable is correlated with the error term. In this case IV estimation will give consistent parameter estimates which will be superior to the inconsistent parameter estimates obtained by OLS – but in small samples or when instruments are weak IV would not necessarily be superior. Note that the instrumental variable must satisfy the basic requirements that it has non-zero covariance with the independent variables and zero covariance with the error term.

i. When the independent variable is subject to measurement error OLS parameter estimates will be inconsistent. IV estimates will be consistent.

ii. In the case of simultaneous equations OLS parameter estimates will be inconsistent whereas IV estimates will be consistent.

Any one of the above two should be illustrated with a simple model.

SECTION B

Answer **three** questions from this section (25 marks each).

Question 2

(a) **An econometrician believes that consumption expenditure is dependent on the last** two **years of disposable income. Assuming data is available** quarterly **write down the resulting model. What econometric problems, if any, are likely to result from your model and why?**

(8 marks)

(b) **Explain how you would test the hypothesis that consumption is dependent on disposable income for the last year only against the alternative hypothesis that consumption is dependent on the last** two **years. Give details of the information you need for this test.**

(8 marks)

(c) **A colleague suggests that you should use an infinite lag model instead of the model you wrote in (a). How would you do this and what are the advantages and disadvantages of this approach?**

(9 marks)

Reading for this question

C. Dougherty, **Introduction to Econometrics**. (third edition) Chapters 3.4 and 6.5.

D. N. Gujarati, **Basic Econometrics**. (fourth edition), Chapters 8.5 and 17.3.

(a) **Approaching the question**

Assuming the data is quarterly then 2 years of data will give a model with 8 lags. The econometric problem is one of multicollinearity, lagged values are likely to be very similar to each other unless the variable is essentially random and this is unlikely to be true for disposable income which is going to be slowly changing.

(b) **Approaching the question**

Model should be specified with lags 0 to 7 and an F-test should be used to test the restrictions. The formula for the F-test of restrictions should be given. Degrees of freedom should be clearly specified. The solution follows.

The 2-year model will be

$$C_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \beta_3 Y_{t-3} + \beta_4 Y_{t-4} + \beta_5 Y_{t-5} + \beta_6 Y_{t-6} + \beta_7 Y_{t-7} + \beta_8 Y_{t-8} + u_t.$$

Note that you could construct the model with lags 0 to 7 and this would be perfectly acceptable and the hypothesis that consumption depends on the last year only will be defined by $\mathrm{H}_0 : \beta_5 = \beta_6 = \beta_7 = \beta_8 = 0$ against the alternative that at least one of the βs are non-zero. This test can be accomplished by an F-test

$$F_{4,T-9} = \frac{(RSS_R - RSS_U)/4}{RSS_U/(T-9)},$$

where T is the number of observations, RSS_R is the restricted residual sum of squares under H_0 and RSS_U is the unrestricted residual sum of squares obtained by running the regression on all the lagged variables.

(c) **Approaching the question**

The infinite lagged variable model would be constructed with geometrically declining weights, for example

$$C_t = \beta_0 + \beta \sum_{j=0}^{\infty} \lambda^j Y_{t-j} + u_t$$

where $|\lambda| < 1$. The advantage of this approach is that there are only two parameters to estimate but against this the weights of successive values of Y_t will be forced to decline geometrically which may not be appropriate and in order to estimate the parameters we shall have to transform the above equation to include a lagged dependent variable. Note that if we lag the above equation by one period and multiply the resulting equation by λ we get

$$\lambda C_{t-1} = \lambda\beta_0 + \beta \sum_{j=0}^{\infty} \lambda^j Y_{t-j-1} + \lambda u_{t-1}$$

so by subtracting

$$C_t - \lambda C_{t-1} = (1-\lambda)\beta_0 + \beta Y_t + u_t - \lambda u_{t-1}$$

which can be easily estimated using a lagged value of C_t (note that the error term is serially correlated but this will not, of itself, give rise to biased parameter estimates).

Question 3

(a) **A researcher is investigating the impact of advertising on sales using cross-section data from firms producing recreational goods. For each firm there**

are data on sales, S, and expenditure on advertising, A, both measured in suitable units, for a recent year. The researcher proposes the following model:

$$\begin{aligned} S_t &= \alpha_1 + \alpha_2 A_t + u_{St} \\ A_t &= \beta_1 + \beta_2 S_t + u_{At} \end{aligned}$$

where u_S and u_A are disturbance terms. The first relationship reflects the positive effect of advertising on sales, and the second the fact that largest firms, as measured by sales, tend to spend most on advertising. Give a mathematical analysis of what would happen if the researcher tried to fit the model using ordinary least squares (OLS).

(15 marks)

(b) **The researcher discovers that last year's advertising budget, A_{t-1}, is also an important determinant of A_t, so that the model is**

$$\begin{aligned} S_t &= \alpha_1 + \alpha_2 A_t + u_{St} \\ A_t &= \beta_1 + \beta_2 S_t + \beta_3 A_{t-1} + u_{At} \end{aligned}$$

Explain how this information could be used to obtain a consistent estimator of α_2, and show that it is consistent.

(10 marks)

Reading for this question

C. Dougherty, **Introduction to Econometrics**. (third edition) Chapter 9.3.

C. Dougherty, **Subject Guide** (2011), Chapter 9.

D. N. Gujarati, **Basic Econometrics**. (fourth edition), Chapters 19.2 and 20.3.

(a) **Approaching the question**

It should be shown that both equations are underidentified by the order condition. There is no way we can have a meaningful estimation of these equations. A detailed working follows.

Reduced forms corresponding to A and S are

$$\begin{aligned} A_t &= \frac{\beta_1 + \alpha_1\beta_2}{1 - \alpha_1\beta_2} + \frac{\beta_2 u_{St} + u_{At}}{1 - \alpha_1\beta_2} = \Pi_A + V_A \\ S_t &= \frac{\alpha_1 + \alpha_2\beta_1}{1 - \alpha_1\beta_2} + \frac{\alpha_2 u_{At} + u_{St}}{1 - \alpha_1\beta_2} = \Pi_S + V_S \end{aligned}$$

where Π_A and Π_S are reduced form coefficients and V_A and V_S are reduced form disturbances. Hence we have four structural coefficients and only two reduced form coefficients. The two reduced form coefficients contain all the four structural coefficients, but there is no way four unknown structural coefficients can be estimated by only two reduced form coefficients.

Another more illustrative way of looking at this problem is the following:

Multiply S_t by w and A_t by $(1 - w)$, where w is a non-zero weight and add the two equations to get

$$wS_t + (1-w)A_t = w\alpha_1 + (1-w)\beta_1 + (1-w)\beta_2 S_t + \alpha_2 w A_t + (1-w)u_{At} + wu_{St}.$$

Rearranging we get

$$\begin{aligned} S_t &= \frac{w\alpha_1 + (1-w)\beta_1}{w + (1-w)\beta_2} + \frac{\alpha_2 w - (1-w)}{w + (1-w)\beta_2} A_t + \frac{(1-w)u_{At} + wu_{St}}{w + (1-w)\beta_2} \\ &= \alpha_1^* + \alpha_2^* A_t + u_{St}^* \end{aligned} \tag{.1}$$

Equation (.1) is observationally equivalent to the Sales equation. If we estimate the Sales equation we do not know whether we have estimated α_1, α_2 or α_1^*, α_2^*, i.e. we do not know whether we are estimating the parameters of the Sales equation or a mixture of the parameters of the Sales and Advertising equation.

The same way it can be demonstrated that if we try to estimate the Advertising equation we do not know whether we are estimating the parameters of the Advertising equation or a mixture of the parameters of the Sales and Advertising equations.

(b) **Approaching the question**

As both equations are exactly identified consistent estimates can be obtained either by applying indirect least squares or two-stage least squares. One of the methods of estimation should be explained.

Method of instrumental variables:

Use A_{t-1} as an instrument for A_t:

$$\begin{aligned} a_2^{IV} &= \frac{\text{Cov}(A_{t-1}, S_t)}{\text{Cov}(A_{t-1}, A_t)} = \frac{\text{Cov}(A_{t-1}, [\alpha_1 + \alpha_2 A_t + u_{St}])}{\text{Cov}(A_{t-1}, A_t)} \\ &= \alpha_2 + \frac{\text{Cov}(A_{t-1}, u_{St})}{\text{Cov}(A_{t-1}, A_t)}. \end{aligned}$$

Hence plim $a_2^{IV} = \alpha_2$, provided that A_{t-1} is exogenous. Clearly A_{t-1} should be highly correlated with A_t and therefore may be a good instrument.

Question 4

(a) **Explain what you understand by a dummy variable. Under what circumstances would you use dummy variables in econometric analysis?**

(10 marks)

(b) **In the model $y_t = \beta x_t + u_t$ for $t = 1, 2, \ldots, T$, where $E(u_t) = 0$, $E(u_t^2) = \sigma^2$, $E(u_t u_s) = 0$ for $s \neq t$, it is known that β changes after a certain point in the sample period, i.e. $\beta = \beta_1$ for $t = 1, 2, \ldots, t_0$ and $\beta = \beta_2$ for $t = t_0 + 1, t_0 + 2, \ldots, T$.**

i. **Explain carefully how you could estimate β_1 and β_2 using dummy variables.**

(5 marks)

ii. **If an econometrician ignores the change in β and simply regresses y_t on x_t for the whole sample $t = 1, 2, \ldots, T$ to produce an estimate b show that the resulting estimator b is a biased estimator of β. Under what conditions, if any, is the bias zero?**

(5 marks)

iii. **Explain in detail how you would use an F-test to test the hypothesis that β does not change at time t_0.**

(5 marks)

Reading for this question

- C. Dougherty, **Introduction to Econometrics**. (third edition) Chapter 5.
- C. Dougherty, **Subject Guide** (2011), Chapter 5.
- D. N. Gujarati, **Basic Econometrics**. (fourth edition), Chapter 9.

(a) **Approaching the question**

To answer this question the definition of dummy variables should be given and also how dummy variables can be used to test for a change in the intercept or a change in the slope. How the change in slope or intercept and also the joint test for a change in intercept and slope can be conducted should be explained. The dummy variable trap should be explained. Technical details should be given.

(b) i. **Approaching the question**

$$y_t = \beta x_t + \lambda D_t x_t + u_t \quad \text{where } D_t = \begin{cases} 1 & \text{if } t > t_0 \\ 0 & \text{otherwise.} \end{cases}$$

The slope coefficients are β and $\beta + \lambda$.

ii. **Approaching the question**

It is required to write the OLS estimator of β. Bias should be derived. Detailed working follows.

$$\hat{\beta} = \frac{\sum x_t y_t}{\sum x_t^2} = \frac{\sum x_t(\beta x_t + \lambda D_t x_t + u_t)}{\sum x_t^2}$$
$$= \beta + \lambda \frac{\sum D_t x_t^2}{\sum x_t^2} + \frac{\sum x_t u_t}{\sum x_t^2}.$$

Hence $E(\hat{\beta}) = \beta + \lambda \frac{\sum D_t x_t^2}{\sum x_t^2}$ so the bias is $\lambda \frac{\sum D_t x_t^2}{\sum x_t^2}$ which is zero if $\lambda = 0$ or if $x_t = 0$ for $t > t_0$.

iii. **Approaching the question**

Run three regressions:

(1) Overall $y_t = \beta x_t + u_t$ for $t = 1, 2, \ldots, T$ to give RSS_R

(2) $t = 1, 2, \ldots, t_0$ to give RSS_1

(3) $t = t_0 + 1, t_0 + 2, \ldots, T$ to give RSS_2.

Now compute $RSS_U = RSS_1 + RSS_2$.

The F-statistic is

$$\frac{(RSS_R - RSS_U)/1}{RSS_U/(T-2)},$$

which has an F-distribution with $(1, T-2)$ degrees of freedom.

Alternatively, one could use one regression with dummy variables included.

Question 5

(a) i. **What do you understand by heteroskedasticity? What are its effects on ordinary least squares (OLS) estimation?**

(4 marks)

ii. **Describe fully the Goldfeld-Quandt test and the White test against heteroskedasticity. Is there a preferred test for heteroskedasticity, and why?**

(8 marks)

iii. **How do you correct for the presence of heteroskedasticity?**

(3 marks)

iv. **Can heteroskedasticity be caused by model mis-specification? Explain.**

(3 marks)

(b) **An accountant is interested in how the level of inventories (I) is related to the cost of borrowing (R) and sales (S). She formulates the following model:**

$$I_i = \beta_0 + \beta_1 S_i + \beta_2 R_i + u_i$$

For a sample of 35 firms in the same industry sector she obtains the OLS results Regn1 in the table below. A colleague suggests that she also test for heteroskedasticity as sales may be correlated across firms in the industry. She sorts the variables appropriately and regresses the top and bottom 14 firms producing Regns 2 and 3 respectively, then Regn 4 is the result if heteroskedasticity is to be accounted for. (t-values are in parentheses and ESS is the error sum of squares).

	Const	S	R	R^2	ESS
Regn1	**−6.17**	**0.20**	**−0.25**	**0.98**	
		(12.39)	**(−2.67)**	$N = 35$	
Regn2	**−2.23**	**0.16**	**−0.22**	**0.94**	**0.908**
		(1.90)	**(−0.81)**	$N = 14$	
Regn3	**16.10**	**0.11**	**−1.40**	**0.96**	**5.114**
		(3.36)	**(−3.35)**	$N = 14$	
Regn4	**−8.45**	**0.21**	**−0.18**	**0.93**	
		(12.34)	**(−2.98)**	$N = 35$	

i. **Perform a test against heteroskedasticity stating clearly the null hypothesis, the distribution of the test statistic and the rejection rule.**

(3 marks)

ii. **What criteria would you expect the accountant to have used to sort the variables?**

(2 marks)

iii. **From your answer in (a) explain conceptually, how Regn1 and Regn 4 differ? State clearly how the constants and slope coefficients are calculated.**

(2 marks)

Reading for this question

C. Dougherty, **Introduction to Econometrics**. (third edition) Chapter 7.

C. Dougherty, **Subject Guide** (2011), Chapter 7.

D. N. Gujarati, **Basic Econometrics**. (fourth edition), Chapter 11.

(a) i. **Approaching the question**

Heteroskedasticity refers to the case in which the variance of the error term is not constant for all values of the independent variable; that is, $E(u_i^2) \neq \sigma_u^2$. This is a violation of the Gauss-Markov assumption (of constant variance in the errors) that supports the OLS technique. The problem occurs most frequently in cross-section data analysis. An appropriate example should be given.

In the presence of heteroskedasticity the OLS parameter estimates remain unbiased and consistent, but they are inefficient, that is, they have larger than minimum variances. In addition the estimated variances of the parameters are biased leading to invalid statistical tests and biased confidence intervals.

ii. **Approaching the question**

Candidates are required to discuss both the Goldfeld-Quandt and White tests of heteroskedasticity. Both tests are discussed below.

Goldfeld-Quandt Test.

This test assumes that $\sigma^2 (= E(u_t^2))$ is proportional to the size of one of the RHS variables (say X_t). The observations are ranked by X and separate regressions run for the first $n_1 (< n/2)$ and the last n_1 observations – the middle $(n-2)$ observations are not used. If heteroskedasticity is present the RSS from the two regressions will differ. Form the test RSS_2/RSS_1 where RSS_1 is the residual sum of squares from the first n_1 observations and RSS_2 is the residual sum of squares from the last n_1 observations. The test statistic will have an F-distribution with $(n_1 - k, n_1 - k)$ degrees of freedom where k is the number of parameters in the equation. If the calculated F is greater than the critical value of F then reject the null of homoskedasticity.

White Test.

The White test looks for more general evidence of association between the variance of the error term and the regressors. Regress the squared residuals from the original regression on the regressors from that model, together with the squares and the cross-products of those variables. The test statistic is nR^2 where n is the sample size and R^2 is the R^2

from the White regression. This test statistic has a chi-square distribution with degrees of freedom equal to the number of regressors. The test assumes the sample size is large.

Each of the tests has strengths and weaknesses: the GQ test is the basic test and requires knowledge of the independent variable that is correlated to the error; the BP test is more general and does not depend on the functional form of the model; the Glesjer test allows closer examination of the nature of the heteroskedasticity as various γ-specifications can be tested.

iii. **Approaching the question**

Weighted least squares should be used. The model has to be transformed in such a way that the transformed disturbance term is homoskedastic. OLS is used to estimate the parameters of the transformed model. The resulting estimator is known as the weighted least squares estimator. The solution follows.

Given

$$Y_i = b_0 + b_1 X_i + u_i,$$

if it is assumed that $\text{Var}(u_i) = CX_i$, where C is a non-zero constant, we can divide (i.e. weight) every term of the regression by the square root of X_i and then re-estimate the regression using the transformed variables. The transformed error term is now homoskedastic. In the two-variable case:

$$\frac{Y_i}{\sqrt{X_i}} = \frac{b_0}{\sqrt{X_i}} + b_1\sqrt{X_i} + \frac{u_i}{\sqrt{X_i}} \tag{.2}$$

The error term is now homoskedastic:

$$\text{Var}\left(\frac{u_i}{\sqrt{X_i}}\right) = \frac{\text{Var}(u_i)}{X_i} = \frac{CX_i}{X_i} = C.$$

As the residuals are now homoskedastic, the OLS estimates from (.2) are not only unbiased and consistent, but also efficient.

In the multiple regression case the model is weighted by the variable that is thought to be associated with the error term.

The original intercept becomes variable whilst the coefficient on the suspect variable becomes the new intercept. It is possible to discover which variable is associated with the disturbance by plotting the variable against the OLS residuals from the original equation.

iv. **Approaching the question**

If an incorrect functional form is used, e.g. the model is linear when it should contain quadratic terms then the effect of the omitted variable will be present in the residual. Since the omitted variable is the square of a variable in the model the error term will possibly display heteroskedastic effects with the variance of the error being proportional to the linear part of the function.

(b) i. **Approaching the question**

Goldfeld-Quandt test has to be used. Null and alternative hypotheses should be clearly specified. The solution follows.

Test H_0 : Homoskedasticity vs. H_1 : Heteroskedasticity.

$ESS_3/ESS_2 = 5.114/0.908 = 5.63$ exceeds $F_{11,11} = 2.82$ at the 5% level of significance. Hence we reject the null hypothesis of homoskedasticity.

ii. **Approaching the question**

All the variables are in the ascending order of the variable suspected of being correlated with the variance of the error term.

iii. **Approaching the question**

Regn1 is a naïve regression without regard to heteroskedasticity and unsorted. Regn4 controls for heteroskedasticity induced by S by dividing all the variables through by S on the assumption suggested by the colleague. $\beta_1 = 0.21$ is now the slope coefficient associated with variable S, (instead of 0.16 before the transformation), the constant is -8.45 (S/S) while $\beta_2 = -0.18$ is the slope coefficient associated with R (R/S).

Question 6

(a) **Describe an adaptive expectations model. Show how to derive an adaptive expectations model. Explain why this model might be used, and how it could be estimated.**

(10 marks)

(b) **Koyck investigated the relationship between investment in railcars and the volume of freight carried on the U.S. railroads using annual data for the period 1884–1939. Assuming that the desired stock of railcars in year t depended on the volume of freight in year $t-1$ and year $t-2$ and a time trend, and assuming that investment in railcars was subject to a partial adjustment process, he fitted the following regression equation using OLS (standard errors and constant term not reported):**

$$\hat{I}_t = 0.077F_{t-1} + 0.017F_{t-2} - 0.0033t - 0.110K_{t-1} \qquad R^2 = 0.85$$

where $I_t = K_t - K_{t-1}$ is investment in railcars in year t (thousands), K_t is the stock of railcars at the end of year t (thousands), and F_t is the volume of freight handled in year t (ton-miles).

i. **Explain how Koyck's model can be derived from an adjustment equation $K_t - K_{t-1} = \lambda(K_t^* - K_{t-1})$ and a behavioural equation using variables F (lagged) and t (time) and whose dependent variable is K_t^*.**

(7 marks)

ii. **Using Koyck's estimated equations estimate the parameters of your behavioural equation. What are the implications for the behaviour in the long run, i.e. as t increases towards infinity?**

(8 marks)

Reading for this question

C. Dougherty, **Introduction to Econometrics**. (third edition) Chapter 11.3.

C. Dougherty, **Subject Guide** (2011), Chapter 11.

D. N. Gujarati, **Basic Econometrics**. (fourth edition), Chapters 17.5 and 17.7.

(a) **Approaching the question**

An adaptive expectations model involves a learning process in which, in each time period, the actual value of the variable is compared with the value that had been expected. If the actual value is greater, the expected value is adjusted upwards for the next period. If it is lower, the expected value is adjusted downwards. The size of the adjustment is hypothesised to be proportional to the discrepancy between the actual and expected value.

If X is the variable in question, and X_t^e is the value expected in time period t given the information available at time period $t-1$, then

$$\begin{aligned} X_{t+1}^e - X_t^e &= \lambda(X_t - X_t^e); \qquad 0 \le \lambda \le 1 \\ \text{or} \qquad X_{t+1}^e &= \lambda X_t + (1-\lambda)X_t^e. \end{aligned}$$

The model is derived to capture the changing nature of expectations formation, often in variables that are also changing with time. It is an attempt at a 'simple learning' solution to model building in order to forecast often macroeconomic variables; such variables include investment, savings and demand for assets.

The model is estimated by repeated substitution for the expected variable, by its lagged variant which has known components of the previous period and the unobserved expectation lagged, until the term on the unobserved expectation $(1-\lambda)s$ is so small as to be ignored resulting in a model in which all the variables are observed, where s is the period lagged and λ is the speed of adjustment of expected and actual and λ is between 0 and 1. Technical details should be given.

(b) i. **Approaching the question**

Given the information in the question, the model may be written:

$$K_t^* = \beta_1 F_{t-1} + \beta_2 F_{t-2} + \beta_3 t + u_t. \tag{.3}$$

It is given that

$$K_t - K_{t-1} = I_t = \lambda(K_{t-1}^* - K_{t-1}).$$

Hence

$$I_t = \lambda\beta_1 F_{t-1} + \lambda\beta_2 F_{t-2} + \lambda\beta_3 t - \lambda K_{t-1} + \lambda u_t.$$

ii. **Approaching the question**

Estimates of the coefficients should be obtained and interpreted. Working and interpretations follow.

From the fitted equation,

$$\hat{\lambda} = 0.110; \quad b_1 = \frac{0.077}{0.110} = 0.70; \quad b_2 = \frac{0.017}{0.110} = 0.15; \quad b_3 = \frac{-0.0033}{0.110} = -0.030.$$

Hence the short-run effect of an increase of one million ton-miles of freight is to increase investment in railcars by 77 one year later and by 17 two years later. It does not make much sense to talk of a short-run effect of a time trend.

In the long-run equilibrium, neglecting the effects of the disturbance term, K_t and K_t^* are both equal to the equilibrium value $\bar{K}$ and F_{t-1} and F_{t-2} are both equal to their equilibrium value $\bar{F}$. Hence, using equation (.3),

$$\bar{K} = (\beta_1 + \beta_2)\bar{F} + \beta_3 t.$$

Thus an increase of one million ton-miles of freight will increase the stock of railcars by 850 and the time trend will be responsible for a secular decline of 33 railcars per year (obsolescence/damage).

Examiners' commentaries 2012

EC2020 Elements of econometrics

Important note

This commentary reflects the examination and assessment arrangements for this course in the academic year 2011–12. The format and structure of the examination may change in future years, and any such changes will be publicised on the virtual learning environment (VLE).

Information about the subject guide

Unless otherwise stated, all cross-references will be to the latest version of the subject guide (2011).

General remarks

Learning outcomes

At the end of this course, and having completed the Essential reading and activities, you should be able to:

- describe and apply the classical regression model and its application to cross-section data
- describe and apply the:
  - Gauss-Markov conditions and other assumptions required in the application of the classical regression model
  - reasons for expecting violations of these assumptions in certain circumstances
  - tests for violations
  - potential remedial measures, including, where appropriate, the use of instrumental variables
- recognise and apply the advantages of logit, probit and similar models over regression analysis when fitting binary choice models
- competently use regression, logit and probit analysis to quantify economic relationships using standard regression programmes (Stata and EViews) in simple applications
- describe and explain the principles underlying the use of maximum likelihood estimation
- apply regression analysis to fit time-series models using stationary time series, with awareness of some of the econometric problems specific to time series applications (for example, autocorrelation) and remedial measures
- recognise the difficulties that arise in the application of regression analysis to nonstationary time series, know how to test for unit roots, and know what is meant by cointegration.

Common mistakes committed by candidates

- A large number of candidates were not able to distinguish between sample variance and covariance, and population variance and covariance (this is happening year after year). They treat them as the same. This results in incorrect analysis and candidates lose significant marks.

  Consider an example: Suppose data is deviation from respective sample means and the regression model is:

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

  The ordinary least squares estimator of β is:

  $$\hat{\beta} = \frac{\sum_{t=1}^T x_t y_t}{\sum_{t=1}^T x_t^2} = \beta + \frac{\sum_{t=1}^T x_t u_t}{\sum_{t=1}^T x_t^2}.$$

  In terms of variances and covariances (a large number of candidates prefer this terminology) this can be written as:

  $$\hat{\beta} = \beta + \frac{\text{Cov}(x, u)}{\text{Var}(x)}.$$

  Here $\text{Cov}(x, u)$ and $\text{Var}(x)$ are sample[$\text{Cov}(x, u)$] and sample[$\text{Var}(x)$].

  Candidates should realise that $\sum_{t=1}^T u_t$, $\sum_{t=1}^T x_t u_t$, $\text{Cov}(x, u)$ and $\text{Var}(x)$ given above are sample moments and as such $\sum_{t=1}^T u_t \neq 0$, $\sum_{t=1}^T x_t u_t \neq 0$ and $\text{Cov}(x, u) \neq 0$. But, if we take expectation, then:

  $$E[u_t] = 0,$$

  by assumption. Then:

  $$E\left[\sum_{t=1}^T x_t u_t\right] = \sum_{t=1}^T x_t[E(u_t)] = 0,$$

  as the x_t are fixed they can be taken out of the expectation, and so:

  $$E[\text{Cov}(x, u)] = E\left[\frac{1}{T}\sum_{t=1}^T x_t u_t\right] = 0,$$

  as previously argued. This makes $E(\hat{\beta}) = \beta$, i.e. $\hat{\beta}$ is an unbiased estimator for β.

  To prove consistency take plim to get:

  $$\begin{aligned} \text{plim}\left(\hat{\beta}\right) &= \beta + \text{plim}\left(\frac{\frac{1}{T}\sum_{t=1}^T x_t u_t}{\frac{1}{T}\sum_{t=1}^T x_t^2}\right) \\ &= \beta + \frac{\text{plim}\left(\frac{1}{T}\sum_{t=1}^T x_t u_t\right)}{\text{plim}\left(\frac{1}{T}\sum_{t=1}^T x_t^2\right)} \\ &= \beta + \frac{\text{plim}(\text{sample Cov}(x, u))}{\text{plim}(\text{sample Var}(x))} \\ &= \beta + \frac{\text{population Cov}(x, u)}{\text{population Var}(x)}. \end{aligned}$$

  By assumption, population $\text{Cov}(x, u) = 0$ and population $\text{Var}(x) > 0$, hence $\text{plim}\left(\hat{\beta}\right) = \beta$, in other words $\hat{\beta}$ is a consistent estimator of β.

  Remember that in general:

  $$\begin{aligned} \text{plim(sample variance)} &= \text{population variance}, \\ \text{plim(sample covariance)} &= \text{population covariance}. \end{aligned}$$

  This concept has been used in many questions. This simple mistake of not distinguishing between sample variance and covariance and population variance and covariance results in a significant loss of marks which might result in the loss of a degree class or even be the difference between pass and fail.

- Candidates struggled to give competent answers to the interpretation of empirical results. When interpreting an empirical result you should discuss the significance of the coefficients, magnitude and sign of the coefficients. Also, you should make sure that the GM conditions hold.
- Just as last year, many candidates did not appear to read the questions carefully enough and often omitted to give answers to parts of questions which asked for details of such things as the assumptions necessary for a particular result to be true.

Key steps to improvement

- Essential reading for this course includes the subject guide and

  Dougherty, C. *Introduction to Econometrics.* (Oxford: Oxford University Press, 2011) fourth edition [ISBN 9780199567089].

  To understand the subject clearly it is important to supplement C. Dougherty, *Introduction to Econometrics* (fourth edition) with the subject guide **EC2020 Elements of econometrics**, especially the chapters on maximum likelihood and panel data.

  Apart from Essential reading you should do some supplementary reading. Two very good books of the same level are:

  Gujarati, Damodar N. *Basic Econometrics.* (Boston; London: McGraw-Hill Education, 2009) fifth edition [ISBN 9780071276252].

  Woolridge, Jeffrey M. *Introductory Econometrics.* (Mason, Ohio: Thomson Learning, 2008) fourth edition [ISBN 9780324788907].
- It is very important to go through the subject guide carefully. The subject guide contains solutions to the questions given in the main textbook and also some additional questions and solutions. Working through these will improve the clarity of the subject.
- The chapter on maximum likelihood in the subject guide (Chapter 10) includes some additional theory which is not covered in the main textbook. It is important to read the additional theory given in the subject guide to have a better understanding of the principles of maximum likelihood and tests based on the likelihood function.
- At the website URL http://econ.lse.ac.uk/ie are PowerPoint slideshows that provide a graphical treatment of the topics covered in the text, data sets, statistical tables and a downloadable copy of the subject guide **EC2020 Elements of econometrics**. Candidates should utilise data sets using standard regression programmes (STATA or EViews). This will help in their understanding of the subject.

Question spotting

Many candidates are disappointed to find that their examination performance is poorer than they expected. This can be due to a number of different reasons and the *Examiners' commentaries* suggest ways of addressing common problems and improving your performance. We want to draw your attention to one particular failing – '**question spotting**', that is, confining your examination preparation to a few question topics which have come up in past papers for the course. This can have very serious consequences.

We recognise that candidates may not cover all topics in the syllabus in the same depth, but you need to be aware that Examiners are free to set questions on **any** aspect of the syllabus. This means that you need to study enough of the syllabus to enable you to answer the required number of examination questions.

The syllabus can be found in the 'Course information sheet' in the section of the VLE dedicated to this course. You should read the syllabus very carefully and ensure that you cover sufficient material in preparation for the examination.

Examiners will vary the topics and questions from year to year and may well set questions that have not appeared in past papers – every topic on the syllabus is a legitimate examination target. So although past papers can be helpful in revision, you cannot assume that topics or specific questions that have come up in past examinations will occur again.

If you rely on a question spotting strategy, it is likely you will find yourself in difficulties when you sit the examination paper. We strongly advise you not to adopt this strategy.

Examiners' commentaries 2012

EC2020 Elements of econometrics

Important note

This commentary reflects the examination and assessment arrangements for this course in the academic year 2011–12. The format and structure of the examination may change in future years, and any such changes will be publicised on the virtual learning environment (VLE).

Information about the subject guide

Unless otherwise stated, all cross-references will be to the latest version of the subject guide (2011).

Comments on specific questions – Zone A

Candidates should answer **FOUR** of the following **SIX** questions: **QUESTION 1** of Section A (25 marks in total) and **THREE** questions from Section B (25 marks each).

Section A

Answer **all** parts of question 1 (25 marks in total).

Question 1

(a) **Why is R^2 meaningless in probit and logit models? What measures of 'goodness of fit' are applicable to probit and logit models?**

(5 marks)

(b) **Suppose the regression model $Y_t = \beta_0 + \beta_1 X_t + u_t$, where $\mathbf{E}(u_t) = 0$, $\mathbf{E}(u_t^2) = \sigma^2$ and $\mathbf{E}(u_s u_t) = 0$ if $s \neq t$; where σ^2 is known, is estimated by ordinary least squares (OLS) using T_1 observations to produce estimates of β_0, β_1 and their standard errors. Suppose now that extra data is measured such that there are now T_2 observations ($T_2 > T_1$) how would the estimates of the parameters and their standard errors change? Explain.**

(5 marks)

(c) **Suppose $y_t = \beta x_t + \epsilon_t$ where ϵ_t has zero mean, constant variance and is not serially correlated. If we define $y_t^* = y_t + y_{t-1}$, $x_t^* = x_t + x_{t-1}$ and $\epsilon_t^* = \epsilon_t + \epsilon_{t-1}$ show that least squares regression of y_t^* on x_t^* will give unbiased estimates of β. Also show that the error term has a constant variance but $\mathbf{E}(\epsilon_t^* \epsilon_{t-1}^*) \neq 0$. What will be the properties of the least squares estimates?**

(5 marks)

(d) **Let $\hat{u}_t$ be the residuals in the least squares fit of y_t against x_t and a constant term for $t = 1, 2, 3, \ldots, T$. Derive the following results:-**

$$\sum_{t=1}^{T} \hat{u}_t = 0 \qquad \textbf{and} \qquad \sum_{t=1}^{T} x_t \hat{u}_t = 0$$

(5 marks)

(e) **Suppose the random variable Y is defined as $Y = \beta + \epsilon$ where ϵ can take values $+1$ or -1 each with probability 1/2. Show that the sample mean of Y is an unbiased estimator of β. If you had a sample of 10 observations on Y to estimate the mean and another of 100 observations to estimate the mean how would expect the two estimates of β to compare?**

(5 marks)

Reading for this question

Subject guide, Chapters 1 and 2.

Dougherty, C. *Introduction to Econometrics* (fourth edition) Chapters 1.4, 2.2, 2.3, 2.5, 2.6 and 10.6.

Gujarati, D.N. *Basic Econometrics* (fifth edition) Chapters 7A.1, 15.2 and 15.8.

(a) **Approaching the question**

It should be discussed that as the dependent variable only takes two values R^2 is meaningless. Pseudo-R^2 and the likelihood ratio test statistic should be discussed. The solution follows.

The definition of R^2 is

$$R^2 = \frac{ESS}{TSS} = \frac{TSS - RSS}{TSS} = 1 - \frac{RSS}{TSS}$$

where TSS is the total sum of squares, ESS is the explained sum of squares and RSS is the residual sum of squares. Under logit and probit the dependent variable only takes two states, 0 and 1, hence TSS will take different values dependent on the coding of 'success' or 'failure' even though the independent variables are the same.

The possibilities for measuring goodness of fit are (i) the pseudo-R^2 defined by

$$1 - \frac{\log L}{\log L_0}$$

where $\log L$ is the log-likelihood and $\log L_0$ is the log-likelihood that would have been obtained with only the intercept in the regression. This has a minimum of 0 but the maximum will be less than 1 and, unlike R^2 it does not have a natural interpretation.

The alternative is (ii) the likelihood ratio statistic defined by

$$2 \log \frac{L}{L_0} = 2 \left(\log L - \log L_0 \right)$$

which has a chi-square distribution with $k-1$ degrees of freedom where $k-1$ is the number of explanatory variables. The null hypothesis is that the coefficients of the variables are all jointly zero.

(b) **Approaching the question**

It should be discussed that $\hat{\beta}$ would change, $\text{var}(\hat{\beta})$ would decrease but estimated $\text{var}(\hat{\beta})$ may increase or decrease. The solution follows.

The OLS estimate of $\hat{\beta} = \sum x_t y_t / \sum x_t^2$ would obviously change with extra data but we cannot predict exactly how it would change, it may increase or decrease. We do, however, know that it is an unbiased estimator and this is true regardless of sample size.

The variance of the OLS estimate is $\text{var}(\hat{\beta}) = \sigma^2 / \sum x_t^2$ and this will reduce as the sample size is increased from T_1 to T_2 so standard errors will decrease in general. Note, however, that we generally have to use $\hat{\sigma}^2$ instead of σ^2 in the variance and it is impossible to predict whether this will increase or decrease with the larger sample size. Finally, note that the smaller variance is **not** a consequence of the estimator being more efficient, it is just that more data are available.

(c) **Approaching the question**

Expectation, variance and covariance of the disturbance term should be derived to check the properties of $\hat{\beta}$. The solution follows.

Applying OLS we get

$$\begin{aligned}\hat{\beta} &= \frac{\sum x_t^* y_t^*}{\sum x_t^{*2}} = \frac{\sum x_t^*(\beta x_t^* + \epsilon_t^*)}{\sum x_t^{*2}} = \beta + \frac{\sum x_t^* \epsilon_t^*}{\sum x_t^{*2}} \\ \mathrm{E}(\hat{\beta}) &= \beta + \frac{\sum x_t^* \mathrm{E}(\epsilon_t^*)}{\sum x_t^{*2}}\end{aligned}$$

where $\mathrm{E}(\epsilon_t^*) = \mathrm{E}(\epsilon_t + \epsilon_{t-1}) = 0$.

Hence $\mathrm{E}(\hat{\beta}) = \beta$ and the estimator is unbiased.

The error term has constant variance since

$$\mathrm{var}(\epsilon_t^*) = \mathrm{E}(\epsilon_t^{*2}) = \mathrm{E}(\epsilon_t^2 + \epsilon_{t-1}^2 + 2\epsilon_t\epsilon_{t-1}) = 2\sigma^2$$

as $\mathrm{E}(\epsilon_t^2) = \mathrm{E}(\epsilon_{t-1}^2) = \sigma^2$ and $\mathrm{E}(\epsilon_t\epsilon_{t-1}) = 0$. But the error terms are serially correlated, as

$$\mathrm{E}(\epsilon_t^*\epsilon_{t-1}^*) = \mathrm{E}(\epsilon_t + \epsilon_{t-1})(\epsilon_{t-1} + \epsilon_{t-2}) = \mathrm{E}(\epsilon_t\epsilon_{t-1}) + \mathrm{E}(\epsilon_t\epsilon_{t-2}) + \mathrm{E}(\epsilon_{t-1}\epsilon_{t-1}) + \mathrm{E}(\epsilon_{t-1}\epsilon_{t-2}) = \sigma^2.$$

Therefore, although the OLS estimates will be unbiased they will not be efficient.

(d) **Approaching the question**

To answer this question first order conditions have to be derived. The solution follows.

To apply OLS we minimise the sum of squares of errors

$$I = \sum_{t=1}^{T} \hat{u}_t^2 = \sum_{t=1}^{T} \left(Y_t - \hat{\beta}_0 - \hat{\beta}_1 X_t\right)^2$$

by differentiating with respect to $\hat{\beta}_0$ and $\hat{\beta}_1$. The resulting equations are

$$\frac{\partial I}{\partial \hat{\beta}_0} = 2\sum_{t=1}^{T} \left(Y_t - \hat{\beta}_0 - \hat{\beta}_1 X_t\right)(-1) = -2\sum_{t=1}^{T} \hat{u}_t = 0$$

and

$$\frac{\partial I}{\partial \hat{\beta}_1} = 2\sum_{t=1}^{T} \left(Y_t - \hat{\beta}_0 - \hat{\beta}_1 X_t\right)(-X_t) = -2\sum_{t=1}^{T} \hat{u}_t X_t = 0$$

which gives the two results.

(e) **Approaching the question**

Derivation of the expectation and variance of the sample mean of Y is required. The solution follows.

$$\begin{aligned}\bar{Y} &= \frac{\sum Y_i}{N} = \frac{\sum(\beta + \epsilon_i)}{N} = \beta + \frac{\sum \epsilon_i}{N} \\ \mathrm{E}(\bar{Y}) &= \beta + \frac{\sum \mathrm{E}(\epsilon_i)}{N} = \beta\end{aligned}$$

as $\mathrm{E}(\epsilon_i) = 0.5 \times 1 + 0.5 \times (-1) = 0$.

$$\mathrm{var}(\bar{Y}) = \mathrm{var}(\beta) + \frac{\sum(\mathrm{var}(\epsilon_i))}{N^2} = \frac{1}{N}$$

as $\mathrm{var}(\beta) = 0$ and $\mathrm{E}(\epsilon_i\epsilon_j) = 0$ if $i \neq j$.

We can conclude that the sample mean of Y is an unbiased estimator of β and it has variance $1/N$ which will become smaller as N increases so the variance of the mean of Y with samples of 10 and 100 will decrease by a factor of 10/100, i.e. the variance of the sample of 10 will be 10 times the variance of the sample of 100.

Section B

Answer **three** questions from this section (25 marks each).

Question 2

Briefly explain the following:

(a) Dickey-Fuller and Augumented Dickey-Fuller tests.

(8 marks)

(b) Error correction model.

(9 marks)

(c) Likelihood ratio test.

(8 marks)

Reading for this question

Subject guide, Chapter 10.

Dougherty, C. *Introduction to Econometrics* (fourth edition) Chapters 10.6, 13.3 and 13.5.

Gujarati, D.N. *Basic Econometrics* (fifth edition) Chapter 21.9.

(a) **Approaching the question**

Dickey-Fuller (DF) and augmented Dickey-Fuller (ADF) tests should be explained. Null and alternative hypotheses should be clearly specified. The solution follows.

The standard test for a unit root is due to Dickey and Fuller and is based on the model

$$Y_t = \beta_1 + \beta_2 Y_{t-1} + \gamma t + u_t$$

which can be rewritten as

$$\Delta Y_t = \beta_1 + (\beta_2 - 1)Y_{t-1} + \gamma t + u_t$$

where $\Delta Y_t = Y_t - Y_{t-1}$. The null hypothesis for stationarity is $\text{H}_0 : \beta_2 - 1 = 0$ and the alternative hypothesis is $\text{H}_1 : \beta_2 - 1 \neq 0$. We cannot use the standard t-test procedure in this case because the distribution of the t-statistic is not a t-distribution so critical values have been computed by Dickey and Fuller using Monte-Carlo techniques.

The test is sensitive to the presence of serial correlation in the error term so we need to take steps to remove the effects of this serial correlation – this is done by including lagged values of Y_t in the regression, in other words

$$Y_t = \beta_1 + \beta_2 Y_{t-1} + \beta_3 Y_{t-2} + \gamma t + u_t$$

for an AR(1) serial correlation. This is more easily tested by using the model

$$\Delta Y_t = \beta_1 + (\beta_2 + \beta_3 - 1)Y_{t-1} - \beta_3 \Delta Y_{t-1} + \gamma t + u_t$$

with null and alternative hypotheses as $\text{H}_0 : \beta_2 + \beta_3 - 1 = 0$ and $\text{H}_1 : \beta_2 + \beta_3 - 1 \neq 0$.

(b) **Approaching the question**

The ADL(1,1) model should be specified and the error correction model (ECM) should be derived. The solution follows.

Consider a simple ADL(1,1) model. [This is also known as ARDL(1,1).]

$$Y_t = \alpha_1 + \alpha_2 Y_{t-1} + \alpha_3 X_t + \alpha_4 X_{t-1} + u_t \qquad \text{(i)}$$

where Y_t and X_t are I(1). Rewrite (i) as

$$\begin{aligned}
Y_t - Y_{t-1} &= \alpha_1 + \alpha_2 Y_{t-1} - Y_{t-1} + \alpha_3 X_t - \alpha_3 X_{t-1} + \alpha_3 X_{t-1} + \alpha_4 X_{t-1} + u_t \\
\Delta Y_t &= \alpha_1 - (1-\alpha_2) Y_{t-1} + \alpha_3 \Delta X_t + (\alpha_3 + \alpha_4) X_{t-1} + u_t \\
\Delta Y_t &= \alpha_3 \Delta X_t - (1-\alpha_2)\left[Y_{t-1} - \frac{\alpha_1}{(1-\alpha_2)} - \frac{(\alpha_3+\alpha_4)}{(1-\alpha_2)} X_{t-1}\right] + u_t \\
\Delta Y_t &= \alpha_3 \Delta X_t - (1-\alpha_2)\left[Y_{t-1} - \beta_1 - \beta_2 X_{t-1}\right] + u_t
\end{aligned}$$

or,

$$\Delta Y_t = \alpha_3 \Delta X_t - \pi \left[Y_{t-1} - \beta_1 - \beta_2 X_{t-1}\right] + u_t \qquad \text{(ii)}$$

where

$$\pi = (1-\alpha_2); \qquad \beta_1 = \frac{\alpha_1}{(1-\alpha_2)} \qquad \text{and} \qquad \beta_2 = \frac{(\alpha_3+\alpha_4)}{(1-\alpha_2)}.$$

Equation (ii) is the ECM.

When the two variables Y and X are cointegrated, the ECM incorporates not only the short-run but also long-run effects. The long run equilibrium $Y_{t-1} - \beta_1 - \beta_2 X_{t-1}$ is included in the model together with the short-run effect captured by the differenced term.

All the terms in the ECM, given by (ii), are stationary. As Y and X are I(1), then ΔX and ΔY are I(0). As Y and X are cointegrtaed their linear combination $u_{t-1} = Y_{t-1} - \beta_1 - \beta_2 X_{t-1}$ is I(0).

The coefficient π provides us with the information about the speed of adjustment in cases of disequilibrium:

i. If $\pi = 1$, then 100% of the adjustment takes place within the period. In other words adjustment is instantaneous and full.

ii. If $\pi = 0.5$, then 50% adjustment takes place each period.

iii. If $\pi = 0$, then there is no adjustment.

(c) Brief description of the likelihood ratio (LR) test is required. Large sample LR test should also be discussed. The solution follows.

Suppose we have to test a simple hypothesis $\mathrm{H}_0 : \theta = \theta_0$ against all possible alternatives. Given a simple random sample $X_1, X_2, \ldots, X_N$, a natural way of judging the acceptability or otherwise of the hypothesis would be to compare the likelihood functions.

Let

L_R = Restricted likelihood (likelihood based on the null hypothesis)

L_U = Unrestricted likelihood (likelihood based on the alternative hypothesis).

If the LR

$$\lambda = \frac{L_R}{L_U}$$

is close to unity then in light of the given sample H_0 would seem highly plausible, on the other hand if this ratio is close to zero H_0 would seem to have little validity. Since λ is a random variable its distribution may be derived and hence we can make probability statements about how close the LR is to unity. A test for H_0 is thus provided by a critical region defined by $\lambda < \lambda_0$, where λ_0 is such that $P(\lambda < \lambda_0 | \mathrm{H}_0) = \alpha$. α is the probability of Type 1 error.

For large N, $-2\ln\lambda$ has approximately a χ^2 distribution with degrees of freedom equal to the number of restrictions imposed by the null hypothesis.

Question 3

(a) **The true model is**

$$y_t = \beta_1 x_{1t} + \beta_2 x_{2t} + u_{1t}$$

but an econometrician estimates

$$y_t = \beta_1 x_{1t} + u_{2t}$$

Show that the ordinary least squares estimate of β_1 is biased in general and explain under what conditions you would expect the bias to be positive.

(10 marks)

(b) **The three earnings regressions given below were obtained for a sample collected in 1972 of 7,000 British male employees aged 15-64 who worked at least one week in the year preceding the interview. The variables are; Y = annual real earnings, S = years of full-time education, EXP = years of work experience [(age) − (years of full-time education) − 5], W = weeks worked during the year.**

The dependent variable is $\ln Y$. Standard errors are given in brackets.

	1	**2**	**3**
constant	**5.199**	**4.094**	**0.444**
S	**0.097**	**0.269**	**0.215**
	(0.003)	**(0.024)**	**(0.017)**
S^2	—	**−0.0064**	**−0.0049**
		(0.0009)	**(0.0006)**
EXP	**0.091**	**0.092**	**0.068**
	(0.002)	**(0.002)**	**(0.001)**
EXP^2	**−0.0015**	**−0.0015**	**−0.0012**
	(0.00004)	**(0.00004)**	**(0.00003)**
$\ln W$			**1.115**
			(0.013)
R^2	**0.316**	**0.321**	**0.665**

The average value of S is 10.3 years over the whole sample. Those with a First degree had an average value of S of 17.7.

i. **Test the hypothesis that the coefficient of $\ln W$ is unity. On what assumptions is this test based? How do you interpret this result?**

(5 marks)

ii. **Explain the role played by the quadratic terms in these equations and interpret their estimated coefficients.**

(5 marks)

iii. **Using your results in (a) on omitted variables bias, account for the difference in the value of the coefficients of S in specifications (1) and (2).**

(5 marks)

Reading for this question

Subject guide, Chapter 2.

Dougherty, C. *Introduction to Econometrics* (fourth edition) Chapters 2.8 and 6.2.

Gujarati, D.N. *Basic Econometrics* (fifth edition) Chapters 7.10 and 13.3.

(a) **Approaching the question**

Omitted variable bias should be discussed. Technical details are required. The solution follows.

If the true model is

$$y_t = \beta_1 x_{1t} + \beta_2 x_{2t} + u_t$$

but you estimate the regression equation

$$y_t = \beta_1 x_{1t} + u_t$$

then the ordinary least squares (OLS) estimator of β_1 is

$$\hat{\beta}_1 = \frac{\sum x_{1t} y_t}{\sum x_t^2} = \frac{\sum x_{1t}(\beta_1 x_{1t} + \beta_2 x_{2t} + u_t)}{\sum x_{1t}^2}$$

hence

$$\hat{\beta}_1 = \beta_1 + \beta_2 \frac{\sum x_{1t} x_{2t}}{\sum x_{1t}^2} + \frac{\sum x_{1t} u_t}{\sum x_{1t}^2}$$

and

$$\mathrm{E}(\hat{\beta}_1) = \beta_1 + \beta_2 \frac{\sum x_{1t} x_{2t}}{\sum x_{1t}^2}$$

since $\mathrm{E}(u_t) = 0$ and x_{1t} is non-stochastic.

The bias will be $\beta_2 \frac{\sum x_{1t} x_{2t}}{\sum x_{1t}^2}$ which will be positive if β_2 is positive and x_{1t} and x_{2t} are positively correlated (or β_2 is negative and x_{1t} and x_{2t} are negatively correlated). This bias is known as the omitted variable bias and occurs because β_1 takes up the effect of the missing x_{2t}.

(b) i. **Approaching the question**

t-test on $\ln W$ gives $(1.115 - 1)/0.013 = 8.85$, which is highly significant, hence $\ln W$ affects $\ln Y$. The size of the coefficient gives the elasticity of W on Y.

ii. **Approaching the question**

The quadratic terms illustrate the non-linear effect of schooling and work experience on earnings. The linear term is positive and the quadratic term is negative in both cases indicating that the positive relationship of schooling and work experience on earnings suffers from diminishing returns as these variables increase.

iii. **Approaching the question**

Sign of the omitted variable bias should be discussed. The solution follows.

From part (a) the omitted variable bias suggests that the bias is proportional to the coefficient of the omitted variable and the correlation between the omitted variable and the included variable. This correlation will be positive for the variables S and EXP since there will be a positive (but not equal to 1) correlation between S and S^2 and EXP and EXP^2 and the coefficient on the omitted variable will be negative for the above reasons, hence the bias will be negative, which means that you would expect the coefficient on S and EXP to be greater than the coefficient on these variables when the quadratic term is present (as we find).

Question 4

(a) Consider a 3 variable regression $Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + u_t$. The variances of the ordinary least squares (OLS) parameter estimates are:-

$$\mathbf{var}\left(\hat{\beta}_1\right) = \frac{\sigma^2}{(1 - r^2)\sum(X_{1t} - \bar{X}_1)^2} \quad \textbf{and} \quad \mathbf{var}\left(\hat{\beta}_2\right) = \frac{\sigma^2}{(1 - r^2)\sum(X_{2t} - \bar{X}_2)^2}$$

where r is the correlation between X_{1t} and X_{2t}. Using this result explain the concept of multicollinearity. What effects will multicollinearity have on the OLS estimates of a linear model? Explain carefully how you would estimate these variances given data on Y_t, X_{1t} and X_{2t}.

(10 marks)

(b) Data on 30 individuals smoking habits contain information on:-

Packs: the number of packs of cigarettes smoked per month.

Inc: annual income before tax

Price: typical price paid per pack

The correlation matrix between these variables is

	Age	Inc	Price	Packs
Age	**1**			
Inc	**0.88**	**1**		
Price	**−0.23**	**−0.09**	**1**	
Packs	**0.94**	**0.90**	**−0.19**	**1**

i. Describe the basic interrelationships suggested by these results.

(5 marks)

ii. The following regressions were run:-

$$\text{Packs} = \underset{(1.53)}{28.6} + \underset{(10.82)}{0.258}\,\text{Inc} - \underset{(1.32)}{11.39}\,\text{Price} \qquad R^2 = 0.83$$

$$\text{Packs} = \underset{(0.78)}{-11.74} + \underset{(5.38)}{0.986}\,\text{Age} + \underset{(2.30)}{0.084}\,\text{Inc} - \underset{(0.08)}{0.531}\,\text{Price} \qquad R^2 = 0.91$$

Interpret the estimates of these two regressions in the light of the correlation matrix given above. Figures in brackets are t values.

(10 marks)

Reading for this question

Subject guide, Chapter 3.

Dougherty, C. *Introduction to Econometrics* (fourth edition) Chapter 3.4.

Gujarati, D.N. *Basic Econometrics* (fifth edition) Chapter 10.5.

(a) **Approaching the question**

r is the estimated correlation between X_{1t} and X_{2t}, in other words

$$r = \frac{\sum(X_{1t} - \bar{X}_1)(X_{2t} - \bar{X}_2)}{\sum(X_{1t} - \bar{X}_1)^2 \sum(X_{2t} - \bar{X}_2)^2}$$

and $\hat{\sigma}^2$ is the estimate of the variance of the error term and is calculated by $\hat{\sigma}^2 = \sum\left(\hat{u}_t^2\right)/T$, where $\hat{u}_t$ is the estimated error. If the variance of the error term is large or the correlation between X_t and Y_t is near 1 or the variance of X_t is small, then the variance of the coefficient is large and the OLS estimates of the coefficients are imprecise.

(b) **Approaching the question**

i. The correlation matrix shows that, at a bivariate level, age and income are highly correlated and they are both highly correlated with packs. Price has a small negative correlation with Inc and Age, so you would conclude on the basis of these bivariate correlations that there would be some multicollinearity effects between Age and Inc but not between Price and the other variables. Note, however, that if all three variables are included in the equation we cannot rely on bivariate correlations to indicate the extent of multicollinearity.

ii. The 95% critical t-value for a sample size of 30 and 3 parameters is 2.052 and for 4 parameters is 2.056. These levels would suggest that Inc and Age are variables whose parameter estimates are both significantly different from zero but this is not true for the intercept and Price in both equations. The results indicate that the number of packs purchased increases with Age and Inc but Price is not a determinant of the number of packs. The R^2 value is high in both regressions which shows that the equations are a good fit. A further point to note is that the coefficient on Inc changes substantially between the two regressions when Age is added. This is to be expected from the results of the correlation matrix, which showed that Age and Inc are highly correlated. The comparison of the two regressions shows the effect of omitted variable bias since the correlation between Age and Inc is positive and the coefficient on Inc is positive, so the bias on the Inc coefficient must have been positive and you would therefore expect the coefficient on Inc to decline when Age is introduced (which it does). Also note that when Age is introduced the coefficients are still well-defined, which is to say that the t-values are high, thus multicollinearity is not a problem here despite the high correlation between Inc and Age. This must be due to the presence of Price which has reduced the multicollinearity effect.

Question 5

(a) Suppose a time series is generated as

$$y_t = \beta y_{t-1} + u_t; \quad t = 1, 2, \ldots T$$

where $\mathbf{E}(u_t) = 0$; $\mathbf{E}(u_t^2) = \sigma_u^2$ and $\mathbf{E}(u_s u_t) = 0$ if $s \neq t$ for all $s, t = 1, 2, \ldots T$. It is given that $\beta = 1 - \delta$, where δ is small enough that terms involving δ^2 may be neglected. Show that, for finite t, the variance of y_t may be approximated as

$$\sigma_y^2 = [1 - (t-1)\delta]\, t\sigma_u^2.$$

Explain the significance of this result. Assume $Y_0 = 0$.

(8 marks)

[Note: If x is small enough then $(1+x)^n$ can be approximated by $1 + nx$. Also, $1 + 2 + \ldots + n = \frac{n(n+1)}{2}$.]

(b) The following equation was estimated by ordinary least squares (OLS):

$$\begin{array}{rcccccc} Y_t = & \mathbf{-15.267} & \mathbf{-0.209}Y_{t-1} & \mathbf{+\ 0.259}\Delta Y_{t-1} & \mathbf{+\ 0.009}t & +\ \hat{u}_t \\ & \mathbf{(6.013)} & \mathbf{(0.084)} & \mathbf{(0.139)} & \mathbf{(0.004)} & \end{array}$$

$R^2 = 0.146, \quad \chi_2^2 = 2.86$

$\hat{u}_t$ is OLS residual, standard errors are given in brackets, χ_2^2 is an LM test statistic to test against second order residual serial correlation.

i. Using the results above, test the null hypotheses that Y_t is non-stationary. [Note: Critical value at 5% level of significance from MacKinnon table is −3.4126.]

(3 marks)

ii. What assumptions have you made? Are they likely to be true? Explain.

(3 marks)

(c) Consider a time series process

$$\ln Y_t = \alpha + \beta t + u_t; \quad t = 1, 2, \ldots, T$$

Examine the order of integration of $\ln Y_t$.

(6 marks)

(d) Consider two time series variables Y_t and X_t, both are integrated of order one. Explain how you will test that Y_t and X_t are cointegrated.

(5 marks)

Reading for this question

Subject guide, Chapter 13.

Dougherty, C. *Introduction to Econometrics* (fourth edition) Chapters 13.1, 13.3 and 13.4.

Gujarati, D.N. *Basic Econometrics* (fifth edition) Chapters 21.6 and 21.9.

(a) **Approaching the question**

Recursive substitution is required. Technical details should be given. The solution follows. By recursive substitution we get

$$Y_t = \beta^{t-1}u_1 + \ldots + u_t.$$

Therefore,

$$\begin{aligned}
\text{var}(Y_t) = \sigma_y^2 &= \left(\beta^{2t-2} + \ldots + \beta^2 + 1\right)\sigma_u^2 \\
&= \left[(1-\delta)^{2t-2} + \ldots + (1-\delta)^2 + 1\right]\sigma_u^2 \\
&= \left[1 - (2t-2)\delta + \ldots + (1-2\delta) + 1\right]\sigma_u^2 \qquad \text{(as } \delta \text{ is small)} \\
&= \left[t - 2\delta(t - 1 + \ldots + 1)\right]\sigma_u^2 \\
&= \left[t - \delta t(t-1)\right]\sigma_u^2 \\
&= \left[1 - (t-1)\delta\right] t\sigma_u^2.
\end{aligned}$$

Hence, for finite t, the variance of Y_t is a function of t. Therefore the series exhibits non-stationary behaviour for finite t, even though it is stationary.

(b) **Approaching the question**

i. For exports the ADF test statistic is $-0.209/0.084 = -2.489$. 5% critical value is -3.41. Thus do not reject the null that exports are non-stationary.

ii. The main assumption is that the error terms in both equations have constant variances and no serial correlation.

LM(2) test indicates no rejection of the null of no serial correlation up to second order. 5% critical value is 5.99.

As there is no information homoscedasticity cannot be tested, but as the series is time series we may assume homoscedasticity.

We also need to assume that the specifications are correct (e.g. no structural breaks).

(c) **Approaching the question**

To establish that $\ln Y_t$ is I(1), the mean, variance and covariances of $\Delta \ln Y_t$ have to be derived. The solution follows.

$$\text{E}(\ln Y_t) = \text{E}(\alpha + \beta t + u_t) = \alpha + \beta t.$$

Mean is a function of t, hence $\ln Y_t$ is non-stationary (not I(0)).

Taking the first difference we get

$$\begin{aligned}
\Delta \ln Y_t &= \ln Y_t - \ln Y_{t-1} \\
&= \alpha + \beta t + u_t - \alpha - \beta(t-1) - u_{t-1} \\
&= \beta + u_t - u_{t-1}
\end{aligned}$$

$$\text{E}(\Delta \ln Y_t) = \beta$$

$$\begin{aligned}
\text{var}(\Delta \ln Y_t) &= \text{var}(u_t - u_{t-1}) \\
&= \text{var}(u_t) + \text{var}(u_{t-1}) \\
&= 2\sigma^2
\end{aligned}$$

$$\begin{aligned}
\text{cov}(\Delta \ln Y_t, \Delta \ln Y_{t-1}) &= \text{cov}(\beta + u_t - u_{t-1}, \beta + u_{t-1} - u_{t-2}) \\
&= \text{E}(u_t - u_{t-1})(u_{t-1} - u_{t-2}) \\
&= \sigma^2.
\end{aligned}$$

$\text{cov}(\Delta \ln Y_t, \Delta \ln Y_{t-s}) = 0$ for all $s > 1$.

As the mean, variance and covariances of $\Delta \ln Y_t$ are not functions of time, $\Delta \ln Y_t$ is stationary (I(0)).

This implies that the order of integration of $\ln Y_t$ is one, which is to say $\ln Y_t$ is I(1).

(d) **Approaching the question**

If Y_t and X_t are I(1) then to establish cointegration it should be checked that the linear combination of Y_t and X_t is I(0). The solution follows.

Once it is established that each of the variables Y_t and X_t are I(1), the following relationship is estimated by OLS:

$$Y_t = \alpha_0 + \alpha_1 X_t + u_t \qquad \text{(i)}$$

It is a simple equation with no lags. Apply OLS to (i) to get $\hat{u}_t$.

To carry out the test for cointegration, estimate the following

$$\Delta \hat{u}_t = \theta \hat{u}_{t-1} + \sum \theta_i \Delta \hat{u}_{t-i} + \epsilon_t$$

and test using the ADF test.

$\text{H}_0 : \theta = 1 \Rightarrow \hat{u}_t$ is non-stationary so is u_t. This means that Y_t and X_t are not cointegrated. For Y_t and X_t to be cointegrated the null has to be rejected.

Question 6

(a) **Consider a model**

$$Y_{it} = \beta_1 + \beta_j \sum_{j=2}^{K} X_{jit} + \alpha_i + u_{it}; \quad i = 1, 2, \ldots, N; \quad t = 1, 2, \ldots, T$$

where Y is the dependent variable, the X_j are observed explanatory variables, α_i is an unobserved effect and u_{it} is the disturbance term assumed to satisfy the usual regression model conditions. The index i refers to cross-section and t refers to the time period.

Explain the differences between the within-groups, first differences and least squares dummy variables versions of the fixed effect model.

(15 marks)

(b) **The following estimates were made as part of a study of whether countries with low income per head had higher growth rates than countries with higher income per head (convergence). Data was obtained over a period of four years on 121 countries.**

The first estimates are pooled ordinary least squares estimates:

$$\underset{}{gr_{it}} = \underset{(0.358)}{\mathbf{-0.129}} \; \underset{(0.004)}{\mathbf{-0.023}pgdp_{it}} + \underset{(0.018)}{\mathbf{0.139}inv_{it}} + e_{1it}; \quad i = 1, 2, \ldots, N \quad t = 1, 2, \ldots, T \qquad \mathbf{(1)}$$

$N = 121, \quad T = 4, \quad R^2 = 0.133.$

The second estimates are fixed effects (least squares dummy variable variant):

$$gr_{it} = \underset{(0.358)}{\mathbf{3.524}} \; \underset{(0.012)}{\mathbf{-0.046}pgdp_{it}} \; \underset{(0.032)}{\mathbf{-0.020}inv_{it}} + e_{2it}; \quad i = 1, 2, \ldots, N \quad t = 1, 2, \ldots, T \qquad \mathbf{(2)}$$

$N = 121, \quad T = 4, \quad R^2 = 0.541.$

Standard errors are in brackets. gr **is the average percentage annual rate of growth of gdp per head,** $pgdp$ **is the gdp per head in constant US dollar,** inv **is the average percentage ratio of investment to gdp.** e_1 **and** e_2 **are the ordinary least squares residuals.** R^2 **is the conventionally calculated coefficient of determination.**

i. **Comment on the differences in** R^2 **for each equation.**

(3 marks)

ii. **Equation (1) is restricted version of equation (2). Test the validity of the restrictions.**

(4 marks)

iii. **From these results , what would you conclude about the convergence of gdp per head? Give details.**

(3 marks)

Reading for this question

Dougherty, C. *Introduction to Econometrics* (fourth edition) Chapters 1.7, 6.5 and 14.2.

Gujarati, D.N. *Basic Econometrics* (fifth edition) Chapter 16.3.

(a) **Approaching the question**

Within-groups fixed effect, first differences fixed effect and least squares dummy variable (LSDV) fixed effect regressions should be derived and the drawbacks of each model should be discussed. A detailed discussion is given below:

- **Within-groups fixed effect**

  Model given is

$$Y_{it} = \beta_1 + \sum_{j=2}^{K} \beta_j X_{jit} + \alpha_i + u_{it}; \quad i = 1, 2, \ldots, n; \quad t = 1, 2, \ldots, T \qquad \text{(i)}$$

  Sum the observations of each cross-sectional unit over the time dimension and divide by T to get

$$\bar{Y}_i = \beta_1 + \sum_{j=2}^{K} \beta_j \bar{X}_{ij} + \alpha_i + \bar{u}_i \qquad \text{(ii)}$$

  Subtracting (ii) from (i), we get

$$Y_{it} - \bar{Y}_i = \sum_{j=2}^{K} \beta_j (X_{jit} - \bar{X}_{ij}) + u_{it} - \bar{u}_i \qquad \text{(iii)}$$

  In (iii) we can see that the unobserved effect (α_i) disappears. This is known as within-groups regression. It explains the variations about the mean of the dependent variable in terms of the variations about the means of the explanatory variables for the group of observations relating to a given individual.

  - Drawbacks
    · The intercept and any explanatory variable that remain constant for each individual will drop out of the model.
    · The variation in $(X_{ij} - \bar{X}_i)$ may be much smaller than the variation in X_j. If this is the case, the impact of the disturbance term may be relatively large, giving rise to imprecise estimates.
    · There is loss of a substantial number of degrees of freedom.

- **First differences fixed effect**

  In this approach, the unobserved heterogeneity is eliminated by subtracting the observation from the previous time period from the observation for the current time period, for all time periods.

  Lag (i) by one period to get

$$Y_{it-1} = \beta_1 + \sum_{j=2}^{K} \beta_j X_{jit-1} + \alpha_i + u_{it-1} \tag{iv}$$

  Subtracting (i) from (iv), we get

$$\Delta Y_{it-1} = \sum_{j=2}^{K} \beta_j \Delta X_{jit-1} + \alpha_i + u_{it} - u_{it-1} \tag{v}$$

  and the unobserved heterogeneity (α_i) disappears.

  - Drawbacks
    - The intercept and any explanatory variable that remain constant for each individual will drop out of the model.
    - n degrees of freedom are lost as the first observation for each individual is not defined.
    - It gives rise to autocorrelation.

- **Least squares dummy variable (LSDV) fixed effect**

  In this approach the unobserved effect is brought explicitly into the model. A set of dummy variables D_i is defined, where D_i is equal to 1 in the case of an observation relating to an individual i and 0 otherwise. The model can be written as

$$Y_{it} = \sum_{j=2}^{K} \beta_j X_{jit} + \sum_{i=1}^{n} \alpha_i D_i + u_{it} \tag{vi}$$

  The unobserved effect is now being treated as the coefficient of the specific individual i. The term $\alpha_i D_i$ represents a fixed effect on the dependent variable Y_i for individual i.

  If we want to keep the intercept in the model then instead of n dummy variables $(n-1)$ dummy variables have to be used, otherwise we will fall into the dummy variable trap.

  It can be shown that the LSDV method is identical to the within-groups method. Hence, the drawbacks are the same.

(b) **Approaching the question**

i. In equation (1) there are two explanatory variables, whereas in equation (2) there are 122 explanatory variables (extra 120 dummy variables are included in the model).

$$R^2 = \frac{\text{Explained sum of squares } (ESS)}{\text{Total sum of square } (TSS)}$$

where TSS is the same in both models but ESS is greater in (2) because of 120 more explanatory variables (the dummies), hence R^2 in (2) is more than R^2 in (1).

ii. Null hypothesis is that the coefficients of all dummies in (2) are equal to zero. This is an F-test for linear restrictions.

$$F_{120,361} = \frac{(0.541 - 0.133)/120}{(1 - 0.541)/361} = 2.677.$$

Critical value of $F_{120,361}$ at 5% level of significance is approximately 1.22. Reject the null hypothesis. The restrictions imposed on (1) are not valid.

iii. Since (1) is rejected we should concentrate on (2).

Estimated coefficient of *inv* has wrong sign but it is insignificant. Estimated coefficient of *pgdp* is negative and significant showing some slow convergence.

Examiners' commentaries 2012

EC2020 Elements of econometrics

Important note

This commentary reflects the examination and assessment arrangements for this course in the academic year 2011–12. The format and structure of the examination may change in future years, and any such changes will be publicised on the virtual learning environment (VLE).

Information about the subject guide

Unless otherwise stated, all cross-references will be to the latest version of the subject guide (2011).

Comments on specific questions – Zone B

Candidates should answer **FOUR** of the following **SIX** questions: **QUESTION 1** of Section A (25 marks in total) and **THREE** questions from Section B (25 marks each).

Section A

Answer **all** parts of question 1 (25 marks in total).

Question 1

(a) **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 $\mathbf{H}_0 : \theta = \theta_0$.**

(5 marks)

(b) **Explain what is meant by consistency in a statistical estimator. Under what conditions the least squares estimate of the slope coefficient in a simple regression of y_t on x_t is consistent?**

(5 marks)

(c) **A simple random sample of size three, X_1, X_2 and X_3 is drawn from population with mean μ and variance σ^2. Consider the following estimators of μ:**

$$\hat{\mu}_1 = 0.02X_1 + 0.5X_2 + 0.8X_3; \quad \hat{\mu}_2 = \frac{X_1 + X_2}{2} \quad \text{and} \quad \hat{\mu}_3 = \frac{X_1 + X_2 + X_3}{3}$$

Which estimator is most efficient? Explain your answer fully.

(5 marks)

(d) **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 $y_t = \beta x_t + u_t$, $\mathbf{E}(u_t) = 0$ and $\mathbf{E}(u_t^2) = \sigma^2$. 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?**

(5 marks)

(e) Explain what you understand by difference-stationary and trend-stationary time series.

(5 marks)

Reading for this question

Subject guide, Chapter 10.

Dougherty, C. *Introduction to Econometrics* (fourth edition) Chapters R.6, R.7, 2.5, 8.3 and 13.1.

Gujarati, D.N. *Basic Econometrics* (fifth edition) Chapter 21 and Appendix A.

(a) **Approaching the question**

Brief description of the Wald test is required. The solution follows.

The Wald test evaluates whether the discrepancy between the maximum likelihood estimate of θ and θ_0 is significant. The test statistic for the null hypothesis $\mathrm{H}_0 : \theta = \theta_0$ is

$$W = \frac{(\hat{\theta} - \theta_0)^2}{\hat{\sigma}^2_{\hat{\theta}}}$$

where $\hat{\theta}$ is the maximum likelihood estimator of θ. $\hat{\sigma}^2_{\hat{\theta}}$ is the estimate of the variance of θ evaluated at the maximum likelihood value. $\hat{\sigma}^2_{\hat{\theta}}$ is obtained as minus the inverse of the second differential of the log-likelihood function evaluated at the maximum likelihood estimate.

Under the null hypothesis that the restriction is valid, the test statistic has a chi-square distribution with degrees of freedom equal to the number of restrictions imposed by H_0.

(b) **Approaching the question**

The definition and sufficient condition of consistency should be given. It should be illustrated with an example. The solution follows.

If $\hat{\theta}$, based on a sample of size T, is a consistent estimator of θ then $\Pr\left(|\hat{\theta} - \theta| > \epsilon\right) \to 0$ as $T \to \infty$ for every $\epsilon > 0$. Another way of expressing this is that $\hat{\theta}$ converges in probability to θ. In short, we can write the above statement as $\operatorname{plim} \hat{\theta} = \theta$, where plim stands for the probability limit. Hence if $\operatorname{plim} \hat{\theta} = \theta$ then $\hat{\theta}$ is a consistent estimator of θ.

The sufficient condition for consistency is

$$\mathrm{E}(\hat{\theta}) = \theta$$

or

$$\lim_{T\to\infty} \mathrm{E}(\hat{\theta}) = \theta \qquad \text{and} \qquad \lim_{T\to\infty} \mathrm{Var}(\hat{\theta}) \to 0.$$

Let the model be

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

$\mathrm{E}(u_t) = 0$, $\mathrm{E}(u_t^2) = \sigma^2$ and $\mathrm{E}(u_s u_t) = 0$ if $s \neq t$, for all $s, t = 1, 2, \ldots, T$.

$$\begin{aligned}\hat{\beta} &= \frac{\sum x_t y_t}{\sum x_t^2} = \beta + \frac{\sum x_t u_t}{\sum x_t^2} \\ \operatorname{plim}(\hat{\beta}) &= \beta + \frac{\operatorname{plim}(\sum x_t u_t)/T}{\operatorname{plim}(\sum x_t^2)/T} = \beta + \frac{\mathrm{cov}(x,u)}{\mathrm{var}(x)} = \beta + \frac{0}{\sigma_x^2} = \beta.\end{aligned}$$

Hence $\hat{\beta}$ is a consistent estimator of β.

Assumptions: The xs are non-stochastic, or $\mathrm{cov}(x, u) = 0$.

(c) **Approaching the question**

To demonstrate efficiency it is required to show that the estimator has the minimum variance among the class of unbiased estimators. The solution follows.

$$\begin{aligned} \mathrm{E}(\hat{\mu}_1) &= (0.2+0.5+0.8)\mu = 1.5\mu \quad \Rightarrow \quad \text{biased} \\ \mathrm{E}(\hat{\mu}_2) &= \mu \quad \Rightarrow \quad \text{unbiased} \\ \mathrm{E}(\hat{\mu}_3) &= \mu \quad \Rightarrow \quad \text{unbiased} \end{aligned}$$

For efficiency, variances of only unbiased estimators are compared. Note the Xs are mutually independent.

$$\begin{aligned} \mathrm{var}(\hat{\mu}_2) &= \frac{\mathrm{var}(X_1+X_2)}{4} = \frac{\mathrm{var}(X_1)+\mathrm{var}(X_2)}{4} = \frac{2\sigma^2}{4} = \frac{\sigma^2}{2} \\ \mathrm{var}(\hat{\mu}_3) &= \frac{\mathrm{var}(X_1+X_2+X_3)}{4} = \frac{\mathrm{var}(X_1)+\mathrm{var}(X_2)+\mathrm{var}(X_3)}{4} = \frac{3\sigma^2}{9} = \frac{\sigma^2}{3}. \end{aligned}$$

Since $\mathrm{var}(\hat{\mu}_2) > \mathrm{var}(\hat{\mu}_3) \Rightarrow \hat{\mu}_3$ is efficient in comparison to $\hat{\mu}_2$.

(d) **Approaching the question**

Assumptions should be clearly stated. The solution follows.

$$\tilde{\beta} = \frac{1}{N}\sum_{t=1}^{N}\frac{y_t}{x_t}; \quad y_t = \beta x_t + u_t; \quad \mathrm{E}(u_t) = 0; \quad \mathrm{E}(u_t^2) = \sigma^2$$

$$\mathrm{E}(\tilde{\beta}) = \frac{1}{N}\sum_{t=1}^{N}\frac{\mathrm{E}(y_t)}{x_t} = \frac{1}{N}\sum_{t=1}^{N}\frac{\beta x_t}{x_t} = \frac{N\beta}{N} = \beta \quad \Rightarrow \quad \text{unbiased}$$

Assuming $\mathrm{E}(u_s u_t) = 0$ if $s \neq t$,

$$\mathrm{var}(\tilde{\beta}) = \mathrm{var}\left[\frac{1}{N}\sum_{t=1}^{N}\frac{y_t}{x_t}\right] = \frac{1}{N^2}\sum_{t=1}^{N}\frac{\mathrm{var}(y_t)}{x_t^2} = \frac{1}{N^2}\sum_{t=1}^{N}\frac{\mathrm{var}(u_t)}{x_t^2} = \frac{\sigma^2}{N^2}\sum_{t=1}^{N}\frac{1}{x_t^2}.$$

Extra assumptions used are:

i. $\mathrm{E}(u_s u_t) = 0$ if $s \neq t$.

ii. The xs are non-stochastic.

(e) **Approaching the question**

The definition of difference-stationary series and trend-stationary series is required. It should also be illustrated with an example. The behaviour of time series in both cases should be explained. The solution follows.

If after removing the trend from a non-stationary series the resulting variable becomes stationary, then the variable is called *trend-stationary*. Let

$$Z_t = X_t - \alpha_1 t = \alpha_0 + u_t$$

where $\mathrm{E}(u_t) = 0$, $\mathrm{var}(u_t) = \sigma^2$ and $\mathrm{E}(u_t u_{t-s}) = 0$ for all s and t. Then

$$\begin{aligned} \mathrm{E}(Z_t) &= \mathrm{E}(\alpha_0 + u_t) \\ \mathrm{var}(Z_t) &= \mathrm{var}(\alpha_0 + u_t) = \sigma^2 \\ \mathrm{cov}(Z_t, Z_{t-s}) &= \mathrm{E}\left((Z_t - \mathrm{E}(Z_t))(Z_{t-s} - \mathrm{E}(Z_{t-s}))\right) = \mathrm{E}(u_t u_{t-s}) = 0. \end{aligned}$$

This means that Z_t has constant mean and variance for all t, and covariance is zero for all s. It implies that the series is trend-stationary.

If a non-stationary process can be transformed into a stationary process by differencing then the series is said to be *difference-stationary*.

Let X_t be a random walk with a drift

$$X_t = \beta_0 + X_{t-1} + \epsilon_t \qquad \text{(i)}$$

where $\mathrm{E}(\epsilon_t) = 0$, $\mathrm{var}(\epsilon_t) = \sigma^2$ and $\mathrm{E}(\epsilon_t \epsilon_s) = 0$ for all s and t, $s \neq t$.

Subtract X_{t-1} from both sides of (i) to get

$$\Delta X_t = X_t - X_{t-1} = \beta_0 + \epsilon_t.$$

It can be easily checked that $\mathrm{E}(\Delta X_t) = \beta_0$, $\mathrm{var}(\Delta X_t) = \sigma_\epsilon^2$ and $\mathrm{cov}(\Delta X_t, \Delta X_{t-s}) = 0$ for all s and t. This means that ΔX_t is stationary. This implies that X_t is difference-stationary.

It is important to know whether a variable is difference- or trend-stationary because for difference-stationary variables shocks have a permanent effect whereas for trend-stationary variables shocks are transitory.

Section B

Answer **three** questions from this section (25 marks each).

Question 2

Briefly explain the following:

(a) Goldfeld-Quandt test.

(7 marks)

(b) Instrumental variable estimation.

(8 marks)

(c) Error correction model.

(10 marks)

Reading for this question

Dougherty, C. *Introduction to Econometrics* (fourth edition) Chapters 7.2, 8.6 and 13.5.

Gujarati, D.N. *Basic Econometrics* (fifth edition) Chapters 11.5, 13.5 and 21.11.

(a) **Approaching the question**

In Dougherty the Goldfeldt-Quandt test is given. The test is explained below.

Suppose the model is

$$Y_t = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i; \qquad i = 1, 2, \ldots, n.$$

H_0 : Homoskedastic disturbances

H_1 : Heteroskedastic disturbances

This test assumes that $\mathrm{var}(u_i)$ is proportional to the size of one of the RHS variables (say X_{1i}). The observations are ranked by X_{1i} and run a separate regression for the first $n_1 < n$ and the last n_1 observations – the middle $(n - 2n_1)$ observations are not used. If heteroskedasticity is present, the RSS from the two regressions will differ. Form the test RSS_2/RSS_1 where RSS_1 is the residual sum of squares from the first n_1 observations and RSS_2 is the residual sum of squares from the last n_1 observations. The test statistic will have an F-distribution with $(n_1 - K, n_1 - K)$ degrees of freedom, where K is the number of parameters in the equation.

(b) **Approaching the question**

With the help of a two variable linear model, IV, the estimator and its consistency should be illustrated. The reason for using IV should also be stated. The working is given below.

Consider the model

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

If X_t is not independently distributed of u_t then the OLS estimator of β will be inconsistent. Consider a variable Z that is correlated with u but not correlated with X. Z can be considered as an instrumental variable. An estimator of β based on Z is known as an instrumental variable (IV) estimator. It is defined as

$$\hat{\beta}_{IV} = \frac{\sum Z_t Y_t}{\sum Z_t X_t}.$$

It can be shown that $\hat{\beta}_{IV}$ is a consistent estimator of β.

$$\begin{aligned}
\hat{\beta}_{IV} &= \frac{\sum Z_t Y_t}{\sum Z_t X_t} = \frac{\sum Z_t(\beta X_t + u_t)}{\sum Z_t X_t} = \beta + \frac{\sum Z_t u_t}{\sum Z_t X_t} \\
\text{plim}(\hat{\beta}_{IV}) &= \beta + \frac{\text{plim}\,(\sum Z_t u_t/T)}{\text{plim}\,(\sum Z_t X_t/T)} = \beta.
\end{aligned}$$

Hence $\hat{\beta}$ is a consistent estimator of β.

[Note: $\text{plim}(\sum Z_t u_t/T) = 0$ and $\text{plim}(\sum Z_t X_t/T) \neq 0$.]

(c) **Approaching the question**

The ADL(1,1) model should be specified and the error correction model (ECM) should be derived. The solution follows.

Consider a simple ADL(1,1) model. [This is also known as ARDL(1,1).]

$$Y_t = \alpha_1 + \alpha_2 Y_{t-1} + \alpha_3 X_t + \alpha_4 X_{t-1} + u_t \qquad \text{(i)}$$

where Y_t and X_t are I(1). Rewrite (i) as

$$\begin{aligned}
Y_t - Y_{t-1} &= \alpha_1 + \alpha_2 Y_{t-1} - Y_{t-1} + \alpha_3 X_t - \alpha_3 X_{t-1} + \alpha_3 X_{t-1} + \alpha_4 X_{t-1} + u_t \\
\Delta Y_t &= \alpha_1 - (1-\alpha_2) Y_{t-1} + \alpha_3 \Delta X_t + (\alpha_3 + \alpha_4) X_{t-1} + u_t \\
\Delta Y_t &= \alpha_3 \Delta X_t - (1-\alpha_2)\left[Y_{t-1} - \frac{\alpha_1}{(1-\alpha_2)} - \frac{(\alpha_3+\alpha_4)}{(1-\alpha_2)} X_{t-1}\right] + u_t \\
\Delta Y_t &= \alpha_3 \Delta X_t - (1-\alpha_2)\left[Y_{t-1} - \beta_1 - \beta_2 X_{t-1}\right] + u_t
\end{aligned}$$

or,

$$\Delta Y_t = \alpha_3 \Delta X_t - \pi\left[Y_{t-1} - \beta_1 - \beta_2 X_{t-1}\right] + u_t \qquad \text{(ii)}$$

where

$$\pi = (1-\alpha_2); \qquad \beta_1 = \frac{\alpha_1}{(1-\alpha_2)} \qquad \text{and} \qquad \beta_2 = \frac{(\alpha_3+\alpha_4)}{(1-\alpha_2)}.$$

Equation (ii) is the ECM.

When the two variables Y and X are cointegrated, the ECM incorporates not only the short-run but also long-run effects. The long run equilibrium $Y_{t-1} - \beta_1 - \beta_2 X_{t-1}$ is included in the model together with the short-run effect captured by the differenced term.

All the terms in the ECM, given by (ii), are stationary. As Y and X are I(1), then ΔX and ΔY are I(0). As Y and X are cointegrtaed their linear combination $u_{t-1} = Y_{t-1} - \beta_1 - \beta_2 X_{t-1}$ is I(0).

The coefficient π provides us with the information about the speed of adjustment in cases of disequilibrium:

i. If $\pi = 1$, then 100% of the adjustment takes place within the period. In other words adjustment is instantaneous and full.

ii. If $\pi = 0.5$, then 50% adjustment takes place each period.

iii. If $\pi = 0$, then there is no adjustment.

Question 3

(a) Explain the meaning of spurious regression.

(6 marks)

(b) The following equations were estimated by ordinary least squares:

$$Y_t = \underset{(\mathbf{0.1305})}{\mathbf{3.0920}} + \underset{(\mathbf{0.0103})}{\mathbf{0.6959}} X_t + \hat{u}_t \qquad (1)$$

$R^2 = 0.99$, $F = 4,523.25$, $s = 0.0236$, $DW = 0.557$, $T = 740$

$$\Delta\hat{u}_t = \mathbf{0.2349}\Delta\hat{u}_{t-1} + \mathbf{0.2029}\Delta\hat{u}_{t-2} - \mathbf{0.2161}\hat{u}_{t-1} + \hat{\epsilon}_t \qquad (2)$$

$R^2 = 0.1799$, $s = 0.0115$, $T = 737$, $LM(2) = 11.28$

s is the standard error of the residuals, T is the number of observations, $\hat{u}_t$ and $\hat{\epsilon}_t$ are OLS residuals, standard errors in brackets. LM(2) is an LM test statistic to test against second order residual serial correlation.

Do the results above indicate that Y_t and X_t are cointegrated? Specify clearly all the assumptions you have made.

(6 marks)

[Note: Critical value at 5% level of significance from MacKinnon table is -3.3377.]

(c) Consider a model

$$y_t = u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2}; \quad t = 1, 2, \ldots, T$$

$\mathbf{E}(u_t) = 0$; $\mathbf{E}(u_t^2) = \sigma^2$ and $\mathbf{E}(u_s u_t) = 0$ if $s \neq t$ for all $s, t = 1, 2, \ldots, T$.

i. Is y_t stationary? Explain in detail.

(8 marks)

ii. Calculate the autocorrelation function of y_t.

(5 marks)

Reading for this question

Subject guide, Chapter 13.

Dougherty, C. *Introduction to Econometrics* (fourth edition) Chapters 13.2, 13.3 and 13.4.

Gujarati, D.N. *Basic Econometrics* (fifth edition) Chapters 21.7, 21.8 and 21.11.

(a) **Approaching the question**

The meaning of spurious regression should be discussed. With the help of simple models it should be demonstrated that although there may be no relationship between the dependent variable and explanatory variables, spurious regression may result in showing that explanatory variables significantly explain variations in the dependent variable. A detailed discussion follows.

Spurious regression was first demonstrated by Granger and Newbold who showed, using Monte Carlo techniques, that a regression involving two non-stationary series could give rise to spurious results in that the t-statistics over-rejected the null hypothesis of a zero coefficient for two independent random walk series.

If Y_t and X_t are non-stationary and we regress Y_t and X_t, as in

$$Y_t = \pi_0 + \pi_1 X_t + v_t$$

then even if there is no relationship between Y and X, the regression will produce a t-ratio which *will reject the null hypothesis* $\mathrm{H}_0 : \pi_1 = 0$.

The reason for this result is that if $\mathrm{H}_0 : \pi_1 = 0$ then

$$Y_t = \pi_0 + v_t.$$

Suppose Y_t is I(1). Since Y_t is I(1) and π_0 is constant, it follows that v_t must be I(1). This violates the standard distributional theory based on the assumption that v_t is stationary, i.e. v_t is I(0). Hence the misleading result.

(b) **Approaching the question**

We test

H_0 : Not cointegrated

H_1 : Cointegrated

The cointegration test statistic is $-0.2161/0.0845 = -2.557$. The 5% critical value given in the Mackinnon table is -3.3377. Thus we cannot reject the null of no cointegration.

The main assumption is that the error terms in both equations have constant variances and no serial correlation. The critical value of χ^2_2 at the 5% level of significance is 5.99. $LM(2)$ gives a test statistic value of 11.28. Hence we reject the null of no autocorrelation. It casts doubt on the result of cointegration. Also we need to assume that the specifications are correct (e.g. no structural breaks).

(c) **Approaching the question**

For part (i) it should be shown that the mean, variance and covariances are independent of time. For part (ii) the first formula for the autocorrelation function should be given. The solution follows.

(i) We have

$$\begin{aligned}
\mathrm{E}(y_t) &= 0 \\
\\
\mathrm{var}(y_t) &= \mathrm{E}(y_t^2) \\
&= \mathrm{E}(u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2})^2 \\
&= \mathrm{E}(u_t^2) + \theta_1^2 \mathrm{E}(u_{t-1}^2) + \theta_2^2 \mathrm{E}(u_{t-2})^2; \qquad \text{as } \mathrm{E}(u_s u_t) = 0 \text{ if } s \neq t \\
&= (1 + \theta_1^2 + \theta_2^2)\sigma^2 \\
\\
\mathrm{cov}(y_t, y_{t-1}) &= \mathrm{E}(y_y y_{t-1}) \\
&= \mathrm{E}\left((u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2})(u_{t-1} + \theta_1 u_{t-2} + \theta_2 u_{t-3})\right) \\
&= (\theta_1^2 + \theta_1\theta_2)\sigma^2 \\
\\
\mathrm{cov}(y_t, y_{t-2}) &= \mathrm{E}(y_t y_{t-2}) \\
&= \mathrm{E}\left((u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2})(u_{t-2} + \theta_1 u_{t-3} + \theta_2 u_{t-4})\right) \\
&= \theta_2\sigma^2 \\
\\
\mathrm{cov}(y_t, y_{t-s}) &= \mathrm{E}(y_t y_{t-s}) = 0 \qquad \text{for all } s > 2.
\end{aligned}$$

Hence as the mean, variance and covariances are constant over time, y_t is weakly stationary. If the us are normally distributed then this also implies strong stationarity.

(ii) The autocorrelation function is defined as

$$\rho_s = \frac{\text{cov}(y_t, y_{t-s})}{\sqrt{\text{var}(y_t)\text{var}(y_{t-s})}} = \frac{\text{cov}(y_t, y_{t-s})}{\text{var}(y_t)},$$

as $\text{var}(y_t) = \text{var}(y_{t-s})$. Hence,

$$\rho_s = \begin{cases} 1 & \text{if } s = 0 \\ \frac{(\theta_1+\theta_1\theta_2)}{(1+\theta_1^2+\theta_2^2)} & \text{if } s = 1 \\ \frac{\theta_2}{(1+\theta_1^2+\theta_2^2)} & \text{if } s = 2 \\ 0 & \text{if } s > 2. \end{cases}$$

Question 4

In a certain bond market, the demand for bonds, B_t, in period t is negatively related to the expected interest rate, I^e_{t+1}, in period $t+1$ as

$$B_t = \beta_1 + \beta_2 I^e_{t+1} + u_t; \quad t = 1, 2, \ldots, T \qquad (1)$$

where $\mathbf{E}(u_t) = 0$; $\mathbf{E}(u_t^2) = \sigma^2$ and $\mathbf{E}(u_s u_t) = 0$ if $s \neq t$ for all $s, t = 1, 2, \ldots, T$. The expected interest rate is determined by an adaptive expectations process:

$$I^e_{t+1} - I^e_t = \lambda(I_t - I^e_t) \qquad (2)$$

where I_t is the actual interest rate in period t. A researcher uses the following model to fit the relationship:

$$B_t = \alpha_1 + \alpha_2 I_t + \alpha_3 B_{t-1} + v_t \qquad (3)$$

where v_t is an unobserved disturbance term.

(a) Show that the model can be derived from the demand function and the adaptive expectations process.

(7 marks)

(b) Explain intuitively why inconsistent estimates of the parameters will be obtained if equation (3) is fitted using ordinary least squares (OLS).

(4 marks)

(c) Describe a method for fitting the model that would yield consistent estimates.

(7 marks)

(d) Suppose that u_t followed a first order autoregressive process:

$$u_t = \rho u_{t-1} + \epsilon_t; \quad t = 1, 2, \ldots, T$$

$\mathbf{E}(\epsilon_t) = 0$; $\mathbf{E}(\epsilon_t^2) = \sigma^2$ and $\mathbf{E}(\epsilon_s \epsilon_t) = 0$ if $s \neq t$ for all $s, t = 1, 2, \ldots, T$. Assume that $\rho + \lambda = 1$. How would this affect the answer given in (b)?

(4 marks)

(e) Suppose the true relationship was actually

$$B_t = \beta_1 + \beta_2 I_t + u_t; \quad t = 1, 2, \ldots, T$$

where $\mathbf{E}(u_t) = 0$; $\mathbf{E}(u_t^2) = \sigma^2$ and $\mathbf{E}(u_s u_t) = 0$ if $s \neq t$ for all $s, t = 1, 2, \ldots, T$. If, on the other hand equation (3) was fitted using OLS how would this affect the regression results?

(3 marks)

Reading for this question

Subject guide, Chapter 11.

Dougherty, C. *Introduction to Econometrics* (fourth edition) Chapters 6.3 and 8.1.

(a) **Approaching the question**

From the adaptive expectations process, I^e_{t+1} should be obtained and substituted in the model. Then the obtained model should be lagged by one period to get $\beta_2 I^e_t$. Detailed working follows.

The demand function given is

$$B_t = \beta_1 + \beta_2 I^e_{t+1} + u_t. \qquad \text{(i)}$$

The adaptive expectations process may be written as

$$I^e_{t+1} = \lambda I_t + (1-\lambda) I^e_t. \qquad \text{(ii)}$$

Substituting (ii) in (i), we obtain

$$\begin{aligned} B_t &= \beta_1 + \beta_2[\lambda I_t + (1-\lambda)I^e_t] + u_t \\ &= \beta_1 + \beta_2 \lambda I_t + \beta_2(1-\lambda) I^e_t + u_t. \qquad \text{(iii)} \end{aligned}$$

Lagging (i) by one period, we get

$$B_{t-1} = \beta_1 + \beta_2 I^e_t + u_{t-1}.$$

This gives

$$\beta_2 I^e_t = B_{t-1} - \beta_1 - u_{t-1}. \qquad \text{(iv)}$$

Substituting (iv) into (iii) gives

$$\begin{aligned} B_t &= \beta_1 \lambda + \beta_2 \lambda I_t + (1-\lambda) B_{t-1} + u_t - (1-\lambda) u_{t-1} \\ &= \alpha_1 + \alpha_2 I_t + \alpha_3 B_{t-1} + v_t. \end{aligned}$$

(b) **Approaching the question**

It should be explained that there is correlation between the disturbance term and one of the explanatory variables. The solution follows.

As shown above, $B_{t-1} = \beta_1 + \beta_2 I^e_t + u_{t-1}$. Hence B_{t-1} is partly determined by u_{t-1}. The disturbance term $v_t = u_t - (1-\lambda)u_{t-1}$, hence v_t also has a component u_{t-1}. Therefore the requirement that the disturbance term and the regressors be distributed independently of each other is violated. This violation will lead to inconsistent OLS estimates.

(c) **Approaching the question**

The result can be obtained by recursive substitution. The solution follows.

From (ii) in part (a), we get

$$I^e_t = \lambda I_{t-1} + (1-\lambda) I^e_{t-1}.$$

Substituting this in (iii) of (a), we get

$$B_t = \beta_1 + \beta_2 \lambda I_t + \beta_2 \lambda (1-\lambda) I_{t-1} + (1-\lambda)^2 I^e_{t-1} + u_t.$$

Doing recursive substitution we get

$$B_t = \beta_1 + \beta_2 \lambda I_t + \beta_2 \lambda (1-\lambda) I_{t-1} + \ldots + \beta_2 \lambda (1-\lambda)^{s-1} I_{t-s+1} + (1-\lambda)^s I^e_{t-s+1} + u_t.$$

For large enough s, the term $(1-\lambda)^s$ will be so small that we can drop the unobservable term with negligible omitted variable bias. The disturbance term u_t is distributed independently of regressors, hence we obtain consistent estimates of the parameters. This should be estimated using non-linear estimation techniques.

(d) **Approaching the question**

In this case,

$$\begin{aligned} v_t &= u_t - (1-\lambda)u_{t-1} \\ &= \rho u_{t-1} + \epsilon_t - (1-\lambda)u_{t-1} \\ &= \epsilon_t - (1-\rho-\lambda)u_{t-1} \\ &= \epsilon_t \end{aligned}$$

as $\rho + \lambda = 1$. Hence there is no violation of the regression assumptions and parameters in (3) can be estimated by OLS.

(e) **Approaching the question**

As B_{t-1} is a redundant variable, this is the case of inclusion of an irrelevant explanatory variable. OLS estimates will be unbiased, but inefficient.

Question 5

(a) Consider a model

$$Y_{it} = \beta_1 + \beta_j \sum_{j=2}^{K} X_{jit} + \alpha_i + u_{it}; \quad i = 1, 2, \ldots, N; \quad t = 1, 2, \ldots, T$$

where Y is the dependent variable, the X_j are observed explanatory variables, α_i is an unobserved effect and u_{it} is the disturbance term assumed to satisfy the usual regression model conditions. The index i refers to cross-section and t refers to the time period.

Explain the differences between the within-groups, first differences and least squares dummy variables versions of the fixed effect model.

(15 marks)

(b) The following estimates were made as part of a study of whether countries with low income per head had higher growth rates than countries with higher income per head (convergence). Data was obtained over a period of four years on 121 countries.

The first estimates are pooled ordinary least squares estimates:

$$\underset{}{gr_{it}} = \underset{(\mathbf{0.358})}{\mathbf{-0.129}} \; \underset{(\mathbf{0.004})}{\mathbf{-0.023}\,pgdp_{it}} \; + \underset{(\mathbf{0.018})}{\mathbf{0.139}\,inv_{it}} \; + e_{1it}; \quad i = 1, 2, \ldots, N \quad t = 1, 2, \ldots, T \qquad \mathbf{(1)}$$

$N = 121$, $T = 4$, $R^2 = 0.133$.

The second estimates are fixed effects (least squares dummy variable variant):

$$gr_{it} = \underset{(\mathbf{0.358})}{\mathbf{3.524}} \; \underset{(\mathbf{0.012})}{\mathbf{-0.046}\,pgdp_{it}} \; \underset{(\mathbf{0.032})}{\mathbf{-0.020}\,inv_{it}} \; + e_{2it}; \quad i = 1, 2, \ldots, N \quad t = 1, 2, \ldots, T \qquad \mathbf{(2)}$$

$N = 121$, $T = 4$, $R^2 = 0.541$.

Standard errors are in brackets. gr is the average percentage annual rate of growth of gdp per head, $pgdp$ is the gdp per head in constant US dollar, inv is the average percentage ratio of investment to gdp. e_1 and e_2 are the ordinary least squares residuals. R^2 is the conventionally calculated coefficient of determination.

i. Comment on the differences in R^2 for each equation.

(3 marks)

ii. Equation (1) is restricted version of equation (2). Test the validity of the restrictions.

(4 marks)

iii. From these results , what would you conclude about the convergence of gdp per head? Give details.

(3 marks)

Reading for this question

Dougherty, C. *Introduction to Econometrics* (fourth edition) Chapters 1.7, 6.5 and 14.2.

Gujarati, D.N. *Basic Econometrics* (fifth edition) Chapter 16.3.

(a) **Approaching the question**

Within-groups fixed effect, first differences fixed effect and least squares dummy variable (LSDV) fixed effect regressions should be derived and the drawbacks of each model should be discussed. A detailed discussion follows.

- **Within-groups fixed effect**

The model given is

$$Y_{it} = \beta_1 + \sum_{j=2}^{K} \beta_j X_{jit} + \alpha_i + u_{it}; \quad i = 1, 2, \ldots, n; \quad t = 1, 2, \ldots, T. \qquad \text{(i)}$$

Sum the observations of each cross-sectional unit over the time dimension and divide by T to get

$$\bar{Y}_i = \beta_1 + \sum_{j=2}^{K} \beta_j \bar{X}_{ij} + \alpha_i + \bar{u}_i. \qquad \text{(ii)}$$

Subtracting (i) from (ii), we get

$$Y_{it} - \bar{Y}_i = \sum_{j=2}^{K} \beta_j (X_{jit} - \bar{X}_{ij}) + u_{it} - \bar{u}_i. \qquad \text{(iii)}$$

In (iii) we can see that the unobserved effect (α_i) disappears. This is known as within-groups regression. It explains the variations about the mean of the dependent variable in terms of the variations about the means of the explanatory variables for the group of observations relating to a given individual.

- Drawbacks
  - · The intercept and any explanatory variable that remain constant for each individual will drop out of the model.
  - · The variation in $(X_{ij} - \bar{X}_i)$ may be much smaller than the variation in X_j. If this is the case, the impact of the disturbance term may be relatively large, giving rise to imprecise estimates.
  - · There is loss of a substantial number of degrees of freedom.

- **First differences fixed effect**

In this approach, the unobserved heterogeneity is eliminated by subtracting the observation from the previous time period from the observation for the current time period, for all time periods.
Lag (i) by one period to get

$$Y_{it-1} = \beta_1 + \sum_{j=2}^{K} \beta_j X_{jit-1} + \alpha_i + u_{it-1}. \qquad \text{(iv)}$$

Subtracting (i) from (iv), we get

$$\Delta Y_{it-1} = \sum_{j=2}^{K} \beta_j \Delta X_{jit-1} + \alpha_i + u_{it} - u_{it-1} \qquad \text{(v)}$$

and the unobserved heterogeneity (α_i) disappears.

- Drawbacks
  - The intercept and any explanatory variable that remain constant for each individual will drop out of the model.
  - n degrees of freedom are lost as the first observation for each individual is not defined.
  - It gives rise to autocorrelation.

- **Least squares dummy variable (LSDV) fixed effect**

  In this approach the unobserved effect is brought explicitly into the model. A set of dummy variables D_i is defined, where D_i is equal to 1 in the case of an observation relating to individual i and 0 otherwise. The model can be written as

$$Y_{it} = \sum_{j=2}^{K} \beta_j X_{jit} + \sum_{i=1}^{n} \alpha_i D_i + u_{it}. \qquad \text{(vi)}$$

  The unobserved effect is now being treated as the coefficient of the specific individual i. The term $\alpha_i D_i$ represents a fixed effect on the dependent variable Y_i for individual i.

  If we want to keep the intercept in the model then instead of n dummy variables $(n-1)$ dummy variables have to be used, otherwise we will fall into the dummy variable trap.

  It can be shown that the LSDV method is identical to the within-groups method. Hence, the drawbacks are the same.

(b) i. **Approaching the question**

The definition of R^2 should be given and an explanation that R^2 is a non-decreasing function of the number of explanatory variables. The solution follows.

In equation (1) there are two explanatory variables whereas in equation (2) there are 122 explanatory variables (an extra 120 dummy variables are included in the model).

$$R^2 = \frac{\text{Explained Sum of Squares } (ESS)}{\text{Total sum of squares } (TSS)}.$$

TSS is the same in both models but ESS is greater in (2) because of 120 more explanatory variables (the dummies), hence R^2 in (2) is more than R^2 in (1).

ii. **Approaching the question**

An F-test for testing restrictions has to be used. The solution follows.

The null hypothesis is that the coefficients of all the dummies in (2) are equal to zero. This is an F-test for linear restrictions.

$$F_{120,361} = \frac{(0.541 - 0.133)/120}{(1 - 0.541)/361} = 2.677.$$

The critical value of $F_{120,361}$ at the 5% level of significance is approximately 1.22. Reject the null hypothesis. Restrictions imposed on (1) are not valid.

iii. **Approaching the question**

Since (1) is rejected we should concentrate on (2).

Estimated coefficient of *inv* has wrong sign but it is insignificant. Estimated coefficient of *pgdp* is negative and significant showing some slow convergence.

Question 6

(a) State the properties of maximum likelihood (ML) estimators.

(5 marks)

(b) Let X be a random variable with probability density function

$$f(X) = \left(2\pi\sigma^2\right)^{-1/2} \exp\left[-\frac{(X-\mu)^2}{2\sigma^2}\right].$$

Let $X_1, X_2, \ldots, X_T$ be an identically independently distributed random sample from X.

i. Derive the ML estimators of μ and σ^2.

(10 marks)

ii. Show how to test the null hypothesis $\mathbf{H}_0 : \mu = \mu_0$, using the large sample likelihood ratio test.

(10 marks)

Reading for this question

Subject guide, Chapter 10.

Dougherty, C. *Introduction to Econometrics* (fourth edition) Chapter 10.6.

Gujarati, D.N. *Basic Econometrics* (fifth edition) Chapters 4.4 and 4.5.

(a) **Approaching the question**

i. MLE is consistent.

ii. MLE is invariant to the transformation of parameters, for example

* If $\hat{\theta}$ is the MLE of θ, then $\hat{\theta}^2$ is the MLE of θ^2.
* If $\hat{\theta}$ is the MLE of θ, then $\exp(\hat{\theta})$ is the MLE of $\exp(\theta)$.

iii. MLE is efficient in large samples in the sense that the variance of the MLE reaches the Cramer-Rao lower bound (CRLB) in large samples.

iv. MLE is asymptotically normally distributed.

v. If a sufficient estimator exists, then MLE is a function of the sufficient estimator.

(b) i. **Approaching the question**

To obtain the MLE, the likelihood function (L) has to be obtained. The estimator which maximises L is the MLE. As L in general is highly non-linear in the parameters, $\ln L$ is always maximised instead of L. Both give the same maximum. Detailed working follows. The log of the likelihood function is

$$\ln L(\mu, \sigma) = -\frac{T}{2}\ln(2\pi) - \frac{T}{2}\ln\sigma^2 - \frac{\sum(X_t-\mu)^2}{2\sigma^2}. \qquad \text{(i)}$$

Differentiating $\ln L$ with respect to μ, we get

$$\frac{\partial \ln L(\mu,\sigma)}{\partial\mu} = \frac{2\sum(X_t-\mu)}{2\sigma^2} = 0.$$

Solving, we get

$$\hat{\mu}_{MLE} = \frac{\sum X_t}{T} = \bar{X}.$$

Differentiating $\ln L$ with respect to σ^2, we get

$$\frac{\partial \ln L(\mu,\sigma)}{\partial\sigma^2} = -\frac{T}{2\sigma^2} + \frac{\sum(X_t-\mu)^2}{2\sigma^4} = 0.$$

Solving, we get

$$\hat{\sigma}^2_{MLE}(\mu) = \frac{\sum(X_t-\mu)^2}{T} \qquad \text{(ii)}$$

and, after substituting μ by $\hat{\mu}$,

$$\hat{\sigma}^2_{MLE} = \frac{\sum(X_t - \hat{\mu})^2}{T} = \frac{\sum(X_t - \bar{X})^2}{T}. \qquad \text{(ii)}$$

ii. **Approaching the question**

First, the concentrated log-likelihood should be obtained. To derive the large sample LR test one has to obtain logs of unrestricted and restricted likelihood functions. The solution follows.

Substituting (ii) in (i), we get the concentrated log-likelihood function (concentrated log-likelihood function is a function of μ only) as

$$\ln L(\mu) = -\frac{T}{2}\ln(2\pi) - \frac{T}{2}\ln\left(\frac{\sum(X_t - \mu)^2}{T}\right) - \frac{T}{2}.$$

The large sample likelihood ratio (LR) test is defined as

$$LR = 2\ln(\ln L_U - \ln L_R) \sim \chi^2_q$$

where q is the number of restrictions imposed by H_0. $\ln L_U$ is the log of the unrestricted likelihood and $\ln L_R$ is the log of the restricted likelihood.

The unrestricted ML estimator of μ is $\bar{X}$ (obtained in part (i)) and the restricted ML estimator of μ under $\mathrm{H}_0 : \mu = \mu_0$ is μ_0. Hence,

$$\begin{aligned} LR &= 2\left[\left(-\frac{T}{2}\ln(2\pi) - \frac{T}{2}\left(\frac{\sum(X_t - \bar{X})^2}{T}\right) - \frac{T}{2}\right) - \left(-\frac{T}{2}\ln(2\pi) - \frac{T}{2}\left(\frac{\sum(X_t - \mu_0)^2}{T}\right) - \frac{T}{2}\right)\right] \\ &= T\left(\ln\sum(X_t - \mu_0)^2 - \ln\sum(X_t - \bar{X})^2\right) \sim \chi^2_1. \end{aligned}$$

Examiners' commentaries 2013

EC2020 Elements of econometrics

Important note

This commentary reflects the examination and assessment arrangements for this course in the academic year 2012–13. In 2014 the format of the examination will change to:

Candidates should answer **EIGHT** of the following **TEN** questions: **ALL FIVE** questions from Section A (8 marks each) and **THREE** questions from Section B (20 marks each).

Section A is intended to examine across the whole syllabus while Section B is designed to examine a selection of topics in greater depth.

The format and structure of the examination may change again in future years, and any such changes will be publicised on the virtual learning environment (VLE).

Information about the subject guide

Unless otherwise stated, all cross-references will be to the latest version of the subject guide (2011). You should always attempt to use the most recent edition of any Essential reading textbook, even if the commentary and/or online reading list and/or subject guide refers to an earlier edition. If different editions of Essential reading are listed, please check the VLE for reading supplements – if none are available, please use the contents list and index of the new edition to find the relevant section.

General remarks

Learning outcomes

At the end of this course, and having completed the Essential reading and activities, you should be able to:

- describe and apply the classical regression model and its application to cross-section data
- describe and apply the:
  - Gauss-Markov conditions and other assumptions required in the application of the classical regression model
  - reasons for expecting violations of these assumptions in certain circumstances
  - tests for violations
  - potential remedial measures, including, where appropriate, the use of instrumental variables
- recognise and apply the advantages of logit, probit and similar models over regression analysis when fitting binary choice models
- competently use regression, logit and probit analysis to quantify economic relationships using standard regression programmes (Stata and EViews) in simple applications

- describe and explain the principles underlying the use of maximum likelihood estimation
- apply regression analysis to fit time-series models using stationary time series, with awareness of some of the econometric problems specific to time series applications (for example, autocorrelation) and remedial measures
- recognise the difficulties that arise in the application of regression analysis to nonstationary time series, know how to test for unit roots, and know what is meant by cointegration.

Common mistakes committed by candidates

- A large number of candidates were not able to distinguish between sample variance and covariance, and population variance and covariance (this is happening year after year). They treat them as the same. This results in incorrect analysis and candidates lose significant marks.

  Consider an example: Suppose data is deviation from respective sample means and the regression model is:

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

  The ordinary least squares estimator of β is:

  $$\hat{\beta} = \frac{\sum_{t=1}^{T} x_t y_t}{\sum_{t=1}^{T} x_t^2} = \beta + \frac{\sum_{t=1}^{T} x_t u_t}{\sum_{t=1}^{T} x_t^2}.$$

  In terms of variances and covariances (a large number of candidates prefer this terminology) this can be written as:

  $$\hat{\beta} = \beta + \frac{\text{Cov}(x, u)}{\text{Var}(x)}.$$

  Here $\text{Cov}(x, u)$ and $\text{Var}(x)$ are sample[$\text{Cov}(x, u)$] and sample[$\text{Var}(x)$].

  Candidates should realise that $\sum_{t=1}^{T} u_t$, $\sum_{t=1}^{T} x_t u_t$, $\text{Cov}(x, u)$ and $\text{Var}(x)$ given above are sample moments and as such $\sum_{t=1}^{T} u_t \neq 0$, $\sum_{t=1}^{T} x_t u_t \neq 0$ and $\text{Cov}(x, u) \neq 0$. But, if we take expectation, then:

  $$E[u_t] = 0,$$

  by assumption. Then:

  $$E\left[\sum_{t=1}^{T} x_t u_t\right] = \sum_{t=1}^{T} x_t[E(u_t)] = 0,$$

  as the x_t are fixed they can be taken out of the expectation, and so:

  $$E[\text{Cov}(x, u)] = E\left[\frac{1}{T}\sum_{t=1}^{T} x_t u_t\right] = 0,$$

  as previously argued. This makes $E(\hat{\beta}) = \beta$, i.e. $\hat{\beta}$ is an unbiased estimator for β.

  To prove consistency take plim to get:

  $$\begin{aligned}
  \text{plim}\left(\hat{\beta}\right) &= \beta + \text{plim}\left(\frac{\frac{1}{T}\sum_{t=1}^{T} x_t u_t}{\frac{1}{T}\sum_{t=1}^{T} x_t^2}\right) \\
  &= \beta + \frac{\text{plim}\left(\frac{1}{T}\sum_{t=1}^{T} x_t u_t\right)}{\text{plim}\left(\frac{1}{T}\sum_{t=1}^{T} x_t^2\right)} \\
  &= \beta + \frac{\text{plim(sample Cov}(x, u))}{\text{plim(sample Var}(x))} \\
  &= \beta + \frac{\text{population Cov}(x, u)}{\text{population Var}(x)}.
  \end{aligned}$$

By assumption, population $\text{Cov}(x, u) = 0$ and population $\text{Var}(x) > 0$, hence $\text{plim}\left(\hat{\beta}\right) = \beta$, in other words $\hat{\beta}$ is a consistent estimator of β.

Remember that in general:

$$\begin{aligned} \text{plim(sample variance)} &= \text{population variance,} \\ \text{plim(sample covariance)} &= \text{population covariance.} \end{aligned}$$

This concept has been used in many questions. This simple mistake of not distinguishing between sample variance and covariance and population variance and covariance results in a significant loss of marks which might result in the loss of a degree class or even be the difference between pass and fail.

- Candidates struggled to give competent answers to the interpretation of empirical results. When interpreting an empirical result you should discuss the significance of the coefficients, magnitude and sign of the coefficients. Also, you should make sure that the GM conditions hold.
- Just as last year, many candidates did not appear to read the questions carefully enough and often omitted to give answers to parts of questions which asked for details of such things as the assumptions necessary for a particular result to be true.

Key steps to improvement

- Essential reading for this course includes the subject guide and

  Dougherty, C. *Introduction to Econometrics.* (Oxford: Oxford University Press, 2011) fourth edition [ISBN 9780199567089].

  To understand the subject clearly, it is important to supplement the textbook with the subject guide, especially the chapters on maximum likelihood and panel data. It is very important to go through the subject guide carefully. It contains solutions to the questions given in the textbook and also some additional questions and solutions. Working through these will improve the clarity of the subject.
- Chapter 10 of the subject guide on maximum likelihood estimation includes some additional theory which is not covered in the textbook. It is important to read the additional theory given in the subject guide to have a better understanding of the principles of maximum likelihood and tests based on the likelihood function.
- Apart from Essential reading you should do some supplementary reading. Two very good books of the same level are:

  Gujarati, Damodar N. *Basic Econometrics.* (Boston; London: McGraw-Hill Education, 2009) fifth edition [ISBN 9780071276252].

  Woolridge, Jeffrey M. *Introductory Econometrics.* (Mason, Ohio: Thomson Learning, 2008) fourth edition [ISBN 9780324788907].
- You should make full use of the resources available in the Online Resource Centre maintained by the textbooks publisher, Oxford University Press: www.oup.com/uk/orc/bin/9780199567089 Here you will find PowerPoint slideshows that provide graphical treatment of the topics covered in the textbook, data sets and statistical tables. Candidates should utilise data sets using standard regression programmes (STATA or EViews). This will help in the understanding of the subject.

Question spotting

Many candidates are disappointed to find that their examination performance is poorer than they expected. This can be due to a number of different reasons and the *Examiners' commentaries* suggest ways of addressing common problems and improving your performance. We want to draw your attention to one particular failing – '**question spotting**', that is, confining your examination preparation to a few question topics which have come up in past papers for the course. This can have very serious consequences.

We recognise that candidates may not cover all topics in the syllabus in the same depth, but you need to be aware that Examiners are free to set questions on **any** aspect of the syllabus. This means that you need to study enough of the syllabus to enable you to answer the required number of examination questions.

The syllabus can be found in the 'Course information sheet' in the section of the VLE dedicated to this course. You should read the syllabus very carefully and ensure that you cover sufficient material in preparation for the examination.

Examiners will vary the topics and questions from year to year and may well set questions that have not appeared in past papers – every topic on the syllabus is a legitimate examination target. So although past papers can be helpful in revision, you cannot assume that topics or specific questions that have come up in past examinations will occur again.

If you rely on a question spotting strategy, it is likely you will find yourself in difficulties when you sit the examination paper. We strongly advise you not to adopt this strategy.

Examiners' commentaries 2013

EC2020 Elements of econometrics

Important note

This commentary reflects the examination and assessment arrangements for this course in the academic year 2012–13. In 2014 the format of the examination will change to:

Candidates should answer **EIGHT** of the following **TEN** questions: **ALL FIVE** questions from Section A (8 marks each) and **THREE** questions from Section B (20 marks each).

Section A is intended to examine across the whole syllabus while Section B is designed to examine a selection of topics in greater depth.

The format and structure of the examination may change again in future years, and any such changes will be publicised on the virtual learning environment (VLE).

Information about the subject guide

Unless otherwise stated, all cross-references will be to the latest version of the subject guide (2011). You should always attempt to use the most recent edition of any Essential reading textbook, even if the commentary and/or online reading list and/or subject guide refers to an earlier edition. If different editions of Essential reading are listed, please check the VLE for reading supplements – if none are available, please use the contents list and index of the new edition to find the relevant section.

Comments on specific questions – Zone A

Candidates should answer **FOUR** of the following **SIX** questions: **QUESTION 1** of Section A (25 marks in total) and **THREE** questions from Section B (25 marks each). **Candidates are strongly advised to divide their time accordingly.**

Section A

Answer question 1 from this section.

Question 1

(a) **A simple random sample of size n, $X_1, X_2, \ldots, X_n$, 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.

(5 marks)

(b) **There are 3 standard ways of measuring the explanatory power of a regression equation: R^2 (adjusted and unadjusted) and the standard F-statistic. Explain**

how these are related to each other, and how they measure the explanatory power of the equation.

(5 marks)

(c) **Explain what is meant by (i) serial correlation and (ii) heteroscedasticity in the disturbances of regression models. Discuss the problems they create for econometric work.**

(5 marks)

(d) **Explain how dummy variables can be used to verify structural changes.**

(5 marks)

(e) **Consider a two equation model:**

$$\begin{aligned} y_t &= \alpha z_t + u_{1t} \\ z_t &= \beta_1 y_t + \beta_2 x_t + u_{2t}; \qquad t = 1, 2, \ldots, T \end{aligned}$$

where y_t and z_t are endogenous variables, x_t is an exogenous variable, u_{1t} and u_{2t} are serially uncorrelated disturbances with zero means, variances σ_1^2 and σ_2^2 and covariance σ_{12} for all t. Write down the reduced form corresponding to y and z. Suggest a method of estimation of the reduced form parameters, which will give unbiased and consistent estimators.

(5 marks)

Reading for this question

Subject guide (2011), Chapters 5, 7, 9 and 12.

Dougherty, C. *Introduction to econometrics* (fourth edition) 3.5 (Goodness of fit: R^2), 5 (Dummy variables), 7 (Heteroscedasticity), 9.1 (Simultaneous equations models: Structural and reduced form coefficients), 12.1 (Definition and consequences of autocorrelation) and 12.2 (Detection of autocorrelation).

Gujarati, D.N. *Basic econometrics* (fifth edition) Chapters 7.8 (R^2 and adjusted R^2), 9.5 (Dummy variable regression models), 12 (Autocorrelation: What happens if error terms are autocorrelated) and 19 (The identification problem).

Approaching the question

(a) Sufficient condition of consistency has to be used. A detailed answer is:

$$\mathrm{E}(\hat{\mu}_1) = \mathrm{E}(X_1) + \frac{\sum_{i=2}^{n} \mathrm{E}(X_i)}{n} = \mu + \frac{(n-1)\mu}{n} = 2\mu - \frac{\mu}{n}.$$

Hence, $\hat{\mu}_1$ is a **biased** estimator of μ, and

$$\lim_{n\to\infty} \mathrm{E}(\hat{\mu}_1) \to 2\mu.$$

Therefore, $\hat{\mu}_1$ is an **asymptotically biased** estimator of μ. The sufficient condition for consistency does not hold. Hence, $\hat{\mu}_1$ is **not a consistent** estimator of μ.

$$\mathrm{E}(\hat{\mu}_2) = \frac{\sum_{i=1}^{n-1} \mathrm{E}(X_i)}{n} = \frac{(n-1)\mu}{n} = \mu - \frac{\mu}{n}.$$

Hence, $\hat{\mu}_2$ is a **biased** estimator of μ, and

$$\lim_{n\to\infty} \mathrm{E}(\hat{\mu}_2) \to \mu.$$

Therefore, $\hat{\mu}_2$ is an **asymptotically unbiased** estimator of μ.

$$\text{Var}(\hat{\mu}_2) = \text{Var}\left(\frac{\sum_{i=1}^{n-1} X_i}{n}\right) = \frac{\sum_{i=1}^{n-1}\text{Var}(X_i)}{n^2} = \frac{(n-1)\sigma^2}{n^2} = \frac{\sigma^2}{n} - \frac{\sigma^2}{n^2},$$

and hence

$$\lim_{n\to\infty} \text{Var}(\hat{\mu}_2) \to 0.$$

The sufficient condition for consistency does hold. Hence, $\hat{\mu}_2$ is a **consistent** estimator of μ.

(b) A definition of R^2 and adjusted R^2 should be given. F-statistic should be given in terms of R^2. The answer is:

$$R^2 = \frac{ESS}{TSS}$$

where ESS is the explained sum of squares and TSS is the total sum of squares.

$$\bar{R}^2 = \frac{ESS/(n-k)}{TSS/(n-1)}$$

where n is the sample size and k is the number of parameters.

$$F = \frac{R^2/(k-1)}{(1-R^2)/(n-k)}.$$

R^2 measures the percentage of the total variation in the dependent variable explained by the regression model, $\bar{R}^2$ is the adjusted value of R^2. The term 'adjusted' means adjusted for the degrees of freedom associated with the sum of squares entering into R^2. The F-statistic tests the null hypothesis that all the slope coefficients are simultaneously equal to zero.

(c) Serial correlation and heteroscedasticity should be defined and their consequences on the ordinary least squares should be explained. Answer is:

Serial correlation exists when successive observations over time are related to one another. There are several reasons why autocorrelation may exist such as inertia, omitted variable and incorrect functional form. The most common reason for serial correlation is that an important explanatory variable has been omitted. The easiest correction is to collect data on the omitted variable and include it in a new formulation of the model. Similarly, when quarterly (or monthly) data are employed the presence of non-systematic seasonal variation, or an incomplete accounting for seasonality by the included variables, will produce seasonal effects in the error terms with the consequence that the fourth-order (or the twelfth-order) autocorrelation will be significant. In the presence of autocorrelation the least squares estimators are still linear and unbiased, but they are not efficient, that is, they do not have minimum variance compared to the procedures that take into account autocorrelation. The estimated variances of ordinary least squares (OLS) estimators are biased. Sometimes they are seriously **underestimated**, inflating t-values. Therefore, the usual t- and F-tests are not generally reliable.

Heteroscedasticity occurs whenever the variance of the conditional distribution of the error term is not constant. In the presence of heteroscedasticity the least squares estimators are still linear unbiased and consistent, but they are no longer efficient.

(d) Dummy variable should be defined. It should then be explained how dummy variables can be used to capture changes in the intercept, slope and both intercept and slope. Tests on the coefficients should be explained. The answer is:

To investigate structural change a dummy variable (D) is defined as $D_t = 1$ after the change and $D_t = 0$ otherwise. If the relationship being investigated is $Y_t = \beta_0 + \beta_1 X_t + u_{1t}$ then the model

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 D_t + \beta_3 D_t X_t + u_{2t}$$

will allow a t-test for a change in intercept by using the null $\text{H}_0 : \beta_2 = 0$ and a change in slope, $\text{H}_0 : \beta_3 = 0$. An alternative is to use an F-test to test the joint hypothesis $\text{H}_0 : \beta_2 = \beta_3 = 0$.

(e) Reduced form should be defined. It should be explained that to estimate reduced form OLS is used as all the explanatory variables are exogenous.

RF corresponding to z_t and y_t are:

$$z_t = \frac{\beta_2}{\alpha - \beta_1} x_t + \frac{u_{2t} - u_{1t}}{\alpha - \beta_1} \qquad \text{(i)}$$

$$\begin{aligned} y_t &= \frac{\alpha\beta_2}{\alpha - \beta_1} x_t + \frac{\alpha(u_{2t} - u_{1t})}{\alpha - \beta_1} + u_{1t} \\ &= \frac{\alpha\beta_2}{\alpha - \beta_1} x_t + \frac{\alpha u_{2t} - \beta u_{1t}}{\alpha - \beta_1} \qquad \text{(ii)} \end{aligned}$$

(i) and (ii) will exist only if $\alpha - \beta_1 \neq 0$.

In the reduced form explanatory variable and the disturbance term are independent, i.e. they are exogenous, hence reduced form parameters can be consistently estimated by OLS.

Section B

Answer three questions from this section.

Question 2

Consider a two variable linear model:

$$Y_i = \beta_0 + \beta_1 x_i + u_i \qquad i = 1, 2, \ldots, n$$

where $\mathbf{E}(u_i) = 0$; $\mathbf{E}(u_i^2) = \sigma^2$ and $\mathbf{E}(u_i u_j) = 0$ if $i \neq j$.

(a) Obtain the ordinary least squares (OLS) estimators of β_0 and β_1.

(5 marks)

(b) Suppose that the fitted line is $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_i$ where $\hat{\beta}_0$ and $\hat{\beta}_1$ are the OLS estimators. Prove that the fitted line must pass through the point $(\bar{Y}, \bar{X})$ representing the mean of the variables in the sample.

(5 marks)

(c) Demonstrate that the fitted values of the dependent variable are uncorrelated with the residuals, u_i, in the simple regression model.

(6 marks)

(d) An investigator correctly believes that the relationship between the variables X and Y is described by the linear model specified above. Given a sample of n observations, the investigator estimate $\hat{\beta}_1$ by calculating it as the average value of Y divided by the average value of X. Discuss the properties of this estimator. What difference would it make if it could be assumed that $\beta_0 = 0$?

(9 marks)

Reading for this question

Subject guide (2011), Chapter 1.

Dougherty, C. *Introduction to econometrics* (fourth edition) Chapter 1 (Simple regression analysis).

Gujarati, D.N. *Basic econometrics* (fifth edition) Chapter 3 (Two-variable regression model: The problem of estimation).

Approaching the question

This question examines the concept of ordinary least squares. The answer is:

(a) The OLS estimators minimise the sum of squared residuals

$$\sum_{i=1}^{n} \hat{u}_i^2 = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 = \sum_{i=1}^{n}\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right)^2.$$

Taking partial derivatives with respect to $\hat{\beta}_0$ and $\hat{\beta}_1$, we obtain

$$\begin{aligned}
\frac{\partial\left(\sum \hat{u}_i^2\right)}{\partial \hat{\beta}_0} &= -2\sum\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right) = -2\sum \hat{u}_i \\
\frac{\partial\left(\sum \hat{u}_i^2\right)}{\partial \hat{\beta}_1} &= -2\sum\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right)X_i = -2\sum \hat{u}_i X_i.
\end{aligned}$$

Setting these equations to zero, after algebraic simplification and manipulation give the estimators:

$$\begin{aligned}
\hat{\beta}_1 &= \frac{\frac{1}{n}\sum X_i Y_i - \bar{X}\bar{Y}}{\frac{1}{n}\sum X_i^2 - (\bar{X})^2} = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sum(X_i - \bar{X})^2} \\
\hat{\beta}_0 &= \bar{Y} - \hat{\beta}_1 \bar{X}.
\end{aligned}$$

(b) Since

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}, \qquad \bar{Y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{X}$$

and so $(\bar{Y}, \bar{X})$ lies on the regression line.

(c) The numerator of the sample correlation coefficient for $\hat{Y}$ and e can be decomposed as follows, using the fact that $\bar{e} = 0$.

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}(\hat{Y} - \bar{\hat{Y}})(e_i - \bar{e}) &= \frac{1}{n}\sum_{i=1}^{n}\left[(\hat{\beta}_0 + \hat{\beta}_1 X_i) - (\hat{\beta}_0 + \hat{\beta}_1 \bar{X})\right]e_i \\
&= \frac{1}{n}\hat{\beta}_1 \sum_{i=1}^{n}(X_i - \bar{X})e_i \\
&= \hat{\beta}_1\left[\frac{1}{n}\sum_{i=1}^{n} X_i e_i - \frac{1}{n}\sum_{i=1}^{n}\bar{X}e_i\right] \\
&= \hat{\beta}_1\left[0 - \bar{X}\frac{1}{n}\sum_{i=1}^{n} e_i\right] \\
&= 0.
\end{aligned}$$

(d) Since

$$\begin{aligned}
Y_i &= \beta_0 + \beta_1 X_i + u_i \\
\bar{Y} &= \beta_0 + \beta_1 \bar{X} + \bar{u}
\end{aligned}$$

and

$$\hat{\beta}_1 = \frac{\bar{Y}}{\bar{X}} = \frac{\beta_0 + \beta_1 \bar{X} + \bar{u}}{\bar{X}} = \frac{\beta_0}{\bar{X}} + \beta_1 + \frac{\bar{u}}{\bar{X}}.$$

Hence, assuming that X is non-stochastic,

$$\mathrm{E}\left(\hat{\beta}_1\right) = \frac{\beta_0}{\bar{X}} + \beta_1 + \frac{1}{\bar{X}}\mathrm{E}(\bar{u}) = \frac{\beta_0}{\bar{X}} + \beta_1$$

since $\mathrm{E}(\bar{u}) = 0$. Thus $\hat{\beta}_1$ is biased unless $\beta_0 = 0$, and the direction of the bias depends on the sign of both β_0 and $\bar{X}$. Since

$$\lim_{n\to\infty} \mathrm{E}\left(\hat{\beta}_1\right) \neq \beta_1,$$

the estimator is not consistent.

Special case: $\beta_0 = 0$.

$\hat{\beta}_1$ is now unbiased and

$$\text{Var}\left(\hat{\beta}_1\right) = \frac{n\sigma^2}{\left(\sum X_i\right)^2} \longrightarrow 0$$

as $n \longrightarrow \infty$. Hence if $\beta_0 = 0$, then $\hat{\beta}_1$ is a consistent estimator of β_1. In comparison to the OLS estimator, $\hat{\beta}_1$ is inefficient. If the Gauss-Markov assumptions hold, the OLS estimator is the most efficient estimator.

Question 3

(a) What is meant by 'weak stationarity' and the 'order of integration' of a random variable? Why is it important to test whether variables used in regression are stationary or not?

(6 marks)

(b) Consider the MA(1) process

$$Y_t = \epsilon_t + \theta_1 \epsilon_{t-1} \qquad t = 1, 2, \ldots, T.$$

$\mathbf{E}(\epsilon_t) = 0$; $\mathbf{E}(\epsilon_t^2) = \sigma_\epsilon^2$ and $\mathbf{E}(\epsilon_t \epsilon_s) = 0$ if $t \neq s$ for all $t, s = 1, 2, \ldots T$.

Is Y_t stationary? Does the result generalize to higher-order MA processes? Explain your answers.

(6 marks)

(c) Derive the autocorrelation function for the MA(1) process specified in part (b).

(5 marks)

(d) Assume that Y_t is a process generated by:

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \epsilon_t$$

where ϵ_t is white noise. Show that the appropriate specification for the Augmented Dickey-Fuller test is given by:

$$\Delta Y_t = \phi_0 + (\phi_1 + \phi_2 - 1) Y_{t-1} + \phi_2 \Delta Y_{t-1} + \epsilon_t.$$

(8 marks)

Reading for this question

Subject guide (2011), Chapter 13.

Dougherty, C. *Introduction to econometrics* (fourth edition) Chapter 13 (Introduction to nonstationary time series).

Gujarati, D.N. *Basic econometrics* (fifth edition) Chapter 21 (Time series econometrics: Some basic concepts).

Approaching the question

In part (a), definition of 'weak stationarity' and 'order of integration' is required. Part (b) requires the derivation of mean, variance and the covariances to check stationarity. Part (c) requires the concept of the autocorrelation function and part (d) is based on the augmented Dickey-Fuller test. A detailed answer is:

(a) The requirements for a series to be weak stationary are that both the mean and the covariance structure of the series be stable over time, and that the variance of the series be finite. Formally,

$$\begin{aligned} \mathrm{E}(y_t) &= \mu \\ \gamma(t,\tau) &= \gamma(\tau) \\ \gamma(0) &< \infty. \end{aligned}$$

The order of integration of a random variable is the number of times a variable needs to be differenced to become stationary. Testing for stationarity is important for several reasons: (1) The stationarity or non-stationarity of a series can strongly influence its behaviour and properties (e.g. the persistence of shocks will be infinite for non-stationary series). (2) If two variables are non-stationary a regression of one on the other could have a high R^2 even if the two variables are completely unrelated (i.e. spurious regression). (3) If the variables in the regression model are non-stationary, then it can be proved that the standard assumptions for asymptotic analysis will not be valid. In other words, the usual t-ratios will not follow a t-distribution, so we cannot validly undertake hypothesis tests about the regression coefficients.

(b) The expected value of Y_t is 0, and therefore independent of time:

$$\mathrm{E}(Y_t) = \mathrm{E}(\epsilon_t + \theta_1\epsilon_{t-1}) = \mathrm{E}(\epsilon_t) + \theta_1\mathrm{E}(\epsilon_{t-1}) = 0.$$

Since ϵ_t and ϵ_{t-1} are uncorrelated,

$$\sigma_Y^2 = \sigma_\epsilon^2 + \theta_1^2\sigma_\epsilon^2 = (1+\theta_1^2)\sigma_\epsilon^2,$$

and this is independent of time. Finally, because

$$Y_{t-1} = \epsilon_{t-1} + \theta_1\epsilon_{t-2},$$

the population covariance of Y_t and Y_{t-1} is given by

$$\sigma_{Y_t,Y_{t-1}} = \theta_1\sigma_\epsilon^2.$$

This is fixed and independent of time. The population covariance between Y_t and Y_{t-s} is 0 for all $s > 1$ because they have no elements in common if $s > 1$. Thus all the conditions for stationarity are satisfied. All MA processes are stationary, the general proof being a simple extension of that for the MA(1) case.

(c) The autocorrelation function can be written as

$$\rho = \frac{\mathrm{E}[(Y_t-\mu_Y)(Y_{t+k}-\mu_Y)]}{\sqrt{\mathrm{E}(Y_t-\mu_Y)^2\mathrm{E}(Y_{t+k}-\mu_Y)^2}} = \frac{\mathrm{E}(Y_tY_{t+k})}{(1+\theta_1^2)\sigma_\epsilon^2} = \frac{\mathrm{E}[(\epsilon_t+\theta_1\epsilon_{t-1})(\epsilon_{t+k}+\theta_1\epsilon_{t+k-1})]}{(1+\theta_1^2)\sigma_\epsilon^2}.$$

Now if $k > 1$, the two factors in the numerator are independent and the numerator is zero. If $k = 1$, the numerator is

$$\mathrm{E}[(\epsilon_t+\theta_1\epsilon_{t-1})(\epsilon_{t+1}+\theta_1\epsilon_t)] = \theta_1\sigma_\epsilon^2,$$

and the autocorrelation is

$$\rho_1 = \frac{\theta_1}{1+\theta_1^2}.$$

(d) Write the AR(1) process for u_t as

$$u_t = \rho u_{t-1} + \epsilon_t.$$

Lagging the process for Y_t one period and multiplying through by ρ, we have

$$\rho Y_{t-1} = \phi_0\rho + \phi_1\rho Y_{t-2} + \rho u_{t-1}.$$

Subtracting this from the equation for Y_t, and rearranging, we obtain

$$Y_t = \phi_0(1-\rho) + (\phi_1+\rho)Y_{t-1} - \phi_1\rho Y_{t-2} + \epsilon_t.$$

A necessary condition for stationarity is that the sum of the coefficients of Y_{t-1} and Y_{t-2}, $(\phi_1 + \rho - \phi_1\rho)$, should be less than one. We can rewrite the model

$$Y_t - Y_{t-1} = \phi_1(1-\rho) + (\phi_1 + \rho - \phi_1\rho - 1)Y_{t-1} + \phi_1\rho Y_{t-1} - \phi_1\rho Y_{t-2} + \epsilon_t.$$

Hence, we can obtain

$$\Delta Y_t = \phi_0 + \phi_1^* Y_{t-1} + \phi_2^* \Delta Y_{t-1} + \epsilon_t,$$

where $\phi_1^* = \phi_1 + \rho - \phi_1\rho - 1$ and $\phi_2^* = \phi_1\rho$. Under the null hypothesis of nonstationarity $\phi_1^0 = 0$ and under the alternative hypothesis of stationarity, $\phi_1^* < 0$.

Question 4

In order to model the demand for motor vehicles, an econometrician proposes the general linear regression model:

$$Y_t = \beta_0 + \beta_P P_t + \beta_E E_t + \beta_B B_t + u_t \qquad t = 1985, 1986, \ldots, 2006$$

where Y is an index of consumer expenditure on motor vehicles, spare and accessories at constant prices, P is a relative price index of motor vehicles, E is the real total household expenditure, B is the relative price index of public road transport, and u is the error term. The dependent and independent variables are in logarithmic terms.

This model was fitted using ordinary least squares (OLS) to annual data covering the period 1985–2006, and the following results were obtained:

Dependent variable: logarithm of consumer expenditure on motor vehicles, spares and accessories (Y).

Regressor	Model A	Model B
P	**−0.705**	**−0.926**
	(0.067)	**(0.347)**
E	—	**1.78**
		(0.644)
B	—	**0.061**
		(0.310)
constant	**6.27**	**−2.05**
	(0.56)	**(3.03)**
RSS	**0.636**	**0.192**
R^2	**0.074**	**0.720**
T	**22**	**22**

Where R^2 is the coefficient of determination, RSS denotes residual sum of squares and estimated standard errors are given in parentheses.

(a) **Test the hypothesis $\mathbf{H}_0 : \beta_P = 0$ in both fitted models (A) and (B). Comment on your results.**

(4 marks)

(b) **Test the individual hypotheses $\mathbf{H}_0 : \beta_E = 0$ and $\mathbf{H}_0 : \beta_B = 0$ and the joint hypothesis $\mathbf{H}_0 : \beta_E = \beta_B = 0$.**

(6 marks)

(c) **Which of the fitted models (A) and (B) is preferable? Explain.**

(5 marks)

(d) **Discuss the economic implications of the fitted model (B). Is there any evidence that public road transport acts as a substitute for private motor travel? Explain.**

(5 marks)

(e) **How might a plot of the OLS residuals from fitted model (B) against time assist you in ascertaining whether or not the model for the demand for motor vehicles is misspecified? Explain.**

(5 marks)

Reading for this question

Subject guide (2011), Chapter 3.

Dougherty, C. *Introduction to econometrics* (fourth edition) Chapter 3 (Multiple regression analysis).

Gujarati, D.N. *Basic econometrics* (fifth edition) Chapters 11 (Heteroscedasticity) and 12 (Autocorrelation).

Approaching the question

This question deals with the concept of testing a single coefficient and testing more than one coefficient in a multiple regression model. It should also be discussed how the plot of residuals can help in detecting autocorrelation, heteroscedasticity and outliers. Answer is:

(a) $\text{H}_0 : \beta_P = 0$ versus $\text{H}_1 : \beta_P < 0$ (note the one-tail alternative hypothesis).

- Model A: $T = 22, k = 2$, $t_{calc} = -0.705/0.0673 = -10.48$, $t_{20,0.05} = -1.725$, hence reject H_0.
- Model B: $T = 22, k = 4$, $t_{calc} = -0.926/0.346 = -2.67$, $t_{18,0.05} = -1.734$, hence reject H_0.

(b) $\text{H}_0 : \beta_E = 0$ versus $\text{H}_1 : \beta_E > 0$ (once again note the one-tail alternative), $t_{calc} = 1.78/0.644 = 2.76$, hence reject H_0.

$\text{H}_0 : \beta_B = 0$ versus $\text{H}_1 : \beta_B > 0$, $t_{calc} = 0.0608/0.310 = 0.196$, hence do not reject H_0.

$\text{H}_0 : \beta_E = \beta_B = 0$ versus $\text{H}_1 : \beta_E$ and/or $\beta_B \neq 0$,

$$F_{calc} = \frac{(0.636 - 0.192)/2}{0.192/18} = 20.81,$$

$F_{2,18,0.05} = 3.49$, hence we reject H_0.

(c) Model B is preferable to Model A. Model A does not include the independent variables real household expenditure and the relative price of public road transport, which are (jointly) statistically significant as indicated by the F-statistic calculated in part (b). Consequently, there is an important difference between their coefficient of determination, which in Model A is only 7.4% while in Model B is 72%.

(d) Very slight (see part (b)). Note that since the model is a log-log model the estimated coefficients are estimates of elasticities. The estimated coefficients have the correct signs according to standard economic relationships.

(e) Plot of the residuals will indicate the presence of heteroscedasticity and autocorrelation. It will also capture outliers. These concepts should be shown with the help of diagrams.

Question 5

(a) **In the following panel data model**

$$\begin{aligned} y_i &= \alpha_0 + \alpha_1 x_{it} + w_{it} \qquad i = 1, 2, \ldots, N; \quad t = 1, 2, \ldots, T \\ w_{it} &= v_i + u_{it} \end{aligned}$$

$\mathbf{E}(v_i|x_{it}) = \mathbf{E}(u_{it}|x_{it}) = 0$, $\mathbf{E}(v_i^2) = \sigma_v^2$, $\mathbf{E}(u_{it}^2) = \sigma_u^2$, $\mathbf{E}(v_i u_{it}) = \mathbf{E}(v_i v_j) = \mathbf{E}(u_{it} u_{js}) = 0$ **for all** $i, j = 1, 2, \ldots, N \quad i \neq j$ **and** $t, s = 1, 2, \ldots, T \quad t \neq s$**, there are** N **cross section observations and** T **time series observations. The** v_i **are cross section random effects.**

If the model was estimated by ordinary least squares (OLS) what would be the properties of the OLS estimators? Explain.

(5 marks)

(b) The following estimates were made of a wage equation on a panel sample from the US. The sample contained 4,165 observations covering 595 individuals of working age for 7 years. The dependent variable was the log wage.

	Estimated Coefficients			
	(1)	**(2)**	**(3)**	**(4)**
exp	**0.036**	**0.027**	**0.113**	**0.111**
	(0.004)	**(0.005)**	**(0.002)**	**(0.003)**
exp2	**−0.066**	**−0.053**	**−0.042**	**−0.004**
	(0.010)	**(0.010)**	**(0.008)**	**(0.005)**
RSS	**607.13**	**475.68**	**82.27**	**81.52**
R^2	**0.315**	**0.437**	**0.907**	**0.908**

where exp = years of labour market experience, exp2 = exp*exp/100. Each regression also includes a constant and 6 regional dummy variables. Standard errors are in brackets.

The estimates from column (1) are pooled ordinary least squares, those from column (2) have time fixed effects, those in column (3) cross-section fixed effects and those in column (4) have both time and cross-section fixed effects.

i. Explain why the values of the residual sum of squares (RSS) and R^2 differ in the four sets of results.

(5 marks)

ii. Test each of these specifications against the most general alternative available.

(10 marks)

iii. The authors of this study intended to include years of education in these equations. This proved impossible for some of these models. Explain why.

(5 marks)

Reading for this question

Subject guide (2011), Chapter 14.

Dougherty, C. *Introduction to econometrics* (fourth edition) Chapter 14.2 (Fixed effects regressions).

Gujarati, D.N. *Basic econometrics* (fifth edition) Chapter 16.3 (Estimation of panel data regression models).

Approaching the question

In part (a) it has to be recognised that the disturbance term is autocorrelated and discuss consequences of autocorrelation on the OLS estimators. Part (b) requires a good concept of fixed estimation method. A detailed answer is:

(a)

$$\mathrm{E}(w_{it}w_{is}) = \mathrm{E}(v_i + u_{it})\mathrm{E}(v_i + u_{is}) = \mathrm{E}(v_i^2) = \sigma_v^2 \neq 0.$$

Hence, OLS estimators are unbiased but inefficient and have wrongly calculated standard errors.

(b) i. The number of observations remains the same for all four models. In calculating the degrees of freedom, the relevant number of coefficients is as follows:

- ∗ In (1) there are two explanatory variables, six regional dummy variables and a constant so $k = 9$.
- ∗ In (2) there are two explanatory variables, six regional dummy variables, a constant and six time-fixed effects, so $k = 15$.
- ∗ In (3) there are two explanatory variables, six regional dummy variables, a constant and 594 cross-section fixed effects, so $k = 603$.
- ∗ In (4) there are two explanatory variables, six regional dummy variables, a constant, six time fixed effects and 594 cross section fixed effects, so $k = 609$.

Thus RSS must fall as k rises, and as

$$R^2 = 1 - \frac{RSS}{TSS},$$

and TSS remains the same for all the models, R^2 rises.

ii. (1) is a restricted version (and therefore nested) of the other three, (2) and (3) are restricted versions (and also nested) of (4) but not with each other.

(1) against (4) gives

$$F = \frac{(607.13 - 81.52)/600}{81.52/3556} = \frac{0.876}{0.0229} = 38.25,$$

and the critical value is $F_{600,3556} = 1.3$ (approximately), hence we reject (1).

(2) against (4) gives

$$F = \frac{(475.68 - 81.52)/594}{81.52/3556} = \frac{0.664}{0.0229} = 28.98,$$

and the critical value is $F_{594,3556} = 1.3$ (approximately), hence we reject (2).

(3) against (4) gives

$$F = \frac{(82.27 - 81.52)/6}{81.52/3556} = \frac{0.125}{0.0229} = 5.46,$$

and the critical value is $F_{6,3556} = 2.1$ (approximately), hence we reject (3).

iii. Since years of full-time education do not generally vary over time once an individual has reached working age, the education variable would be collinear with the cross-section fixed effects. Thus models (3) and (4) could not be estimated.

Question 6

(a) Consider the model

$$y_t = \alpha x_t + u_t; \qquad t = 1, 2, \ldots, T$$

where $\mathbf{E}(u_t) = 0$; $\mathbf{E}(u_t^2) = \sigma^2 x_t^2$; $\mathbf{E}(u_s u_t) = 0$ if $s \neq t$, for all s and t. x_t is an observed non-random variable.

The density function of u_t is

$$f(u_t) = \left(2\pi\sigma^2 x_t^2\right)^{-1/2} \exp\left[-\frac{1}{2}\left(\frac{u_t}{\sigma x_t}\right)^2\right].$$

i. Derive the maximum likelihood (ML) estimators of α and σ^2. Show that $\hat{\alpha}$ the ML estimator of α is unbiased.

(8 marks)

ii. Compare the ML estimator of α, with the weighted least squares estimator of α.

(7 marks)

(b) Consider a model

$$Y_t = \alpha + \beta X_t + u_t; \qquad t = 1, 2, \ldots, T$$

u_t is normally distributed with mean 0 and variance σ^2 and $\mathbf{E}(u_s u_t) = 0$ for all s and t. α and β have been estimated by maximum likelihood.

Explain how the hypothesis that the coefficients are jointly equal to zero will be tested.

(10 marks)

Reading for this question

Subject guide (2011), Chapter 10.

Dougherty, C. *Introduction to econometrics* (fourth edition) Chapter 10.6 (An introduction to maximum likelihood estimation).

Gujarati, D.N. *Basic econometrics* (fifth edition) Chapters Appendix 4A (Maximum likelihood estimation of two-variable regression model) and Appendix 8A (Likelihood ratio (LR) test).

Approaching the question

In part (a), derive the likelihood function. Differentiate the likelihood function with respect to the parameters and equate to zero to get the first-order conditions. Solution of the first-order conditions will give the maximum likelihood estimators. In part (b), pseudo-R^2 and likelihood ratio test should be discussed. A detailed answer is:

(a)

i. The log-likelihood function is

$$\ln L = -\frac{T}{2}\log 2\pi - \frac{1}{2}\log \sum \sigma^2 x_t^2 - \frac{1}{2\sigma^2}\sum \left(\frac{y_t}{x_t} - \alpha\right)^2.$$

The first-order conditions are:

$$\begin{aligned}
\frac{\partial \ln L}{\partial \alpha} &= -\frac{1}{2\sigma^2}\sum \left(\frac{y_t}{x_t} - \alpha\right) = 0 \qquad \text{(i)} \\
\frac{\partial \ln L}{\partial \sigma^2} &= -\frac{T}{2\sigma^2} + \frac{1}{2\sigma^4}\sum \left(\frac{y_t}{x_t} - \alpha\right)^2 = 0. \qquad \text{(ii)}
\end{aligned}$$

Solving (i) and (ii), the ML estimators of α and σ^2 are obtained as:

$$\begin{aligned}
\hat{\alpha}_{MLE} &= \frac{1}{T}\sum \left(\frac{y_t}{x_t}\right) \qquad \text{and} \\
\hat{\sigma}^2_{MLE} &= \frac{1}{T}\sum \left(\frac{y_t}{x_t} - \alpha\right)^2.
\end{aligned}$$

To show $\hat{\alpha}_{MLE}$ is an unbiased estimator,

$$\mathrm{E}\left(\hat{\alpha}_{MLE}\right) = \frac{1}{T}\sum \left(\frac{\mathrm{E}(y_t)}{x_t}\right) = \frac{1}{T}\sum \left(\frac{\alpha x_t}{x_t}\right) = \frac{T\alpha}{T} = \alpha.$$

Hence $\hat{\alpha}_{MLE}$ is an unbiased estimator of α.

ii. To obtain the weighted least squares (WLS) estimator, the model is divided by x_t to get

$$\frac{y_t}{x_t} = \alpha + \frac{u_t}{x_t}; \qquad u_t = y_t - \alpha x_t.$$

To obtain the WLS estimator of α,

$$\sum \left(\frac{u_t}{x_t}\right)^2 = \sum \left(\frac{y_t}{x_t} - \alpha\right)^2$$

has to be minimised.

To obtain the ML estimator, $\ln L$ is maximised, which is equivalent to minimising

$$\sum \left(\frac{y_t}{x_t} - \alpha \right)^2 .$$

Hence WLS estimators and ML estimators are the same.

(b) All the slope coefficients are equal to zero and can be tested using the likelihood ratio statistic $2(\log L - \log L_0)$. This is asymptotically distributed as a chi-square with $k-1$ degrees of freedom. $k-1$ is the number of explanatory variables in the model. In this case $k-1=2$.

$\log L$ is the log of the unrestricted likelihood and $\log L_0$ is the log of restricted likelihood. $\log L_0$ has been obtained with only the intercept in the regression.

Pseudo-R^2 should also be discussed.

Examiners' commentaries 2013

EC2020 Elements of econometrics

Important note

This commentary reflects the examination and assessment arrangements for this course in the academic year 2012–13. In 2014 the format of the examination will change to:

Candidates should answer **EIGHT** of the following **TEN** questions: **ALL FIVE** questions from Section A (8 marks each) and **THREE** questions from Section B (20 marks each).

Section A is intended to examine across the whole syllabus while Section B is designed to examine a selection of topics in greater depth.

The format and structure of the examination may change again in future years, and any such changes will be publicised on the virtual learning environment (VLE).

Information about the subject guide

Unless otherwise stated, all cross-references will be to the latest version of the subject guide (2011). You should always attempt to use the most recent edition of any Essential reading textbook, even if the commentary and/or online reading list and/or subject guide refers to an earlier edition. If different editions of Essential reading are listed, please check the VLE for reading supplements – if none are available, please use the contents list and index of the new edition to find the relevant section.

Comments on specific questions – Zone B

Candidates should answer **FOUR** of the following **SIX** questions: **QUESTION 1** of Section A (25 marks in total) and **THREE** questions from Section B (25 marks each). **Candidates are strongly advised to divide their time accordingly.**

Section A

Answer question 1 from this section.

Question 1

(a) **Prove that in the demand function of the form $Q = \beta_0 P^{\beta_1} Y^{\beta_2} e^u$ where $Q =$ quantity demanded, $P =$ Price and $Y =$ income, β_1 is the price elasticity of demand and β_2 is the income elasticity of demand.**

(5 marks)

(b) **Explain how dummy variables can be used to verify structural change?**

(5 marks)

(c) **Consider the Data Generating Process**

$$Y_t = \alpha + \beta X_t + u_t; \qquad t = 1, \ldots, T;$$

$\mathbf{E}[u_t] = 0$

where X is fixed in repeated samples. Consider a possible estimator for the slope parameter

$$\hat{\beta} = T^{-1} \sum_{t=1}^{T} \left(\frac{Y_t - \bar{Y}}{X_t - \bar{X}} \right)$$

where $\bar{Y} \equiv T^{-1} \sum_{t=1}^{T} Y_t$ and $\bar{X} \equiv T^{-1} \sum_{t=1}^{T} X_t$. Show that $\hat{\beta}$ is an unbiased estimator of β.

(5 marks)

(d) **What is meant by a common factor test in the context of a linear model with an autocorrelated error? How would you perform a common factor test and what hypothesis would you be testing?**

(5 marks)

(e) **Let the probability density function of a population X be:**

$$f(x) = p^x (1-p)^{1-x}; \qquad x = 0, 1$$

where p is the probability of success. Let $X_1, X_2, \ldots, X_n$ be a simple random sample from X. Obtain the maximum likelihood estimator of p.

(5 marks)

Reading for this question

Subject guide (2011), Chapter 10.

Dougherty, C. *Introduction to econometrics* (fourth edition) Chapters R.6 (Unbiasedness and efficiency), 4.2 (Logarithmic transformations), 5 (Dummy variable), 10.1 (Linear Probability Model) and 10.6 (An Introduction to Maximum Likelihood Estimation).

Gujarati, D.N. *Basic econometrics* (fifth edition) Chapters 4.4 (Method of maximum likelihood estimation), 9.5 (Dummy variable regression models) and 15.2 (The linear probability model (LPM) and Appendix A.7 (Statistical Inference: Estimation).

Approaching the question

(a) Define income elasticity and price elasticity of demand. Differentiate the quantity demanded with respect to price and income to obtain the elasticities. The answer is:

The definition of the price elasticity of demand is

$$\eta_P = \frac{\mathrm{d}Q}{\mathrm{d}P} \frac{P}{Q}.$$

The derivative of Q with respect to P is

$$\frac{\mathrm{d}Q}{\mathrm{d}P} = \beta_1 \left(\beta_0 P^{\beta_1 - 1} Y^{\beta_2} e^u \right) = \beta_1 \left(\beta_0 P^{\beta_1} Y^{\beta_2} e^u \right) P^{-1} = \beta_1 \frac{Q}{P}.$$

Substituting these values in η_P, we get

$$\eta_P = \frac{\mathrm{d}Q}{\mathrm{d}P} \frac{P}{Q} = \beta_1 \frac{Q}{P} \frac{P}{Q} = \beta_1.$$

The definition of income elasticity of demand is

$$\eta_Y = \frac{\mathrm{d}Q}{\mathrm{d}Y} \frac{Y}{Q}.$$

The derivative of Q with respect to Y is

$$\frac{\mathrm{d}Q}{\mathrm{d}Y} = \beta_2 \left(\beta_0 P^{\beta_1} Y^{\beta_2 - 1} e^u \right) = \beta_2 \left(\beta_0 P^{\beta_1} Y^{\beta_2} e^u \right) Y^{-1} = \beta_2 \frac{Q}{Y}.$$

Substituting these values in η_Y, we get

$$\eta_Y = \frac{\mathrm{d}Q}{\mathrm{d}Y}\frac{Y}{Q} = \beta_2 \frac{Q}{Y}\frac{Y}{Q} = \beta_2.$$

(b) Dummy variables should be defined. It should be explained how dummy variables can be used to capture changes in the intercept, slope and both intercept and slope. Tests on the coefficients of the dummy variables should be explained. The answer is:

To investigate structural change a dummy variable (D) is defined as $D_t = 1$ after the change and $D_t = 0$ otherwise. If the relationship being investigated is $Y_t = \beta_0 + \beta_1 X_t + u_{1t}$ then the model $Y_t = \beta_0 + \beta_1 X_t + \beta_2 D_t + \beta_3 D_t X_t + u_{2t}$ will allow a t-test for a change in intercept by using the null $\mathrm{H}_0 : \beta_2 = 0$ and a change in slope, $\mathrm{H}_0 : \beta_3 = 0$. An alternative is to use an F-test to test the joint hypothesis $\mathrm{H}_0 : \beta_2 = \beta_3 = 0$.

(c) To show unbiasedness, expectation of $\hat{\beta}$ should be taken. If $\mathrm{E}\left(\hat{\beta}\right) = \beta$ then $\hat{\beta}$ is an unbiased estimator of β. A detailed answer is:

$$\begin{aligned}
\hat{\beta} &= T^{-1}\sum_{t=1}^{T}\left(\frac{Y_t - \bar{Y}}{X_t - \bar{X}}\right) \\
&= T^{-1}\sum_{t=1}^{T}\left(\frac{\alpha + \beta X_t + u_t - \alpha - \beta \bar{X} - \bar{u}}{X_t - \bar{X}}\right) \\
&= T^{-1}\sum_{t=1}^{T}\left(\frac{\beta(X_t - \bar{X})}{X_t - \bar{X}}\right) + T^{-1}\sum_{t=1}^{T}\left(\frac{u_t - \bar{u}}{X_t - \bar{X}}\right) \\
&= \beta + T^{-1}\sum_{t=1}^{T}\left(\frac{u_t - \bar{u}}{X_t - \bar{X}}\right).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathrm{E}\left(\hat{\beta}\right) &= \beta + \mathrm{E}\left[T^{-1}\sum_{t=1}^{T}\left(\frac{u_t - \bar{u}}{X_t - \bar{X}}\right)\right] \\
&= \beta + T^{-1}\mathrm{E}\left[\sum_{t=1}^{T}\left(\frac{u_t - \bar{u}}{X_t - \bar{X}}\right)\right] \\
&= \beta \\
&\Rightarrow \text{ unbiased.}
\end{aligned}$$

(d) It has to be shown that the AR(1) model is a restricted version of the ADL(1,1) model. Restrictions should be explicitly derived and the test statistic defined. The answer is:

Given the equations $Y_t = \beta_1 + \beta_2 X_t + u_t$ and $u_t = \rho u_{t-1} + v_t$, where v has zero mean, constant variance and zero autocovariance, combine the two equations to give

$$Y_t = \lambda_1 + \lambda_2 Y_{t-1} + \lambda_3 X_t + \lambda_4 X_{t-1} + v_t,$$

which is the restricted version of the general form (an (ADL(1,1)) model)

$$Y_t = \lambda_1 + \lambda_2 Y_{t-1} + \lambda_3 X_t + \lambda_4 X_{t-1} + v_t,$$

and is subject to the restriction $\lambda_4 = -\lambda_2\lambda_3$. The test of this restriction is the common factor test.

Note that the usual F-test of the restriction is not appropriate because the restriction is non-linear so we have to use the test statistic

$$n \log\left(\frac{RSS_R}{RSS_U}\right) \sim \chi^2_1,$$

where RSS_R and RSS_U are the residual sum of squares from the restricted and unrestricted models respectively. n is the sample size and the test statistic is asymptotically chi-square with one degree of freedom.

(e) The likelihood function should be derived and differentiated with respect to p. Solution of the first-order condition will give the maximum likelihood estimator. The answer is:

The likelihood function is

$$L = p^{x_1}(1-p)^{1-x_1} \cdots p^{x_n}(1-p)^{1-x_n} = p^{\sum x_i}(1-p)^{n-\sum x_i}.$$

The log-likelihood function is

$$\ln L = \left(\sum x_i\right) \ln p + \left(n - \sum x_i\right) \ln(1-p).$$

The first-order condition, setting equal to zero, is

$$\frac{\mathrm{d} \ln L}{\mathrm{d}p} = \frac{\sum x_i}{\hat{p}} - \frac{n - \sum x_i}{1-\hat{p}} = 0$$

or

$$\frac{(1-\hat{p})\sum x_i - \hat{p}\,(n - \sum x_i)}{\hat{p}(1-\hat{p})} = 0.$$

Solving, we get

$$\hat{p} = \frac{\sum x_i}{n};$$

this is the maximum likelihood estimate.

The corresponding maximum likelihood estimator (MLE) of p is

$$\hat{p}_{MLE} = \frac{\sum X_i}{n} = \bar{X}.$$

Section B

Answer three questions from this section.

Question 2

Write brief notes on the following:

(a) Linear probability model.

(15 marks)

(b) Likelihood ratio test and Wald test.

(10 marks)

Reading for this question

Subject guide (2011), Chapter 10.

Dougherty, C. *Introduction to econometrics* (fourth edition) Chapter 10 (Binary choice and limited dependent variable models).

Gujarati, D.N. *Basic econometrics* (fifth edition) Chapters 15.2 (The linear probability model) and Appendix 8A (Likelihood ratio (LR) test).

Approaching the question

In part (a) as the weightage of the question is high, it is required to discuss the linear probability model (LPM) in detail. Part (b) requires a brief discussion of the concepts of the likelihood ratio and the Wald tests. A detailed answer is:

(a) LPM is used to denote a model in which the dependent variable is binary which takes the value 1 if the event occurs and 0 if it does not. It is estimated by the ordinary least squares (OLS).

Let the model be:

$$Y_i = \beta_0 + \beta_1 X_i + u_i; \qquad i = 1, 2, \ldots, n \qquad \text{(i)}$$

$$Y_i = \begin{cases} 1 & \text{if event occurs} \\ 0 & \text{if not.} \end{cases}$$

Assume $\mathrm{E}(u_i) = 0$, then

$$\mathrm{E}[Y_i|X_i] = \beta_0 + \beta_1 X_i. \qquad \text{(ii)}$$

Also,

$$\mathrm{E}[Y_i|X_i] = 1 \cdot P(Y_i = 1) + 0 \cdot P(Y_i = 0) = P(Y_i = 1) = p_i. \qquad \text{(iii)}$$

From (ii) and (iii),

$$\mathrm{E}[Y_i|X_i] = \beta_0 + \beta_1 X_i = p_i$$

hence we can interpret $\mathrm{E}[Y_i|X_i] = \beta_0 + \beta_1 X_i$ as the probability that the event will occur, given X_i.

If we denote $\hat{\beta}_0$ and $\hat{\beta}_1$ as estimates of β_0 and β_1, respectively, then we can write

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i = \hat{p}_i \qquad \text{(iv)}$$

as the estimated probability that the event will occur.

As Y_i takes only two values, 1 or 0, therefore u_i can take only two values: $1 - \beta_0 - \beta_1 X_i$ when $Y_i = 1$, and $\beta_0 - \beta_1 X_i$ when $Y_i = 0$. Based on this we can write the probability distribution of u_i as

Y_i	u_i	$f(u_i)$
1	$1 - \beta_0 - \beta_1 X_i$	$\beta_0 + \beta_1 X_i$
0	$-\beta_0 - \beta_1 X_i$	$1 - \beta_0 - \beta_1 X_i$

This probability distribution also satisfies the assumption that

$$\mathrm{E}(u_i) = (1 - \beta_0 - \beta_1 X_i)(\beta_0 + \beta_1 X_i) + (-\beta_0 - \beta_1 X_i)(1 - \beta_0 - \beta_1 X_i) = 0.$$

We can write $\mathrm{Var}(u_i)$ as

$$\begin{aligned}
\mathrm{Var}(u_i) = \mathrm{E}(u_i^2) &= (1 - \beta_0 - \beta_1 X_i)^2(\beta_0 + \beta_1 X_i) + (-\beta_0 - \beta_1 X_i)^2(1 - \beta_0 - \beta_1 X_i) \\
&= (1 - \beta_0 - \beta_1 X_i)(\beta_0 + \beta_1 X_i)[(1 - \beta_0 - \beta_1 X_i) + (\beta_0 + \beta_1 X_i)] \\
&= (\beta_0 + \beta_1 X_i)(1 - \beta_0 - \beta_1 X_i) \\
&= \mathrm{E}(Y_i)[1 - \mathrm{E}(Y_i)] \\
&= p_i(1 - p_i)
\end{aligned}$$

for all $i = 1, 2, \ldots, n$.

Hence the disturbance term is heteroscedastic. This will make OLS estimators inefficient.

In many cases the estimated probability $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$ will be negative or greater than 1.

We can use weighted least squares to solve the problem of heteroscedasticity.

We can see from (iv), that the estimator of p_i is $\hat{P}_i = \hat{Y}_i$, therefore $\hat{Y}_i(1 - \hat{Y}_i)$ can be used as an estimator of

$$\mathrm{Var}(u_i) = \mathrm{E}(Y_i)[1 - \mathrm{E}(Y_i)] = p_i(1 - p_i).$$

Weights can be obtained as

$$W_i = \left[\hat{Y}_i(1-\hat{Y}_i)\right]^{1/2}.$$

Divide (i) by W_i and apply OLS to

$$\frac{Y_i}{W_i} = \frac{\beta_0}{W_i} + \beta_1 \frac{X_i}{W_i} + \frac{u_i}{W_i}; \qquad i = 1, 2, \ldots, n$$

to obtain the WLS estimators of β_0 and β_1. This will give an efficient estimator.

Problem:

i. In practice the estimated variance of u_i, $\hat{Y}_i(1-\hat{Y}_i)$ may be negative as again the estimated probability $\hat{Y}_i$ may be negative or greater than 1.

ii. The distribution of the disturbance term is not continuous and normal. This implies that standard errors and the usual test statistics are not valid.

(b) **Likelihood ratio test:**

Suppose we have to test a simple hypothesis $\mathrm{H}_0 : \theta = \theta_0$ against all possible alternatives. Given a simple random sample $X_1, X_2, \ldots, X_N$, a natural way of judging the acceptability or otherwise of the hypothesis would be to compare the likelihood functions

Let

$$\begin{aligned} L_R &= \text{Restricted likelihood (likelihood based on the null hypothesis)} \\ L_U &= \text{Unrestricted likelihood (likelihood based on the alternative hypothesis).} \end{aligned}$$

If the likelihood ratio (LR)

$$\lambda = \frac{L_R}{L_U}$$

is close to unity then in light of the given sample H_0 would seem highly plausible, on the other hand if this ratio is close to zero H_0 would seem to have little validity. Since λ is a random variable its distribution may be derived and hence we can make probability statements about how close the LR is to unity. A test for H_0 is thus provided by a critical region defined by $\lambda < \lambda_0$, where λ_0 is such that $P(\lambda < \lambda_0 | \mathrm{H}_0) = \alpha$. α is the probability of a Type 1 error.

For large N, $-2 \ln \lambda$ has approximately a χ^2 distribution with degrees of freedom equal to the number of **restrictions** imposed by the null hypothesis.

Wald test:

The Wald test evaluates whether the discrepancy between the maximum likelihood estimate of θ and θ_0 is significant. The test statistic for the null hypothesis $\mathrm{H}_0 : \theta = \theta_0$ is

$$W = \frac{(\hat{\theta} - \theta_0)^2}{\hat{\sigma}^2_{\hat{\theta}}},$$

where $\hat{\theta}$ is the maximum likelihood estimator of θ. $\hat{\sigma}^2_{\hat{\theta}}$ is the estimate of the variance of θ evaluated at the maximum likelihood value. $\hat{\sigma}^2_{\hat{\theta}}$ is obtained as minus the inverse of the second differential of the log likelihood function evaluated at the maximum likelihood estimate.

Under the null hypothesis that the restriction is valid, the test statistic has a chi-square distribution with degrees of freedom equal to the number of restriction imposed by H_0.

Question 3

The following ordinary least squares estimates were made on a sample of 3,356 employed male workers in Britain.

	total sample (i)	(ii)	under 40 (iii)	40 and over (iv)
ed	0.333	—	0.208	0.386
	(0.047)		(0.067)	(0.067)
ed2	−0.159	—	−0.553	−1.087
	(0.007)		(0.036)	(0.228)
age	0.142	0.151	0.262	0.128
	(0.006)	(0.006)	(0.021)	(0.036)
age2	−0.159	−0.175	−0.362	−0.138
	(0.007)	(0.007)	(0.036)	(0.035)
constant	0.002	2.647	−0.705	−0.277
	(0.357)	(0.110)	(0.558)	(1.016)
n	3,356	3,356	1,807	1,549
R^2	0.233	0.180	0.312	0.135
RSS	1,235.88	1,610.19	607.89	608.73

The dependent variable is the log of weekly earnings. The estimates in the first two columns use the whole sample. The estimates in the third column use only observations on men younger than 40, the estimates in the fourth column use only observations on men aged 40 and over; ed is years of full time education, ed2 is ed*ed/100, age is age in years; age2 is age*age/100. n is the sample size. RSS is the residual sum of squares. Standard errors are given in brackets.

(a) Using the model in (ii), briefly explain the method of ordinary least squares estimation.

(5 marks)

(b) Discuss the role of the quadratic terms in the models estimated above.

(3 marks)

(c) Test the hypothesis that the coefficients of ed and ed2 are jointly zero in full sample.

(6 marks)

(d) Test the hypothesis that the coefficients of the model are constant between men under 40 and men aged 40 and over.

(6 marks)

(e) On what assumptions are your tests based? Are they likely to be true in this sample? Give details.

(5 marks)

Reading for this question

Subject guide (2011), Chapter 3.

Dougherty, C. *Introduction to Econometrics* (fourth edition) Chapter 3.2 (Derivation and interpretation of the multiple regression coefficients).

Gujarati, D.N. *Basic econometrics* (fifth edition) Chapters 3.1 (The method of ordinary least squares), 4 (Classical normal regression model (CNLRM)) and 7.10 (Polynomial regression model).

Approaching the question

It has to be explained that OLS estimators minimise the residual sum of squares. This should be illustrated with the help of a simple model. This question tests the knowledge of the estimation, testing and assumptions of the OLS. The answer is:

(a) The model is

$$\text{earning}_t = \beta_0 + \beta_1 \text{age}_t + \beta_2 \text{age2} + u_t; \qquad t = 1, 2, \ldots, T.$$

OLS estimators minimise the residual sum of squares (RSS). Let

$$\hat{u}_t = \text{earning}_t - \hat{\beta}_0 - \hat{\beta}_1 \text{age}_t - \hat{\beta}_2 \text{age2}_t$$

be the residual, then

$$\sum \hat{u}_t^2 = \sum \left(\text{earning}_t - \hat{\beta}_0 - \hat{\beta}_1 \text{age}_t - \hat{\beta}_2 \text{age2}_t \right)^2$$

is the RSS. Minimising RSS with respect to $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\beta}_2$, we get the first-order conditions

$$\frac{\partial \sum \hat{u}_t^2}{\partial \hat{\beta}_0} = 0; \qquad \frac{\partial \sum \hat{u}_t^2}{\partial \hat{\beta}_1} = 0 \qquad \text{and} \qquad \frac{\partial \sum \hat{u}_t^2}{\partial \hat{\beta}_2} = 0.$$

Solving these first-order conditions we obtain the OLS estimators $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\beta}_2$.

(b) The quadratic terms show that in every case both education and age have a positive effect on earnings (the positive estimated coefficients of the linear terms) the rate of increase in earnings declines as age and education increase (the negative quadratic terms). Note all the coefficients (both linear and quadratic) of education and age are significant.

(c) We test

$$\begin{aligned} &\text{H}_0 : \text{Coefficients are jointly equal to 0} \\ &\text{H}_1 : \text{Not all coefficients are equal to 0.} \end{aligned}$$

Then

$$F = \frac{(1610.19 - 1235.88)/2}{1235.88/3351} = \frac{187.155}{0.3688} = 507.47.$$

Since the 5% critical value is $F_{2,3351,0.05} = 2.996$ (approximately), we reject H_0.

(d) We test

$$\begin{aligned} &\text{H}_0 : \text{The coefficients are constant} \\ &\text{H}_1 : \text{The coefficients are not constant.} \end{aligned}$$

Then

$$F = \frac{(1235.88 - (607.89 + 608.73))/5}{(607.89 + 608.73)/3346} = \frac{3.852}{0.3636} = 10.59.$$

Since the 5% critical value is $F_{5,3346,0.05} = 2.214$ (approximately), we reject H_0.

(e) The tests are based on the following assumptions.

- The model is correctly specified
- $\text{E}(u_i) = 0$
- $\text{E}(u_i^2) = \sigma^2$
- $\text{E}(u_i u_j) = 0$ for all $i \neq j$

where u_i is an unobserved random disturbance.

Question 4

(a) **In the following panel data model**

$$\begin{aligned} y_i &= \alpha_0 + \alpha_1 x_{it} + w_{it} \qquad i = 1, 2, \ldots, N; \quad t = 1, 2, \ldots, T \\ w_{it} &= v_i + u_{it} \end{aligned}$$

$\mathbf{E}(v_i|x_{it}) = \mathbf{E}(u_{it}|x_{it}) = 0$, $\mathbf{E}(v_i^2) = \sigma_v^2$, $\mathbf{E}(u_{it}^2) = \sigma_u^2$, $\mathbf{E}(v_i u_{it}) = \mathbf{E}(v_i v_j) = \mathbf{E}(u_{it} u_{js}) = 0$ for all $i, j = 1, 2, \ldots, N \quad i \neq j$ and $t, s = 1, 2, \ldots, T \quad t \neq s$, there are N cross section observations and T time series observations. The v_i are cross section random effects.

If the model was estimated by ordinary least squares (OLS) what would be the properties of the OLS estimators? Explain.

(5 marks)

(b) **The following estimates were made of a wage equation on a panel sample from the US. The sample contained 4,165 observations covering 595 individuals of working age for 7 years. The dependent variable was the log wage.**

Estimated Coefficients

	(1)	(2)	(3)	(4)
exp	0.036	0.027	0.113	0.111
	(0.004)	(0.005)	(0.002)	(0.003)
exp2	−0.066	−0.053	−0.042	−0.004
	(0.010)	(0.010)	(0.008)	(0.005)
RSS	607.13	475.68	82.27	81.52
R^2	0.315	0.437	0.907	0.908

where exp = years of labour market experience, exp2 = exp*exp/100. Each regression also includes a constant and 6 regional dummy variables. Standard errors are in brackets.

The estimates from column (1) are pooled ordinary least squares, those from column (2) have time fixed effects, those in column (3) cross-section fixed effects and those in column (4) have both time and cross-section fixed effects.

i. **Explain why the values of the residual sum of squares (RSS) and R^2 differ in the four sets of results.**

(5 marks)

ii. **Test each of these specifications against the most general alternative available.**

(10 marks)

iii. **The authors of this study intended to include years of education in these equations. This proved impossible for some of these models. Explain why.**

(5 marks)

Reading for this question

Subject guide (2011), Chapter 14.

Dougherty, C. *Introduction to econometrics* (fourth edition) Chapter 14.2 (Fixed effects regressions).

Gujarati, D.N. *Basic econometrics* (fifth edition) Chapter 16.3 (Estimation of panel data regression models).

Approaching the question

In part (a) it has to be recognised that the disturbance term is autocorrelated and discuss consequences of autocorrelation on the OLS estimators. Part (b) requires a good concept of fixed estimation method. A detailed answer is:

(a)

$$\mathrm{E}(w_{it}w_{is}) = \mathrm{E}(v_i + u_{it})\mathrm{E}(v_i + u_{is}) = \mathrm{E}(v_i^2) = \sigma_v^2 \neq 0.$$

Hence, OLS estimators are unbiased but inefficient and have wrongly calculated standard errors.

(b) i. The number of observations remains the same for all four models. In calculating the degrees of freedom, the relevant number of coefficients is as follows:

- In (1) there are two explanatory variables, six regional dummy variables and a constant so $k = 9$.
- In (2) there are two explanatory variables, six regional dummy variables, a constant and six time-fixed effects, so $k = 15$.
- In (3) there are two explanatory variables, six regional dummy variables, a constant and 594 cross-section fixed effects, so $k = 603$.
- In (4) there are two explanatory variables, six regional dummy variables, a constant, six time fixed effects and 594 cross section fixed effects, so $k = 609$.

Thus RSS must fall as k rises, and as

$$R^2 = 1 - \frac{RSS}{TSS},$$

and TSS remains the same for all the models, R^2 rises.

ii. (1) is a restricted version (and therefore nested) of the other three, (2) and (3) are restricted versions (and also nested) of (4) but not with each other.

(1) against (4) gives

$$F = \frac{(607.13 - 81.52)/600}{81.52/3556} = \frac{0.876}{0.0229} = 38.25,$$

and the critical value is $F_{600,3556} = 1.3$ (approximately), hence we reject (1).

(2) against (4) gives

$$F = \frac{(475.68 - 81.52)/594}{81.52/3556} = \frac{0.664}{0.0229} = 28.98,$$

and the critical value is $F_{594,3556} = 1.3$ (approximately), hence we reject (2).

(3) against (4) gives

$$F = \frac{(82.27 - 81.52)/6}{81.52/3556} = \frac{0.125}{0.0229} = 5.46,$$

and the critical value is $F_{6,3556} = 2.1$ (approximately), hence we reject (3).

iii. Since years of full-time education do not generally vary over time once an individual has reached working age, the education variable would be collinear with the cross-section fixed effects. Thus models (3) and (4) could not be estimated.

Question 5

(a) **Explain what do you understand by the order condition of identification in the context of simultaneous equations.**

(5 marks)

(b) **Consider a set of simultaneous equation models**

$$
\begin{aligned}
y_{1t} &= \alpha y_{2t} + u_{1t} \\
y_{2t} &= \beta_1 x_t + \beta_2 y_{1t} + u_{2t}; \qquad t = 1, 2, \ldots, T.
\end{aligned}
$$

Where x_t is an exogenous variable, y_{1t} and y_{2t} are endogeneous variables.

i. **Examine the identifiability of the above given equations.**

(5 marks)

ii. **Explain in detail problems which arise if the parameter α is estimated by the ordinary least squares.**

(10 marks)

iii. **Explain how a consistent estimator of α can be obtained.**

(5 marks)

Reading for this question

Subject guide (2011), Chapter 9.

Dougherty, C. *Introduction to econometrics* (fourth edition) Chapters 9.2 (Simultaneous equation bias) and 9.3 (Instrumental variable estimation).

Gujarati, D.N. *Basic econometrics* (fifth edition) Chapters 18.3 (Simultaneous equation bias), 20.3 (Estimation of a just identified equation: The method of indirect least squares) and 20.4 (Estimation of an overidentified equation: The method of two-stage least squares).

Approaching the question

Part (a), requires the conditions of the order condition of identification. In part (b), the concept of simultaneous equation bias, with technical detail, should be explained. In part (c), as the equation is exactly identified indirect least squares and two-stage least squares will give consistent estimator of α. A detailed answer is:

(a) Order condition:

$$R \geq G - 1,$$

where R is the number of restrictions imposed on the equation. G is the number of endogenous variables in the complete model which is also equal to the number of equations.

$$
\begin{aligned}
R > G - 1 &\Rightarrow \text{overidentified} \\
R = G - 1 &\Rightarrow \text{exactly identified} \\
R < G - 1 &\Rightarrow \text{underidentified.}
\end{aligned}
$$

(b) i. First equation:

$$R = 1; \qquad G - 1 = 2 - 1 = 1,$$

hence exactly identified.

Second equation:

$$R = 0; \qquad G - 1 = 2 - 1 = 1,$$

hence, as $R < G - 1$, it is underidentified.

ii. We have

$$\begin{aligned} y_1 &= \alpha y_2 + u_1 \qquad \text{(i)} \\ y_2 &= \beta_1 x + \beta_2 y + u_2 \qquad \text{(ii)} \end{aligned}$$

and so substituting (ii) in (i), we get:

$$\begin{aligned} y_1 &= \alpha(\beta_1 x + \beta_2 y + u_2) \\ &= \alpha\beta_1 x + \alpha\beta_2 y_1 + \alpha u_2 + u_1 \end{aligned}$$

or,

$$\begin{aligned} (1-\alpha\beta_2)y_1 &= \alpha\beta_1 x + \alpha u_2 + u_1 \\ \Rightarrow \qquad y_1 &= \frac{\alpha\beta_1 x}{1-\alpha\beta_2} + \frac{\alpha u_2 + u_1}{1-\alpha\beta_2}. \qquad \text{(iii)} \end{aligned}$$

Similarly,

$$y_2 = \frac{\beta_1 x}{1-\alpha\beta_2} + \frac{\beta_2 u_1 + u_2}{1-\alpha\beta_2}. \qquad \text{(iv)}$$

(ii) and (iv) are reduced forms corresponding to y_1 and y_2. These will exist only when $\alpha\beta_2 \neq 1$.
The OLS estimator of α is

$$\hat{\alpha} = \frac{\sum y_1 y_2}{\sum y_2^2} = \frac{\sum y_2(\alpha y_2 + u_1)}{\sum y_2^2} = \alpha + \frac{\sum y_2 u_1}{\sum y_2^2},$$

and

$$\begin{aligned} \operatorname{plim} \hat{\alpha} &= \alpha + \frac{\operatorname{plim} \frac{1}{T}\sum y_2 u_1}{\operatorname{plim} \frac{1}{T}\sum y_2^2} = \alpha + \frac{\operatorname{Cov}(y_2, u_1)}{\operatorname{Var}(y_2)} \\ \operatorname{Cov}(y_2, u_1) &= \operatorname{E}[y_2 - \operatorname{E}(y_2)]u_1 \\ &= \operatorname{E}\left[\frac{\beta_2 u_1 + u_2}{1-\alpha\beta_2}\right] \qquad \text{from (iv)} \\ &= \frac{\beta_2\sigma_1^2 + \sigma_{12}}{1-\alpha\beta_2} \\ &\neq 0. \end{aligned}$$

Hence

$$\operatorname{plim} \hat{\alpha} = \alpha + \frac{\left(\frac{\beta_2\sigma_1^2+\sigma_{12}}{1-\alpha\beta_2}\right)}{\operatorname{Var}(y_2)} \neq \alpha \qquad \Rightarrow \qquad \text{inconsistency.}$$

iii. Equation (i) is exactly identified hence ILS or 2SLS can be used. A brief description of any method is required.

Question 6

(a) A stationary AR(1) process

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \epsilon_t$$

where $|\phi_1| < 1$ and ϵ_t is identically and independently distributed with zero mean and finite variance, has initial value Y_0, where Y_0 is defined as

$$Y_0 = \frac{\phi_0}{1-\phi_1} + \sqrt{\frac{1}{(1-\phi_1^2)}}\epsilon_0.$$

Show that Y_0 is a random draw from the ensemble distribution for Y. Explain the advantages of defining the starting point Y_0 in this way.

(12 marks)

(b) **Consider two variables Y_t and X_t, where**

$$\begin{aligned} Y_t &= \alpha + Y_{t-1} + \epsilon_t \\ X_t &= \beta + X_{t-1} + v_t \end{aligned}$$

where ϵ_t and v_t are uncorrelated white noise processes. A researcher regresses Y_t on X_t, and tests the significance of the slope coefficient. Discuss in detail the result she will get.

(7 marks)

(c) **Explain what is meant when variables are cointegrated. Why is this considered to be important?**

(6 marks)

Reading for this question

Subject guide (2011), Chapter 13.

Dougherty, C. *Introduction to econometrics* (fourth edition) Chapter 13 (Introduction to no-stationary time series).

Gujarati, D.N. *Basic econometrics* (fifth edition) Chapter 21 (Time series econometrics: Some basic concepts).

Approaching the question

In part (a), by recursive substitution get the value of Y_t in terms of Y_0. Substitute the value of Y_0 from the question in the derived equation of Y_t and derive the expectation and variance of Y_t. Part (b) is based on the concept of spurious regression. Part (c) requires discussion of cointegration. A detailed answer is:

(a) Lagging and substituting t times,

$$\begin{aligned} Y_t &= \phi_1^t Y_0 + \phi_0 \left(\phi_1^{t-1} + \ldots + \phi_1^2 + \phi_1\right) + \phi_1^{t-1}\epsilon_1 + \ldots + \phi_1^2\epsilon_{t-2} + \phi_1\epsilon_{t-1} + \epsilon_t \\ &= \phi_1^t Y_0 + \phi_0 \frac{1-\phi_1^t}{1-\phi_1} + \phi_1^{t-1}\epsilon_1 + \ldots + \phi_1^2\epsilon_{t-2} + \phi_1\epsilon_{t-1} + \epsilon_t. \end{aligned}$$

With the stochastic definition of Y_0, we now have

$$\begin{aligned} Y_t &= \phi_1^t \left(\frac{\phi_0}{1-\phi_1} + \sqrt{\frac{1}{1-\phi_1^2}}\epsilon_0\right) + \phi_0 \frac{1-\phi_1^t}{1-\phi_1} + \phi_1^{t-1}\epsilon_1 + \ldots + \phi_1^2\epsilon_{t-2} + \phi_1\epsilon_{t-1} + \epsilon_t \\ &= \frac{\phi_0}{1-\phi_1} + \phi_1^t \sqrt{\frac{1}{1-\phi_1^2}}\epsilon_0 + \phi_1^{t-1}\epsilon_1 + \ldots + \phi_1^2\epsilon_{t-2} + \phi_1\epsilon_{t-1} + \epsilon_t. \end{aligned}$$

Hence,

$$\mathrm{E}(Y_t) = \frac{\phi_0}{1-\phi_1}$$

and

$$\begin{aligned}\mathrm{Var}(Y_t) &= \mathrm{Var}\left(\phi_1^t\sqrt{\frac{1}{1-\phi_1^2}}\epsilon_0 + \phi_1^{t-1}\epsilon_1 + \ldots + \phi_1^2\epsilon_{t-2} + \phi_1\epsilon_{t-1} + \epsilon_t\right) \\ &= \frac{\phi_1^{2t}}{1-\phi_1^2}\sigma_\epsilon^2 + \left(\phi_1^{2t-2} + \ldots + \phi_1^4 + \phi_1^2 + 1\right)\sigma_\epsilon^2 \\ &= \frac{\phi_1^{2t}}{1-\phi_1^2}\sigma_\epsilon^2 + \frac{1-\phi_1^{2t}}{1-\phi_1^2}\sigma_\epsilon^2 \\ &= \frac{\sigma_\epsilon^2}{1-\phi_1^2}.\end{aligned}$$

Given the generating process for Y_0, one has

$$\mathrm{E}(Y_t) = \frac{\phi_0}{1-\phi_1} \qquad \text{and} \qquad \mathrm{Var}(Y_t) = \frac{\sigma_\epsilon^2}{1-\phi_1^2}.$$

Hence Y_0 is a random draw from the ensemble distribution. Implicitly it has been assumed that the distribution of the error term and Y_0 are both normal. If we determine the starting point Y_0 in this way we can get rid of the transient time-dependent initials effects associated with Y_0, therefore, the expectation and variance of the process Y_t both become strictly independent of time.

(b) Regression of this type is known as *spurious regression.* If we regress Y_t and X_t so that

$$Y_t = \pi_0 + \pi_1 X_t + v_t,$$

Granger and Newbold have shown that although there is no relationship between Y and X, the regression will produce a t-ratio which will reject the null hypothesis $\mathrm{H}_0 : \pi_1 = 0$.

The reason for this result is that if $\mathrm{H}_0 : \pi_1 = 0$, then

$$Y_t = \pi_0 + v_t$$

and since Y_t is I(1) and π_0 is constant, it follows that v_t must be I(1). This violates the standard distributional theory based on the assumption that v_t is stationary (i.e. v_t is I(0)). Hence the misleading result.

(c) In general a linear combination of two time series will be non-stationary if one or more of them is non-stationary. The degree of integration of the combination will be equal to that of most highly integrated individual series. For example, a combination of I(1) and I(0) series will be I(1) and a combination of I(1) and I(1) series will be I(1). If long-run relationships exist between the time series then the result may be different. Suppose Y_t and X_t are both I(1). A linear combination of Y_t and X_t may be written as $u_t = Y_t - \beta X_t$. If the linear combination u_t is I(0), then Y_t and X_t are said to be cointegrated.

If Y_t and X_t are cointegrated then it implies that long-run relationship exists between Y_t and X_t. This concept can be generalised. Consider a general linear model

$$Y_t = \beta_1 + \beta_2 X_{2t} + \ldots + \beta_K X_{Kt} + u_t.$$

Then the disturbance term u_t can be thought of as measuring the deviation between components of the model. In the short run the divergence between the components will fluctuate, but if the model is correctly specified there will be a limit to the divergence. Hence though Ys and Xs are non-stationary, 'u' will be stationary. If there are K variables in the model the maximum number of cointergrating relationships will be $K - 1$.

Cointegration is an overriding requirement for any economic model using non-stationary time series data. If the variables do not cointegrate then we have a problem of spurious regression and econometric work becomes almost meaningless. If a cointegrating relationship exists then OLS estimators can be shown to be superconsistent.

Examiners' commentaries 2014

EC2020 Elements of econometrics

Important note

This commentary reflects the examination and assessment arrangements for this course in the academic year 2013–14. The format of the examination for 2014–15 remains the same as last year. However, the course syllabus of the examination for 2014–15 has changed. Panel data models (fixed effect and random effect models) and limited dependent variable models (tobit and sample selection bias) have been excluded from the course. These changes will also be publicised on the virtual learning environment (VLE).

Information about the subject guide

Unless otherwise stated, all cross-references will be to the latest version of the subject guide (2014). You should always attempt to use the most recent edition of any Essential reading textbook, even if the commentary and/or online reading list and/or subject guide refers to an earlier edition. If different editions of Essential reading are listed, please check the VLE for reading supplements – if none are available, please use the contents list and index of the new edition to find the relevant section.

General remarks

Learning outcomes

At the end of this course, and having completed the Essential reading and activities, you should be able to:

- describe and apply the classical regression model and its application to cross-section data
- describe and apply the:
  - Gauss-Markov conditions and other assumptions required in the application of the classical regression model
  - reasons for expecting violations of these assumptions in certain circumstances
  - tests for violations
  - potential remedial measures, including, where appropriate, the use of instrumental variables
- recognise and apply the advantages of logit, probit and similar models over regression analysis when fitting binary choice models
- competently use regression, logit and probit analysis to quantify economic relationships using standard regression programmes (Stata and EViews) in simple applications
- describe and explain the principles underlying the use of maximum likelihood estimation
- apply regression analysis to fit time-series models using stationary time series, with awareness of some of the econometric problems specific to time series applications (for example, autocorrelation) and remedial measures

- recognise the difficulties that arise in the application of regression analysis to nonstationary time series, know how to test for unit roots, and know what is meant by cointegration.

Common mistakes committed by candidates

- A large number of candidates were not able to distinguish between sample variance and covariance, and population variance and covariance (this is happening year after year). They treat them as the same. This results in incorrect analysis and candidates lose significant marks.

  Consider an example: Suppose data is deviation from respective sample means and the regression model is:

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

  The ordinary least squares estimator of β is:

  $$\hat{\beta} = \frac{\sum_{t=1}^{T} x_t y_t}{\sum_{t=1}^{T} x_t^2} = \beta + \frac{\sum_{t=1}^{T} x_t u_t}{\sum_{t=1}^{T} x_t^2}.$$

  In terms of variances and covariances (a large number of candidates prefer this terminology) this can be written as:

  $$\hat{\beta} = \beta + \frac{\text{Cov}(x, u)}{\text{Var}(x)}.$$

  Here $\text{Cov}(x, u)$ and $\text{Var}(x)$ are sample[$\text{Cov}(x, u)$] and sample[$\text{Var}(x)$].

  Candidates should realise that $\sum_{t=1}^{T} u_t$, $\sum_{t=1}^{T} x_t u_t$, $\text{Cov}(x, u)$ and $\text{Var}(x)$ given above are sample moments and as such $\sum_{t=1}^{T} u_t \neq 0$, $\sum_{t=1}^{T} x_t u_t \neq 0$ and $\text{Cov}(x, u) \neq 0$. But, if we take expectation, then:

  $$E[u_t] = 0,$$

  by assumption. Then:

  $$E\left[\sum_{t=1}^{T} x_t u_t\right] = \sum_{t=1}^{T} x_t [E(u_t)] = 0,$$

  as the x_t are fixed they can be taken out of the expectation, and so:

  $$E[\text{Cov}(x, u)] = E\left[\frac{1}{T}\sum_{t=1}^{T} x_t u_t\right] = 0,$$

  as previously argued. This makes $E(\hat{\beta}) = \beta$, i.e. $\hat{\beta}$ is an unbiased estimator for β.

  To prove consistency take plim to get:

  $$\begin{aligned}
  \text{plim}\left(\hat{\beta}\right) &= \beta + \text{plim}\left(\frac{\frac{1}{T}\sum_{t=1}^{T} x_t u_t}{\frac{1}{T}\sum_{t=1}^{T} x_t^2}\right) \\
  &= \beta + \frac{\text{plim}\left(\frac{1}{T}\sum_{t=1}^{T} x_t u_t\right)}{\text{plim}\left(\frac{1}{T}\sum_{t=1}^{T} x_t^2\right)} \\
  &= \beta + \frac{\text{plim(sample Cov}(x, u))}{\text{plim(sample Var}(x))} \\
  &= \beta + \frac{\text{population Cov}(x, u)}{\text{population Var}(x)}.
  \end{aligned}$$

  By assumption, population $\text{Cov}(x, u) = 0$ and population $\text{Var}(x) > 0$, hence $\text{plim}\left(\hat{\beta}\right) = \beta$, in other words $\hat{\beta}$ is a consistent estimator of β.

Remember that in general:

$$\begin{aligned}\text{plim(sample variance)} &= \text{population variance,} \\ \text{plim(sample covariance)} &= \text{population covariance.}\end{aligned}$$

This concept has been used in many questions. This simple mistake of not distinguishing between sample variance and covariance and population variance and covariance results in a significant loss of marks which might result in the loss of a degree class or even be the difference between pass and fail.

- Candidates struggled to give competent answers to the interpretation of empirical results. When interpreting an empirical result you should discuss the significance of the coefficients, magnitude and sign of the coefficients. Also, you should make sure that the GM conditions hold.
- Just as last year, many candidates did not appear to read the questions carefully enough and often omitted to give answers to parts of questions which asked for details of such things as the assumptions necessary for a particular result to be true.

Key steps to improvement

Essential reading for this course includes the subject guide and

> Dougherty, C. *Introduction to Econometrics*. (Oxford: Oxford University Press, 2011) fourth edition [ISBN 9780199280964].

Apart from Essential reading you should do some supplementary reading. Two very good books of the same level are:

> Gujarati, Damodar N. *Basic Econometrics*. (Boston; London: McGraw-Hill Education, 2009) fifth edition [ISBN 9780073375779].
>
> Woolridge, Jeffrey M. *Introductory Econometrics*. (Mason, Ohio: Thomson Learning, 2008) fifth edition [ISBN 9781408093757].

To understand the subject clearly it is important to supplement C. Dougherty, Introduction to econometrics(third edition) with subject guide EC2020 *Elements of econometrics* (2014), especially the chapter on the maximum likelihood.

It is very important to go through the subject guide carefully. It contains solutions to the questions given in the main book and also some additional questions and solutions. Working through these will improve your understanding of the subject.

The chapter on the maximum likelihood in the subject guide (Chapter 10) includes some additional theory which has not been covered in the main book. It is important to read the additional theory given in the subject guide to have a better understanding of the principles of maximum likelihood and tests based on the likelihood function.

At the website URL http://econ.lse.ac.uk/ie are PowerPoint slideshows that provide graphical treatment of the topics covered in the text, data sets, statistical tables and downloadable copy of the subject guide EC2020 *Elements of Econometrics* (2011). Candidates should utilise data sets using standard regression programmes (STATA or EViews). This will help in the understanding of the subject.

Question spotting

Many candidates are disappointed to find that their examination performance is poorer than they expected. This can be due to a number of different reasons and the *Examiners' commentaries* suggest ways of addressing common problems and improving your performance. We want to draw your attention to one particular failing – '**question spotting**', that is, confining your examination preparation to a few question topics which have come up in past papers for the course. This can have very serious consequences.

We recognise that candidates may not cover all topics in the syllabus in the same depth, but you need to be aware that Examiners are free to set questions on **any** aspect of the syllabus. This means that you need to study enough of the syllabus to enable you to answer the required number of examination questions.

The syllabus can be found in the 'Course information sheet' in the section of the VLE dedicated to this course. You should read the syllabus very carefully and ensure that you cover sufficient material in preparation for the examination.

Examiners will vary the topics and questions from year to year and may well set questions that have not appeared in past papers – every topic on the syllabus is a legitimate examination target. So although past papers can be helpful in revision, you cannot assume that topics or specific questions that have come up in past examinations will occur again.

If you rely on a question spotting strategy, it is likely you will find yourself in difficulties when you sit the examination paper. We strongly advise you not to adopt this strategy.

Examiners' commentaries 2014

EC2020 Elements of econometrics

Important note

This commentary reflects the examination and assessment arrangements for this course in the academic year 2013–14. The format of the examination for 2014–15 remains the same as last year. However, the course syllabus of the examination for 2014–15 has changed. Panel data models (fixed effect and random effect models) and limited dependent variable models (tobit and sample selection bias) have been excluded from the course. These changes will also be publicised on the virtual learning environment (VLE).

Information about the subject guide

Unless otherwise stated, all cross-references will be to the latest version of the subject guide (2014). You should always attempt to use the most recent edition of any Essential reading textbook, even if the commentary and/or online reading list and/or subject guide refers to an earlier edition. If different editions of Essential reading are listed, please check the VLE for reading supplements – if none are available, please use the contents list and index of the new edition to find the relevant section.

Comments on specific questions – Zone A

Candidates should answer **EIGHT** of the following **TEN** questions: **ALL** of the questions in Section A (8 marks each) and **THREE** questions from Section B (20 marks each). **Candidates are strongly advised to divide their time accordingly.**

Section A

Answer **ALL** questions from this section.

Question 1

Show that the infinite distributed lag model $Y_t = \alpha + \beta \sum_0^{\infty} \lambda^j X_{t-j} + \varepsilon_t$, where $|\beta| < 1$, can be written in terms of X_t and Y_{t-1}. What problems may occur when estimating this model?

(8 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fourth edition) Chapter 11 (Models using time series data).

Gujarati, D.N. *Basic econometrics* (fifth edition) Chapter 17.4 (The Koyck approach to distributed lag models).

Approaching the question

This question is based upon Koyck transformation of the infinite lag model. Answer is:

$$Y_t = \alpha + \beta X_t + \beta\lambda X_{t-1} + \beta\lambda^2 X_{t-2} + \beta\lambda^3 X_{t-3} + \cdots + \varepsilon_t. \qquad (1)$$

If we multiply this equation through by λ and lag we get:

$$\lambda Y_{t-1} = \alpha\lambda + \beta\lambda X_{t-1} + \beta\lambda^2 X_{t-2} + \beta\lambda^3 X_{t-3} + \beta\lambda^4 X_{t-4} + \cdots + \varepsilon_{t-1}. \qquad (2)$$

Now subtract (2) from (1) to give:

$$\begin{aligned} Y_t - \lambda Y_{t-1} &= \alpha(1-\lambda) + \beta X_t + \varepsilon_t - \lambda\varepsilon_{t-1} \\ Y_t &= \alpha(1-\lambda) + \beta X_t + \lambda Y_{t-1} + (\varepsilon_t - \lambda\varepsilon_{t-1}) \end{aligned}$$

which we could estimate by OLS except that Y_{t-1} and ε_{t-1} are correlated hence a RHS variable is correlated with the error term so OLS produces inconsistent parameter estimates.

Question 2

Let X be a random variable distributed with mean 0 and variances σ^2. Let $X_1, X_2, \ldots, X_T$ be an identically and independently distributed random sample from the distribution of X. Show that

$$S^2 = \frac{\sum_{t=1}^{T}(X_t - \bar{X})^2}{T}; \text{ where } \bar{X} = \frac{\sum_{t=1}^{T} X_t}{T}$$

is a biased estimator of, $\sigma^2 < \infty$.

(8 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fourth edition) Chapters R.6 (Unbiasedness and efficiency) and Appendix R.1 (Unbiased estimators of the population covariance and variance).

Subject guide, Chapter 2 (Properties of the regression coefficients and hypothesis testing).

Gujarati, D.N. *Basic econometrics* (fifth edition) Appendix A.7 (Statistical inference: estimation).

Approaching the question

You are required to take expectation of S^2 and show that it is not equal to σ^2. The answer is:

$$\mathrm{E}(S^2) = \frac{\mathrm{E}\sum(X_t - \bar{X})^2}{T} = \frac{\sum \mathrm{E}(X_t - \bar{X})^2}{T}.$$

We can write:

$$\begin{aligned} \sum \mathrm{E}(X_t - \bar{X})^2 &= \sum \mathrm{E}(X_t^2) - T\mathrm{E}(\bar{X}^2) \\ \mathrm{E}(X_t^2) &= \mathrm{var}(X_t) + [\mathrm{E}(X)]^2 = \sigma^2;\ \mathrm{E}(X) = 0 \\ \mathrm{E}(\bar{X}^2) &= \mathrm{var}(\bar{X}) + [\mathrm{E}(\bar{X})]^2 = \frac{\sigma^2}{T};\ \mathrm{E}(\bar{X}) = 0. \end{aligned}$$

Hence:

$$\sum \mathrm{E}(X_t - \bar{X})^2 = \sum \sigma^2 - T\left[\frac{\sigma^2}{T}\right] = T\sigma^2 - \sigma^2 = (T-1)\sigma^2.$$

Therefore:

$$\mathrm{E}(S^2) = \frac{(T-1)\sigma^2}{T} \neq \sigma^2 \;\Rightarrow\; S^2 \text{ is a biased estimator of } \sigma^2.$$

Question 3

Consider the following model

$$\boldsymbol{Y_t = \alpha X_t + u_t;\;\; t = 1, 2, \ldots, T}$$

where $\mathrm{E}(u_t) = 0$, $\mathrm{E}(u_t^2) = \sigma^2 X_t^2$ and $\mathrm{E}(u_s u_t) = 0$ if $s \neq t$ for all $s, t = 1, 2, \ldots, T$. The values of X_t are fixed in repeated samples.

(a) Derive the weighted least squares (WLS) estimator $\hat{\alpha}$ of α, and also derive the variance of $\hat{\alpha}$.

(b) Is the WLS estimator of a consistent? Explain your answer.

(8 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fourth edition) Chapter 7.3 (What can you do about hetroscedasticity?).

Subject guide, Chapter 7 (Heteroscedasticity).

Gujarati, D.N. *Basic econometrics* (fifth edition) Chapters 3 A.7 (Consistency of least-squares estimator) and 11.3 (The method of generalized least squares (GLS)).

Approaching the question

For part (a) transform the model by dividing it by X_t and then apply OLS to get the WLS estimators. For part (b) use sufficient condition for consistency. Answer is:

(a) Weighted least squares estimator is derived by estimating the following equation by OLS:

$$\frac{y_t}{x_t} = \alpha + \frac{u_t}{x_t}$$

and:

$$\mathrm{E}\left(\frac{u_t}{x_t}\right) = 0;\; \mathrm{var}\left(\frac{u_t}{x_t}\right) = \sigma^2 \text{ and } \mathrm{E}\left(\frac{u_s}{x_s}\frac{u_t}{x_t}\right) = 0 \text{ if } s \neq t.$$

Applying OLS we get the weighted least squares estimator as:

$$\hat{\alpha} = \frac{1}{T}\sum \frac{y_t}{x_t} = \frac{1}{T}\sum \frac{\alpha x_t + u_t}{x_t} = \alpha + \frac{1}{T}\sum \frac{u_t}{x_t}.$$

(b) Use sufficient conditions for consistency:

$$\mathrm{E}(\hat{\alpha}) = \alpha.$$

Variance of $\hat{\alpha}$ is:

$$\mathrm{E}(\hat{\alpha}-\alpha)^2 = \mathrm{E}\left(\frac{1}{T}\sum \frac{u_t}{x_t}\right)^2 = \frac{1}{T^2}\mathrm{E}\left[\sum \left(\frac{u_t}{x_t}\right)^2\right] = \frac{1}{T^2}\mathrm{E}\left(\sum \frac{u_s}{x_s}\frac{u_t}{x_t}\right) = \frac{1}{T^2}\sum \left(\frac{\sigma^2 x_t^2}{x_t^2}\right) = \frac{\sigma^2}{T}.$$

$\mathrm{E}(\hat{\alpha}) = \alpha$ and $\lim \mathrm{var}(\hat{\alpha}) \to 0$ as $T \to \infty$.

This implies that $\hat{\alpha}$ is a consistent estimator of α.

Question 4

Standard hypothesis testing on a linear regression model is achieved by using t tests on the parameter estimates, and an F test, which is 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. Explain potential reasons why would you obtain:

(a) statistically significant coefficient(s) according to the t test(s), but a statistically insignificant F test;

(b) no statistically significant coefficients according to the t tests, but a statistically significant F test;

(c) no statistically significant coefficients according to the t tests and a statistically insignificant F test.

(8 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fourth edition) Chapter 2.6 (Testing hypothesis relating to the regression coefficients).

Gujarati, D.N. *Basic econometrics* (fifth edition) Chapters 5.7 (Hypothesis testing: The test of significance approach) and 8.4 (Testing the overall significance of the sample regression).

Approaching the question

This question tests the basic concept of t and F statistics. Answer is:

(a) Statistically significant coefficient(s) according to the t test(s), but statistically insignificant coefficient(s) according to the F test is not possible. For example, in the 2 variables case the F statistic is the square of the t statistics – both statistics will tell the same story.

(b) Statistically insignificant coefficients according to the t tests but statistically significant coefficients according to the F test would suggest multicollinearity – correlation between variables leads to larger standard errors and hence low t values. The F test is not affected.

(c) Statistically insignificant coefficients according to the t tests and statistically insignificant coefficients according to the F statistic tell the same story – the model is not correct.

Question 5

Explain the concept of the likelihood function and hence the maximum likelihood estimator. Illustrate your answer with an example.

(8 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fourth edition) Chapter 10.6 (An introduction to maximum likelihood estimation).

Subject guide,Chapter 10 (Binary choice and limited dependent variable models, and maximum likelihood estimation).

Gujarati, D.N. *Basic econometrics* (fifth edition) Chapter 4 A.1 (Maximum likelihood estimation of two-variable linear model).

Approaching the question

Required to explain the concept of likelihood function and the maximum likelihood estimator. Both concepts should be explained with an example. Answer is:

Likelihood function

Definition: Let the probability density function (pdf) of X be $f(x;\,\theta)$ where θ is a parameter, then the likelihood function of a SRS $X_1, X_2, \ldots, X_n$ from X is the product of the individual densities of X_is taken as a function of θ.

The joint pdf of $X_1, X_2, \ldots, X_n$ is:

$$f(x_1, x_2, \ldots, x_n;\,\theta) = f(x_1;\,\theta)f(x_2;\,\theta)\cdots f(x_n;\,\theta).$$

The likelihood function $L(\theta;\,x_1, x_2, \ldots, x_n)$ has the same formulation as the joint pdf, but now it is a function of θ.

Maximum likelihood estimator (MLE)

Let $X \sim f(x;\,\theta)$. MLE $\hat{\theta}$ of the parameter θ is an estimator that maximises the likelihood function.

To obtain MLE, set:

$$\frac{\partial L}{\partial \theta} = 0 \text{ and solve for } \hat{\theta}.$$

It is easier to take logs and then maximise, as in both the situations we will arrive at the same maximum. As:

$$\frac{\partial \ln L}{\partial \theta} = \frac{1}{L}\frac{\partial L}{\partial \theta} = 0 \;\Rightarrow\; \frac{\partial L}{\partial \theta} = 0.$$

Example:

Let the probability density function of a population X be:

$$f(x) = p^x(1-p)^{1-x};\; x = 0, 1$$

where p is the probability of success. Let $X_1, X_2, \ldots, X_n$ be a simple random sample from X. Obtain the likelihood function of the sample observations and also the MLE of p.

Answer

Likelihood function is:

$$\begin{aligned} L &= f(x_1)f(x_2)\cdots f(x_n) \\ &= p^{x_1}(1-p)^{1-x_1}\cdots p^{x_n}(1-p)^{1-x_n} \\ &= p^{\sum x_i}(1-p)^{n-\sum x_i}. \end{aligned}$$

Hence:

$$\ln L = \left(\sum x_i\right)\ln p + \left(n - \sum x_i\right)\ln(1-p)$$

and so:

$$\frac{\partial \ln L}{\partial p} = \frac{\sum x_i}{\hat{p}} + \frac{n - \sum x_i}{1-\hat{p}}(-1) = 0$$

or:

$$\frac{(1-\hat{p})\sum x_i - \hat{p}\,(n - \sum x_i)}{\hat{p}(1-\hat{p})} = 0.$$

Solving, we get:

$$\hat{p} = \frac{\sum x_i}{n}$$

which is the maximum likelihood estimate. The corresponding maximum likelihood estimator (MLE) of p is:

$$\hat{p}_{MLE} = \frac{\sum X_i}{n} = \bar{X}.$$

Section B

Answer **THREE** questions from this section.

Question 6

The natural log of expenditure on beer at 1995 prices (beer_t) is regressed on the natural log of total household expenditure at 1995 prices (exp_t), the natural log of the price of beer relative to all consumer prices (pb_t) and the natural log of the price of alcoholic drinks excluding beer relative to all consumer prices (pa_t), gave the following results:

$$\begin{array}{rcl} \text{beer}_t & = & -5.272 + 1.266\text{exp}_t - 0.989\text{pb}_t - 0.412\text{pa}_t + e_t \\ & & \quad (1.387) \quad (0.114) \qquad (0.096) \qquad (0.134) \end{array}$$

where e_t is the estimated residual, standard errors are in brackets, the sample size is 45, and $R^2 = 0.906$.

(a) Interpret the estimated equation.

(4 marks)

(b) Test the null hypothesis that the coefficient on exp_t is unity at the 5% level of significance. What is the relevance of this hypothesis?

(5 marks)

(c) Construct a 95% confidence interval for the coefficient of pb_t. Explain why the confidence interval is a useful construct.

(5 marks)

(d) Assuming that the variable pb_t is positively correlated with pa_t, explain what would happen to the parameter estimate of pa_t if the variable pb_t is dropped from the regression.

(6 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fourth edition) Chapters 2.8 (Testing hypothesis relating to the regression coefficients), 2.9 (Confidence intervals), 3.2 (Derivation and interpretation of the multiple regression coefficients) and 6.2 (The effect of omitted variable that ought to be included).

Subject guide,Chapters 4 (Transformation of variables) and 6 (Specification of regression variables).

Gujarati, D.N. *Basic econometrics* (fifth edition) Chapters 5.7 (Testing hypothesis relating to the regression coefficients), 6.5 (How to measure elasticity: The log-linear model) and 13.3 (Consequences of model specification errors).

Approaching the question

As it is a double log model the coefficients are elasticities. Part (d) is based on omitted variable bias. Answer is:

(a) The t values are 3.80 (constant) 11.10 (exp) 10.30 (pb) and 3.07 (pa). The degrees of freedom are 41 and the 95% critical value is 2.02. All the coefficients are significant and show that (i) as total household expenditure increases the expenditure on beer will increase, (ii) as the price of beer increases the expenditure on beer will decrease and (iii) as the price of alcoholic drinks excluding beer increase the expenditure on beer decreases.These results, except for the last one, are all in agreement with economic theory.

(b) The t test statistic is $(1.266 - 1)/0.114 = 2.333$ which is greater than the critical t value hence we reject H_0. If the hypothesis was true then expenditure would have unit elasticity which implies that as household expenditure increases expenditure on beer has an equivalent increase. This is now not the case since we rejected H_0.

(c) The 95% confidence interval is $-0.989 \pm 2.02 \times 0.096 = [-0.795, \ -11.83]$. The confidence interval gives an interval in which we are 95% sure of containing the true value of the parameter.

(d) If a variable is dropped from the 'true' model then OLS will be subject to 'omitted variable bias' on the remaining parameters. The bias will depend on the sign and size of the parameter on the omitted variable and the sign and size of the correlation coefficient between the omitted variable and the variable in question. We know that the coefficient on pb was significantly different from zero hence there will be bias in the parameter on pa. We know that the coefficient on pb is negative and we would suspect that the correlation between pa and pb is positive (they are likely to move together) hence the omitted variable bias will be negative(i.e. the resulting estimate of the coefficient on pa will be an underestimate.)

Question 7

A simple model of supply and demand for a consumer good is:

$$\begin{aligned} D_t &= \alpha_0 + \alpha_1 P_t + \alpha_2 Y_t + \alpha_3 \bar{P}_t + u_{1t} \\ S_t &= \beta_0 + \beta_1 P_t + \beta_2 P_{t-1} + u_{2t} \\ D_t &= S_t \end{aligned}$$

where D_t is the demand for the good, S_t is the supply of the good, P_t is the price of the good, $\bar{P}_t$ is an index of retail prices, Y_t is consumers' income, and u_{1t} and u_{2t} are random disturbances.

(a) What do you understand by exogenous and endogenous variables? Which variables in the above model would you consider to be endogenous and which exogenous? Explain your answer.

(4 marks)

(b) Examine the identification of each equation in the model.

(5 marks)

(c) Explain why indirect least squares (ILS) is an inappropriate estimation method for the supply equation. When can indirect least squares (ILS) be used?

(6 marks)

(d) Suppose the equations were estimated using two-stage least squares. Explain this method of estimation and describe the conditions under which this method can be used.

(5 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fourth edition) Chapters 9.1 (Simultaneous equation models: Structural and reduced form equations), 9.2 (Simultaneous equation bias) and 9.3 (The order condition of identification and instrumental variable estimation and indirect least squares (Box: 9.2)).

Subject guide, Chapter 9 (Simultaneous equation estimation).

Gujarati, D.N. *Basic econometrics* (fifth edition) Chapters 18.3 (Simultaneous equation bias), 20.3 (Estimation of a just identified equation: The method of indirect least squares) and 20.4 (Estimation of an overidentified equation: The method of two-stage least squares).

Approaching the question

In part (a) concept of endogenous and exogenous variables should be explained. Part (b) is based on order conditions of identification. Part (c) deals with the concept of indirect least squares and in part (d) brief discussion of two-stage least squares method estimation is required. Detailed answer is:

(a) Endogenous variables are determined by the model, exogenous variables are determined outside the model.In this model the endogenous variables are D, S and P since they are effectively the left-hand side (LHS) variables in the model. The exogenous variables are Y, $\bar{P}$ and P_{t-1} since they do not appear as a right hand side (RHS) variables.

(b) An equation is said to be identified if there are enough instruments (i.e. exogenous variables not in the equation) in the model to estimate the parameters on the RHS endogenous variables. In the first equation the parameter on P has to be estimated by Instrumental Variables and there is only P_{t-1} available to act as instruments (both Y and are already present in the equation) hence the equation is exactly identified. In the second equation the parameter on P has to be estimated by Instrumental Variables and there are two valid instruments (Y and $\bar{P}$) for one parameter hence the equation is overidentified. To put this in more formal terms the order condition for identification compares the number of omitted exogenous or predetermined variables with the number of included endogenous $-$ 1. In equation (1) the number of omitted variables is 1 (i.e. P_{t-1}) and the number of included endogenous $-$ 1 is $2 - 1 = 1$ hence the equation is exactly identified. The second equation is overidentified by the same reasoning.

(c) Indirect Least Squares is an estimation procedure whereby the parameter estimates for the reduced form are solved to estimate the parameters of the structural form. If an equation is overidentified there are more equations to solve than there are unknown parameters hence the estimation procedure does not lead to unique estimates. If an equation is exactly identified a unique estimate of the structural form is obtained. The second equation is overidentified and hence cannot be uniquely estimated by indirect least squares.

(d) In two-stage least squares stage one is performed by using OLS on the reduced form. Stage two uses the 'prediction' of the endogenous variable from stage one (which is a linear combination of exogenous and predetermined variables and hence itself is exogenous) to replace the endogenous variables on the RHS. The equation can now be estimated by OLS without incurring simultaneous equation bias.

Question 8

Consider a model

$$y_t = \theta y_{t-1} + u_t; \; t = 1, 2, \ldots, T$$

where $y_0 = 0$, $E(u_t) = 0$, $E(u_t^2) = \sigma^2$ and $E(u_s u_t) = 0$ when $s \neq t$, for all $s, t = 1, 2, \ldots, T$.

(a) Derive the mean and variance of y_t when $\theta = 1$ and comment on the result.

(5 marks)

(b) Describe in detail the Dickey-Fuller procedure for testing for the order of integration of a time series variable. Give the assumptions the test requires and discuss the advantages and disadvantages of the procedure.

(5 marks)

(c) Consider a model

$$\boldsymbol{y_t = \alpha_1 + \alpha_2 y_{t-1} + u_t;\ t = 1, 2, \ldots, T}$$
$$\boldsymbol{u_t = \rho u_{t-1} + \varepsilon_t}$$

where $\mathbf{E}(\varepsilon_t) = 0$; $\mathbf{E}(\varepsilon_t^2) = \sigma^2$ **and** $\mathbf{E}(\varepsilon_s \varepsilon_t) = 0$ **when** $s \neq t$ **for all** $s, t = 1, 2, \ldots, T$. **Derive the specification for the Augmented Dickey-Fuller test.**

(5 marks)

(d) Consider a model

$$\boldsymbol{Y_t = \beta X_t + u_t;\ t = 1, 2, \ldots, T}$$
$$\boldsymbol{u_t = \theta e_{t-1} + e_t \quad |\theta| < 1}$$

where $\mathbf{E}(e_t) = 0$; $\mathbf{E}(e_t^2) = \sigma^2$ **and** $\mathbf{E}(e_s e_t) = 0$ **when** $s \neq t$ **for all** $s, t = 1, 2, \ldots, T$. **Are** Y_t **and** X_t **cointegrated? Explain your answer.**

(5 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fourth edition) Chapter (Introduction to nonstationarity time series).

Subject guide, Chapter 13 (Introduction to nonstationary time series).

Gujarati, D.N. *Basic econometrics* (fifth edition) Chapters 21.3 (Stochastic process), 21.9 (The unit root test) and 21.11 (Cointegration: Regression of a unit root time series on another unit root time series).

Approaching the question

Part (a) is based on the concept of nonstationary processes. Part(b) and (c) are based on Dickey-Fuller and augumented Dickey-Fuller tests. For part (d) concept of cointegration is required. Answer is:

(a) Model is:

$$Y_t = Y_{t-1} + u_t;\ t = 1, 2, \ldots, T. \qquad (1.1)$$

We can write for:

$$\begin{array}{ll} t = 1 & Y_1 = u_1 \\ t = 2 & Y_2 = Y_1 + u_2 = u_1 + u_2 \\ t = 3 & Y_3 = Y_2 + u_3 = u_1 + u_2 + u_3 \\ \vdots & \end{array}$$

Doing these recursive substitutions, we can write:

$$Y_t = Y_{t-1} + u_t = u_t + u_{t-1} + \cdots + u_1 = \sum_{t=1}^{t} u_t.$$

$\mathrm{E}(Y_t) = 0$ and $\mathrm{var}(Y_t) = \mathrm{var}\left(\sum_{t=1}^{T} u_t\right) = t\sigma^2 \Rightarrow Y_t$ is non-stationary as variance of Y_t is dependent on time. Y_t is a random walk.

(b) The standard test for a unit root is due to Dickey and Fuller and is based on the model

$$Y_t = \beta_1 + \beta_2 Y_{t-1} + \lambda t + u_t$$

which can be re-written as:

$$\Delta Y_t = \beta_1 + (1 - \beta_2) Y_{t-1} + \lambda t + u_t$$

where:

$$\Delta Y_t = Y_t - Y_{t-1}.$$

The hypothesis to be tested is:

$$\mathrm{H}_0 : 1 - \beta_2 = 0 \quad \text{vs.} \quad \mathrm{H}_1 : 1 - \beta_2 \neq 0.$$

Under the alternative hypothesis the process is stationary. We cannot use the standard t test procedure in this case because the distribution of the t statistic is not a t distribution so critical values have been computed by Dickey and Fuller using Monte-Carlo techniques. The test is sensitive to the presence of serial correlation in the error term so we need to take steps to remove the effects of this serial correlation, this is done by including lagged values of Y_t in the regression, i.e.:

$$Y_t = \beta_1 + \beta_2 Y_{t-1} + \beta_3 Y_{t-2} + \lambda t + u_t$$

for an AR(1) serial correlation. This is more easily tested by using the model:

$$Y_t = \beta_1 + (1 - \beta_2 - \beta_3) Y_{t-1} + \beta_3 \Delta Y_{t-2} + \lambda t + u_t$$

with null hypothesis $\mathrm{H}_0 : 1 - \beta_2 - \beta_3 = 0$, using Dickey-Fuller tables.

The model is:

$$\begin{aligned} Y_t &= \alpha_1 + \alpha_2 Y_{t-1} + u_t \\ u_t &= \rho u_{t-1} + \varepsilon_t ;\ \varepsilon \text{ is } I(0) \qquad \text{(i)} \end{aligned}$$

Lag (i) by one period and multiply it by ρ to get:

$$\rho Y_{t-1} = \alpha_1 \rho + \alpha_2 \rho Y_{t-2} + u_{t-1} \qquad \text{(ii)}$$

Subtract (ii) from (i) and rearrange to get:

$$Y_t = \alpha_1(1-\rho) + (\alpha_2 + \rho) Y_{t-1} - \alpha_2 \rho Y_{t-2} + (u_t - \rho u_{t-1})$$

or:

$$Y_t = \alpha_1(1-\rho) + (\alpha_2 + \rho) Y_{t-1} - \alpha_2 \rho Y_{t-2} + \varepsilon_t.$$

This can be written as:

$$Y_t = \beta_1 + \beta_2 Y_{t-1} + \beta_3 Y_{t-2} + \varepsilon_t$$

If we subtract Y_{t-1} from both sides and add and subtract $\beta_3 Y_{t-1}$ from the right side we get:

$$Y_t - Y_{t-1} = \beta_1 + \beta_2 Y_{t-1} - Y_{t-1} + \beta_3 Y_{t-1} + \beta_3 Y_{t-2} + \varepsilon_t$$

or:

$$\Delta Y_t = \beta_1 + (\beta_2 + \beta_3 - 1) Y_{t-1} - \beta_3 \Delta Y_{t-1} + \varepsilon_t.$$

To test for unit root we test the coefficient of Y_{t-1}, i.e.:

$$\mathrm{H}_0 : \beta_2 + \beta_3 - 1 = 0 \quad \text{vs.} \quad \mathrm{H}_1 : \beta_2 + \beta_3 - 1 < 0.$$

(c) Y_t and X_t are cointegrated if a linear combination of Y_t and X_t is $I(0)$. Hence, we have to examine the stationarity of u_t.

$$\begin{aligned} \mathrm{E}(u_t) &= \mathrm{E}(e_t + \theta e_{t-1}) = 0 \\ \mathrm{E}(u_t^2) &= \mathrm{E}(e_t^2) + \theta^2 \mathrm{E}(e_{t-1}^2) + 2\theta \mathrm{E}(e_t e_{t-1}) \\ &= (1+\theta^2)\sigma^2 \end{aligned}$$

since $\mathrm{E}(e_t e_{t-1}) = 0$.

$$\begin{aligned} \mathrm{E}(u_t u_{t-1}) &= \mathrm{E}(e_t + \theta e_{t-1})(e_{t-1} + \theta e_{t-2}) \\ &= \mathrm{E}(e_t e_{t-1}) + \theta \mathrm{E}(e_t e_{t-2}) + \theta \mathrm{E}(e_{t-1}^2) + \theta^2 \mathrm{E}(e_{t-1} e_{t-2}) \\ &= \theta \sigma^2 \end{aligned}$$

since all terms in $\mathrm{E}(e_t e_{t-s}) = 0$, $s > 0$.

$$\begin{aligned} \mathrm{E}(u_t u_{t-2}) &= \mathrm{E}(e_t + \theta e_{t-1})(e_{t-2} + \theta e_{t-3}) \\ &= \mathrm{E}(e_t e_{t-2}) + \theta \mathrm{E}(e_t e_{t-3}) + \theta \mathrm{E}(e_{t-1} e_{t-2}) + \theta^2 \mathrm{E}(e_{t-2} e_{t-3}) \\ &= 0. \end{aligned}$$

Thus both first and second moments are independent of t, therefore, u_t must be (weakly) stationary. This implies that Y_t and X_t are cointegrated.

Question 9

(a) Consider a model

$$\boldsymbol{Y_i = \beta_1 + \beta_2 X_i + u_i;\ i = 1, 2, \ldots, n}$$

where Y_i is a binary variable that takes the value of 1 if the event takes place and 0 otherwise, and $\mathrm{E}(u_i) = 0$ for $i = 1, 2, \ldots, n$.

i. Explain fully the problems which arise if the above model is estimated by ordinary least squares (OLS).

(5 marks)

ii. How would you estimate the model by weighted least squares where the weights are the estimated standard deviation? Discuss the advantages and disadvantages of this procedure.

(5 marks)

(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, and $Y_i = 0$ otherwise.

She estimated the model defining $Y_i = f(S_i, E_i) + u_i$, where S_i is the years spent by the head of the i-th household in full time education, E_i is the average earnings of the head of the i-th household, and u_i is an unobserved disturbance term. The model was estimated by logit with the following results:

$$\begin{aligned} \hat{Y}_i &= \underset{(1.57)}{-2.56} + \underset{(0.10)}{0.521} S_i + \underset{(0.012)}{0.067} E_i;\ \log L_U = -321.25 \quad \log L_R = -416.01 \end{aligned}$$

where asymptotic standard errors are in brackets, $\log L_U$ is the log likelihood from the unrestricted model, and $\log L_R$ is the log likelihood of the model where all the slope coefficients are restricted to zero.

i. Explain how the coefficients were estimated.

(6 marks)

ii. Test the null hypothesis that all the slope coefficients are jointly equal to zero at the 5% level of significance.

(4 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fourth edition) Chapters 10.1 (The linear probability model), 10.2 (Logit analysis) and 10.6 (An introduction to maximum likelihood estimation).

Subject guide,Chapter 10 (Binary choice and limited dependent variable models, and maximum likelihood estimation).

Gujarati, D.N. *Basic econometrics* (fifth edition) Chapters 15.2 (The linear probability model (LPM) and 15.5 (The logit model).

Approaching the question

In part (a) the linear probability model should be discussed. For part b(i) a brief discussion of the logit model is required and part b(ii) is based on the concept of the likelihood ratio test. Detailed answer is:

(a) i. LPM is used to denote a model in which dependent variable is binary which takes value 1 if the event occurs and 0 if it does not. It is estimated by the ordinary least squares (OLS).

Let the model be:

$$Y_i = \beta_0 + \beta_1 X_i + u_i;\ i = 1, 2, \ldots, n \qquad \text{(i)}$$

where:

$$Y_i = \begin{cases} 1 \text{ if event occurs} \\ 0 \text{ if not.} \end{cases}$$

Assume $\mathrm{E}(u_i) = 0$, then:

$$\mathrm{E}[Y_i \mid X_i] = \beta_0 + \beta_1 X_i. \qquad \text{(ii)}$$

Also:

$$\mathrm{E}[Y_i \mid X_i] = 1 \cdot P(Y_i = 1) + 0 \cdot P(Y_i = 0) = P(Y_i = 1) = P_i \qquad \text{(iii)}$$

From (ii) and (iii):

$$\mathrm{E}[Y_i \mid X_i] = \beta_0 + \beta_1 X_i = P_i$$

hence, we can interpret $\mathrm{E}[Y_i \mid X_i] = \beta_0 + \beta_1 X_i$ as the probability that the event will occur given X_i.

If we denote $\hat{\beta}_0$ and $\hat{\beta}_1$ as an estimate of β_0 and β_1, then we can write:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i = \hat{P}_i \qquad \text{(iv)}$$

as the estimated probability that the event will occur.

As Y_i takes only two values 1 or 0, therefore u_i can take only two values $1 - \beta_0 - \beta_1 X_i$ when $Y_i = 1$ and $\beta_0 - \beta_1 X_i$ when $Y_i = 0$. Based on this we can write the probability distribution of u_i as:

Y_i	u_i	$f(u_i)$
1	$1 - \beta_0 - \beta_1 X_i$	$\beta_0 + \beta_1 X_i$
0	$-\beta_0 - \beta_1 X_i$	$1 - \beta_0 - \beta_1 X_i$

This probability distribution also satisfies the assumption that:

$$\mathrm{E}(u_i) = (1 - \beta_0 - \beta_1 X_i)(\beta_0 + \beta_1 X_i) + (-\beta_0 - \beta_1 X_i)(1 - \beta_0 - \beta_1 X_i) = 0.$$

We can write $\mathrm{var}(u_i)$ as:

$$\begin{aligned}
\mathrm{var}(u_i) &= \mathrm{E}(u_i^2) \\
&= (1 - \beta_0 - \beta_1 X_i)^2 (\beta_0 + \beta_1 X_i) + (-\beta_0 - \beta_1 X_i)^2 (1 - \beta_0 - \beta_1 X_i) \\
&= (1 - \beta_0 - \beta_1 X_i)(\beta_0 + \beta_1 X_i)[(1 - \beta_0 - \beta_1 X_i) + (\beta_0 + \beta_1 X_i)] \\
&= (\beta_0 + \beta_1 X_i)(1 - \beta_0 - \beta_1 X_i) \\
&= \mathrm{E}(Y_i)[1 - \mathrm{E}(Y_i)] \\
&= P_i(1 - P_i);\ \text{for all } i = 1, 2, \ldots, n.
\end{aligned}$$

Hence the disturbance term is heteroscedastic. This will make OLS estimators inefficient. In many cases the estimated probability $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$ will be negative or greater than 1.

ii. Weighted least squares:

We can see from (iv), that estimator of P_i is $\hat{P}_i = \hat{Y}_i$, therefore $\hat{Y}_i(1 - \hat{Y}_i)$ can be used as an estimator of:

$$\mathrm{var}(u_i) = \mathrm{E}(Y_i)[1 - \mathrm{E}(Y_i)] = P_i(1 - P_i).$$

Weights can be obtained as:

$$W_i = [\hat{Y}_i(1 - \hat{Y}_i)]^{1/2}.$$

Divide (i) by W_i and apply OLS to:

$$\frac{Y_i}{W_i} = \frac{\beta_0}{W_i} + \beta_1 \frac{X_i}{W_i} + \frac{u_i}{W_i};\ i = 1, 2, \ldots, n$$

obtain the WLS estimator of β_0 and β_1. This will give an efficient estimator.

Problem:

In practice the estimated variance of u_i, $\hat{Y}_i(1 - \hat{Y}_i)$ may be negative as again the estimated probability $\hat{Y}_i$ may be negative or greater than 1.The obvious correction of the estimated negative probability is to constrain estimated probabilities within [0, 1] interval. If we do this we might predict an occurrence with probability 1, when it is possible that it might not occur or we might predict an occurrence with probability 0 when it might actually occur. The estimation process may give unbiased estimates but predictions obtained from it will be biased.

(b) i. Logit model uses the cumulative standardised logistic distribution. Maximum likelihood technique is used to obtain the estimates of the parameters. Estimates have the standard maximum likelihood properties i.e. the estimators are consistent, asymptotically efficient and asymptotically normally distributed. [For technical details see Dougherty (third edition) Ch. 10.2].

ii. H_0 : all slope coefficients are equal to zero.
$-2[\log L_R - \log L_U] \sim \chi^2_q$, where q is the number of restrictions imposed by H_0.
$\Rightarrow -2[-416.01 - (-321.25)] = 189.52 \sim \chi^2_2$.
Critical value of chi-square with 2 df at 5% level of significance is = 5.991, hence reject H_0.

Question 10

(a) Consider a model

$$Y_{it} = \beta_1 + \sum_{j=2}^{K} \beta_j X_{jit} + \alpha_i + u_{it};\ i = 1, 2, \ldots, n;\ t = 1, 2, \ldots, T$$

where Y is the dependent variable, X_j, $j = 2, \ldots, K$, are observed explanatory variables, α_i is an unobserved fixed effect and u_{it} is the disturbance term assumed to satisfy the usual regression model conditions. The index i refers to cross-section and the index t refers to the time period.

In the context of the model above briefly explain the least squares dummy variables method of estimation and its drawback.

(8 marks)

(b) The following estimates were obtained using an annually recorded US data panel of 550 individuals' wages over a period of 7 years. The dependent variable is the natural log of wage.

Independent variables	Pooled OLS	Random Effects	Fixed Effects
Years of full time education	0.091 (0.005)	0.092 (0.011)	
Black	−0.139 (0.024)	−0.139 (0.048)	
Hispanic	0.016 (0.021)	0.022 (0.043)	
Work experience	0.067 (0.014)	0.106 (0.015)	
Experience squared	−0.0024 (0.0008)	−0.0047 (0.0007)	−0.0052 (0.0007)
Married	0.108 (0.016)	0.064 (0.017)	0.047 (0.018)
Union membership	0.182 (0.017)	0.106 (0.018)	0.080 (0.019)

Six year dummy variables and a constant term were included in all three equations but the results are not reported. Black, Hispanic, Married and Union membership are dummy variables which take the value of one if the respondent has the relevant characteristic and are zero otherwise. The numbers in parentheses are standard errors of the coefficient estimates.

i. **Explain why there are no fixed effects estimated coefficients for the first four explanatory variables in the table.**

(3 marks)

ii. **What interpretation would you give to the unobserved effects in a wage equation of this kind?**

(3 marks)

iii. **The OLS coefficients for Union membership and Married are higher than for the other two estimates. What does this suggest about the correlation between being married and being a member of a union and the unobserved effects? Explain your answer.**

(3 marks)

iv. **If there is a significant correlation between these two explanatory variables and the unobserved effects, what does this indicate about the properties of these estimates?**

(3 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fourth edition) Chapter 14 (Introduction to panel data models).

Subject guide, Chapter 14 (Introduction to panel data).

Gujarati, D.N. *Basic econometrics* (fifth edition) Chapter 16.3 (Estimation of panel data regression models).

Approaching the question

Part (a) is a discussion of least squares dummy variable method of estimation. Part (b) requires a good concept of fixed effect and random effect estimation methods. Answer is:

(a) In this approach the unobserved effect is brought explicitly in the model. A set of dummy variables D_i is defined, where D_i is equal to 1 in the case of an observation relating to an individual i and 0 otherwise. The model can be written as:

$$Y_{it} = \sum_{j=2}^{K} \beta_j X_{jit} + \sum_{i=1}^{n} \alpha_i D_i + u_{it}. \qquad \text{(vi)}$$

The unobserved effect is now being treated as the coefficient of the individual specific dummy variable. The term $\alpha_i D_i$ represents a fixed effect on the dependent variable Y_i for individual i.

If we want to keep the intercept in the model then instead of n dummy variables $(n-1)$ dummy variables has to be used, otherwise we will fall into the dummy variable trap.

It can be shown that LSDV method is identical to the within groups method.

Drawbacks:

- The intercept and any explanatory variable that remain constant for each individual will drop out of the model.
- The variation in $(X_{ij} - \bar{X}_i)$ may be much smaller than the variation in X_j. If this is the case, the impact of the disturbance term may be relatively large, giving rise to imprecise estimates.

- There is loss of substantial number of degrees of freedom.

(b) i. The first three variables (education, black and Hispanic) do not vary over time. Thus they will be perfectly collinear with the fixed effects and their coefficients cannot be estimated. The work experience variable in most cases will increase by one each year and will thus be collinear with the time dummies. (Note this will not apply to the quadratic term)

ii. Unobserved effects in this kind of equation are usually thought to represent talent, ability, capacity to work hard etc. If so, random effects models may not give consistent estimates

iii. The OLS estimates make no allowance for the unobserved effects. Thus the higher coefficients for union and Married suggest that these variables are picking up some of the unobserved effects. The OLS estimates probably suffer from omitted variable bias as a result of the omission of the unobserved effects. If the unobserved effects (ability) are positively correlated with earnings, this suggests that Union and Marries are positively correlated with the unobserved effects (as their coefficients appear to be biased upwards in the OLS estimates).

(iv) It suggests that the OLS estimates are biased and inconsistent (omitted variable bias). It suggests that the Random Effects estimates are also inconsistent (they are biased anyway) as the random effects are correlated with explanatory variables.

Examiners' commentaries 2014

EC2020 Elements of econometrics

Important note

This commentary reflects the examination and assessment arrangements for this course in the academic year 2013–14. The format of the examination for 2014–15 remains the same as last year. However, the course syllabus of the examination for 2014–15 has changed. Panel data models (fixed effect and random effect models) and limited dependent variable models (tobit and sample selection bias) have been excluded from the course. These changes will also be publicised on the virtual learning environment (VLE).

Information about the subject guide

Unless otherwise stated, all cross-references will be to the latest version of the subject guide (2014). You should always attempt to use the most recent edition of any Essential reading textbook, even if the commentary and/or online reading list and/or subject guide refers to an earlier edition. If different editions of Essential reading are listed, please check the VLE for reading supplements – if none are available, please use the contents list and index of the new edition to find the relevant section.

Comments on specific questions – Zone B

Candidates should answer **EIGHT** of the following **TEN** questions: **ALL** of the questions in Section A (8 marks each) and **THREE** questions from Section B (20 marks each). **Candidates are strongly advised to divide their time accordingly.**

Section A

Answer **ALL** questions from this section.

Question 1

In the model

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

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

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

And $\mathrm{E}(u_t) = \mathrm{E}(v_t) = 0$, $\mathrm{E}(u_t v_t) = \mathrm{E}(x_t u_t) = 0$. y_t, x_t and x_t^* have zero arithmetic 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.

(8 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fourth edition) Chapter 8.4 (The consequences of measurement errors).

Subject guide(2011) Chapter 8 (Stochastic regressors and measurement errors).

Gujarati, D.N. *Basic econometrics* (fifth edition) Chapter 13.5 (Errors of measurement).

Approaching the question

To examine consistency plim of $\hat{\beta}$ has to be derived. Detailed answer is:

$$\begin{aligned}
y_t &= \beta x_t + u_t \quad \text{where} \quad x_t^* = x_t + v_t \\
\hat{\beta} &= \frac{\sum x_t^* y_t}{\sum x_t^{*2}} = \frac{\sum (x_t + v_t)(\beta x_t + u_t)}{\sum (x_t + v_t)^2} \\
&= \frac{\beta \sum x_t^2 + \sum x_t u_t + \beta \sum x_t v_t + \sum v_t u_t}{\sum x_t^2 + \sum v_t^2 + 2 \sum x_t v_t} \\
\text{plim } \hat{\beta} &= \frac{\text{plim}[\beta \sum x_t^2 + \sum x_t u_t + \beta \sum x_t v_t + \sum v_t u_t]/T}{\text{plim}[\sum x_t^2 + \sum v_t^2 + 2 \sum x_t v_t]/T}
\end{aligned}$$

or:

$$\text{plim } \hat{\beta} = \frac{\beta \sigma_x^2}{\sigma_x^2 + \sigma_v^2} \neq \beta \quad \Rightarrow \quad \text{Inconsistency.}$$

Question 2

Show that the infinite distributed lag model $Y_t = \alpha + \beta \sum_0^\infty \lambda^j X_{t-j} + \varepsilon_t$, where $|\beta| < 1$, can be written in terms of X_t and Y_{t-1}. What problem will occur when estimating this model?

(8 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fourth edition) Chapter 11 (Models using time series data).

Gujarati, D.N. *Basic econometrics* (fifth edition) Chapter 17.4 (The Koyck approach to distributed lag models).

Approaching the question

This question is based upon Koyck transformation of the infinite lag model. Answer is:

$$Y_t = \alpha + \beta X_t + \beta\lambda X_{t-1} + \beta\lambda^2 X_{t-2} + \beta\lambda^3 X_{t-3} + \cdots + \varepsilon_t. \tag{1}$$

If we multiply this equation through by λ and lag we get:

$$\lambda Y_{t-1} = \alpha\lambda + \beta\lambda X_{t-1} + \beta\lambda^2 X_{t-2} + \beta\lambda^3 X_{t-3} + \beta\lambda^4 X_{t-4} + \cdots + \varepsilon_{t-1}. \tag{2}$$

Now subtract (2) from (1) to give:

$$\begin{aligned}
Y_t - \lambda Y_{t-1} &= \alpha(1-\lambda) + \beta X_t + \varepsilon_t - \lambda\varepsilon_{t-1} \\
Y_t &= \alpha(1-\lambda) + \beta X_t + \lambda Y_{t-1} + (\varepsilon_t - \lambda\varepsilon_{t-1})
\end{aligned}$$

which we could estimate by OLS except that Y_{t-1} and ε_{t-1} are correlated hence a RHS variable is correlated with the error term so OLS produces inconsistent parameter estimates.

Question 3

In the linear regression model

$$\boldsymbol{Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t;\ t = 1, 2, \ldots, T.}$$

Prove that

(a) $\boldsymbol{\sum_{t=1}^{T} \hat{u}_t = 0}$**, where** $\boldsymbol{\hat{u}_t}$ **is the residual defined as** $\boldsymbol{\hat{u}_t = Y_t - \hat{\beta}_1 - \hat{\beta}_2 X_{2t} - \hat{\beta}_3 X_{3t}}$**.** $\boldsymbol{\hat{\beta}_1}$**,** $\boldsymbol{\hat{\beta}_2}$ **and** $\boldsymbol{\hat{\beta}_3}$ **are the ordinary least squares estimators of** $\boldsymbol{\beta_1}$**,** $\boldsymbol{\beta_2}$ **and** $\boldsymbol{\beta_3}$ **respectively.**

(4 marks)

(b) $\boldsymbol{\sum_{t=1}^{T} X_{2t}\hat{u}_t = 0 = \sum_{t=1}^{T} X_{3t}\hat{u}_t}$**.**

(4 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fourth edition) Chapter 3.2 (Derivation and interpretation of the multiple regression coefficients).

Gujarati, D.N. *Basic econometrics* (fifth edition) Chapter 7A.1 (Derivation of OLS estimators).

Approaching the question

Differentiate the residual sum of squares with respect to $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\beta}_3$ and equate them to zero to get the result. Answer is:

To apply OLS we minimise the sum of squares of errors:

$$I = \sum_{t=1}^{T} \hat{u}_t^2 = \sum_{t=1}^{T} [Y_t - \hat{\beta}_1 - \hat{\beta}_2 X_{2t} - \hat{\beta}_3 X_{3t}]^2$$

by differentiating with respect to $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\beta}_3$.

The resulting equations are:

(a)

$$\frac{\partial I}{\partial \hat{\beta}_0} = 2\sum_{t=1}^{T} \left(Y_t - \hat{\beta}_1 - \hat{\beta}_2 X_{2t} - \hat{\beta}_3 X_{3t}\right)(-1) = -2\sum_{t=1}^{T} \hat{u}_t = 0 \ \Rightarrow\ \sum \hat{u}_t = 0.$$

(b)

$$\frac{\partial I}{\partial \hat{\beta}_2} = 2\sum_{t=1}^{T} \left(Y_t - \hat{\beta}_1 - \hat{\beta}_2 X_{2t} - \hat{\beta}_3 X_{3t}\right)(-X_{2t}) = -2\sum_{t=1}^{T} \hat{u}_t X_{2t} = 0$$

and

$$\frac{\partial I}{\partial \hat{\beta}_3} = 2\sum_{t=1}^{T} \left(Y_t - \hat{\beta}_1 - \hat{\beta}_2 X_{2t} - \hat{\beta}_3 X_{3t}\right)(-X_{3t}) = -2\sum_{t=1}^{T} \hat{u}_t X_{3t} = 0$$

which gives the two results.

Question 4

Let X be a random variable with probability density function

$$f(X) = (2\pi\sigma^2)^{-1/2} \exp\left[-\frac{(X-\mu)^2}{2\sigma^2}\right].$$

Let $X_1, X_2, \ldots, X_T$ be an identically and independently distributed random sample from the distribution of X. Derive the maximum likelihood estimators of μ and σ^2.

(8 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fourth edition) Chapter 10.6 (An introduction to maximum likelihood estimation).

Subject guide Chapter 10 (Binary choice and limited dependent variable models, and maximum likelihood estimation).

Gujarati, D.N. *Basic econometrics* (fifth edition) Chapter 4 A.1 (Maximum likelihood estimation of two variable regression model).

Approaching the question

Obtain the log of the likelihood function and maximise it with respect to μ and σ^2. Answer is:

The log of the likelihood function is:

$$\ln L(\mu, \sigma) = -\frac{T}{2}\ln(2\pi) - \frac{T}{2}\ln\sigma^2 - \frac{\sum(X_t-\mu)^2}{2\sigma^2}. \qquad \text{(i)}$$

Differentiating $\ln L$ with respect to μ, we get:

$$\frac{\partial L(\mu,\sigma)}{\partial\mu} = -\frac{2\sum(X_t-\mu)}{2\sigma^2} = 0.$$

Solving we get:

$$\hat{\mu}_{MLE} = \frac{\sum X_t}{T} = \bar{X}.$$

Differentiating $\ln L$ with respect to σ^2, we get:

$$\frac{\partial L(\mu,\sigma)}{\partial\sigma^2} = -\frac{T}{2\sigma^2} + \frac{\sum(X_t-\mu)^2}{2\sigma^4} = 0.$$

Solving we obtain:

$$\hat{\sigma}^2_{MLE}(\mu) = \frac{\sum(X_t-\mu)^2}{T} \qquad \text{(ii)}$$

and:

$$\hat{\sigma}^2_{MLE} = \frac{\sum(X_t-\hat{\mu})^2}{T}$$

(after substituting μ by $\hat{\mu}$).

Question 5

Let the regression equation be:

$$\boldsymbol{Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t;\ t = 1, 2, \ldots, T.}$$

(a) Outline briefly, how you would test

i. $\mathbf{H_0 : \beta_2 = 1,}$
$\mathbf{H_1 : \beta_2 \neq 1.}$

(2 marks)

ii. $\mathbf{H_0 : \beta_2, \beta_3 = 0,}$
$\mathbf{H_1 : \beta_2, \beta_3 \neq 0.}$

(3 marks)

(b) Specify the assumptions required for these tests to be valid.

(3 marks)

Reading for this question

Dougherty, C. Introduction to econometrics (fourth edition) Chapter 2.6 (Testing hypothesis relating to the regression coefficients).

Gujarati, D.N. Basic econometrics (fifth edition)Chapters 5.7 (Hypothesis testing: the test of significance approach) and 8.4 (Testing the overall significance of the sample regression).

Approaching the question

It is required to specify the t statistic for a (i) and the F statistic for a (ii). Degrees of freedom in both the cases must be specified. For part (b), explicitly state the GM assumptions. Simply writing that the G-M assumptions are required is not a valid answer. Detailed answer is:

(a) i. This is a standard two-tail t test of the form:

$$t = \frac{\hat{\beta}_2 - 1}{\text{se}_{\hat{\beta}_2}}.$$

The degrees of freedom for the t test are $T - 3$.

ii. The test of the null hypothesis $\text{H}_0 : \beta_2 = \beta_3 = 0$ against the alternative hypothesis $\text{H}_1 : \beta_2 \neq 0$ and/or $\beta_3 \neq 0$ is achieved by applying the F test where:

$$F = \frac{R^2/(k-1)}{(1-R^2)/(T-k)}$$

where T is the sample size and k is the number of parameters in the regression, i.e. 3. R^2 is the coefficient of determination. The F test has $2, T - 3$ degrees of freedom.

(b) Assumptions required are that disturbance term should not be autocorrelated and should be homoscedastic otherwise these tests won't be valid. Disturbance term should be normally distributed or the sample size should be large.

Section B

Answer **THREE** questions from this section.

Question 6

(a) Outline the Goldfeld-Quandt test for heteroscedasticity and explain when this is a sensible test to use.

(8 marks)

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

(12 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fourth edition) Chapters 7.2 (Detection of heteroscedasticity) and 11.3 (The adaptive expectation model).

Subject guide Chapters 7 (Heteroscedasticity) and 12 (Properties of regression models with time series data).

Gujarati, D.N. *Basic econometrics* (fifth edition) Chapters 11.5 (Detection of heteroscedasticity) and 17.5 (Rationalization of the Koyck model: The adaptive expectation model).

Approaching the question

For part (a) description of Goldfeld-Quandt test is required. Students should specify which distribution it follows and also must specify the degrees of freedom. For part (b) adaptive expectation model should be derived with emphasis on short run and long run relationship between the dependent variable and the independent variable. Answer is:

(a) This test assumes that σ^2 = variance of the disturbance term, is proportional to the size of one of the RHS variables (say X_t). The observations are ranked by X and run separate regression for the first n_1 ($< n/2$) and the last n_1 observations – the middle ($n - 2n_1$) observations are not used. If heteroskedasticity is present the RSS from the two regressions will differ. Form the test $\text{RSS}_2/\text{RSS}_1$ where RSS_1 is the residual sum of squares from the first n_1 observations and RSS_2 is the residual sum of squares from the last n_1 observations. The test statistic will have an F distribution with $(n_1 - k, n_1 - k)$ degrees of freedom where k is the number of parameters in the equation. If the calculated F is greater than the critical value of F then reject the null of homoscedasticity.

(b) An adaptive expectation model involves a learning process in which, in each time period, the actual value of the variable is compared with the value that has been expected. If the actual value is greater, the expected value is adjusted upwards for the next period. If it is lower, the expected value is adjusted downwards. The size of the adjustment is hypothesised to be proportional to the discrepancy between the actual and expected value.

If X is the variable in question, and X_t^e is the value expected in time period t given the information available at time period $t - 1$, then:

$$X_{t+1}^e - X_t^e = \lambda(X_t - X_t^e); \quad 0 \leq \lambda \leq 1 \quad [1]$$

or:

$$X_{t+1}^e = \lambda X_t + (1 - \lambda)X_t^e \qquad [2].$$

The model is derived to capture the changing nature of expectations formation, often in variables that are also changing with time. It is an attempt at a 'simple learning' solution to model building in order to forecast often macroeconomic variables, such variables include investment, savings and demand for assets.

The model is estimated by repeated substitution for the expected variable, by its lagged variant which has known components of the previous period and the unobserved expectation lagged, until the term on the unobserved expectation $(1 - \lambda)s$ is so small as to be ignored resulting in a model with all the variables are observed. Where s is the period lagged and λ is the speed of adjustment of expected and actual and λ is between 0 and 1.

The long term relationship is measured by putting in equilibrium values for all variables to find the relationship between the equilibrium values of X and Y. The procedure is described in Dougherty (Section 11.3, p. 333).

Question 7

(a) Explain what is meant by a trend stationary series and a difference stationary series. Why is it important to differentiate between the two types of stationarity?

(5 marks)

(b) Let the model be

$$\boldsymbol{Y_t = \theta Y_{t-1} + u_t;\ t = 1, 2, \ldots, T}$$

where

$\boldsymbol{\mathrm{E}(u_t) = 0;\ \mathrm{E}(u_t^2) = \sigma^2}$ **and** $\boldsymbol{\mathrm{E}(u_s u_t) = 0}$ **if** $\boldsymbol{s \neq t}$ **for all** $\boldsymbol{s, t = 1, 2, \ldots, T}$ **and** $\boldsymbol{Y_0 = 0}$**.**

Explain how you would test the $\boldsymbol{\mathrm{H}_0 : |\theta| = 1}$**.**

(5 marks)

(c) Consider an ADL(1, 1) model

$$\boldsymbol{Y_t = \alpha_1 + \alpha_2 Y_{t-1} + \alpha_3 X_t + \alpha_4 X_{t-1} + u_t}$$

where both $\boldsymbol{Y_t}$ **and** $\boldsymbol{X_t}$ **are I(1).** $\boldsymbol{u_t}$ **is the disturbance term assumed to satisfy the usual regression model conditions. Express the ADL(1, 1) model in an error correction form and interpret the coefficients of the error correction model. Discuss the advantages of the error correction form compared to the ADL form.**

(10 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fourth edition) Chapter (Introduction to nonstationarity time series).

Subject guide Chapter 13 (Introduction to nonstationary time series).

Gujarati, D.N. *Basic econometrics* (fifth edition) Chapters 21.5 (Trend stationary (TS) and difference stationary (DS) stochastic process), 21.9 (The unit root test) and 21.11 (Cointegration: Regression of a unit root time series on another unit root time series).

Approaching the question

For part (a) trend stationary and difference stationary series should be explained with an example and consequences of the two types of stationarity should be discussed. Part (b) is based on Dickey-Fuller test and part (c) requires a good concept of the error correction model. Detailed answer is:

(a) If after removing the trend from a nonstationary series the resulting variables becomes stationary, then the variable is called *trend stationary.* Let

$$Z_t = X_t - \alpha_1 t = \alpha_0 + u_t$$

where $\mathrm{E}(u_t) = 0$; $\mathrm{var}(u_t) = \sigma^2$ and $\mathrm{E}(u_t u_{t-s}) = 0$ for all s and t. Then:

$$\begin{aligned}
\mathrm{E}(Z_t) &= \mathrm{E}(\alpha_0 + u_t) = \alpha_0 \\
\mathrm{var}(Z_t) &= \mathrm{var}(\alpha_0 + u_t) = \sigma^2 \\
\mathrm{cov}(Z_t, Z_{t-s}) &= \mathrm{E}[Z_t - \mathrm{E}(Z_t)][Z_{t-s} - \mathrm{E}(Z_{t-s})] = \mathrm{E}(u_t u_{t-s}) = 0.
\end{aligned}$$

This means that Z_t has constant mean and variance for all t, and covariance is zero for all s. It implies that the series is trend-stationary.

If a nonstationary process can be transformed into stationary process by differencing then the series is said to be *difference-stationary.*

Let X_t be a random walk with a drift:

$$X_t = \beta_0 + X_{t-1} + \varepsilon_t \qquad \text{(ii)}$$

where $\mathrm{E}(\varepsilon_t) = 0$; $\mathrm{var}(\varepsilon_t) = \sigma^2$ and $\mathrm{E}(\varepsilon_t \varepsilon_s) = 0$ for all s and t, $s \neq t$.

Subtract X_{t-1} from both sides of (ii) to get:

$$\Delta X_t = X_t - X_{t-1} = \beta_0 + \varepsilon_t.$$

It can be easily checked that $\mathrm{E}(\Delta X_t) = \beta_0$; $\mathrm{var}(\Delta X_t) = \sigma_\varepsilon^2$ and $\mathrm{cov}(\Delta X_t, \Delta X_{t-s}) = 0$ for all s and t. This means that ΔX_t is stationary. This implies that X_t is difference stationary.

It is important to know whether a variable is difference or trend stationary because for difference stationary variables shocks have a permanent effect whereas for trend stationary variables shocks are transitory.

(b) The standard test for a unit root is due to Dickey and Fuller. The model is:

$$Y_t = \theta Y_{t-1} + u_t$$

which can be re-written as:

$$\Delta Y_t = (1 - \theta) Y_{t-1} + u_t$$

where $\Delta Y_t = Y_t - Y_{t-1}$. The hypothesis to be tested is $\mathrm{H}_0 : 1 - \theta = 0$ vs. $\mathrm{H}_1 : 1 - \theta \neq 0$. Under the alternative hypothesis the process is stationary. We cannot use the standard t test procedure in this case because the distribution of the t statistic is not a t distribution so critical values have been computed by Dickey and Fuller using Monte-Carlo techniques. The test is sensitive to the presence of serial correlation in the error term so we need to take steps to remove the effects of this serial correlation, this is done by including lagged values of Y_t in the regression, i.e.:

$$Y_t = \theta Y_{t-1} + \theta_1 Y_{t-2} + u_t$$

for an AR(1) serial correlation. This is more easily tested by using the model:

$$Y_t = (1 - \theta - \theta_1) Y_{t-1} + \theta_1 \Delta Y_{t-2} + u_t$$

with null hypothesis $\mathrm{H}_0 : 1 - \theta - \theta_1 = 0$, using Dickey-Fuller tables.

(c) Consider a simple ADL(1,1) [This is also known as ARDL(1, 1)] model

$$Y_t = \alpha_1 + \alpha_2 Y_{t-1} + \alpha_3 X_t + \alpha_4 X_{t-1} + u_t. \qquad \text{(i)}$$

Rewrite (i) as:

$$\begin{aligned}
Y_t - Y_{t-1} &= \alpha_1 + \alpha_2 Y_{t-1} - Y_{t-1} + \alpha_3 X_t - \alpha_3 X_{t-1} + \alpha_3 X_{t-1} + \alpha_4 X_{t-1} + u_t \\
\Delta Y_t &= \alpha_1 - (1 - \alpha_2) Y_{t-1} + \alpha_3 \Delta X_t + (\alpha_3 + \alpha_4) X_{t-1} + u_t \\
\Delta Y_t &= \alpha_3 \Delta X_t - (1 - \alpha_2) \left[Y_{t-1} - \frac{\alpha_1}{(1 - \alpha_2)} - \frac{(\alpha_3 + \alpha_4)}{(1 - \alpha_2)} X_{t-1} \right] + u_t \\
\Delta Y_t &= \alpha_3 \Delta X_t - (1 - \alpha_2)[Y_{t-1} - \beta_1 - \beta_2 X_{t-1}] + u_t
\end{aligned}$$

or:

$$\Delta Y_t = \alpha_3 \Delta X_t - \pi [Y_{t-1} - \beta_1 - \beta_2 X_{t-1}] + u_t \qquad \text{(ii)}$$

where:

$$\pi = (1 - \alpha_2); \quad \beta_1 = \frac{\alpha_1}{(1 - \alpha_2)} \quad \text{and} \quad \beta_2 = \frac{(\alpha_3 + \alpha_4)}{(1 - \alpha_2)}.$$

Equation (ii) is the ECM.

When the two variables Y and X are cointegrated the ECM incorporates not only the short-run but also log-run effects. The long run equilibrium:

$$Y_{t-1} - \beta_1 - \beta_2 X_{t-1}$$

is included in the model together with the short-run effect captured by the differenced term.

All the terms in the ECM, given by (ii), are stationary. As Y and X are I(1), then ΔX and ΔY are I(0). As Y and X are cointegrated their linear combination:

$$u_{t-1} = Y_{t-1} - \beta_1 - \beta_2 X_{t-1} \sim I(0).$$

The coefficient π provides us with the information about the speed of adjustment in cases of disequilibrium:

- If $\pi = 1$ then 100% of the adjustment takes place within the period. In other words adjustment is instantaneous and full.
- If $\pi = 0.5$ then 50% adjustment takes place each period.
- If $\pi = 0$ then there is no adjustment.

Question 8

(a) Consider a model

$$Y_{it} = \beta_1 + \sum_{j=2}^{K} \beta_j X_{jit} + \alpha_i + u_{it};\ i = 1, 2, \ldots, n;\ t = 1, 2, \ldots, T$$

where Y is the dependent variable, X_j, $j = 2, \ldots, K$, are observed explanatory variables, α_i is an unobserved fixed effect and u_{it} is the disturbance term assumed to satisfy the usual regression model conditions. The index i refers to cross-section and the index t refers to the time period.

In the context of the model above briefly explain the least squares dummy variables method of estimation and its drawback.

(8 marks)

(b) The following estimates were obtained using a US data panel of 550 individuals' wages over a period of 7 years. The dependent variable is the log of wage.

Dependent variable: log of wage

Independent variables	Pooled OLS	Random Effects	Fixed Effects
Years of full time education	0.091 (0.005)	0.092 (0.011)	
Black	−0.139 (0.024)	−0.139 (0.048)	
Hispanic	0.016 (0.021)	0.022 (0.043)	
Work experience	0.067 (0.014)	0.106 (0.015)	
Experience squared	−0.0024 (0.0008)	−0.0047 (0.0007)	−0.0052 (0.0007)
Married	0.108 (0.016)	0.064 (0.017)	0.047 (0.018)
Union membership	0.182 (0.017)	0.106 (0.018)	0.080 (0.019)

Six year dummy variables and a constant term were included in all three equations but the results are not reported. Black, Hispanic, Married and Union membership are dummy variables which take the value of one if the respondent has the relevant characteristic and are zero otherwise. The numbers in parentheses are standard errors of the coefficient estimates.

i. Explain why there are no fixed effects estimated coefficients for the first four explanatory variables in the table.

(3 marks)

ii. What interpretation would you give to the unobserved effects in a wage equation of this kind?

(3 marks)

iii. The OLS coefficients for Union and Married are higher than for the other two estimates. What does this suggest about the correlation between being married and being a member of a union and the unobserved effects? Explain your answer.

(3 marks)

iv. **If there is a significant correlation between these two explanatory variables and the unobserved effects, what does this indicate about the properties of these estimates?**

(3 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fourth edition) Chapter 14 (Introduction to panel data models).

Subject guide, Chapter 14 (Introduction to panel data).

Gujarati, D.N. *Basic econometrics* (fifth edition) Chapter 16.3 (Estimation of panel data regression models).

Approaching the question

Part (a) is a discussion of least squares dummy variable method of estimation. Part (b) requires a good concept of fixed effect and random effect estimation methods. Answer is:

(a) In this approach the unobserved effect is brought explicitly in the model. A set of dummy variables D_i is defined, where D_i is equal to 1 in the case of an observation relating to an individual i and 0 otherwise. The model can be written as:

$$Y_{it} = \sum_{j=2}^{K} \beta_j X_{jit} + \sum_{i=1}^{n} \alpha_i D_i + u_{it}. \qquad \text{(vi)}$$

The unobserved effect is now being treated as the coefficient of the individual specific dummy variable. The term $\alpha_i D_i$ represents a fixed effect on the dependent variable Y_i for individual i.

If we want to keep the intercept in the model then instead of n dummy variables $(n-1)$ dummy variables has to be used, otherwise we will fall into the dummy variable trap.

It can be shown that LSDV method is identical to the within groups method.

Drawbacks:

- The intercept and any explanatory variable that remain constant for each individual will drop out of the model.
- The variation in $(X_{ij} - \bar{X}_i)$ may be much smaller than the variation in X_j. If this is the case, the impact of the disturbance term may be relatively large, giving rise to imprecise estimates.
- There is loss of substantial number of degrees of freedom.

(b) i. The first three variables (education, black and Hispanic) do not vary over time. Thus they will be perfectly collinear with the fixed effects and their coefficients cannot be estimated. The work experience variable in most cases will increase by one each year and will thus be collinear with the time dummies. (Note this will not apply to the quadratic term)

ii. Unobserved effects in this kind of equation are usually thought to represent talent, ability, capacity to work hard etc. If so, random effects models may not give consistent estimates

iii. The OLS estimates make no allowance for the unobserved effects. Thus the higher coefficients for union and Married suggest that these variables are picking up some of the unobserved effects. The OLS estimates probably suffer from omitted variable bias as a result of the omission of the unobserved effects. If the unobserved effects (ability) are positively correlated with earnings, this suggests that Union and Marries are positively correlated with the unobserved effects (as their coefficients appear to be biased upwards in the OLS estimates).

(iv) It suggests that the OLS estimates are biased and inconsistent (omitted variable bias). It suggests that the Random Effects estimates are also inconsistent (they are biased anyway) as the random effects are correlated with explanatory variables.

Question 9

(a) Consider a model

$$Y_i = \beta_0 + \beta_1 X_i + u_i; \ i = 1, 2, \dots, n$$

where

$$\begin{aligned} Y_i &= 1 \text{ if the event occurs} \\ &= 0 \text{ otherwise} \end{aligned}$$

and $E(u_i) = 0$, $\mathrm{Var}(u_i) = \sigma^2$ for $i = 1, 2, \dots, n$.

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

(10 marks)

(b) A researcher wants to examine the determinants of household decisions to buy alcohol. For this purpose she defines

$$\begin{aligned} Y &= 1 \text{ if the household purchased alcohol} \\ &= 0 \text{ otherwise} \end{aligned}$$

A random sample of households is available and the following logit estimates of the coefficients of variables were obtained:

Dependent variable: Y

Independent variables	Estimated Coefficients	Estimated Asymptotic Standard Errors
Income of the household (000's Pounds)	158.39	48.86
Square of income	−76.00	25.14
Number of adults in the household	1.16	0.23
Number of children in the household	0.49	0.12
1 if no worker in the household, 0 otherwise	−0.11	0.28
1 if head of the household is male, 0 otherwise	0.93	0.26
Constant	−83.94	23.77

$$\log L = -321.25 \qquad \log L_0 = -416.01$$

$\log L$ and $\log L_0$ are the log of the likelihood from the unrestricted model and the log of the likelihood of the model where all the slope coefficients are restricted to zero, respectively.

i. Explain how the logit estimates were calculated.

(5 marks)

ii. Test the hypothesis that all the slope coefficients are jointly equal to zero.

(5 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fourth edition) Chapters 10.1 (The linear probability model), 10.2 (Logit analysis) and 10.6 (An introduction to maximum likelihood estimation).

Subject guide Chapter 10 (Binary choice and limited dependent variable models, and maximum likelihood estimation).

Gujarati, D.N. *Basic econometrics* (fifth edition) Chapters 15.2 (The linear probability model (LPM) and 15.5 (The logit model).

Approaching the question

For part (a) the linear probability model should be discussed in detail. Part (b) requires a brief discussion of the logit model and its estimation by the maximum likelihood method. Part (c) is based upon the likelihood ratio test. Answer is:

(a) The model is:

$$Y_i = \beta_0 + \beta_1 X_i + u_i; \; i = 1, 2, \ldots, n \qquad \text{(i)}$$

where:

$$Y_i = \begin{cases} 1 \text{ if event occurs} \\ 0 \text{ if not.} \end{cases}$$

Assume $\mathrm{E}(u_i) = 0$, then:

$$\mathrm{E}[Y_i \mid X_i] = \beta_0 + \beta_1 X_i. \qquad \text{(ii)}$$

Also:

$$\mathrm{E}[Y_i \mid X_i] = 1 \cdot P(Y_i = 1) + 0 \cdot P(Y_i = 0) = P(Y_i = 1) = P_i \qquad \text{(iii)}$$

From (ii) and (iii):

$$\mathrm{E}[Y_i \mid X_i] = \beta_0 + \beta_1 X_i = P_i$$

hence, we can interpret $\mathrm{E}[Y_i \mid X_i] = \beta_0 + \beta_1 X_i$ as the probability that the event will occur given X_i.

As Y_i takes only two values 1 or 0, therefore u_i can take only two values $1 - \beta_0 - \beta_1 X_i$ when $Y_i = 1$ and $\beta_0 - \beta_1 X_i$ when $Y_i = 0$. Based on this we can write the probability distribution of u_i as:

Y_i	u_i	$f(u_i)$
1	$1 - \beta_0 - \beta_1 X_i$	$\beta_0 + \beta_1 X_i$
0	$-\beta_0 - \beta_1 X_i$	$1 - \beta_0 - \beta_1 X_i$

We can write $\mathrm{var}(u_i)$ as:

$$\begin{aligned}
\mathrm{var}(u_i) &= \mathrm{E}(u_i^2) \\
&= (1 - \beta_0 - \beta_1 X_i)^2 (\beta_0 + \beta_1 X_i) + (-\beta_0 - \beta_1 X_i)^2 (1 - \beta_0 - \beta_1 X_i) \\
&= (1 - \beta_0 - \beta_1 X_i)(\beta_0 + \beta_1 X_i)[(1 - \beta_0 - \beta_1 X_i) + (\beta_0 + \beta_1 X_i)] \\
&= (\beta_0 + \beta_1 X_i)(1 - \beta_0 - \beta_1 X_i) \\
&= \mathrm{E}(Y_i)[1 - \mathrm{E}(Y_i)] \\
&= P_i(1 - P_i); \; \text{for all } i = 1, 2, \ldots, n.
\end{aligned}$$

Hence the disturbance term is heteroscedastic. This will make OLS estimators inefficient. Weighted least squares can be used to obtain efficient estimates of β_0 and β_1.

Problems:

- As the distribution of the disturbance term only takes two values, it is not continuous. This implies that usual test statistics are invalidated.
- In many cases the estimated probability $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$ will be negative or greater than 1.

(b) i. Logit model uses the cumulative standardised logistic distribution. Maximum likelihood technique is used to obtain the estimates of the parameters. Estimates have the standard maximum likelihood properties i.e. the estimators are consistent, asymptotically efficient and asymptotically normally distributed. [For technical details see Dougherty (fourth edition) Chapter10.3].

ii. As the model has been estimated by maximum likelihood, an F test cannot be used. In these situations likelihood ratio test is used. The likelihood ratio test statistic is:

$$LR = 2(\ln L - \ln L_0) = 2(-321.25 - (-416.01)) = 189.52.$$

Critical value of χ^2_6 at 5% level of significance is = 12.592.
Critical value of χ^2_6 at 1% level of significance is = 16.812.
Hence we reject H_0.

Question 10

The following ordinary least squares (OLS) estimates were obtained of the demand for labour in manufacturing using UK quarterly, seasonally adjusted data. The dependent variable is the change in unemployment, Δemp_t.

Dependent Variabel: Δemp_t

Independent variables	Estimated coefficients	Estimated coefficients	Estimated coefficients
Δy_t	0.0823 (5.24)	0.0779 (4.72)	0.1503 (7.95)
y_{t-1}	−0.0972 (9.52)	0.1002 (7.26)	0.1864 (11.16)
Δsemp_t	−0.8219 (9.40)	−0.5041 (4.11)	−0.8000 (6.57)
semp_t	−0.2176 (8.70)	−0.2327 (4.48)	−0.3988 (11.90)
emp_{t-1}	−0.2173 (8.80)	−0.2532 (4.37)	−0.3293 (12.81)
const	0.5534 (7.25)	0.7123 (3.08)	0.6603 (10.82)
R^2	0.89	0.93	0.96
S_u	0.0028	0.0024	0.0016
DW	0.58	1.01	1.39
Sample	1980:3–2001:3	1980:3–1989:4	1990:1–2001:3
N	85	38	47

Where emp_t is the logarithm of employment in manufacturing, y_t is the logarithm of manufacturing output, semp_t is the logarithm of difference between total employment and manufacturing employment, Δ indicates the first difference operator e.g., $\Delta x_t = x_t - x_{t-1}$. t values are in parentheses and N is the sample size.

(a) In the regression results above explain what is meant by R^2, S_u and DW. What, if anything, can be inferred from these statistics?

(6 marks)

(b) Test the hypothesis that the structure of the model has not changed over the two periods for which data is available. What assumptions does your test require? Are they likely to be true in this case? Explain.

(8 marks)

(c) Evaluate these estimates of the demand for labour. Do the regression results suggest there are problems with the model? If so, how would you rectify the problems?

(6 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fourth edition) Chapters 1.6(Goodness of fit:), 2.5 (Precision of the regression coefficients), 5.4 (The Chow test) and 12 (Detection and definition of Autocorrelation).

Subject guide Chapter 3 (Multiple regression analysis).

Gujarati, D.N. *Basic econometrics* (fifth edition) Chapters 3.3 (Precision or standard error of least square estimates), 3.5 (The coefficient of determination: A measure of 'Goodness of Fit'), 8.7 (Testing for structural or parameter stability of regression models: The Chow test) and 12.6 (Detecting autocorrelation).

Approaching the question

This question is based on the precision of estimates and structural stability. Answer is:

(a) R^2 high $\Rightarrow$ there is a relationship.

S is the standard error of the disturbance. Small values $\Rightarrow$ the equation fits well (agrees with the R^2 value).

DW is the Durbin-Watson statistic which will test the hypothesis that $\rho = 0$ in the relationship:

$$u_t = \rho u_{t-1} + v_t.$$

N	Variables	Critical value	Sample value	Decision
85	5	1.53, 1.74	0.58	Reject H_0
38	5	1.20, 1.79	1.01	Reject H_0
47	5	1.30, 1.77	1.39	Inconclusive

There appears to be serial correlation in the full sample and over the first period. The second period is inconclusive.

(b) To test the stability of the function over the two periods use an F test.

$$\begin{aligned} \text{SSE}_R &= (0.0028)^2 \times 85 = 0.000666 \\ \text{SSE}_{U_1} &= (0.0024)^2 \times 38 = 0.000219 \\ \text{SSE}_{U_2} &= (0.0016)^2 \times 47 = 0.000120. \end{aligned}$$

$$F_{k,N-2k} = \frac{(\text{SSE}_R - \text{SSE}_U)/k}{\text{SSE}_U/(N-2k)} = \frac{(0.000666 - 0.000339)/6}{0.000339/(85-12)} = 11.7.$$

The critical value for $F_{k-1,N-k} = F_{5,79} = 2.33$.

Hence we reject H_0: parameters are constant over the two periods, i.e. there has been a change in structure.

Assumptions needed are:

- Disturbance term has a normal distribution.
- The model is correctly specified.
- There is no autocorrelation.
- Disturbance terms are homoscedastic.

From the above it is clear that there is strong evidence of serial correlation hence the t and F tests are invalid.

(c) The regressions show that the change in unemployment is +ve related to change in income and −ve related to the difference between total employment and manufacturing employment, i.e. the employment in service and the public sector. These results are as expected. There are also links with lagged and differenced terms but it is not possible to specify the direction of the relationship in general for these variables.

As pointed out above there is the problem of serial correlation which can be addressed by the use of Cochrane-Orcutt or other non-linear estimators.

Examiners' commentaries 2015

EC2020 Elements of econometrics

Important note

This commentary reflects the examination and assessment arrangements for this course in the academic year 2014–15. The format and structure of the examination may change in future years, and any such changes will be publicised on the virtual learning environment (VLE).

Information about the subject guide and the Essential reading references

Unless otherwise stated, all cross-references will be to the latest version of the subject guide (2014). You should always attempt to use the most recent edition of any Essential reading textbook, even if the commentary and/or online reading list and/or subject guide refer to an earlier edition. If different editions of Essential reading are listed, please check the VLE for reading supplements – if none are available, please use the contents list and index of the new edition to find the relevant section.

General remarks

Learning outcomes

At the end of this course, and having completed the Essential reading and activities, you should be able to:

- describe and apply the classical regression model and its application to cross-section data
- describe and apply the:
  - Gauss–Markov conditions and other assumptions required in the application of the classical regression model
  - reasons for expecting violations of these assumptions in certain circumstances
  - tests for violations
  - potential remedial measures, including, where appropriate, the use of instrumental variables
- recognise and apply the advantages of logit, probit and similar models over regression analysis when fitting binary choice models
- competently use regression, logit and probit analysis to quantify economic relationships using standard regression programmes (Stata and EViews) in simple applications
- describe and explain the principles underlying the use of maximum likelihood estimation
- apply regression analysis to fit time-series models using stationary time series, with awareness of some of the econometric problems specific to time series applications (for example, autocorrelation) and remedial measures
- recognise the difficulties that arise in the application of regression analysis to nonstationary time series, know how to test for unit roots, and know what is meant by cointegration.

Common mistakes committed by candidates

- A large number of candidates were not able to distinguish between sample variance and covariance, and population variance and covariance (this is happening year after year). They treat them as the same. This results in incorrect analysis and candidates lose significant marks.

  Consider an example: Suppose data are deviations from the respective sample means and the regression model is:

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

  The ordinary least squares estimator of β is:

  $$\widehat{\beta} = \frac{\sum_{t=1}^{T} x_t y_t}{\sum_{t=1}^{T} x_t^2} = \beta + \frac{\sum_{t=1}^{T} x_t u_t}{\sum_{t=1}^{T} x_t^2}.$$

  In terms of variances and covariances (a large number of candidates prefer this terminology), this can be written as:

  $$\widehat{\beta} = \beta + \frac{\operatorname{Cov}(x, u)}{\operatorname{Var}(x)}.$$

  Here $\operatorname{Cov}(x, u)$ and $\operatorname{Var}(x)$ are sample[$\operatorname{Cov}(x, u)$] and sample[$\operatorname{Var}(x)$].

  Candidates should realise that $\sum_{t=1}^{T} u_t$, $\sum_{t=1}^{T} x_t u_t$, $\operatorname{Cov}(x, u)$ and $\operatorname{Var}(x)$ given above are sample moments and as such $\sum_{t=1}^{T} u_t \neq 0$, $\sum_{t=1}^{T} x_t u_t \neq 0$ and $\operatorname{Cov}(x, u) \neq 0$. However, if we take the expectation, then:

  $$\operatorname{E}(u_t) = 0$$

  by assumption. Then:

  $$\operatorname{E}\left[\sum_{t=1}^{T} x_t u_t\right] = \sum_{t=1}^{T} x_t[\operatorname{E}(u_t)] = 0$$

  as the x_ts are fixed so they can be taken out of the expectation, and so:

  $$\operatorname{E}[\operatorname{Cov}(x, u)] = \operatorname{E}\left[\frac{1}{T}\sum_{t=1}^{T} x_t u_t\right] = 0$$

  as previously argued. This makes $\operatorname{E}(\widehat{\beta}) = \beta$, i.e. $\widehat{\beta}$ is an unbiased estimator for β.

  To prove consistency take the plim to get:

  $$\begin{aligned}
  \operatorname{plim}\left(\widehat{\beta}\right) &= \beta + \operatorname{plim}\left(\frac{\frac{1}{T}\sum_{t=1}^{T} x_t u_t}{\frac{1}{T}\sum_{t=1}^{T} x_t^2}\right) \\
  &= \beta + \frac{\operatorname{plim}\left(\frac{1}{T}\sum_{t=1}^{T} x_t u_t\right)}{\operatorname{plim}\left(\frac{1}{T}\sum_{t=1}^{T} x_t^2\right)} \\
  &= \beta + \frac{\operatorname{plim}(\text{sample } \operatorname{Cov}(x, u))}{\operatorname{plim}(\text{sample } \operatorname{Var}(x))} \\
  &= \beta + \frac{\text{population } \operatorname{Cov}(x, u)}{\text{population } \operatorname{Var}(x)}.
  \end{aligned}$$

By assumption, population $\text{Cov}(x, u) = 0$ and population $\text{Var}(x) > 0$, hence $\text{plim}\left(\widehat{\beta}\right) = \beta$, in other words $\widehat{\beta}$ is a consistent estimator of β.

Remember that in general:

$$\text{plim(sample variance)} = \text{population variance}$$

and:

$$\text{plim(sample covariance)} = \text{population covariance.}$$

This concept has been used in many questions. This simple mistake of not distinguishing between sample variance and covariance, and population variance and covariance, results in a significant loss of marks which might result in the loss of a degree class or even be the difference between pass and fail.

- Candidates struggled to give competent answers to the interpretation of empirical results. When interpreting an empirical result you should discuss the significance of the coefficients, magnitude and sign of the coefficients. Also, you should make sure that the Gauss–Markov conditions hold.
- Just as last year, many candidates did not appear to read the questions carefully enough and often omitted to give answers to parts of questions which asked for details of such things as the assumptions necessary for a particular result to be true.

Key steps to improvement

Essential reading for this course includes the subject guide and the following.

> Dougherty, C. *Introduction to econometrics.* (Oxford: Oxford University Press, 2011) 4th edition [ISBN 9780199567089];
> http://global.oup.com/uk/orc/busecon/economics/dougherty4e/

Apart from Essential readings you should do some supplementary readings. One very good book of the same level is:

> Gujarati, D.N. and D.C. Porter *Basic econometrics.* (McGraw–Hill, 2009, International edition) 5th edition [ISBN 9780071276252].

To understand the subject clearly it is important to supplement C. Dougherty, *Introduction to econometrics* (fourth edition) with the subject guide **EC2020 Elements of econometrics** (2014), especially Chapter 10 which covers maximum likelihood.

It is very important to carefully go through the subject guide. The subject guide contains solutions to the questions given in the main textbook and also some additional questions and solutions. Working through these will improve your understanding of the subject.

The chapter in the subject guide on maximum likelihood (Chapter 10) includes some additional theory which has not been covered in the main textbook. It is important to read the additional theory given in the subject guide to have a better understanding of the principles of maximum likelihood and tests based on the likelihood function.

Please check the VLE course page for resources for this subject such as a downloadable copy of the subject guide **EC2020 Elements of econometrics** (2014), PowerPoint slideshows that provide graphical treatment of the topics covered in the textbook, datasets and statistical tables. Candidates should utilise datasets using standard regression programmes (STATA or EViews). This will help in the understanding of the subject.

Examination revision strategy

Many candidates are disappointed to find that their examination performance is poorer than they expected. This may be due to a number of reasons. The *Examiners' commentaries* suggest ways of addressing common problems and improving your performance. One particular failing is '**question spotting**', that is, confining your examination preparation to a few questions and/or topics which have come up in past papers for the course. This can have serious consequences.

We recognise that candidates may not cover all topics in the syllabus in the same depth, but you need to be aware that examiners are free to set questions on **any aspect** of the syllabus. This means that you need to study enough of the syllabus to enable you to answer the required number of examination questions.

The syllabus can be found in the Course information sheet in the section of the VLE dedicated to each course. You should read the syllabus carefully and ensure that you cover sufficient material in preparation for the examination. Examiners will vary the topics and questions from year to year and may well set questions that have not appeared in past papers. Examination papers may legitimately include questions on any topic in the syllabus. So, although past papers can be helpful during your revision, you cannot assume that topics or specific questions that have come up in past examinations will occur again.

If you rely on a question-spotting strategy, it is likely you will find yourself in difficulties when you sit the examination. We strongly advise you not to adopt this strategy.

Examiners' commentaries 2015

EC2020 Elements of econometrics

Important note

This commentary reflects the examination and assessment arrangements for this course in the academic year 2014–15. The format and structure of the examination may change in future years, and any such changes will be publicised on the virtual learning environment (VLE).

Information about the subject guide and the Essential reading references

Unless otherwise stated, all cross-references will be to the latest version of the subject guide (2014). You should always attempt to use the most recent edition of any Essential reading textbook, even if the commentary and/or online reading list and/or subject guide refer to an earlier edition. If different editions of Essential reading are listed, please check the VLE for reading supplements – if none are available, please use the contents list and index of the new edition to find the relevant section.

Comments on specific questions – Zone A

Candidates should answer **EIGHT** of the following **TEN** questions: **ALL** of the questions in Section A (8 marks each) and **THREE** questions from Section B (20 marks each). **Candidates are strongly advised to divide their time accordingly.**

Section A

Answer all questions from this section.

Question 1

Explain the concept of consistency of an estimator. Show that in a simple regression model of Y_i on X_i, the ordinary least squares estimate of the slope is consistent.

(8 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (4th edition) Chapters R.14 (Probability limits and consistency) and 8.3 (Asymptotic properties of OLS regression estimators).

Gujarati, D.N. and D.C. Porter *Basic econometrics.* (5th edition) [ISBN 9780071276252], Chapter 3A.7 (Consistency of least-squares estimators).

Approaching the question

The definition of consistency is required, and the sufficient condition of consistency should also be given. The probability limit or sufficient condition of consistency should be used to show the consistency of the OLS estimator of the slope. The solution is as follows:

Definition:

$\widehat{\beta}$ is a consistent estimator of β if:

$$\lim_{T\to\infty} P(|\widehat{\beta}-\beta|>\varepsilon)\to 0$$

where ε is an arbitrarily small positive number. In short, plim $\widehat{\beta}=\beta$.

The sufficient condition for consistency comprises:

i. $\mathrm{E}(\widehat{\beta})=\beta$ or Asy. $\mathrm{E}(\widehat{\beta})\to\beta$.
ii. $\mathrm{Var}(\widehat{\beta})\to 0$ as $T\to\infty$, where T is the sample size.

If the sufficient condition holds, then the definition holds.

We now examine the consistency of the estimator of the slope parameter:

Let the model be:

$$Y_t=\beta_1+\beta_2X_t+u_t;\quad t=1,2,\ldots,T.$$

The OLS estimator of β_2 is:

$$\widehat{\beta}_2=\frac{\sum_{t=1}^{T}(X_t-\bar{X})(Y_t-\bar{Y})}{\sum_{t=1}^{T}(X_t-\bar{X})^2}=\beta_2+\frac{\sum_{t=1}^{T}(X_t-\bar{X})(u_t-\bar{u})}{\sum_{t=1}^{T}(X_t-\bar{X})^2}$$

and:

$$\text{plim }\widehat{\beta}_2=\beta_2+\frac{\text{plim}\frac{1}{T}\sum_{t=1}^{T}(X_t-\bar{X})(u_t-\bar{u})}{\text{plim}\frac{1}{T}\sum_{t=1}^{T}(X_t-\bar{X})^2}=\beta_2+\frac{\sigma_{Xu}}{\sigma_X^2}=\beta_2+\frac{0}{\sigma_X^2}=\beta_2\ \Rightarrow\ \text{consistent.}$$

σ_{Xu} is the population covariance between X and u, which by assumption is zero. σ_X^2 is the population variance of X, and it is >0.

Question 2

In the model

$$\boldsymbol{y_t=\alpha x_t+u_t;\quad t=1,2,\ldots,T}$$

x_t is an explanatory variable which can be regarded as fixed in repeated samples.

u_t is an unobserved disturbance for which it is assumed that

$$\begin{aligned}\mathbf{E}(u_t)&=0\\ \mathbf{E}(u_su_t)&=\sigma^2\text{ if }s=t\\ &=0\text{ if }s\neq t\end{aligned}$$

An estimator of α is $\frac{1}{T}\sum_{t=1}^{T}\left(\frac{y_t}{x_t}\right)$.

Under the assumptions above show that the estimator is unbiased and consistent. Comment briefly on the efficiency of the estimator.

(8 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics.* (4th edition) Chapters R.6 (Unbiasedness and efficiency), R.14 (Probability limits and consistency) and 8.3 (Asymptotic properties of OLS regression estimators).

Gujarati, D.N. and D.C. Porter. *Basic econometrics.* (5th edition) [ISBN 9780071276252], Chapter 3 A.7 (Consistency of least-squares estimators).

Approaching the question

The sufficient condition of consistency should be used to show the consistency of the given estimator. The solution is as follows:

We have:

$$\widehat{\alpha} = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{y_t}{x_t}\right) = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{\alpha x_t + u_t}{x_t}\right) = \alpha + \frac{1}{T}\sum_{t=1}^{T}\left(\frac{u_t}{x_t}\right).$$

Hence:

$$\mathrm{E}(\widehat{\alpha}) = \alpha + \frac{1}{T}\sum_{t=1}^{T}\frac{\mathrm{E}(u_t)}{x_t} = \alpha \ \Rightarrow \ \text{unbiased.}$$

To show consistency, the sufficient condition of consistency will be used. We have:

$$\mathrm{Var}(\widehat{\alpha}) = \mathrm{E}\left((\widehat{\alpha}-\alpha)^2\right) = \mathrm{E}\left(\frac{1}{T}\sum_{t=1}^{T}\left(\frac{u_t}{x_t}\right)\right)^2 = \sigma^2\frac{1}{T^2}\sum_{t=1}^{T}\left(\frac{1}{x_t^2}\right)$$

which will tend to zero as $T \to \infty$.

As the estimator is unbiased and also as the variance of the estimator tends to zero as $T \to \infty$, the sufficient condition of consistency holds, hence the estimator is consistent. The given estimator is not efficient because under the assumptions above the ordinary least squares estimator is the most efficient estimator.

Question 3

(a) **If a random variable X has a distribution with probability density function $f(x) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$, show that the maximum likelihood (ML) estimator of the mean (μ) of the random variable X is the sample mean.**

(4 marks)

(b) **State the statistical properties of the ML estimators.**

(4 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics.* (4th edition) Chapter 10.6 (An introduction to maximum likelihood estimation).

Dougherty, C. Subject guide, Chapter 10 (Binary choice and limited dependent variable models, and maximum likelihood estimation).

Approaching the question

(a) The log-likelihood function should be derived, and it should be differentiated with respect to μ and equated to zero to obtain the ML estimator. The solution is as follows:

The likelihood function is:

$$L = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(X_i-\mu)^2}{2\sigma^2}\right) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{\sum(X_i-\mu)^2}{2\sigma^2}\right).$$

Hence, taking logs, we obtain the log-likelihood function:

$$\ln L = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2.$$

To maximise $\ln L$, differentiate with respect to μ, and set the partial derivative equal to zero. We have:

$$\frac{\partial \ln L}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu) = 0$$

which gives:

$$\tilde{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}.$$

(b) The properties of ML estimators should be discussed as follows:

- ML estimators are consistent.
- ML estimators are invariant to the transformation of parameters. For example, if $\widehat{\theta}$ is the ML estimator of θ, then $\widehat{\theta}^2$ is the ML estimator of θ^2. Similarly, if $\widehat{\theta}$ is the ML estimator of θ, then $\exp(\widehat{\theta})$ is the ML estimator of $\exp(\theta)$.
- ML estimators are efficient in large samples in the sense that the variance of ML estimators reaches the Cramer–Rao lower bound (CRLB) in large samples.
- ML estimators are asymptotically normally distributed.
- If a sufficient estimator exists, then the ML estimator is a function of the sufficient estimator.

Question 4

Suppose that business expenditure for a new plant (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. Derive a linear relationship that can be used to estimate α and β, using the observable variables Y_t and X_t.

(8 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics.* (4th edition) Chapter 11.4 (Models with lagged dependent variable).

Gujarati, D.N. and D.C. Porter. *Basic econometrics.* (5th edition) [ISBN 9780071276252], Chapter 17.5 (Rationalization of the Koyck model: The adaptive expectations model).

Approaching the question

In order to get an estimable equation we need to eliminate the expected values from the equation. First, multiply through by $(1-\gamma)$ and lag to get a new equation, then subtract the new equation from the original equation to get the result. The solution is as follows:

We have:

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

To get an estimable equation we need to eliminate the expected values from the equation. We multiply through by $(1-\gamma)$ and lag to get:

$$(1-\gamma)\ln(Y_{t-1}) = (1-\gamma)\alpha + (1-\gamma)\beta\ln(X_{t-1}^*) + (1-\gamma)u_{t-1}.$$

Now subtract this from the original equation:

$$\begin{aligned}\ln(Y_t) - (1-\gamma)\ln(Y_{t-1}) &= \alpha\gamma + \beta(\ln(X_t^*) - (1-\gamma)\ln(X_{t-1}^*)) + (u_t - (1-\gamma)u_{t-1}) \\ &= \alpha\gamma + \beta(\gamma\ln(X_{t-1})) + (u_t - (1-\gamma)u_{t-1})\end{aligned}$$

or:

$$\ln(Y_t) = \alpha\gamma + \beta\gamma\ln(X_{t-1}) + (1-\gamma)\ln(Y_{t-1}) + (u_t - (1-\gamma)_{t-1}).$$

The parameters are estimated by non-linear techniques. If these procedures are not available, then a grid search can be used where γ is given values between 0 and 1 in steps of 0.1, and the remaining parameters are estimated using OLS.

Question 5

Discuss how dummy variables can be used to test

(a) change in intercept,

(3 marks)

(b) change in slope and,

(3 marks)

(c) changes in both intercept and slope.

(2 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics.* (4th edition) Chapters 5.1 (Illustration of the use of a dummy variable) and 5.3 (Slope dummy variables).

Dougherty, C. Subject guide, Chapter 5 (Dummy variables).

Gujarati, D.N. and D.C. Porter. *Basic econometrics.* (5th edition) [ISBN 9780071276252], Chapter 9 (Dummy variable regression models).

Approaching the question

(a) **Only intercept has changed.**

Specify the model as:

$$Y_t = \beta_0 + \beta_1 X_t + \alpha Z_t + u_t; \quad t = 1, 2, \ldots, T \qquad \text{(i)}$$

where Z_t is a dummy variable defined as:

$$Z_t = \begin{cases} 1 & \text{for war period} \\ 0 & \text{for peace period.} \end{cases}$$

Estimating (i) by OLS we get:

$$Y_t = \widehat{\beta}_0 + \widehat{\beta}_1 X_t + \widehat{\alpha} Z_t + \widehat{u}_t; \quad t = 1, 2, \ldots, T. \qquad \text{(ii)}$$

From (ii), we can write two separate regressions for two different periods as:

$$Y_t = \begin{cases} (\widehat{\beta}_0 + \widehat{\alpha}) + \widehat{\beta}_1 X_t + \widehat{u}_t & \text{(war period)} \qquad \text{(iii)} \\ \widehat{\beta}_0 + \widehat{\beta}_1 X_t + \widehat{u}_t & \text{(peace period).} \qquad \text{(iv)} \end{cases}$$

To test whether the intercept has changed or not, the hypotheses are:

$$\mathrm{H}_0 : \alpha = 0 \qquad \text{(intercept has not changed).}$$

$$\mathrm{H}_1 : \alpha \neq 0 \qquad \text{(intercept has changed).}$$

This can be tested by a t test. If we do not reject H_0, then we can apply OLS to $Y_t = \beta_0 + \beta_1 X_t + u_t$ and get the estimated parameters. If H_0 is rejected, then our estimated equations for the two different periods are given by (iii) and (iv).

(b) **Only slope has changed.**

Specify the model as:

$$Y_t = \beta_0 + \beta_1 X_t + \alpha X_t Z_t + u_t; \quad t = 1, 2, \ldots, T \qquad \text{(v)}$$

where Z_t is a dummy variable defined as:

$$Z_t = \begin{cases} 1 & \text{for war period} \\ 0 & \text{for peace period.} \end{cases}$$

Estimating (v) by OLS we get:

$$Y_t = \widehat{\beta}_0 + \widehat{\beta}_1 X_t + \widehat{\alpha} X_t Z_t + \widehat{u}_t; \quad t = 1, 2, \ldots, T. \qquad \text{(vi)}$$

From (vi), we can write two separate regressions for two different periods as:

$$Y_t = \begin{cases} \widehat{\beta}_0 + (\widehat{\beta}_1 + \widehat{\alpha}) X_t + \widehat{u}_t & \text{(war period)} \qquad \text{(vii)} \\ \widehat{\beta}_0 + \widehat{\beta}_1 X_t + \widehat{u}_t & \text{(peace period)} \qquad \text{(viii)} \end{cases}$$

To test whether the slope has changed or not, the hypotheses are:

$$\mathrm{H}_0 : \alpha = 0 \qquad \text{(slope has not changed).}$$

$$\mathrm{H}_1 : \alpha \neq 0 \qquad \text{(slope has changed).}$$

This can be tested by a t test. If we do not reject H_0, then we can apply OLS to $Y_t = \beta_0 + \beta_1 X_t + u_t$ and get the estimated parameters. If H_0 is rejected, then our estimated equations for the two different periods are given by (vii) and (viii).

(c) **Intercept and slope both have changed.**

Specify the model as:

$$Y_t = \beta_0 + \beta_1 X_t + \alpha_1 Z_t + \alpha_2 X_t Z_t + u_t; \quad t = 1, 2, \ldots, T \qquad \text{(ix)}$$

where Z_t is a dummy variable defined as:

$$Z_t = \begin{cases} 1 & \text{for war period} \\ 0 & \text{for peace period.} \end{cases}$$

Estimating (ix) by OLS we get:

$$Y_t = \widehat{\beta}_0 + \widehat{\beta}_1 X_t + \widehat{\alpha}_1 Z_t + \widehat{\alpha}_2 X_t Z_t + \widehat{u}_t; \quad t = 1, 2, \ldots, T. \qquad \text{(x)}$$

From (x), we can write two separate regressions for two different periods as:

$$Y_t = \begin{cases} (\widehat{\beta}_0 + \widehat{\alpha}_1) + (\widehat{\beta}_1 + \widehat{\alpha}_2)X_t + \widehat{u}_t & \text{(war period)} \quad \text{(xi)} \\ \widehat{\beta}_0 + \widehat{\beta}_1 X_t + \widehat{u}_t & \text{(peace period)} \quad \text{(xii)} \end{cases}$$

To test jointly whether both intercept and slope has changed or not hypotheses are:

$$\text{H}_0 : \alpha_1, \alpha_2 = 0 \qquad \text{(both intercept and slope have not changed).}$$

$$\text{H}_1 : \alpha_1, \alpha_2 \neq 0 \qquad \text{(both intercept and slope have changed).}$$

This can be tested by an F test. If we do not reject H_0, then we can apply OLS to $Y_t = \beta_0 + \beta_1 X_t + u_t$ and get the estimated parameters. If H_0 is rejected, then our estimated equations for two different periods are given by (xi) and (xii).

Section B

Answer three questions from this section.

Question 6

The Cobb–Douglas production function can be written as follows:

$$\ln Y_t = \alpha_0 + \alpha_1 \ln L_t + \alpha_2 \ln K_t + u_t; \quad t = 1, 2, \ldots, T \qquad \text{(i)}$$

where Y_t is real output, L_t is a measure of labour input, K_t is a measure of real capital input, and u_t is an unobserved random disturbance with $\text{E}(u_t) = 0$.

The following estimates of (i) were obtained by ordinary least squares (OLS) using 15 annual observations from the Taiwanese agricultural sector.

$$\begin{aligned} \ln Y_t \; &= \; \underset{(2.44)}{-3.329} + \underset{(0.54)}{1.498} \ln L_t + \underset{(0.10)}{0.489} \ln K_t + e_t \qquad \text{(ii)} \end{aligned}$$

where e_t are OLS residuals, standard errors are in parentheses and $R^2 = 0.89$.

(a) **Give an economic interpretation of the estimated coefficients. Are the estimated slope parameters of the expected sign? Explain.**

(4 marks)

(b) **Test the slope parameters for significance, and explain what assumptions your tests require in order to be valid.**

(6 marks)

(c) **If the Taiwanese agricultural sector has constant returns to scale then $\alpha_1 + \alpha_2 = 1$. Discuss whether the estimates in (ii) support this restriction.**

(3 marks)

(d) **The equation was also estimated in the following restricted form**

$$[\ln Y_t - \ln L_t] \; = \; \underset{(0.42)}{1.712} + \underset{(0.09)}{0.612}[\ln K_t - \ln L_t] + \nu_t$$

where ν_t are OLS residuals, standard errors are in parentheses and $R^2 = 0.77$.

Test the restriction(s) in (iii), and show that (iii) incorporates the restriction of constant returns to scale.

(7 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics.* (4th edition) Chapters 2.6 (Testing hypotheses relating to the regression coefficients) and 6.5 (Testing linear restriction).

Dougherty, C. Subject guide, Additional exercise sections A6.9 in Chapter 6 (Specification of regression variables).

Gujarati, D.N. and D.C. Porter *Basic econometrics.* (5th edition) [ISBN 9780071276252], Chapters 5.8 (Hypothesis testing: Some practical aspects) and 7.9 (Cobb–Douglas production function: More on functional form).

Approaching the question

(a) The interpretation of α_1, α_2 and $\alpha_1 + \alpha_2$ should be given. The solution is as follows:

The properties of the Cobb–Douglas production function are well-known.

α_1 is the (partial) elasticity of output with respect to the labour input. Hence, a 1 per cent increase in the labour input, holding capital constant, will increase output by 1.498 per cent, on average.

Similarly, α_2 is the (partial) elasticity of output with respect to the capital input, holding the labour input constant. Hence, a 1 per cent increase in the capital input, holding labour constant, will increase output by 0.489 per cent, on average.

The sum $\alpha_1 + \alpha_2$ gives information about the returns to scale; that is, the response of output to a proportionate change in the inputs. The estimated slope parameters are of the expected sign as both are positive – more input should produce more output.

(b) The significance of both slope parameters should be tested, and the assumptions should be explicitly provided. The solution is as follows:

The t statistics are:

$$t_{\alpha_1} = \frac{1.498}{0.54} = 2.77 \qquad \text{and} \qquad t_{\alpha_2} = \frac{0.489}{0.10} = 4.89.$$

The 5 per cent critical values for the two-sided t distribution with 12 degrees of freedom are ± 2.179. Hence, reject the null hypothesis in both cases. The test requires the following assumptions:

- The model is linear in the parameters and correctly specified.
- There is some variation in the regressor in the sample.
- The disturbance term has zero expectation, i.e. $\mathrm{E}(u_t) = 0$ for all t.
- The disturbance term is homoscedastic, i.e. $\mathrm{E}(u_t^2) = \sigma_u^2$ for all t.
- The values of the disturbance term are independent, i.e. $\mathrm{E}(u_i u_j) = 0$ for $i \neq j$.
- The disturbance term has a normal distribution.

(c) It should be mentioned that not enough information is given to test this restriction. The solution is as follows:

The estimated coefficients sum to 1.987, which gives the value of the returns to scale. The results suggest that, over the period of estimation, the Taiwanese agricultural sector was characterised by increasing returns to scale. However, we do not know whether 1.987 is significantly different from 1 without a formal statistical test, in which case we would need a measure of the standard error of the sum of the coefficients, or use an F test.

(d) The F test for linear restriction should be used. The solution is as follows:

The F test is given by:

$$F = \frac{(R_U^2 - R_R^2)/q}{(1 - R_U^2)/(n - k - 1)} = \frac{(0.89 - 0.77)/1}{(1 - 0.89)/12} = 13.09.$$

The 5 per cent critical value for $F_{1,12}$ is 4.75. Therefore, we reject the null hypothesis of constant returns to scale, and so 1.987 is significantly different from 1.

Equation (iii) incorporates the restriction of constant returns to scale as:

$$\begin{aligned}\ln Y_t &= \alpha_0 + \alpha_1 \ln L_t + \alpha_2 \ln K_t + \nu_t \\ &= \alpha_0 + (1-\alpha_2)\ln L_t + \alpha_2 \ln K_t + \nu_t.\end{aligned}$$

Re-arranging gives:

$$(\ln Y_t - \ln L_t) = \alpha_0 + \alpha_2(\ln K_t - \ln L_t) + \nu_t$$

which is the same as (iii).

Question 7

(a) **Explain the problem of identification in the context of simultaneous equation models.**

(3 marks)

(b) **In the model**

$$\begin{aligned}y_{1t} &= \alpha y_{2t} + u_{1t} \\ y_{2t} &= \beta_1 x_t + \beta_2 y_{1t} + u_{2t} \quad t = 1, 2, \ldots, T\end{aligned}$$

where x_t is an exogenous variable.

i. **Examine the identification of both equations.**

(4 marks)

ii. **Obtain the ordinary least squares estimator of α, and examine its consistency.**

(7 marks)

iii. **Derive the two-stage least squares of α and also prove its consistency stating carefully any assumptions you need.**

(3 marks)

(c) **What is meant by indirect least squares? Explain.**

(3 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics.* (4th edition) Chapters 9.2 (Simultaneous equations bias) and 9.3 (Instrumental variable estimation).

Dougherty, C. Subject guide, Chapter 9 (Simultaneous equation estimation).

Gujarati, D.N. and D.C. Porter *Basic econometrics.* (5th edition) [ISBN 9780071276252], Chapter 19.3 (Rules of identification) and 20.4 (Estimation of an overidentified equation: The method of two-stage least squares (2SLS)).

Approaching the question

(a) To answer this question, the order condition of identification should be used. The solution is as follows:

Order Condition of Identification (Necessary condition of identification):

$$R \geq G - 1$$

where:

$$\begin{aligned}R &= \text{the number of restrictions imposed on the equation under consideration} \\ &= \text{in our case, the number of variables excluded from the equation.}\end{aligned}$$

$$\begin{aligned}G &= \text{the number of jointly dependent variables in the model} \\ &= \text{the number of equations in the model.}\end{aligned}$$

If:

$R = G - 1$	the equation under consideration is exactly identified
$R > G - 1$	the equation under consideration is over identified
$R < G - 1$	the equation under consideration cannot be identified.

(b) i. The concept of the order of identification given in part (a) should be used. The solution is as follows:

In the first equation, $R = 0$ and $G - 1 = 1$, hence $R < G - 1$. Therefore, the equation is under identified.

In the second equation, $R = 1$ and $G - 1 = 1$, hence $R = G - 1$. Therefore, the equation is exactly identified.

ii. It is necessary to derive the OLS estimator. Also, the probability limit should be used to examine consistency. The solution is as follows:

The OLS estimator of α is given by:

$$\widehat{\alpha} = \frac{\sum y_{1t}y_{2t}}{\sum y_{2t}^2}.$$

To show consistency we consider:

$$\text{plim } (\widehat{\alpha}) = \frac{\text{plim } \left(\frac{1}{T}\sum y_{1t}y_{2t}\right)}{\text{plim } \left(\frac{1}{T}\sum y_{2t}^2\right)} = \alpha + \frac{\text{plim } \left(\frac{1}{T}\sum y_{2t}u_{1t}\right)}{\text{plim } \frac{1}{T}\sum y_{2t}^2} \neq \alpha$$

since:

$$\text{plim } \left(\frac{1}{T}\sum y_{2t}u_{1t}\right) = \text{plim } \left(\frac{1}{T}\sum(\beta_1 x_t + \beta_2 y_{1t} + u_{2t})u_{1t}\right) \neq 0$$

and:

$$\text{plim } \left(\frac{1}{T}\sum y_{2t}^2\right) \neq 0.$$

This implies $\widehat{\alpha}_{OLS}$ is an inconsistent estimator of α.

iii. It is necessary to derive the 2SLS estimator. Again, the probability limit should be used to examine consistency. The solution is as follows:

The two-stage estimator of α is:

$$\tilde{\alpha} = \frac{\sum y_{1t}z_t}{\sum y_{2t}z_t}$$

where z_t is the linear combination of instruments (in this case x_t only). To show consistency, we have:

$$\begin{aligned} \text{plim } (\tilde{\alpha}) = \text{plim } \frac{\sum y_{1t}z_t}{\sum y_{2t}z_t} = \text{plim } \frac{\sum(\alpha y_{2t} + u_{1t})z_t}{\sum y_{2t}z_t} &= \alpha + \text{plim } \left(\frac{\frac{1}{T}\sum z_t u_{1t}}{\frac{1}{T}\sum y_{2t}z_t}\right) \\ &= \alpha + \frac{\text{Cov}(z_t, u_{1t})}{\text{Cov}(y_{2t}, z_t)} \\ &= \alpha \end{aligned}$$

since z_t is correlated with y_{2t}, but uncorrelated with u_{1t}. The covariances given are the population covariances.

(d) A brief discussion of the ILS estimator is required. The solution is as follows:

- Obtain the reduced form from the given simultaneous equation model. There is a relationship between the reduced form (RF) parameters and the structural parameters.
- Estimate the RF parameters by OLS. The estimates will be consistent as in the RF all explanatory variables are exogenous.
- As the RF parameters and the structural parameters are related, once the RF parameters have been estimated, the estimates of the structural parameters can be obtained. These estimates will be consistent.

Question 8

A study of applications for home mortgages used the linear probability model

$$MORT_i = \beta_0 + \beta_1 INC_i + \beta_2 AGE_i + \beta_3 PROP_i + u_i; \quad i = 1, 2, \ldots, 700$$

where

$MORT_i$ = 1 if a mortgage is granted to the i-th applicant: 0 otherwise

INC_i = income of the i-th applicant in thousands of pounds

AGE_i = age of the i-th applicant in years

$PROP_i$ = age of the property for which the mortgage is being applied.

(a) The estimated coefficient for INC_i was 1.02 with standard error 0.51. What is the interpretation of this coefficient?

(4 marks)

(b) Why is R^2 meaningless in probit and logit models? What measures of 'goodness of fit' are applicable to probit and logit models?

(7 marks)

(c) Using a two variable linear model, show that the ordinary least squares estimator will be heteroscedastic if the dependent variable takes only values 0 and 1.

(9 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics.* (4th edition) Chapters 10.1 (Linear probability model) and 10.2 (Logit analysis).

Gujarati, D.N. and D.C. Porter *Basic econometrics.* (5th edition) [ISBN 9780071276252], Chapter 15.2 (The linear probability model (LPM)).

Approaching the question

(a) It is important to discuss that in the linear probability model the estimated probability of an event occurring may be greater than one or less than zero. The solution is as follows:

$t = 1.02/0.51 = 2$ which is significantly different from 0. As income increases by £1000, $MORT$ increases by 1.02 units, but since the estimated $MORT$ can be interpreted as a probability, the prediction is likely to lie outside [0, 1].

(b) It should be discussed that as the dependent variable takes only two values, R^2 is meaningless. A brief discussion of the likelihood ratio test and pseudo-R^2 should be given. The solution is as follows:

The definition of R^2 is:

$$R^2 = \frac{\text{ESS}}{\text{TSS}} = \frac{\text{TSS} - \text{RSS}}{\text{TSS}} = 1 - \frac{\text{RSS}}{\text{TSS}}$$

where TSS is the total sum of squares, ESS is the explained sum of squares, and RSS is the residual sum of squares. Under logit and probit the dependent variable only takes two values, 0 and 1, hence TSS will take different values dependent on the coding of 'success' or 'failure' even though the independent variables are the same.

The possibilities for measuring goodness of fit are (i) the pseudo-R^2 defined by $1 - (\ln L / \ln L_0)$, where $\ln L$ is the unrestricted log-likelihood and $\ln L_0$ is the log-likelihood that would have been obtained with only the intercept in the regression. This has a minimum of 0, but the maximum will be less than 1 and, unlike R^2, it does not have a natural interpretation.

The alternative is (ii) the likelihood ratio statistic defined by $2\ln(L/L_0) = 2[\ln L - \ln L_0]$, which has in large samples a chi-squared distribution with q degrees of freedom, where q is the number of restrictions imposed by the null hypothesis. The null hypothesis is that the coefficients of the variables are all jointly zero.

The parameters of $\ln L$ are estimated by maximum likelihood under the alternative hypothesis, and $\ln L_0$ is estimated by maximum likelihood under the null hypothesis.

(c) It is necessary to derive the variance of the disturbance term. The solution is as follows:

Let the model be:

$$Y_i = \beta_0 + \beta_1 X_i + u_i; \quad i = 1, 2, \ldots, n \qquad \text{(i)}$$

where:

$$Y_i = \begin{cases} 1 & \text{if the event occurs} \\ 0 & \text{if not.} \end{cases}$$

As Y_i takes only two values, 1 or 0, u_i can take only two values: $1 - \beta_0 - \beta_1 X_i$ when $Y_i = 1$, and $-\beta_0 - \beta_1 X_i$ when $Y_i = 0$. Based on this we can write the probability distribution of u_i as:

Y_i	u_i	$f(u_i)$
1	$1 - \beta_0 - \beta_1 X_i$	$\beta_0 + \beta_1 X_i$
0	$-\beta_0 - \beta_1 X_i$	$1 - \beta_0 - \beta_1 X_i$

This probability distribution also satisfies the assumption that:

$$\mathrm{E}(u_i) = (1 - \beta_0 - \beta_1 X_i)(\beta_0 + \beta_1 X_i) + (-\beta_0 - \beta_1 X_i)(1 - \beta_0 - \beta_1 X_i) = 0.$$

We can write $\mathrm{Var}(u_i)$ as:

$$\begin{aligned} \mathrm{Var}(u_i) = \mathrm{E}(u_i^2) &= (1 - \beta_0 - \beta_1 X_i)^2(\beta_0 + \beta_1 X_i) + (-\beta_0 - \beta_1 X_i)^2(1 - \beta_0 - \beta_1 X_i) \\ &= (1 - \beta_0 - \beta_1 X_i)(\beta_0 + \beta_1 X_i)[(1 - \beta_0 - \beta_1 X_i) + (\beta_0 + \beta_1 X_i)] \\ &= (\beta_0 + \beta_1 X_i)(1 - \beta_0 - \beta_1 X_i) \\ &= \mathrm{E}(Y_i)[1 - \mathrm{E}(Y_i)] \\ &= P_i(1 - P_i), \quad \text{for all } i = 1, 2, \ldots, n. \end{aligned}$$

Hence the disturbance term is heteroscedastic. This will make the OLS estimators inefficient.

Question 9

(a) Explain the meaning of spurious regression.

(4 marks)

(b) The following equations were estimated by ordinary least squares.

$$\begin{aligned} \boldsymbol{Y_t} &= \mathbf{3.0920 + 0.6959}\boldsymbol{X_t} + \boldsymbol{\widehat{u}_t} \qquad \mathbf{(1)} \\ & \quad\ \mathbf{(0.1305)}\ \ \mathbf{(0.0103)} \end{aligned}$$

$$\boldsymbol{R^2} = \mathbf{0.99},\ \boldsymbol{F} = \mathbf{4523.25},\ \boldsymbol{s} = \mathbf{0.0236},\ \mathbf{DW = 0.557},\ \boldsymbol{T} = \mathbf{740.}$$

$$\begin{aligned} \boldsymbol{\Delta\widehat{u}_t} &= \mathbf{-0.2161}\boldsymbol{\widehat{u}_{t-1}} + \mathbf{0.2349}\boldsymbol{\Delta\widehat{u}_{t-1}} + \mathbf{0.2029}\boldsymbol{\Delta\widehat{u}_{t-2}} + \boldsymbol{\widehat{\varepsilon}_t} \qquad \mathbf{(2)} \\ & \quad\ \mathbf{(0.0845)} \qquad\quad \mathbf{(0.1592)} \qquad\quad \mathbf{(0.1631)} \end{aligned}$$

$$\boldsymbol{R^2} = \mathbf{0.1799};\ \boldsymbol{s} = \mathbf{0.0115};\ \boldsymbol{T} = \mathbf{737.}$$

Where s is the standard error of the residuals, T is the number of observations, $\widehat{u}_t$ and $\widehat{\varepsilon}_t$ are OLS residuals, and standard errors are in parentheses.

Do the results above indicate that Y_t and X_t are cointegrated? Specify clearly all the assumptions you have made.

(6 marks)

[Note: Critical value at 5% significant level from MacKinnon table is −3.3377].

(c) Consider a model

$$y_t = u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2}; \quad t = 1, 2, \ldots, T$$

$$\mathrm{E}(u_t) = 0;\ \mathrm{E}(u_t^2) = \sigma^2;\ \text{and } \mathrm{E}(u_s u_t) = 0 \text{ for all } s, t = 1, 2, \ldots, T.$$

i. Is y_t stationary? Explain.

(5 marks)

ii. Calculate the autocorrelation function of y_t.

(5 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics.* (4th edition) Chapters 13.1 (Stationarity and nonstationarity), 13.2.(Spurious regressions) 13.3 (Graphical techniques for detecting nonstationarity) and 13.5 (Cointegration).

Dougherty, C. Subject guide, Chapter 13 (Introduction to nonstationary time series).

Gujarati, D.N. and D.C. Porter *Basic econometrics.* (5th edition) [ISBN 9780071276252], Chapter 21.3 (Stochastic Processes), 21.8 (Tests for stationarity) and 21.11 (Cointegration: Regression of a unit root time series on another unit root time series).

Approaching the question

(a) The concept of spurious regression should be explained with a simple model. The solution is as follows:

Spurious regression was first demonstrated by Granger and Newbold who showed, using Monte Carlo techniques, that a regression involving 2 non-stationary series could give rise to spurious results, in that the t statistics over-rejected the null hypothesis of a zero coefficient for 2 independent random walk series.

If Y_t and X_t are non-stationary and we regress Y_t on X_t, that is:

$$Y_t = \pi_0 + \pi_1 X_t + v_t$$

then even if there is no relationship between Y_t and X_t, the regression will produce a t ratio which will reject the null hypothesis $\mathrm{H}_0 : \pi_1 = 0$.

The reason for this result is that if $\mathrm{H}_0 : \pi_1 = 0$ then:

$$Y_t = \pi_0 + v_t.$$

Suppose Y_t is $I(1)$. Since Y_t is $I(1)$ and π_0 is constant, it follows that v_t must be $I(1)$. This violates the standard distributional theory based on the assumption that v_t is stationary, i.e. v_t is $I(0)$. Hence the misleading result.

(b) A clear concept of cointegration is required, and the assumptions should be stated explicitly. The solution is as follows:

We test:

$$\mathrm{H}_0 : \text{No cointegration} \qquad \text{vs.} \qquad \mathrm{H}_1 : \text{Cointegration.}$$

The cointegration test statistic is $-0.2161/0.0845 = -2.557$.

The 5 per cent critical value given in the MacKinnon table is -3.3377. Therefore, we cannot reject the null hypothesis of no cointegration.

The main assumption is that the error terms in both equations have constant variances and no serial correlation. We also need to assume that the specifications are correct (for example, no structural breaks).

(c) It should be shown that the mean, variance and covariances are independent of time. Part (ii) is based on part (i). The solution is as follows:

i. We have:

$$\begin{aligned}
\mathrm{E}(y_t) &= 0 \\
\mathrm{Var}(y_t) &= \mathrm{E}(y_t^2) \\
&= \mathrm{E}(u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2})^2 \\
&= \mathrm{E}(u_t^2) + \theta_1^2 \mathrm{E}(u_{t-1}^2) + \theta_2^2 \mathrm{E}(u_{t-2}^2); \quad \text{as } \mathrm{E}(u_s u_t) = 0 \text{ if } s \neq t \\
&= (1 + \theta_1^2 + \theta_2^2)\sigma^2 \\
\mathrm{Cov}(y_1, y_{t-1}) &= \mathrm{E}(y_t y_{t-1}) \\
&= \mathrm{E}(u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2})(u_{t-1} + \theta_1 u_{t-2} + \theta_2 u_{t-3}) \\
&= (\theta_1 + \theta_1\theta_2)\sigma^2
\end{aligned}$$

$$\begin{aligned}
\mathrm{Cov}(y_1, y_{t-2}) &= \mathrm{E}(y_t y_{t-2}) \\
&= \mathrm{E}(u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2})(u_{t-2} + \theta_1 u_{t-3} + \theta_2 u_{t-4}) \\
&= \theta_2 \sigma^2 \\
\mathrm{Cov}(y_t, y_{t-s}) &= \mathrm{E}(y_t y_{t-s}) = 0 \ \text{ for all } s > 2.
\end{aligned}$$

Hence as the mean, variance and covariances are constant over time, y_t is weakly stationary. If the u_ts are normally distributed then this also implies strong stationarity.

(ii) The autocorrelation function is defined as:

$$\rho_s = \frac{\mathrm{Cov}(y_t, y_{t-s})}{\sqrt{\mathrm{Var}(y_t)}\sqrt{\mathrm{Var}(y_{t-s})}} = \frac{\mathrm{Cov}(y_t, y_{t-s})}{\mathrm{Var}(y_t)}; \quad \text{as } \mathrm{Var}(y_t) = \mathrm{Var}(y_{t-s}).$$

Hence:

$$\rho_s = \begin{cases} 1 & \text{if } s = 0 \\ \frac{\theta_1 + \theta_1\theta_2}{1+\theta_1^2+\theta_2^2} & \text{if } s = 1 \\ \frac{\theta_2}{1+\theta_1^2+\theta_2^2} & \text{if } s = 2 \\ 0 & \text{if } s > 2. \end{cases}$$

Question 10

Let the regression equation be

$$\boldsymbol{Y_t = \beta_1 + \beta_2 X_t + u_t; \quad t = 1, 2, \ldots, T}$$

where

$$\begin{aligned}
\boldsymbol{u_t} &= \boldsymbol{\rho u_{t-1} + \varepsilon_t} \quad \textbf{for all } \boldsymbol{t}; \quad \boldsymbol{|\rho| < 1} \\
\mathbf{E}(\boldsymbol{\varepsilon_t}) &= \mathbf{0} \\
\mathbf{E}(\boldsymbol{\varepsilon_s \varepsilon_t}) &= \boldsymbol{\sigma_s^2} \ \textbf{if } \boldsymbol{s = t} \\
&= \mathbf{0} \ \textbf{if } \boldsymbol{s \neq t}
\end{aligned}$$

(a) Derive

i. the variance of u_t; and

ii. $\mathbf{E}(u_t u_{t-1})$.

(7 marks)

(b) Explain the consequences of this model specification on ordinary least squares estimators for β_1 and β_2.

(3 marks)

(c) Explain how would you test the null hypothesis $H_0 : \rho = 0$ against the alternative $H_1 : \rho \neq 0$. Specify all the assumptions needed for this test.

(5 marks)

(d) Discuss in detail a method of estimation which gives best linear unbiased estimates of β_1 and β_2.

(5 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics.* (4th edition) Chapters 12.1 (Definition and consequences of autocorrelation), 12.2 (Detection of autocorrelation) and 12.3 (Fitting a model subject to AR(1) autocorrelation).

Dougherty, C. Subject guide, Chapter 12 (Properties of regression models with time series data).

Gujarati, D.N. and D.C. Porter *Basic econometrics.* (5th edition) [ISBN 9780071276252], Chapters 12.4 (Consequences of using OLS in the presence of autocorrelation) and 12.6 (Detecting autocorrelation).

Approaching the question

(a) i. It is necessary to derive the variance and covariance of u_t. For finding these it has to be shown that $\mathrm{E}(u_t) = 0$. We have:

$$u_t = \rho u_{t-1} + \varepsilon_t$$

which can be written in lag operator form as:

$$(1 - \rho L)u_t = \varepsilon_t$$

or:

$$u_t = (1 - \rho L)^{-1}\varepsilon_t = (1 + \rho L + \rho^2 L^2 + \cdots)\varepsilon_t = \varepsilon_t + \rho\varepsilon_{t-1} + \rho^2\varepsilon_{t-2} + \cdots .$$

Therefore, the variance of u_t is:

$$\mathrm{Var}(u_t) = (1 + \rho^2 + \rho^4 + \cdots)\sigma_\varepsilon^2 = \frac{\sigma_\varepsilon^2}{1 - \rho^2}.$$

ii. We have:

$$\begin{aligned} u_t u_{t-1} &= (\varepsilon_t + \rho\varepsilon_{t-1} + \rho^2\varepsilon_{t-2} + \cdots)(\varepsilon_{t-1} + \rho\varepsilon_{t-2} + \rho^2\varepsilon_{t-3} + \cdots) \\ &= [\varepsilon_t + \rho(\varepsilon_{t-1} + \rho\varepsilon_{t-2} + \cdots)](\varepsilon_{t-1} + \rho\varepsilon_{t-2} + \rho^2\varepsilon_{t-3} + \cdots) \\ &= \varepsilon_t(\varepsilon_{t-1} + \rho\varepsilon_{t-2} + \rho^2\varepsilon_{t-3} + \cdots) + \rho(\varepsilon_{t-1} + \rho\varepsilon_{t-2} + \rho^2\varepsilon_{t-3} + \cdots)^2 \end{aligned}$$

Hence:

$$\mathrm{E}(u_t u_{t-1}) = \rho(1 + \rho^2 + \rho^4 + \cdots)\sigma_\varepsilon^2 = \frac{\rho\sigma_\varepsilon^2}{1 - \rho^2}.$$

(b) It should be explained for which properties OLS holds and also the properties for which OLS does not hold. The solution is as follows:

The effect of serially correlated errors is to produce unbiased and consistent, but inefficient, parameter estimates. The standard errors are incorrect leading to invalid t tests. Hence the estimates of β_1 and β_2 will be unbiased and consistent, but inefficient.

(c) This question is based on the Durbin–Watson test, whose assumptions should be stated clearly. The solution is as follows:

Durbin–Watson (DW) Statistic

The DW statistic is defined as:

$$\text{DW} = \frac{\sum_{t=2}^{T} (\widehat{u}_t - \widehat{u}_{t-1})^2}{\sum_{t=1}^{T} \widehat{u}_t^2}.$$

The Durbin–Watson statistic can be applied only if:

- the disturbance term follows an AR(1) process
- the model has an intercept term
- there is no lagged dependent variable as an explanatory variable.

In the Durbin–Watson table (which can be found in the appendix of any econometrics book) there are two values:

$$\begin{aligned} d_L &= \text{lower limit} \\ d_U &= \text{upper limit.} \end{aligned}$$

We have:

$$\begin{aligned} \text{DW} < d_L &\Rightarrow \text{positive autocorrelation} \\ \text{DW} > 4 - d_L &\Rightarrow \text{negative autocorrelation} \\ d_L \leq \text{DW} \leq d_U &\Rightarrow \text{no conclusion} \\ 4 - d_U \leq \text{DW} \leq 4 - d_L &\Rightarrow \text{no conclusion.} \end{aligned}$$

(d) As ρ is unknown, the Cochrane–Orcutt method of estimation, or Prais–Winstein method of estimation, should be used. The solution is as follows:

Cochrane–Orcutt method of estimation

For simplicity assume the model is:

$$Y_t = \beta_0 + \beta_1 X_t + u_t; \quad t = 1, 2, \ldots, T \qquad \text{(i)}$$

and:

$$u_t = \rho u_{t-1} + \varepsilon_t$$

with $\text{E}(\varepsilon_t) = 0$, $\text{E}(\varepsilon_t^2) = \sigma_\varepsilon^2$ and $\text{E}(\varepsilon_s \varepsilon_t) = 0$ for $s \neq t$.

This means that the disturbance term, u_t, follows an AR(1) process.

Lag (i) by one period and multiply by ρ to get:

$$\rho Y_{t-1} = \rho\beta_0 + \rho\beta_1 X_{t-1} + \rho u_{t-1}. \qquad \text{(ii)}$$

Subtract (ii) from (i), to get:

$$Y_t - \rho Y_{t-1} = (1-\rho)\beta_0 + \beta_1(X_t - \rho X_{t-1}) + u_t - \rho u_{t-1}. \qquad \text{(iii)}$$

The disturbance term in (iii) is $u_t - \rho u_{t-1} = \varepsilon_t$, which is a well-behaved disturbance term. Hence if ρ is *known*, OLS can be applied to (iii) to obtain the best linear unbiased estimators of β_0 and β_1.

If ρ is *not known*, (iii) cannot be estimated as such. Estimation of the parameters requires the following steps:

- Apply OLS to (i) to obtain $\widehat{u}_t$.

- Apply OLS to $\widehat{u}_t = \rho\widehat{u}_{t-1} + \varepsilon_t$ to obtain the OLS estimator of ρ:

$$\widehat{\rho} = \frac{\sum_{t=2}^{T} \widehat{u}_t \widehat{u}_{t-1}}{\sum_{t=2}^{T} \widehat{u}_{t-1}^2}.$$

  Replace ρ in (iii) by $\widehat{\rho}$ and apply OLS to get $\widehat{\beta}_0$ and $\widehat{\beta}_1$.

- Obtain a new set of residuals by replacing β_0 and β_1 in (i) by $\widehat{\beta}_0$ and $\widehat{\beta}_1$ as:

$$\tilde{u} = Y_t - \widehat{\beta}_0 - \widehat{\beta}_1 X_t.$$

- Repeat steps (b) to (d). Keep on doing so until the estimate of ρ converges, which will be the final estimate of ρ. Denote this as $\widehat{\rho}_F$.
- Replace ρ in (iii) by the final estimate of ρ, i.e. by $\widehat{\rho}_F$. Apply OLS to (iii) to obtain the final estimates of β_0 and β_1.

Examiners' commentaries 2015

EC2020 Elements of econometrics

Important note

This commentary reflects the examination and assessment arrangements for this course in the academic year 2014–15. The format and structure of the examination may change in future years, and any such changes will be publicised on the virtual learning environment (VLE).

Information about the subject guide and the Essential reading references

Unless otherwise stated, all cross-references will be to the latest version of the subject guide (2014). You should always attempt to use the most recent edition of any Essential reading textbook, even if the commentary and/or online reading list and/or subject guide refer to an earlier edition. If different editions of Essential reading are listed, please check the VLE for reading supplements – if none are available, please use the contents list and index of the new edition to find the relevant section.

Comments on specific questions – Zone B

Candidates should answer **EIGHT** of the following **TEN** questions: **ALL** of the questions in Section A (8 marks each) and **THREE** questions from Section B (20 marks each). **Candidates are strongly advised to divide their time accordingly.**

Section A

Answer all questions from this section.

Question 1

Let the regression equation be:

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

Where $\mathrm{E}(u_t) = 0$; $\mathrm{E}(u_t^2) = \sigma^2$ and $\mathrm{E}(u_s u_t) = 0$ if $s \neq t$. x values are fixed in repeated samples.

Obtain the ordinary least squares estimator (OLS) of β. Show that the OLS estimator of β is linear and unbiased.

(8 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics.* (4th edition) Chapters R.6 (Unbiasedness and efficiency), 1.3 (Derivation of the regression coefficients) and 2.3 (The random components and Unbiasedness of the OLS regression coefficients).

Dougherty, C. Subject guide, Chapter 2 (Properties of the regression coefficients and hypothesis testing).

Gujarati, D.N. and D.C. Porter *Basic econometrics.* (5th edition) [ISBN 9780071276252], Chapter 3A.2 (Linearity and unbiasedness properties of least-squares estimators).

Approaching the question

It is necessary to obtain the residual sum of squares (RSS) and minimise the RSS with respect to $\widehat{\beta}$ to obtain the OLS estimator. To show unbiasedness, it has to be shown that $\mathrm{E}(\widehat{\beta}) = \beta$. The solution is as follows:

We have:

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

The residual sum of squares (RSS) is:

$$\sum \widehat{u}_t^2 = \sum (y_t - \widehat{\beta} x_t)^2$$

and minimising RSS with respect to $\widehat{\beta}$ we get:

$$\frac{\mathrm{d} \sum \widehat{u}_t^2}{\mathrm{d}\widehat{\beta}} = 2 \sum (y_t - \widehat{\beta} x_t)(-x_t) = 0.$$

Solving we get:

$$\widehat{\beta}_{OLS} = \frac{\sum x_t y_t}{\sum x_t^2}.$$

To show linearity

Define w_t as:

$$w_t = \frac{x_t}{\sum x_t^2}$$

then:

$$\widehat{\beta} = \frac{\sum x_t y_t}{\sum x_t^2} = \sum w_t y_t.$$

This shows that $\widehat{\beta}$ is linear in the y_ts.

To show unbiasedness

We have:

$$\widehat{\beta} = \sum w_t y_t = \sum w_t (\beta x_t + u_t) = \beta \sum w_t x_t + \sum w_t u_t = \beta + \sum w_t u_t$$

as $\sum w_t x_t = 1$. Hence:

$$\mathrm{E}(\widehat{\beta}) = \beta + \sum w_t \mathrm{E}(u_t) = \beta \;\Rightarrow\; \text{unbiased}$$

since $\mathrm{E}(u_t) = 0$.

Question 2

Suppose that business expenditure for a new plant (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. Derive a linear relationship that can be used to estimate α and β, using the observable variables Y_t and X_t.

(8 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics.* (4th edition) Chapter 11.4 (Models with lagged dependent variable).

Gujarati, D.N. and D.C. Porter. *Basic econometrics.* (5th edition) [ISBN 9780071276252], Chapter 17.5 (Rationalization of the Koyck model: The adaptive expectations model).

Approaching the question

In order to get an estimable equation we need to eliminate the expected values from the equation. First, multiply through by $(1-\gamma)$ and lag to get a new equation, then subtract the new equation from the original equation to get the result. The solution is as follows:

We have:

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

To get an estimable equation we need to eliminate the expected values from the equation. We multiply through by $(1-\gamma)$ and lag to get:

$$(1-\gamma)\ln(Y_{t-1}) = (1-\gamma)\alpha + (1-\gamma)\beta\ln(X_{t-1}^*) + (1-\gamma)u_{t-1}.$$

Now subtract this from the original equation:

$$\begin{aligned}\ln(Y_t) - (1-\gamma)\ln(Y_{t-1}) &= \alpha\gamma + \beta(\ln(X_t^*) - (1-\gamma)\ln(X_{t-1}^*)) + (u_t - (1-\gamma)u_{t-1}) \\ &= \alpha\gamma + \beta(\gamma\ln(X_{t-1})) + (u_t - (1-\gamma)u_{t-1})\end{aligned}$$

or:

$$\ln(Y_t) = \alpha\gamma + \beta\gamma\ln(X_{t-1}) + (1-\gamma)\ln(Y_{t-1}) + (u_t - (1-\gamma)_{t-1}).$$

The parameters are estimated by non-linear techniques. If these procedures are not available, then a grid search can be used where γ is given values between 0 and 1 in steps of 0.1, and the remaining parameters are estimated using OLS.

Question 3

Let the probability density function of a random variable X be $f(x;\theta)$. Explain the procedure to use likelihood ratio test for testing the null hypothesis $\mathrm{H}_0 : \theta = \theta_0$.

(8 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics.* (4th edition) Chapter 10.6 (An introduction to maximum likelihood estimation).

Dougherty, C. Subject guide, Chapter 10 (Binary choice and limited dependent variable models, and maximum likelihood estimation).

Approaching the question

The likelihood ratio test should be discussed. The solution is as follows:

Suppose we have to test a simple hypothesis $H_0 : \theta = \theta_0$ against all possible alternatives. Given a simple random sample $\{X_1, X_2, \ldots, X_T\}$, a natural way of judging the acceptability, or otherwise, of the hypothesis would be to compare the likelihood functions.

Let:

L_R = the restricted likelihood (i.e. the likelihood based on the null hypothesis)

and:

L_U = the unrestricted likelihood (i.e. the likelihood based on the alternative hypothesis).

$\widehat{\theta}_{MLE}$ is used.

If the likelihood ratio (LR):

$$\lambda = \frac{L_R}{L_U}$$

is close to unity, then in light of the given sample H_0 would seem highly plausible. On the other hand if this ratio is close to zero, H_0 would seem to have little validity. Since λ is a random variable, its distribution may be derived, and hence we can make probability statements about how close the LR is to unity. Therefore, a test for H_0 is provided by a critical region defined by $\lambda < \lambda_0$, where λ_0 is such that $P(\lambda < \lambda_0 \,|\, H_0) = \alpha$, where α is the probability of a Type 1 error.

For large N, $-2 \ln \lambda$ has approximately a chi-squared distribution with degrees of freedom equal to the number of *restrictions* imposed by the null hypothesis.

Question 4

What is meant by a common factor test in the context of a linear model with an autocorrelated error? How would you perform a common factor test and what hypothesis would you be testing?

(8 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics.* (4th edition) Chapter 12.3 (Fitting a model subject to AR (1) autocorrelation).

Dougherty, C. Subject guide, Additional Exercises sections A12.3–A12.5 in Chapter 12 (Properties of regression models with time series data).

Approaching the question

The common factor test is used to test whether an autoregressive model or an autoregressive distributed lag model is the correct specification. The test should be specified. The solution is as follows:

We have:

$$Y_t = \beta_1 + \beta_2 X_t + u_t$$

and:

$$u_t = \rho u_{t-1} + v_t; \quad t = 1, 2, \ldots, T$$

where v_t has zero mean, constant variance and zero autocovariance. Combine the two equations to give a new equation, which has the general form:

$$Y_t = \beta_1(1-\rho) + \rho Y_{t-1} + \beta_2 X_t - \beta_2 \rho X_{t-1} + v_t.$$

It is a restricted version of the ADL(1,1) model:

$$Y_t = \lambda_1 + \lambda_2 Y_{t-1} + \lambda_3 X_t + \lambda_4 X_{t-1} + \varepsilon_t$$

and is subject to the restriction $\lambda_4 = -\lambda_2\lambda_3$. The test of this restriction is the common factor test. We test:

$$\mathrm{H}_0 : \lambda_4 = -\lambda_2\lambda_3 \qquad \text{vs.} \qquad \mathrm{H}_1 : \lambda_4 \neq -\lambda_2\lambda_3.$$

Note that the usual F test of the restriction is not appropriate because the restriction is non-linear, so we have to use the test statistic $T\ln(\mathrm{RSS}_R/\mathrm{RSS}_U)$, where RSS_R and RSS_U are the residual sum of squares from the restricted and unrestricted models, respectively. T is the sample size and the test statistic is asymptotically chi-squared with 1 degree of freedom. If the restriction is rejected, we conclude that the AR(1) specification is correct. If the restriction is not rejected, then the unrestricted ADL(1,1) model is the correct specification.

Question 5

Consider the model

$$\boldsymbol{y_t = u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2}; \quad t = 1, 2, \ldots, T.}$$

$\mathbf{E}(\boldsymbol{u_t}) = \mathbf{0}$**;** $\mathbf{E}(\boldsymbol{u_t^2}) = \boldsymbol{\sigma^2}$**; and** $\mathbf{E}(\boldsymbol{u_s u_t}) = \mathbf{0}$ **if** $\boldsymbol{s \neq t}$ **for all** $\boldsymbol{s, t = 1, 2, \ldots, T}$**.**

Is $\boldsymbol{y_t}$ **stationary? Explain in detail.**

(8 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics.* (4th edition) Chapter 13.1 (Stationarity and nonstationarity).

Dougherty, C. Subject guide, Chapter 13 (Introduction to nonstationary time series).

Gujarati, D.N. and D.C. Porter *Basic econometrics.* (5th edition) [ISBN 9780071276252], Chapter 21.3 (Stochastic Processes).

Approaching the question

It should be shown that the mean, variance and covariances are independent of time. The solution is as follows:

We have:

$$\begin{aligned}
\mathrm{E}(y_t) &= 0 \\
\mathrm{Var}(y_t) &= \mathrm{E}(y_t^2) \\
&= \mathrm{E}(u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2})^2 \\
&= \mathrm{E}(u_t^2) + \theta_1^2\mathrm{E}(u_{t-1}^2) + \theta_2^2\mathrm{E}(u_{t-2}^2); \quad \text{as } \mathrm{E}(u_s u_t) = 0 \text{ if } s \neq t \\
&= (1 + \theta_1^2 + \theta_2^2)\sigma^2 \\
\mathrm{Cov}(y_1, y_{t-1}) &= \mathrm{E}(y_t y_{t-1}) \\
&= \mathrm{E}(u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2})(u_{t-1} + \theta_1 u_{t-2} + \theta_2 u_{t-3}) \\
&= (\theta_1 + \theta_1\theta_2)\sigma^2
\end{aligned}$$

$$\begin{aligned}
\text{Cov}(y_1, y_{t-2}) &= \text{E}(y_t y_{t-2}) \\
&= \text{E}(u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2})(u_{t-2} + \theta_1 u_{t-3} + \theta_2 u_{t-4}) \\
&= \theta_2 \sigma^2 \\
\text{Cov}(y_t, y_{t-s}) &= \text{E}(y_t y_{t-s}) = 0 \quad \text{for all } s > 2.
\end{aligned}$$

Hence as the mean, variance and covariances are constant over time, y_t is weakly stationary. If the u_ts are normally distributed then this also implies strong stationarity.

Section B

Answer three questions from this section.

Question 6

Let the model be:

$$Y_t = \beta_0 + \beta_1 X_{t1} + \beta_2 X_{t2} + u_t; \quad t = 1, 2, \ldots, T$$

$\text{E}(u_t) = 0$ for all t. A researcher suspects that the variance of the disturbance term is $\text{Var}(u_t) = \sigma^2 X_{t1}$.

(a) Explain how to test the null hypothesis $\text{H}_0 : \text{Var}(u_t) = \sigma^2$ against the alternative hypothesis $\text{H}_1 : \text{Var}(u_t) = \sigma^2 X_{t1}$, for all $t = 1, 2, \ldots, T$.

(6 marks)

(b) Suggest in detail an estimation procedure which gives best linear unbiased estimates of the parameters.

(6 marks)

(c) Consider the model given by

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

where $\text{E}(u_t) = 0$; $\text{E}(u_t^2) = \sigma^2 x_t^2$; $\text{E}(u_s u_t) = 0$ if $s \neq t$, for all s and t. x_t is an observed non-random variable.

The density function of u_t is

$$f(u_t; x, t) = \left(2\pi\sigma^2 x_t^2\right)^{-1/2} \exp\left[-\frac{1}{2}\left(\frac{u_t}{\sigma x_t}\right)^2\right].$$

Derive the maximum likelihood estimators of α and σ^2.

(8 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics.* (4th edition) Chapters 7.2 (Detection of heteroscedasticity) and 7.3 (Remedies for heteroscedasticity).

Dougherty, C. Subject guide, Chapter 7 (Heteroscedasticity).

Gujarati, D.N. and D.C. Porter *Basic econometrics.* (5th edition) [ISBN 9780071276252], Chapters 4 A.1 (Maximum likelihood estimation of two-variable regression model), 11.5 (Detection of heteroscedasticity) and 11.6 (Remedial measures).

Approaching the question

(a) In Dougherty, the Goldfeldt–Quandt test and White test are given. Candidates are required to describe either one of them. Both tests are explained below:

Goldfeld–Quandt test

This test assumes that $\text{Var}(u_t)$ is proportional to the size of one of the right-hand side variables (say X_1). The observations are ranked by X_1. Run separate regression for the first n_1 ($< n/2$) and the last n_1 observations – the middle $(n - 2n_1)$ observations are not used. If heteroscedasticity is present, the RSS from the two regressions will differ. Form the test statistic $\text{RSS}_2/\text{RSS}_1$, where RSS_1 is the residual sum of squares from the first n_1 observations and RSS_2 is the residual sum of squares from the last n_1 observations. The test statistic will have an $F_{n_1-k,\, n_1-k}$ distribution, where k is the number of parameters in the equation. The null hypothesis is that the variances are homoscedastic.

The power of the test depends upon the choice of n_1 in relation to n. The assumptions are that there is no autocorrelation and the disturbance term is normally distributed.

White test

The White test looks for more general evidence of association between the variance of the error term and the regressors. Regress the squared residuals from the original regression on the regressors from that model, together with the squares and the cross-products of those variables. The test statistic is nR^2, where n is the sample size and R^2 is the R^2 from the White regression. This test statistic has a chi-squared distribution with degrees of freedom equal to the number of regressors. The test assumes the sample size is large. Normality of the disturbance term is not required.

(b) Weighted least squares should be used. The solution is as follows:

The model is:

$$Y_t = \beta_0 + \beta_1 X_{t1} + \beta_2 X_{t2} + u_t; \quad t = 1, 2, \ldots, T. \qquad \text{(i)}$$

Divide (i) by $\sqrt{X_{t1}}$ to get:

$$\frac{Y_t}{\sqrt{X_{t1}}} = \frac{\beta_0}{\sqrt{X_{t1}}} + \beta_1\sqrt{X_{t1}} + \frac{\beta_2 X_{t2}}{\sqrt{X_{t1}}} + \frac{u_t}{\sqrt{X_{t1}}}; \quad t = 1, 2, \ldots, T. \qquad \text{(ii)}$$

Equation (ii) is the transformed model, where:

$$\text{E}\left(\frac{u_t}{\sqrt{X_{t1}}}\right) = 0; \quad \text{as } \text{E}(u_t) = 0$$

and:

$$\text{Var}\left(\frac{u_t}{\sqrt{X_{t1}}}\right) = \frac{\text{Var}(u_t)}{X_{t1}} = \frac{\sigma^2 X_{t1}}{X_{t1}} = \sigma^2.$$

Also:

$$\text{E}\left(\frac{u_s}{\sqrt{X_{s1}}}\frac{u_t}{\sqrt{X_{t1}}}\right) = 0; \quad \text{as } \text{E}(u_s u_t) = 0.$$

All the assumptions required to apply OLS are satisfied. Hence, OLS can be applied to (ii) to obtain the best linear unbiased estimators of β_0, β_1 and β_2.

(c) Taking into account that $\text{Var}(u_t) = \sigma^2 X_{t1}^2$, the log-likelihood function should be derived. To get the answer, the log-likelihood function has to be differentiated with respect to α and σ^2 and equated to zero. The solution is as follows:

The log-likelihood function is:

$$\ln L = -\frac{T}{2}\ln 2\pi - \frac{T}{2}\ln\sigma^2 - \frac{1}{2}\sum \ln x_t^2 - \frac{1}{2\sigma^2}\sum\left(\frac{y_t}{x_t} - \alpha\right)^2.$$

The first-order conditions are:

$$\frac{\partial \ln L}{\partial \alpha} = \frac{1}{\sigma^2} \sum \left(\frac{y_t}{x_t} - \alpha \right) = 0 \qquad \text{(iii)}$$

and:

$$\frac{\partial \ln L}{\partial \sigma^2} = -\frac{T}{2\sigma^2} + \frac{1}{2\sigma^4} \sum \left(\frac{y_t}{x_t} - \alpha \right)^2 = 0 \qquad \text{(iv)}.$$

Solving (iii) and (iv), the ML estimators of α and σ^2 are obtained as:

$$\widehat{\alpha}_{MLE} = \frac{1}{T} \sum \left(\frac{y_t}{x_t} \right)$$

and:

$$\widehat{\sigma}^2_{MLE} = \frac{1}{T} \sum \left(\frac{y_t}{x_t} - \widehat{\alpha}_{MLE} \right)^2.$$

Question 7

(a) **Let the model be**

$$\boldsymbol{Y_t = \beta_0 + \beta_1 X_t + \beta_2 Z_t + u_t; \quad t = 1, 2, \ldots, T.}$$

Suppose the relevant variable Z_t was omitted from the model. Discuss the effects of this omission on unbiasedness and consistency of the ordinary least squares estimators.

(5 marks)

(b) **The following regression was estimated by ordinary least squares (OLS) on 745 observations of annual data from the UK.**

$$\begin{aligned} q_t &= \underset{(1.035)}{18.210} - \underset{(0.087)}{0.718} y_t + e_{1t}; \quad t = 1, 2, \ldots, 745 \qquad \text{(i)} \end{aligned}$$

$R^2 = 0.613$, $F = 67.97$, dw $= 1.68$

where q_t is the log of the quantity of cigarettes purchased, y_t is the log of real disposable income. Standard errors are in brackets. e_{1t} is an OLS residual. dw is the Durbin–Watson statistic.

Test the null hypothesis that the coefficient of y_t is -1. Comment on the result of your test. Do you think the estimates in (i) are plausible? Explain.

(5 marks)

(c) **A second regression was estimated using the same data sample**

$$\begin{aligned} q_t &= \underset{(2.214)}{0.399} - \underset{(0.041)}{0.343} p_t + \underset{(0.181)}{0.737} y_t + e_{2t}; \quad t = 1, 2, \ldots, 745 \qquad \text{(ii)} \end{aligned}$$

$R^2 = 0.855$, $F = 124.28$, dw $= 2.01$

where the variables are defined as before, p_t is the log of the relative prices of cigarettes, e_{2t} is an OLS residual. Standard errors in brackets. dw is the Durbin–Watson statistic.

Test the joint hypothesis that both slope coefficients in (ii) are zero.

(2 marks)

(d) **Compare the results in (ii) with those in (i). Provide an explanation in terms of omitted variable bias of the difference between the estimates of the coefficient of y_t in the two equations.**

(8 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics.* (4th edition) Chapters 2.6 (Testing hypotheses relating to the regression coefficients), 2.7 (The F test of goodness of fit) and 6.2 (The effect of omitting a variable that ought to be included).

Dougherty, C. Subject guide, Chapter 6.4 and 6.7 (Specification of regression variables).

Gujarati, D.N. and D.C. Porter *Basic econometrics.* (5th edition) [ISBN 9780071276252], Chapters 5.8 (Hypothesis testing: Some practical aspects), 8.4 (Testing the overall significance of the sample regression) and 13.3 (Consequences of model specification errors).

Approaching the question

(a) It is required to derive the omitted variable bias. The sign and size of the bias should be discussed. The solution is as follows:

If the true model is:

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 Z_t + u_t$$

but you estimate the regression equation:

$$Y_t = \beta_0 + \beta_1 X_t + v_t$$

then the ordinary least squares (OLS) estimator of β_1 is:

$$\begin{aligned}
\widehat{\beta}_1 &= \frac{\sum(X_t - \bar{X})(Y_t - \bar{Y})}{\sum(X_t - \bar{X})^2} \\
&= \frac{\sum(X_t - \bar{X})[(\beta_0 + \beta_1 X_t + \beta_2 Z_t + u_t) - (\beta_0 + \beta_1 \bar{X} + \beta_2 \bar{Z} + \bar{u})]}{\sum(X_t - \bar{X})^2} \\
&= \beta_1 + \beta_2 \frac{\sum(X_t - \bar{X})(Z_t - \bar{Z})}{\sum(X_t - \bar{X})^2} + \frac{\sum(X_t - \bar{X})(u_t - \bar{u})}{\sum(X_t - \bar{X})^2}
\end{aligned}$$

and:

$$\mathrm{E}(\widehat{\beta}_1) = \beta_1 + \beta_2 \frac{\sum(X_t - \bar{X})(Z_t - \bar{Z})}{\sum(X_t - \bar{X})^2}$$

since $\mathrm{E}(u_t) = 0$ and X_t is non-stochastic.

The bias will be:

$$\beta_2 \frac{\sum(X_t - \bar{X})(Z_t - \bar{Z})}{\sum(X_t - \bar{X})^2}$$

which will be positive if β_2 is positive and X_t and Z_t are positively correlated (or β_2 is negative and X_t and Z_t are negatively correlated). This bias is known as the omitted variable bias and occurs because β_1 takes up the effect of the missing Z_t. We have:

$$\mathrm{plim}(\widehat{\beta}_1) = \beta_1 + \beta_2 \frac{\mathrm{plim}\ \frac{1}{T}\sum(X_t - \bar{X})(Z_t - \bar{Z})}{\mathrm{plim}\ \frac{1}{T}\sum(X_t - \bar{X})^2} + \frac{\mathrm{plim}\ \frac{1}{T}\sum(X_t - \bar{X})(u_t - \bar{u})}{\mathrm{plim}\ \frac{1}{T}\sum(X_t - \bar{X})^2}.$$

If X and Z are uncorrelated in the limit then:

$$\mathrm{plim}\ \left(\frac{1}{T}\sum(X_t - \bar{X})(Z_t - \bar{Z})\right) = 0$$

and:

$$\mathrm{plim}\ \left(\frac{1}{T}\sum(X_t - \bar{X})^2\right) = \sigma_X^2.$$

Hence $\mathrm{plim}(\widehat{\beta}_1) = \beta_1$, and so $\widehat{\beta}_1$ is a consistent estimator of β_1.

Due to the omitted variable, the standard errors of coefficients and test statistics are in general invalidated.

If X and Z are orthogonal, then $\widehat{\beta}_1$ is both unbiased and consistent.

(b) A t test is required, and the null and alternative hypotheses should be clearly specified. The solution is as follows:

$t = (-0.718 + 1)/0.087 = 3.24$. The 5 per cent critical values (two sided) are ± 1.96 (approximately), hence we reject the null hypothesis. The null hypothesis is that the income elasticity of demand for cigarettes is minus one. This is rejected as the point estimate of the income elasticity is significantly different from -1. Since cigarettes are known to be bad for health this might be a plausible estimate.

(c) An F test should be used to test the joint significance of both slope coefficients. The solution is as follows:

Use the F statistic, which is given as 124.28. The 5 per cent critical value for $F_{2,\,742} = 2.96$ (approximately), hence we reject the null hypothesis.

(d) The sign of the bias and the correlation between y_t and p_t should be discussed. The solution is as follows:

The estimated coefficient of y_t has gone from -0.718 to 0.737 as a result of the inclusion of p_t in (ii). p_t has a highly significant coefficient in (ii). The estimates in (i) suffer from omitted variable bias. The bias on the coefficient of y_t in (i) is negative. Since the coefficient of p_t in (ii) is negative, this shows that the correlation between p_t and y_t is positive. This is clearly true for recent UK data. Incomes have risen and so has the relative price of cigarettes as the tax on cigarettes has risen.

Question 8

Consider the simple linear regression model

$$Y_t = \beta_0 + \beta_1 X_t + u_t; \quad t = 1, 2, \ldots, T$$

where β_0 and β_1 are unknown coefficients. A "goodness of fit" measure for ordinary least squares estimation of the above model is defined by

$$R^2 = \frac{\text{Explained Sum of Squares (ESS)}}{\text{Total Sum of Squares (TSS)}}$$

(a) Prove that

$$R^2 = 1 - \frac{\text{RSS}}{\text{TSS}},$$

where the residual sum of squares RSS $= \sum_{t=1}^{T} \widehat{u}_t^2$, $\widehat{u} = Y_t - \widehat{\beta}_0 - \widehat{\beta}_1 X_t$;

TSS $= \sum_{t=1}^{T} (Y_t - \bar{Y})^2$; and ESS $= \sum_{t=1}^{T} (\widehat{Y}_t - \bar{Y}_t)^2$ for all t.

(8 marks)

(b) Hence, or otherwise, show that

$$0 \leq R^2 \leq 1.$$

(4 marks)

(c) Briefly explain what happens to R^2 if an extra explanatory variable is added to the regression model.

(4 marks)

(d) Has R^2 any drawbacks? Explain.

(4 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics.* (4th edition) Chapter 1.6 (Goodness of fit: R^2).

Gujarati, D.N. and D.C. Porter *Basic econometrics.* (5th edition) [ISBN 9780071276252], Chapter 3.5 (The coefficient of determination R^2: A measure of 'goodness of fit').

Approaching the question

(a) It is important to discuss that $\sum \widehat{u}_t X_t = 0$. The solution is as follows:

After running a regression we know:

$$Y_t = \widehat{Y}_t + e_t$$

hence:

$$\text{Var}(Y) = \text{Var}(\widehat{Y}) + \text{Var}(e) + 2\text{Cov}(\widehat{Y}, e).$$

At this stage we need to show that the covariance term vanishes hence:

$$\text{Cov}(\widehat{Y}, e) = \text{Cov}(\widehat{\beta}_1 + \widehat{\beta}_2 X, e) = \text{Cov}(\widehat{\beta}_1, e) + \text{Cov}(\widehat{\beta}_2 X, e) = 0 + \widehat{\beta}_2 \text{Cov}(X, e).$$

However, $\text{Cov}(X, e) = 0$ from the minimisation of the residual sum of squares hence:

$$\text{Var}(Y) = \text{Var}(\widehat{Y}) + \text{Var}(e)$$

from which it follows that $\text{TSS} = \text{ESS} + \text{RSS}$.

The result $R^2 = 1 - \text{RSS}/\text{TSS}$ immediately follows.

(b) It should be discussed that as both RSS and TSS are sums of squares, they are non-negative. The solution is as follows:

$$R^2 = 1 - \frac{\text{RSS}}{\text{TSS}}$$

and $\text{RSS} \leq \text{TSS}$.

RSS and TSS are sums of squares, therefore they are non-negative.

If all the variations in the dependent variable are explained by the explanatory variables, then $\text{RSS} = 0$ and:

$$R^2 = 1 - \frac{0}{\text{TSS}} = 1.$$

If the explanatory variables do not explain any variation in the dependent variable, then $\text{RSS} = \text{TSS}$ and:

$$R^2 = 1 - \frac{\text{RSS}}{\text{TSS}} = 1 - \frac{\text{TSS}}{\text{TSS}} = 0.$$

Hence, $0 \leq R^2 \leq 1$.

(c) It should be discussed that TSS does not change with the addition of extra explanatory variables. The solution is as follows:

As extra variables are added to an equation the RSS must reduce if the estimated coefficient for that variable is not zero. Hence R^2 must increase, since TSS will remain the same.

(d) A discussion of the situations under which R^2 should not be used is required. The solution is as follows:

The main drawback to R^2 is explained in (c) but it is also not a good measure of the fit of a regression if there is no constant term in the regression (assuming that the true intercept is not zero), and it cannot be used to compare regressions where the dependent variable is not the same; that is, one cannot say that one regression is better than another on the basis of the R^2 value if the regressions have different dependent variables.

Question 9

A study of applications for home mortgages used the linear probability model

$$MORT_i = \beta_0 + \beta_1 INC_i + \beta_2 AGE_i + \beta_3 PROP_i + u_i; \quad i = 1, 2, \ldots, 700$$

where

$MORT_i = 1$ if a mortgage is granted to the i-th applicant: 0 otherwise

$INC_i =$ income of the i-th applicant in thousands of pounds

$AGE_i =$ age of the i-th applicant in years

$PROP_i =$ age of the property for which the mortgage is being applied.

(a) **The estimated coefficient for INC_i was 1.02 with standard error 0.51. What is the interpretation of this coefficient?**

(4 marks)

(b) **Why is R^2 meaningless in probit and logit models? What measures of 'goodness of fit' are applicable to probit and logit models?**

(7 marks)

(c) **Using a two variable linear model, show that the ordinary least squares estimator will be heteroscedastic if the dependent variable takes only values 0 and 1.**

(9 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics.* (4th edition) Chapters 10.1 (Linear probability model) and 10.2 (Logit analysis).

Gujarati, D.N. and D.C. Porter *Basic econometrics.* (5th edition) [ISBN 9780071276252], Chapter 15.2 (The linear probability model (LPM)).

Approaching the question

(a) It is important to discuss that in the linear probability model the estimated probability of an event occurring may be greater than one or less than zero. The solution is as follows:

$t = 1.02/0.51 = 2$ which is significantly different from 0. As income increases by £1000, $MORT$ increases by 1.02 units, but since the estimated $MORT$ can be interpreted as a probability, the prediction is likely to lie outside [0, 1].

(b) It should be discussed that as the dependent variable takes only two values, R^2 is meaningless. A brief discussion of the likelihood ratio test and pseudo-R^2 should be given. The solution is as follows:

The definition of R^2 is:

$$R^2 = \frac{\text{ESS}}{\text{TSS}} = \frac{\text{TSS} - \text{RSS}}{\text{TSS}} = 1 - \frac{\text{RSS}}{\text{TSS}}$$

where TSS is the total sum of squares, ESS is the explained sum of squares, and RSS is the residual sum of squares. Under logit and probit the dependent variable only takes two values, 0 and 1, hence TSS will take different values dependent on the coding of 'success' or 'failure' even though the independent variables are the same.

The possibilities for measuring goodness of fit are (i) the pseudo-R^2 defined by $1 - (\ln L / \ln L_0)$, where $\ln L$ is the unrestricted log-likelihood and $\ln L_0$ is the log-likelihood that would have been obtained with only the intercept in the regression. This has a minimum of 0, but the maximum will be less than 1 and, unlike R^2, it does not have a natural interpretation.

The alternative is (ii) the likelihood ratio statistic defined by $2\ln(L/L_0) = 2[\ln L - \ln L_0]$, which has in large samples a chi-squared distribution with q degrees of freedom, where q is the number of restrictions imposed by the null hypothesis. The null hypothesis is that the coefficients of the variables are all jointly zero.

The parameters of $\ln L$ are estimated by maximum likelihood under the alternative hypothesis, and $\ln L_0$ is estimated by maximum likelihood under the null hypothesis.

(c) It is necessary to derive the variance of the disturbance term. The solution is as follows:

Let the model be:

$$Y_i = \beta_0 + \beta_1 X_i + u_i; \quad i = 1, 2, \ldots, n \qquad \text{(i)}$$

where:

$$Y_i = \begin{cases} 1 & \text{if the event occurs} \\ 0 & \text{if not.} \end{cases}$$

As Y_i takes only two values, 1 or 0, u_i can take only two values: $1 - \beta_0 - \beta_1 X_i$ when $Y_i = 1$, and $-\beta_0 - \beta_1 X_i$ when $Y_i = 0$. Based on this we can write the probability distribution of u_i as:

Y_i	u_i	$f(u_i)$
1	$1 - \beta_0 - \beta_1 X_i$	$\beta_0 + \beta_1 X_i$
0	$-\beta_0 - \beta_1 X_i$	$1 - \beta_0 - \beta_1 X_i$

This probability distribution also satisfies the assumption that:

$$\mathrm{E}(u_i) = (1 - \beta_0 - \beta_1 X_i)(\beta_0 + \beta_1 X_i) + (-\beta_0 - \beta_1 X_i)(1 - \beta_0 - \beta_1 X_i) = 0.$$

We can write $\mathrm{Var}(u_i)$ as:

$$\begin{aligned} \mathrm{Var}(u_i) = \mathrm{E}(u_i^2) &= (1 - \beta_0 - \beta_1 X_i)^2(\beta_0 + \beta_1 X_i) + (-\beta_0 - \beta_1 X_i)^2(1 - \beta_0 - \beta_1 X_i) \\ &= (1 - \beta_0 - \beta_1 X_i)(\beta_0 + \beta_1 X_i)[(1 - \beta_0 - \beta_1 X_i) + (\beta_0 + \beta_1 X_i)] \\ &= (\beta_0 + \beta_1 X_i)(1 - \beta_0 - \beta_1 X_i) \\ &= \mathrm{E}(Y_i)[1 - \mathrm{E}(Y_i)] \\ &= P_i(1 - P_i), \quad \text{for all } i = 1, 2, \ldots, n. \end{aligned}$$

Hence the disturbance term is heteroscedastic. This will make the OLS estimators inefficient.

Question 10

Consider a two equation linear model

$$\begin{aligned} q_t &= \beta_1 p_t + \beta_2 y_t + u_t \qquad \text{(i)} \qquad \text{demand equation} \\ q_t &= \alpha p_t + e_t \qquad \text{(ii)} \qquad \text{supply equation} \end{aligned}$$

$t = 1, 2, \ldots, T$; $\mathrm{E}(u_t) = \mathrm{E}(e_t) = 0$; $\mathrm{E}(u_t^2) = \sigma_u^2$; $\mathrm{E}(e_t^2) = \sigma_e^2$; $\mathrm{E}(u_t e_t) = \sigma_{ue}$; $\mathrm{E}(u_s e_t) = 0$ if $s \neq t$, for all $s, t = 1, 2, \ldots, T$ and variables are defined as:

$$\begin{aligned} q_t &= \text{demand for good} \\ p_t &= \text{price for good} \\ y_t &= \text{personal disposable income.} \end{aligned}$$

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.

(6 marks)

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

(10 marks)

(c) Discuss, without derivation, what is the indirect least squares estimator of α.

(4 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics.* (4th edition) Chapter 9.3 (Instrumental variable estimation).

Dougherty, C. Subject guide, Chapter 9 (Simultaneous equation estimation).

Gujarati, D.N. and D.C. Porter *Basic econometrics.* (5th edition) [ISBN 9780071276252], Chapter 19.3 (Rules of identification) and 20.4 (Estimation of an overidentified equation: The method of two-stage least squares (2SLS)).

Approaching the question

(a) To answer this question the order condition of identification should be used. The solution is as follows:

Only the order condition of identification is required.

The order condition is $R \geq G - 1$, where R is the number of restrictions imposed on the equation under consideration and G is the number of jointly dependent variables, or the total number of equations, in the complete model.

1st equation:

$R = 0$ and $G - 1 = 2 - 1 = 1 \;\Rightarrow\; R < G - 1$, hence the equation is underidentified.

2nd equation:

$R = 1$ and $G - 1 = 2 - 1 = 1 \;\Rightarrow\; R = G - 1$, hence the equation is exactly identified.

(b) The two-stage least squares estimator should be derived and to prove consistency it should be shown that $\text{plim}(\widehat{\alpha}) = \alpha$. The solution is as follows:

The model is:

$$\begin{aligned} q_t &= \beta_1 p_t + \beta_2 y_t + u_t && \text{(i)} \quad \text{demand equation} \\ q_t &= \alpha p_t + e_t && \text{(ii)} \quad \text{supply equation} \end{aligned}$$

$t = 1, 2, \ldots, T$. T is the sample size. u and e are disturbance terms.

$\text{E}(u_t) = \text{E}(e_t) = 0$; $\text{E}(u_t^2) = \sigma_u^2$; $\text{E}(e_t^2) = \sigma_e^2$; $\text{E}(u_t e_t) = \sigma_{ue}$; $\text{E}(u_s e_t) = 0$ if $s \neq t$, for all s, $t = 1, 2, \ldots, T$.

q_t and p_t are endogenous variables.

In 2SLS, the right-hand side (RHS) endogenous variable (in this case p_t), is replaced by its estimated value (in this case $\widehat{p_t}$). After this we apply OLS to the equation.

We want to estimate α. Estimation involves two stages.

First stage:

Write down the reduced form (RF) corresponding to the RHS endogenous variable, in this case p_t. Let it be:

$$p_t = \pi y_t + v_t. \qquad \text{(iii)}$$

As RF parameters can always be consistently estimated by OLS, apply OLS to (iii) to get:

$$\widehat{p}_t = \widehat{\pi} y_t; \qquad \widehat{\pi} = \frac{\sum_{t=1}^{T} p_t y_t}{\sum_{t=1}^{T} y_t^2}. \qquad \text{(iv)}$$

We can also write p_t as:

$$p_t = \widehat{p}_t + \widehat{v}_t \qquad \text{(v)}$$

where $\widehat{v}_t$ is the OLS estimate of v_t.

Second stage:

In the second stage instead of p_t write $\widehat{p}_t + \widehat{v}_t$ in (ii) to get:

$$q_t = \alpha \widehat{p}_t + (\alpha \widehat{v}_t + e_t). \qquad \text{(vi)}$$

Apply OLS to (vi) to get the 2SLS estimator of α as:

$$\begin{aligned}
\widehat{\alpha}_{2SLS} &= \frac{\sum_{t=1}^{T} \widehat{p}_t q_t}{\sum_{t=1}^{T} \widehat{p}_t^2} \\
&= \frac{\widehat{\pi} \sum_{t=1}^{T} q_t y_t}{\widehat{\pi}^2 \sum_{t=1}^{T} y_t^2}; \qquad \widehat{p}_t = \widehat{\pi} y_t \text{ from (iv)} \\
&= \frac{\sum_{t=1}^{T} q_t y_t}{\sum_{t=1}^{T} y_t^2} \frac{\sum_{t=1}^{T} y_t^2}{\sum_{t=1}^{T} p_t y_t}; \qquad \widehat{\pi} = \frac{\sum_{t=1}^{T} p_t y_t}{\sum_{t=1}^{T} y_t^2} \text{ from (iv)} \\
&= \frac{\sum_{t=1}^{T} q_t y_t}{\sum_{t=1}^{T} p_t y_t}.
\end{aligned}$$

Consistency of the 2SLS estimator:

We have:

$$\widehat{\alpha}_{2SLS} = \frac{\sum_{t=1}^{T} q_t y_t}{\sum_{t=1}^{T} p_t y_t} = \frac{\sum_{t=1}^{T} y_t(\alpha p_t + e_t)}{\sum_{t=1}^{T} p_t y_t} = \alpha \frac{\sum_{t=1}^{T} p_t y_t}{\sum_{t=1}^{T} p_t y_t} + \frac{\sum_{t=1}^{T} y_t e_t}{\sum_{t=1}^{T} p_t y_t} = \alpha + \frac{\sum_{t=1}^{T} y_t e_t}{\sum_{t=1}^{T} p_t y_t}.$$

Also:

$$\text{plim}(\widehat{\alpha}_{2SLS}) = \alpha + \frac{\text{plim}\left(\frac{1}{T} \sum_{t=1}^{T} y_t e_t\right)}{\text{plim}\left(\frac{1}{T} \sum_{t=1}^{T} p_t y_t\right)} = \alpha + \frac{\text{Cov}(y_t, e_t)}{\text{Cov}(p_t, y_t)} = \alpha$$

since $\text{Cov}(y_t, e_t) = 0$ by definition, as y_t is exogenous.

(c) This question is based upon the fact that the given equation is exactly identified. The solution is as follows:

The ILS and 2SLS estimators are same as the equation under consideration is exactly identified.

Examiners' commentaries 2016

EC2020 Elements of econometrics

Important note

This commentary reflects the examination and assessment arrangements for this course in the academic year 2015–16. The format and structure of the examination may change in future years, and any such changes will be publicised on the virtual learning environment (VLE).

Information about the subject guide and the Essential reading references

Unless otherwise stated, all cross-references will be to the latest version of the subject guide (2011). You should always attempt to use the most recent edition of any Essential reading textbook, even if the commentary and/or online reading list and/or subject guide refer to an earlier edition. If different editions of Essential reading are listed, please check the VLE for reading supplements – if none are available, please use the contents list and index of the new edition to find the relevant section.

General remarks

Learning outcomes

At the end of this course, and having completed the Essential reading and activities, you should be able to:

- describe and apply the classical regression model and its application to cross-section data
- describe and apply the:
  - Gauss–Markov conditions and other assumptions required in the application of the classical regression model
  - reasons for expecting violations of these assumptions in certain circumstances
  - tests for violations
  - potential remedial measures, including, where appropriate, the use of instrumental variables
- recognise and apply the advantages of logit, probit and similar models over regression analysis when fitting binary choice models
- competently use regression, logit and probit analysis to quantify economic relationships using standard regression programmes (Stata and EViews) in simple applications
- describe and explain the principles underlying the use of maximum likelihood estimation
- apply regression analysis to fit time-series models using stationary time series, with awareness of some of the econometric problems specific to time series applications (for example, autocorrelation) and remedial measures
- recognise the difficulties that arise in the application of regression analysis to nonstationary time series, know how to test for unit roots, and know what is meant by cointegration.

Common mistakes committed by candidates

- A large number of candidates were not able to distinguish between sample variance and covariance, and population variance and covariance (this is happening year after year). They treat them as the same. This results in incorrect analysis and candidates lose significant marks.

  Consider an example: Suppose data are deviations from the respective sample means and the regression model is:

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

  The ordinary least squares estimator of β is:

  $$\widehat{\beta} = \frac{\sum\limits_{t=1}^{T} x_t y_t}{\sum\limits_{t=1}^{T} x_t^2} = \beta + \frac{\sum\limits_{t=1}^{T} x_t u_t}{\sum\limits_{t=1}^{T} x_t^2}.$$

  In terms of variances and covariances (a large number of candidates prefer this terminology), this can be written as:

  $$\widehat{\beta} = \beta + \frac{\mathrm{Cov}(x, u)}{\mathrm{Var}(x)}.$$

  Here $\mathrm{Cov}(x, u)$ and $\mathrm{Var}(x)$ are sample[$\mathrm{Cov}(x, u)$] and sample[$\mathrm{Var}(x)$].

  Candidates should realise that $\sum\limits_{t=1}^{T} u_t$, $\sum\limits_{t=1}^{T} x_t u_t$, $\mathrm{Cov}(x, u)$ and $\mathrm{Var}(x)$ given above are sample moments and as such $\sum\limits_{t=1}^{T} u_t \neq 0$, $\sum\limits_{t=1}^{T} x_t u_t \neq 0$ and $\mathrm{Cov}(x, u) \neq 0$. However, if we take the expectation, then:

  $$\mathrm{E}(u_t) = 0$$

  by assumption. Therefore:

  $$\mathrm{E}\left[\sum_{t=1}^{T} x_t u_t\right] = \sum_{t=1}^{T} x_t[\mathrm{E}(u_t)] = 0$$

  as the x_ts are fixed so they can be taken out of the expectation, and so:

  $$\mathrm{E}[\mathrm{Cov}(x, u)] = \mathrm{E}\left[\frac{1}{T}\sum_{t=1}^{T} x_t u_t\right] = 0$$

  as previously argued. This makes $\mathrm{E}(\widehat{\beta}) = \beta$, i.e. $\widehat{\beta}$ is an unbiased estimator of β.

  To prove consistency take the plim to get:

  $$\begin{aligned}
  \mathrm{plim}\left(\widehat{\beta}\right) &= \beta + \mathrm{plim}\left(\frac{\frac{1}{T}\sum\limits_{t=1}^{T} x_t u_t}{\frac{1}{T}\sum\limits_{t=1}^{T} x_t^2}\right) \\
  &= \beta + \frac{\mathrm{plim}\left(\frac{1}{T}\sum\limits_{t=1}^{T} x_t u_t\right)}{\mathrm{plim}\left(\frac{1}{T}\sum\limits_{t=1}^{T} x_t^2\right)} \\
  &= \beta + \frac{\mathrm{plim}(\text{sample Cov}(x, u))}{\mathrm{plim}(\text{sample Var}(x))} \\
  &= \beta + \frac{\text{population Cov}(x, u)}{\text{population Var}(x)}.
  \end{aligned}$$

By assumption, population $\text{Cov}(x, u) = 0$ and population $\text{Var}(x) > 0$, hence $\text{plim}\left(\widehat{\beta}\right) = \beta$, in other words $\widehat{\beta}$ is a consistent estimator of β.

Remember that in general:

$$\text{plim(sample variance)} = \text{population variance}$$

and:

$$\text{plim(sample covariance)} = \text{population covariance.}$$

This concept has been used in many questions. This simple mistake of not distinguishing between sample variance and covariance, and population variance and covariance, results in a significant loss of marks which might result in the loss of a degree class or even be the difference between pass and fail.

- Candidates struggled to give competent answers to the interpretation of empirical results. When interpreting an empirical result you should discuss the significance of the coefficients, magnitude and sign of the coefficients. Also, you should make sure that the Gauss–Markov conditions hold. Gauss–Markov conditions have to be explicitly specified. Only writing that the Gauss–Markov conditions hold is not sufficient.
- Just as last year, many candidates did not appear to read the questions carefully enough and often omitted to give answers to parts of questions which asked for details of such things as the assumptions necessary for a particular result to be true.

Key steps to improvement

Essential reading for this course includes the subject guide and the following.

> Dougherty, C. *Introduction to econometrics.* (Oxford: Oxford University Press, 2011) 4th edition [ISBN 9780199567089];
> http://global.oup.com/uk/orc/busecon/economics/dougherty4e/

Apart from Essential readings you should do some supplementary readings. One very good book of the same level is:

> Gujarati, D.N. and D.C. Porter *Basic econometrics.* (McGraw–Hill, 2009, International edition) 5th edition [ISBN 9780071276252].

To understand the subject clearly it is important to supplement C. Dougherty, *Introduction to econometrics* (fourth edition) with the subject guide **EC2020 Elements of econometrics** (2014), especially Chapter 10 which covers maximum likelihood.

It is very important to carefully go through the subject guide. The subject guide contains solutions to the questions given in the main textbook and also some additional questions and solutions. Working through these will improve your understanding of the subject.

The chapter in the subject guide on maximum likelihood (Chapter 10) includes some additional theory which has not been covered in the main textbook. It is important to read the additional theory given in the subject guide to have a better understanding of the principles of maximum likelihood and tests based on the likelihood function.

Please check the VLE course page for resources for this subject such as a downloadable copy of the subject guide **EC2020 Elements of econometrics** (2014), PowerPoint slideshows that provide graphical treatment of the topics covered in the textbook, datasets and statistical tables. Candidates should utilise datasets using standard regression programmes (STATA or EViews). This will help in the understanding of the subject.

Examination revision strategy

Many candidates are disappointed to find that their examination performance is poorer than they expected. This may be due to a number of reasons. The *Examiners' commentaries* suggest ways of addressing common problems and improving your performance. One particular failing is '**question spotting**', that is, confining your examination preparation to a few questions and/or topics which have come up in past papers for the course. This can have serious consequences.

We recognise that candidates may not cover all topics in the syllabus in the same depth, but you need to be aware that the examiners are free to set questions on **any aspect** of the syllabus. This means that you need to study enough of the syllabus to enable you to answer the required number of examination questions.

The syllabus can be found in the Course information sheet in the section of the VLE dedicated to each course. You should read the syllabus carefully and ensure that you cover sufficient material in preparation for the examination. Examiners will vary the topics and questions from year to year and may well set questions that have not appeared in past papers. Examination papers may legitimately include questions on any topic in the syllabus. So, although past papers can be helpful during your revision, you cannot assume that topics or specific questions that have come up in past examinations will occur again.

If you rely on a question-spotting strategy, it is likely you will find yourself in difficulties when you sit the examination. We strongly advise you not to adopt this strategy.

Examiners' commentaries 2016

EC2020 Elements of econometrics

Important note

This commentary reflects the examination and assessment arrangements for this course in the academic year 2015–16. The format and structure of the examination may change in future years, and any such changes will be publicised on the virtual learning environment (VLE).

Information about the subject guide and the Essential reading references

Unless otherwise stated, all cross-references will be to the latest version of the subject guide (2016). You should always attempt to use the most recent edition of any Essential reading textbook, even if the commentary and/or online reading list and/or subject guide refer to an earlier edition. If different editions of Essential reading are listed, please check the VLE for reading supplements – if none are available, please use the contents list and index of the new edition to find the relevant section.

Comments on specific questions – Zone A

Candidates should answer **EIGHT** of the following **TEN** questions: **ALL** of the questions in Section A (8 marks each) and **THREE** questions from Section B (20 marks each). Candidates are strongly advised to divide their time accordingly.

Section A

Answer all questions from this section.

Question 1

A random variable Y has unknown population mean μ_Y and population variance σ_Y^2. A sample of n observations $\{Y_1, \ldots, Y_n\}$ has been generated. Consider the following estimator:

$$\tilde{Y} = \frac{1}{n}\left(\frac{1}{2}Y_1 + \frac{3}{2}Y_2 + \frac{1}{2}Y_3 + \frac{3}{2}Y_4 + \cdots + \frac{1}{2}Y_{n-1} + \frac{3}{2}Y_n\right)$$

where the number of observations n is assumed to be even for convenience. Show that $\tilde{Y}$ is a consistent estimator of μ_Y.

(8 marks)

Reading for this question

Dougherty, C. *Introduction to Econometrics* (fourth edition) Chapter R.14 (Probability limits and consistency).

Gujarati, D.N. and D.C. Porter. *Basic Econometrics* (fifth edition) (ISBN 9780071276252) Chapter A.7 (Statistical Inference: Estimation).

Approaching the question

The definition of consistency should be given. To answer this question, the sufficient condition of consistency should be used. The answer is as follows.

$\tilde{Y}$ is a consistent estimator of μ_Y if $\tilde{Y} \xrightarrow{p} \mu_Y$. $\tilde{Y}$ is constructed by applying a weight of 1/2 to the $n/2$ 'odd' observations, and a weight of 3/2 to the remaining $n/2$ observations. We have:

$$\begin{aligned}
\mathrm{E}(\tilde{Y}) &= \frac{1}{n}\left(\frac{1}{2}\mathrm{E}(Y_1)+\frac{3}{2}\mathrm{E}(Y_2)+\frac{1}{2}\mathrm{E}(Y_3)+\frac{3}{2}\mathrm{E}(Y_4)+\cdots+\frac{1}{2}\mathrm{E}(Y_{n-1})+\frac{3}{2}\mathrm{E}(Y_n)\right) \\
&= \frac{1}{n}\left(\frac{1}{2}\times\frac{n}{2}\times\mu_Y+\frac{3}{2}\times\frac{n}{2}\times\mu_Y\right) \\
&= \mu_Y
\end{aligned}$$

and:

$$\begin{aligned}
\mathrm{var}(\tilde{Y}) &= \frac{1}{n^2}\left(\frac{1}{4}\mathrm{var}(Y_1)+\frac{9}{4}\mathrm{var}(Y_2)+\cdots+\frac{1}{4}\mathrm{var}(Y_{n-1})+\frac{9}{4}\mathrm{var}(Y_n)\right) \\
&= \frac{1}{n^2}\left(\frac{1}{4}\times\frac{n}{2}\times\sigma_Y^2+\frac{9}{4}\times\frac{n}{2}\times\sigma_Y^2\right) \\
&= 1.25\,\frac{\sigma_Y^2}{n}.
\end{aligned}$$

Therefore, $\tilde{Y}$ is an unbiased estimator of μ_Y, and because $\mathrm{var}(\tilde{Y}) \to 0$ as $n \to \infty$, $\tilde{Y}$ is consistent.

Question 2

Consider a two variable linear model:

$$\boldsymbol{Y_i = \beta_0 + \beta_1 X_i + u_i \qquad i = 1, 2, \ldots, n}$$

where $\mathbf{E}(u_i \mid X) = 0$ for all i; $\mathbf{E}(u_i^2) = \sigma^2$; and $\mathbf{E}(u_i u_j) = 0$ if $i \neq j$. Show that the ordinary least squares (OLS) estimators $\widehat{\beta}_0$ and $\widehat{\beta}_1$ are unbiased estimators of β_0 and β_1, respectively.

(8 marks)

Reading for this question

Dougherty, C. *Introduction to Econometrics* (fourth edition) 2.3 (The random components and unbiasedness of the OLS regression coefficients).

Dougherty, C. Subject guide Chapter 2 (Properties of the regression coefficients and hypothesis testing).

Gujarati, D.N. and D.C. Porter. *Basic Econometrics* (fifth edition) (ISBN 9780071276252) Chapter 3A.2 (Linearity and Unbiasedness Properties of Least-Squares Estimators).

Approaching the question

The derivation of the OLS estimators is not required. To show unbiasedness, candidates should take expectations of the estimators. The answer is as follows.

The OLS estimator of the slope β_1 is:

$$\begin{aligned}\widehat{\beta}_1 &= \frac{\sum(X_i-\bar{X})(Y_i-\bar{Y})}{\sum(X_i-\bar{X})^2}\\ &= \frac{\sum(X_i-\bar{X})([\beta_0+\beta_1X_i+u_i]-[\beta_0+\beta_1\bar{X}+\bar{u}])}{\sum(X_i-\bar{X})^2}\\ &= \beta_1+\frac{\sum(X_i-\bar{X})(u_i-\bar{u})}{\sum(X_i-\bar{X})^2}.\end{aligned}$$

Also:

$$\begin{aligned}\sum(X_i-\bar{X})(u_i-\bar{u}) &= \sum(X_i-\bar{X})u_i-\bar{u}\sum(X_i-\bar{X})\\ &= \sum(X_i-\bar{X})u_i-\bar{u}\left(\sum X_i-n\bar{X}\right)\\ &= \sum(X_i-\bar{X})u_i\end{aligned}$$

since $\sum X_i = n\bar{X}$. Hence:

$$\widehat{\beta}_1=\beta_1+\frac{\sum(X_i-\bar{X})u_i}{\sum(X_i-\bar{X})^2}=\beta_1+\sum a_iu_i$$

where $a_i=(X_i-\bar{X})/\sum(X_i-\bar{X})^2$. Therefore:

$$\mathrm{E}(\widehat{\beta}_1)=\mathrm{E}(\beta_1)+\mathrm{E}\left(\sum a_iu_i\right)=\beta_1+\sum a_i\,\mathrm{E}(u_i)=\beta_1$$

since $\mathrm{E}(u_i)=0$ by assumption.

The OLS estimator of the intercept β_0 is:

$$\widehat{\beta}_0=\bar{Y}-\widehat{\beta}_1\bar{X}$$

where $\widehat{\beta}_1$ is the OLS estimator of the slope β_1. Taking the expectation of $\widehat{\beta}_0$, we have:

$$\begin{aligned}\mathrm{E}(\widehat{\beta}_0)=\mathrm{E}(\bar{Y}-\widehat{\beta}_1\bar{X}) &= \mathrm{E}\left[\left(\beta_0+\beta_1\bar{X}+\frac{1}{n}\sum u_i\right)-\widehat{\beta}_1\bar{X}\right]\\ &= \beta_0+\mathrm{E}(\beta_1-\widehat{\beta}_1)\bar{X}+\frac{1}{n}\sum \mathrm{E}(u_i\,|\,X_i)\\ &= \beta_0\end{aligned}$$

where the third equality in the above equation has used the facts that $\widehat{\beta}_1$ is unbiased so $\mathrm{E}(\beta_1-\widehat{\beta}_1)=0$, and $\mathrm{E}(u_i\,|\,X_i)=0$.

Question 3

Show that the R^2 in the regression of Y on X (with an intercept) is the squared value of the sample correlation between X and Y (i.e. $R^2=r^2_{XY}$).

(8 marks)

Reading for this question

Dougherty, C. *Introduction to Econometrics* (fourth edition) Chapter 1.6 (Goodness of fit: R^2).

Dougherty, C. Subject guide Chapter 1 (Simple regression analysis).

Approaching the question

R^2 is given by:

$$R^2 = \frac{\text{Explained Sum of Squares (ESS)}}{\text{Total Sum of Squares (TSS)}}.$$

Simple algebraic manipulations of ESS and TSS is required to answer this question. The answer is as follows.

The coefficient of determination, R^2, is given by:

$$R^2 = \frac{\text{ESS}}{\text{TSS}} = \frac{\sum(\widehat{Y}_i - \bar{Y})^2}{\sum(Y_i - \bar{Y})^2}.$$

Given that:

$$\widehat{\beta}_0 = \bar{Y} - \widehat{\beta}_1 \bar{X} \qquad \text{and} \qquad \widehat{\beta}_1 = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sum(X_i - \bar{X})^2}$$

we have:

$$\begin{aligned} \text{ESS} = \sum(\widehat{Y}_i - \bar{Y})^2 &= \sum(\widehat{\beta}_0 + \widehat{\beta}_1 X_i - \bar{Y})^2 \\ &= \sum\left[\widehat{\beta}_1(X_i - \bar{X})\right]^2 \\ &= \widehat{\beta}_1^2 \sum(X_i - \bar{X})^2 \\ &= \frac{\left[\sum(X_i - \bar{X})(Y_i - \bar{Y})\right]^2}{\sum(X_i - \bar{X})^2}. \end{aligned}$$

Hence:

$$\begin{aligned} R^2 = \frac{\text{ESS}}{\sum(Y_i - \bar{Y})^2} &= \frac{\left[\sum(X_i - \bar{X})(Y_i - \bar{Y})\right]^2}{\sum(X_i - \bar{X})^2 \sum(Y_i - \bar{Y})^2} \\ &= \left[\frac{\frac{1}{n-1}\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\frac{1}{n-1}\sum(X_i - \bar{X})^2}\sqrt{\frac{1}{n-1}\sum(Y_i - \bar{Y})^2}}\right]^2 \\ &= \left[\frac{s_{XY}}{s_X s_Y}\right]^2 \\ &= r_{XY}^2. \end{aligned}$$

Question 4

Explain how you would estimate the parameters α and β for the model $(M_t/P_t) = \alpha r_t^{\beta}$ given data on money (M_t), prices (P_t) and the nominal interest rate (r_t). Interpret the results.

(8 marks)

Reading for this question

Dougherty, C. *Introduction to Econometrics* (fourth edition) Chapter 4.2 (Logarithmic transformations).

Dougherty, C. Subject guide Chapter 4 (Transformation of variables).

Gujarati, D.N. and D.C. Porter. *Basic Econometrics* (fifth edition) (ISBN 9780071276252) Chapter 6.4 (Functional Forms of Regression Models).

Approaching the question

To answer this question, a logarithmic transformation of the given model is required. The answer is as follows.

We transform the equation into an estimable form by taking logs to give:

$$\log\left(\frac{M_t}{P_t}\right) = \log(\alpha) + \beta\log(r_t)$$

that is, a log-log model. Now we can estimate the parameters by using OLS. The estimate of α is given by exp(constant), and the slope is the estimate of β. In this model β is the interest rate elasticity of demand for money (the percentage change in the demand for money for a 1 per cent change in the interest rate). Alternatively, use a non-linear estimation process.

Question 5

What do you understand by an instrumental variable (IV)? How you would estimate $\boldsymbol{\beta}$ in the model $\boldsymbol{y_t = \beta x_t + u_t}$, where $\boldsymbol{x_t}$ and $\boldsymbol{u_t}$ are correlated, using $\boldsymbol{z_t}$, $\boldsymbol{t = 1, 2, \ldots, T}$ as an IV? Examine the consistency of the IV estimator.

(8 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fourth edition): Chapter 8.5 (Instrumental variables).

Dougherty, C. Subject guide (2011): Chapter 8.5 (Stochastic regressor and measurement errors).

Approaching the question

Candidates should define an instrumental variable (IV) and when it is used. It is required to prove that the IV estimator is consistent. The answer is as follows.

Consider the model:

$$y_t = \beta x_t + u_t, \qquad t = 1, 2, \ldots, T.$$

If x_t is not independently distributed of u_t then the OLS estimator of β will be inconsistent. Consider a variable z that is correlated with x but not correlated with u. z can be considered as an instrumental variable. An estimator of β based on z is known as an instrumental variable (IV) estimator. It is defined as:

$$\widehat{\beta}_{IV} = \frac{\sum z_t y_t}{\sum z_t x_t}.$$

It can be shown that $\widehat{\beta}_{IV}$ is a consistent estimator of β.

$$\widehat{\beta}_{IV} = \frac{\sum z_t y_t}{\sum z_t x_t} = \frac{\sum z_t(\beta x_t + u_t)}{\sum z_t x_t} = \beta + \frac{\sum z_t u_t}{\sum z_t x_t}$$

and:

$$\text{plim}(\widehat{\beta}_{IV}) = \beta + \frac{\text{plim}\left(\sum z_t u_t / T\right)}{\text{plim}\left(\sum z_t x_t / T\right)} = \beta$$

hence $\widehat{\beta}_{IV}$ is a consistent estimator of β.

Note that:

$$\text{plim}\left(\sum z_t u_t / T\right) = 0 \qquad \text{and} \qquad \text{plim}\left(\sum z_t x_t / T\right) \neq 0.$$

Section B

Answer three questions from this section.

Question 6

A researcher wants to test the hypothesis that drug companies practice price discrimination against less developed countries. To do this she estimates two demand for pharmaceuticals equations using data from 32 countries.

Equation A:

$$\underset{(6.27)}{P_i = 32.60} + \underset{(0.24)}{1.66} GDPN_i - \underset{(0.25)}{0.88} CD_i + e_{1i} \qquad R^2 = 0.70 \quad i = 1, \ldots, 32$$

Equation B:

$$\ln(P_i) = \underset{(0.24)}{2.26} + \underset{(0.13)}{0.95} \ln(GDPN_i) - \underset{(0.09)}{0.45} \ln(CD_i) + e_{2i} \quad R^2 = 0.71 \; i = 1 \ldots, 32$$

where P_i is the price index for pharmaceuticals in each country, $GDPN_i$ is the gross domestic product (GDP) per capita for each country, and CD_i is the per capita consumption of pharmaceutical drugs. $\ln(P_i)$, $\ln(GDPN_i)$, $\ln(CD_i)$ are the logarithms of these quantities. Figures in parentheses are the estimated standard errors.

(a) For each of the estimated equations, are the estimated partial regression coefficients individually statistically significant at the 5 percent level of significance? Specify and justify the alternative hypothesis in each test.

(3 marks)

(b) Test the hypothesis that the slope coefficients are all simultaneously zero in each equation. What do you conclude from your results?

(3 marks)

(c) Show that in a log-log model, such as (B), the slope parameters are elasticities.

(4 marks)

(d) The researcher claims that: (i) as the population of a country becomes richer the cost of drugs decreases, and hence the demand becomes less elastic; and (ii) as people consume more drugs the market becomes more developed, and hence more competitive. To what extent, if at all, do the results in (A) and (B) support these claims?

(6 marks)

(e) Are serial correlation and/or heteroscedasticity likely to be a problem in the linear and log-linear regression models (A) and (B)? Explain.

(4 marks)

Reading for this question

Dougherty, C. *Introduction to Econometrics* (fourth edition) Chapters 2.6 (Testing hypotheses relating to the regression coefficients), 3.5 (Goodness of fit: R^2) and 4.2 (Logarithmic transformations).

Dougherty, C. Subject guide Chapter 2 (Properties of the regression coefficients and hypothesis testing).

Approaching the question

In part (a) individual coefficients are being tested, the t test is appropriate for this. Part (b) involves a test of a joint hypothesis. To conduct the test, an F test should be used. In part (d) the sign and significance of the coefficients should be explained. The answer is as follows.

(a) The critical t value for both equation (A) and equation (B) is 2.045 (degrees of freedom 29, 5% significance level for a 2-tailed test). The estimated t values are:

$$\text{Equation (A)} \qquad t = \frac{32.6}{6.27} = 5.2, \qquad t = \frac{1.66}{0.24} = 6.92 \qquad \text{and} \qquad t = -\frac{0.88}{0.25} = -3.52$$

$$\text{Equation (B)} \qquad t = \frac{2.26}{0.24} = 9.42, \qquad t = \frac{0.95}{0.13} = 7.31 \qquad \text{and} \qquad t = -\frac{0.45}{0.09} = -5.$$

These are all larger in absolute terms than the critical value. Hence we can say that all the coefficients are significantly different from zero. The estimated parameters show that as $GDPN$ increases so does the price of pharmaceuticals, the income elasticity of the price for pharmaceuticals is 0.95 (equation (B)). The estimated parameters also show that as CD increases the price for pharmaceuticals decreases with elasticity -0.45.

(b) To test the hypothesis that all coefficients are simultaneously zero we need to use an F test, which would normally be based on the residual sum of squares (RSS). Here the only information we have is the R^2, which contains the same information. The F test is:

$$F = \frac{R^2/k}{(1-R^2)/(n-k-1)} = \frac{0.70/2}{(1-0.70)/(32-2-1)} = 33.83$$

for equation (A), and the F critical value is 3.33 with $(2, 29)$ degrees of freedom. Hence we reject the null hypothesis of zero coefficients. Equation (B) is similar: replace the value of R^2 with 0.71, and repeat the computation. The answer is similar.

(c) If the equation is $\log(Y_i) = \beta_0 + \beta_1 \log(X_i) + u_i$, and we differentiate it we get:

$$\frac{\partial Y_i}{\partial X_i}\frac{1}{Y_i} = \beta_1 \frac{1}{X_i} \qquad \text{or} \qquad \frac{\partial Y_i}{\partial X_i}\frac{X_i}{Y_i} = \beta_1 = \text{elasticity.}$$

(d) The hypothesis might suggest that as the income per capita increases the price of drugs falls. It also says that as the income of the population increases the slope coefficient on equation (B) falls. We have shown that the first statement is not correct – the coefficient on $GDPN$ is positive and significant. The second statement is not covered by the regressions, which simply model fixed elasticities.

The second hypothesis says that as the consumption of drugs rises the market becomes more competitive. There is evidence that this is supported by the equations; the coefficient on CD (the consumption of drugs) is negative and significantly different from zero, which suggests that as the consumption of drugs increases the price falls (i.e. the market becomes more competitive).

(e) The data are cross-sectional, hence there is no possibility of serial correlation, but one might expect heteroskedasticity since countries are going to be of very different sizes, and hence the error term is likely to mirror some of these values.

Question 7

The following estimates were calculated from a sample of 7,634 women respondents from the General Household Survey 1995. The dependent variable takes the value 1 if the woman was in paid employment, and 0 otherwise.

	OLS	Logit	Probit
high	0.093	0.423	0.259
	(0.015)	(0.071)	(0.043)
noqual	−0.210	−0.898	−0.554
	(0.013)	(0.056)	(0.035)
age	0.038	0.173	0.107
	(0.003)	(0.124)	(0.008)
age2	−0.051	−0.230	−0.142
	(0.003)	(0.069)	(0.009)
mar	0.024	0.103	0.063
	(0.009)	(0.057)	(0.035)
Constant	−0.068	−2.587	−1.593
	(0.049)	(0.225)	(0.137)

Where high is one if the respondent has a higher educational qualification, zero otherwise; noqual is one if the respondent has no qualifications, zero otherwise; age is age in years; age2 is (age × age) 100; mar is one if married, zero otherwise. Conventionally calculated standard errors are given in brackets for the ordinary least squares (OLS) results and asymptotic standard errors in brackets elsewhere.

(a) **Explain briefly how Probit estimates are calculated when the model has no intercept and only one explanatory variable.**

(6 marks)

(b) **For all three sets of estimates, test the null hypothesis that the coefficient of mar is zero. Which test statistics would you consider more reliable? Explain.**

(6 marks)

(c) **Using the OLS and Probit estimates, calculate the estimated probabilities of being in employment for a married woman aged 40 with a higher educational qualification. Comment on your results.**

(5 marks)

(d) **Test the null hypothesis that all the slope coefficients of the probit model are jointly equal to zero, given that:**

$$\ln L_R = -416.01$$

$$\ln L_U = -321.25$$

where $\ln L_R$ and $\ln L_U$ are the logs of the likelihood from the restricted and the unrestricted probit models, respectively.

(3 marks)

Reading for this question

Dougherty, C. *Introduction to Econometrics* (fourth edition) Chapters 10.1 (The linear probability model), 10.3 (Probit analysis) and 10.6 (An introduction to maximum likelihood estimation).

Dougherty, C. Subject guide (2011) Chapter 10 (Binary choice and limited dependent variable models, and maximum likelihood estimation).

Gujarati, D.N. and D.C. Porter *Basic Econometrics* (fifth edition): (ISBN 9780071276252), Chapters 15.2 (The Linear Probability Model (LPM)), 15.5 (The Logit Model) and 15.9 (The Probit Model).

Approaching the question

(a) The probit model uses the cumulative standardised normal distribution. The maximum likelihood technique is used to obtain the estimates of the parameters. Estimates have the standard maximum likelihood properties, i.e. the estimators are consistent, asymptotically efficient and asymptotically normally distributed.

[For technical details see Dougherty (fourth edition) Section 10.3].

(b) To test the null hypothesis the t test should be used. t tables are attached with the examination paper. Candidates should know how to look at the critical values. The null and alternative hypotheses should be clearly stated:

$$\mathrm{H}_0 : \text{coefficient of mar} = 0$$

$$\mathrm{H}_1 : \text{coefficient of mar} \neq 0.$$

t test statistics are 2.67 (OLS), 1.81 (logit), and 1.8 (probit).

The test based on the OLS estimates gives outright rejection. However, the standard errors for this test are wrongly calculated because of the heteroskedasticity of the error term. Therefore, we prefer probit or logit estimates.

(c) A large number of candidates wrongly thought that to obtain the probability in the case of probit a sophisticated calculator is needed. The probability in the case of probit can be read directly from the supplied statistical tables. We have:

$$0.093 + 0.038 \times 40 - 0.051 \times 16 + 0.024 \times 0.068 = 0.753 \qquad \text{(OLS)}$$

$$0.259 + 0.107 \times 40 - 0.142 \times 16 + 0.063 - 1.593 = 0.737 \qquad \text{(probit).}$$

The statistical tables give a probability of 0.77, approximately.

They are all fairly close. The OLS estimates do not fall outside the probability bounds.

(d) To test the null hypothesis, the large sample likelihood-ratio test should be used. In large samples $-2(\ln L_R - \ln L_U)$ is distributed as a chi-squared distribution with degrees of freedom equal to the number of restrictions imposed by the null hypothesis. The null and alternative hypotheses should be clearly stated as:

$$\mathrm{H}_0 : \text{all the slope coefficients are} = 0$$

$$\mathrm{H}_1 : \text{at least one slope coefficient is} \neq 0.$$

We have:

$$-2(\ln L_R - \ln L_U) \sim \chi^2_5$$

and:

$$-2(-416.01 - (-321.25)) = 189.92.$$

The critical value of χ^2_5 at the 5% significance level is 11.07, hence we reject H_0.

Question 8

The relationship between a dependent variable Y and an explanatory variable X is given by the linear model

$$\boldsymbol{Y_i = \alpha + \beta X_i + u_i; \qquad i = 1, 2, \ldots, n}$$

for an i.i.d. sample $\{Y_i, X_i\}$, but where the errors u_i are correlated with the explanatory variable, $\mathrm{E}(X_i u_i) = \sigma_{Xu} \neq 0$.

(a) **Assuming that plim $\left(\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2\right)=\sigma_{XX}$, where $\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$, show that the OLS estimator of β is inconsistent.**

(4 marks)

(b) **Three possible instrumental variables z_1, z_2 and z_3 are available in the data set. What properties do these instruments need to have in order to be able to use them to estimate β consistently by the Instrumental Variable, or the Two Stage Least Squares (2SLS), estimator?**

(4 marks)

(c) **Using all three instruments, describe in detail the 2SLS estimator.**

(4 marks)

(d) **Assuming conditional homoskedasticity of the errors, $\mathrm{E}(u_i^2 \mid z_i)=\sigma^2$, explain in detail how you would test whether $\mathrm{E}(z_iu_i)=0$. What are the consequences for the 2SLS estimator if the test result implies a rejection of this hypothesis?**

(4 marks)

(e) **Explain in detail how would you test whether $\mathrm{E}(X_iu_i)=0$ (Durbin–Wu–Hausman test). Which estimator of β would you use if this hypothesis does not get rejected?**

(4 marks)

Reading for this question

Dougherty, C. *Introduction to Econometrics* (fourth edition) Chapters 8.3 (Asymptotic properties of the OLS regression coefficients), 8.5 (Instrumental variables) and 9.3 (Instrumental variable estimation).

Dougherty, C. Subject guide Chapters 8 (Stochastic regressors and measurement errors) and 9 (Simultaneous equations estimation).

Approaching the question

The plim should be used to verify the consistency of the OLS estimators. Properties of the instrumental variables should be clearly explained. The answer is as follows.

(a) We have:

$$\mathrm{plim}(\widehat{\beta}_{OLS}-\beta)=\mathrm{plim}\left(\frac{\sum(X_i-\bar{X})u_i}{\sum(X_i-\bar{X})^2}\right) = \mathrm{plim}\left(\frac{\mathrm{plim}\left(\frac{1}{n}\left(\sum_i X_iu_i-u_i\frac{1}{n}\sum_j X_j\right)\right)}{\mathrm{plim}\left(\frac{1}{n}\sum(X_i-\bar{X})^2\right)}\right)$$

$$=\frac{\left(1-\frac{1}{n}\right)\sigma_{Xu}}{\sigma_{XX}}\neq 0.$$

(b) Candidates need to explain that z_i must be correlated with X_i but not with u_i, and does not already appear in the equation in its own right. Hence in the reduced form:

$$X_i=\pi_0+\pi_1z_{i1}+\pi_2z_{i2}+\pi_3z_{i3}+\omega_i$$

π_1, π_2 and π_3 should all not be zero (if one is zero then it should be dropped from the list). Furthermore, $\mathrm{E}(z_iu_i)=0$.

(c) Estimate by OLS:

$$\widehat{X}_i=\widehat{\pi}_0+\widehat{\pi}_1z_{i1}+\widehat{\pi}_2z_{i2}+\widehat{\pi}_3z_{i3}.$$

Next, estimate by OLS:

$$Y_i=\alpha+\beta\widehat{X}_i+\varepsilon_i$$

in the second stage. $\widehat{X}$ only utilises exogenous variation due to the instruments, and hence identifies the causal effect β.

(d) nR^2 test: Regress the IV residuals:

$$\widehat{u}_{IV} = Y_i - \widehat{\alpha}_{IV} - \widehat{\beta}_{IV} X_i$$

on the instruments:

$$\widehat{u}_{IV} = \delta_0 + \delta_1 z_{i1} + \delta_2 z_{i2} + \delta_3 z_{i3} + \zeta_i.$$

nR^2 in this regression follows a chi-squared distribution with 2 degrees of freedom under the null hypothesis that $\mathrm{E}(z_i u_i) = 0$. If the test statistic is large, then reject the null hypothesis. If the test rejects the null hypothesis, then the IV estimator is not consistent.

(e) Add the first-stage residuals to the model:

$$Y_i = \alpha + \beta X_i + \gamma \widehat{\omega}_i + \xi_i.$$

Testing for exogeneity is a test of $\mathrm{H}_0 : \gamma = 0$. A t test is appropriate as standard errors are correct under the null hypothesis. If the test does not reject the null hypothesis use the OLS estimator of β.

Question 9

Answer the following questions:

(a) Explain what a trend stationary series and what a difference stationary series are. What is an important difference between the two types of stationarity?

(5 marks)

(b) Let the model be:

$$\boldsymbol{Y_t = \phi Y_{t-1} + u_t; \qquad t = 1, 2, \ldots, T}$$

where u_t are independently and identically distributed as $N(0, \sigma^2)$. Explain how would you test $\mathrm{H}_0 : |\phi| = 1$ against the alternative of a zero mean, covariance stationary AR(1) process. Give the assumptions this test requires.

(5 marks)

(c) Consider an ADL(1,1) model:

$$\boldsymbol{Y_t = \alpha_1 + \alpha_2 Y_{t-1} + \alpha_3 X_t + \alpha_4 X_{t-1} + u_t}$$

where both Y_t and X_t are I(1). Express the ADL(1,1) model in an error correction form and interpret the coefficients of the error correction model. Discuss the advantages of the error correction form.

(10 marks)

Reading for this question

Dougherty, C. *Introduction to Econometrics* (fourth edition) Chapters 13.1 (Stationarity and nontationarityy), 13.4 (Tests of nonstationarity) and 13.6 (Fitting models with nonstationary time series).

Dougherty, C. Subject guide Chapter 13 (Introduction to nonstationary time series).

Gujarati, D.N. and D.C. Porter *Basic Econometrics* (fifth edition) (ISBN 9780071276252): Chapters 21.5 (Trend Stationary (TS) and Difference Stationary (DS) Stochastic Processes) and 21.9 (The Unit Root Test).

Approaching the question

In part (a) the meaning of a trend stationary series and a difference stationaary series should be explained. Part (b) requires a discussion of the Dickey–Fuller test, and in part (c) the error correction model (ECM) should be derived and the interpretation of the coefficients of the ECM should be given. The answer is as follows.

(a) If after removing the trend from a nonstationary series the resulting variable becomes stationary, then the variable is called *trend stationary.* Let:

$$Z_t = X_t - \alpha_1 t = \alpha_0 + u_t$$

where $\text{E}(u_t) = 0$, $\text{var}(u_t) = \sigma^2$ and $\text{E}(u_t u_{t-s}) = 0$ for all s and t, then:

$$\begin{aligned} \text{E}(Z_t) &= \text{E}(\alpha_0 + u_t) = \alpha_0 \\ \text{var}(Z_t) &= \text{Var}(\alpha_0 + u_t) = \sigma^2 \\ \text{cov}(Z_t, Z_{t-s}) &= \text{E}[(Z_t - \text{E}(Z_t))(Z_{t-s} - \text{E}(Z_{t-s}))] = \text{E}(u_t u_{t-s}) = 0. \end{aligned}$$

This means that Z_t has constant mean and variance for all t, and covariance is zero for all $s > 0$. It implies that the series is trend stationary.

If a nonstationary process can be transformed into a stationary process by differencing then the series is said to be *difference stationary.*

Let X_t be a random walk with a drift:

$$X_t = \beta_0 + X_{t-1} + \varepsilon_t \qquad \text{(i)}$$

where $\text{E}(\varepsilon_t) = 0$, $\text{var}(\varepsilon_t) = \sigma^2$ and $\text{E}(\varepsilon_t \varepsilon_s) = 0$ for all s and t, $s \neq t$.

Subtract X_{t-1} from both sides of (i) to get:

$$\Delta X_t = X_t - X_{t-1} = \beta_0 + \varepsilon_t.$$

It can be easily checked that $\text{E}(\Delta X_t) = \beta_0$, $\text{var}(\Delta X_t) = \sigma_\varepsilon^2$ and $\text{cov}(\Delta X_t, \Delta X_{t-s}) = 0$ for all s and t, $s \neq t$. This means that ΔX_t is stationary. This implies that X_t is difference stationary. It is important to know whether a variable is difference stationary or trend stationary because for difference stationary variables shocks have a permanent effect whereas for trend stationary variables shocks are transitory.

(b) The standard test for a unit root is due to Dickey and Fuller. In order to test the null hypothesis of a random walk without drift against the alternative of a zero mean covariance stationary AR(1) process:

$$y_t = \phi y_{t-1} + \varepsilon_t$$

subtract y_{t-1} from both sides:

$$\begin{aligned} y_t - y_{t-1} &= \phi y_t - y_{t-1} + \varepsilon_t \\ \Delta y_t &= (\phi - 1) y_{t-1} + \varepsilon_t = \rho y_{t-1} + \varepsilon_t. \end{aligned}$$

Test $\text{H}_0 : \rho = 0$ using:

$$\widehat{\tau} = \frac{\widehat{\rho} - 0}{\text{s.e.}(\widehat{\rho})}.$$

We cannot use the standard t test procedure in this case because the distribution of the statistic is not a t distribution, so critical values have to be computed by Dickey and Fuller using Monte Carlo techniques (Dickey–Fuller tables). The test is sensitive to the presence of serial correlation in the error term so we need to take steps to remove the effects of this serial correlation – this is done by including lagged values of Δy_t in the regression. The statistic has the same asymptotic distribution as $\widehat{\tau}$.

(c) Consider a simple ADL(1,1) model (this is also known as ARDL(1, 1)):

$$Y_t = \alpha_1 + \alpha_2 Y_{t-1} + \alpha_3 X_t + \alpha_4 X_{t-1} + u_t \qquad \text{(i)}$$

Rewrite (i) as:

$$\begin{aligned} Y_t - Y_{t-1} &= \alpha_1 + \alpha_2 Y_{t-1} - Y_{t-1} + \alpha_3 X_t - \alpha_3 X_{t-1} + \alpha_3 X_{t-1} + \alpha_4 X_{t-1} + u_t \\ \Delta Y_t &= \alpha_1 - (1 - \alpha_2) Y_{t-1} + \alpha_3 \Delta X_t + (\alpha_3 + \alpha_4) X_{t-1} + u_t \\ \Delta Y_t &= \alpha_3 \Delta X_t - (1 - \alpha_2) \left[Y_{t-1} - \frac{\alpha_1}{1 - \alpha_2} - \frac{\alpha_3 + \alpha_4}{1 - \alpha_2} X_{t-1} \right] + u_t \\ \Delta Y_t &= \alpha_3 \Delta X_t - (1 - \alpha_2)[Y_{t-1} - \beta_1 - \beta_2 X_{t-1}] + u_t \end{aligned}$$

or:

$$\Delta Y_t = \alpha_3 \Delta X_t - \pi[Y_{t-1} - \beta_1 - \beta_2 X_{t-1}] + u_t \qquad \text{(ii)}$$

where:

$$\pi = 1 - \alpha_2, \qquad \beta_1 = \frac{\alpha_1}{1-\alpha_2} \qquad \text{and} \qquad \frac{\alpha_3 + \alpha_4}{1-\alpha_2}.$$

Equation (ii) is the ECM.

When the two variables Y and X are cointegrated, the ECM incorporates not only the short-run effects but also the long-run effects. The long-run equilibrium $Y_t - \beta_1 - \beta_2 X_{t-1}$ is included in the model together with the short-run effect captured by the differenced term.

All the terms in the ECM, given by (ii), are stationary. As Y and X are I(1), then ΔX and ΔY are I(0). As Y and X are cointegrated their linear combination:

$$u_t = Y_t - \beta_1 - \beta_2 X_{t-1}$$

is I(0).

The coefficient π provides us with the information about the speed of adjustment in cases of disequilibrium.

- If $\pi = 1$ then 100% of the adjustment takes place within the period. In other words adjustment is instantaneous and full.
- If $\pi = 0.5$ then 50% adjustment takes place each period.
- If $\pi = 0$ then there is no adjustment.

Question 10

(a) **Consider a model:**

$$\boldsymbol{Y_t = \beta X_t + u_t; \qquad t = 1, 2, \ldots, T}$$

where $E(u_t) = 0$; $E(u_t^2) = \sigma^2$ and $E(u_s u_t) = 0$ if $s \neq t$ for all $s, t = 1, 2, \ldots, T$. X_t is fixed in repeated samples. The density function of u_t is given by:

$$\boldsymbol{f(u_t) = (2\pi\sigma^2)^{-1/2} \exp\left(-\frac{u_t^2}{2\sigma^2}\right).}$$

i. **Derive the likelihood function.**

(6 marks)

ii. **Obtain the maximum likelihood estimators of β and σ^2.**

(8 marks)

(b) **Consider a model:**

$$\boldsymbol{Y_t = \alpha + \beta X_t + u_t; \qquad t = 1, 2, \ldots, T}$$

where $E(u_t) = 0$; $E(u_t^2) = \sigma^2$ and $E(u_s u_t) = 0$ if $s \neq t$ for all $s, t = 1, 2, \ldots, T$. u is normally distributed. The parameters α and β have been estimated by maximum likelihood.

Explain how the hypothesis that the coefficients are jointly equal to zero can be tested.

(6 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fourth edition) Chapter 10.6 (An introduction to maximum likelihood estimation).

Dougherty, C. Subject guide Chapter 10 (Binary choice and limited dependent variable models, and maximum likelihood estimation).

Approaching the question

Part (a) involves derivation of the likelihood function. To obtain the maximum likelihood estimators the likelihood function should be differentiated with respect to the parameters and equated to zero. Part (b) involves the likelihood-ratio test. The solution is as follows.

(a) i. The likelihood function is:

$$L = (2\pi\sigma^2)^{-T/2} \exp\left(-\frac{\sum(Y_t - \beta X_t)^2}{2\sigma^2}\right).$$

ii. The log-likelihood function is:

$$\log L = -\frac{T}{2}\ln(2\pi) - \frac{T}{2}\ln(\sigma^2) - \frac{\sum(Y_t - \beta X_t)^2}{2\sigma^2}.$$

Differentiating, we have:

$$\frac{\partial \log L}{\partial \beta} = -\frac{2\sum(Y_t - \beta X_t)(-X_t)}{2\sigma^2} = 0 \qquad \Rightarrow \qquad \widehat{\beta}_{MLE} = \frac{\sum X_t Y_t}{\sum X_t^2}$$

and:

$$\frac{\partial \log L}{\partial \sigma^2} = -\frac{T}{2\sigma^2} + \frac{\sum(Y_t - \beta X_t)^2}{2\sigma^4} = 0 \qquad \Rightarrow \qquad \widehat{\sigma}^2_{MLE} = \frac{\sum(Y_t - \widehat{\beta}_{MLE} X_t)^2}{T}.$$

(b) This can be tested using the likelihood-ratio statistic $2(\log L - \log L_0)$. This is asymptotically distributed as a chi-squared distribution with degrees of freedom equal to the number of restrictions imposed by the null hypothesis, 2 in the present case.

Examiners' commentaries 2016

EC2020 Elements of econometrics

Important note

This commentary reflects the examination and assessment arrangements for this course in the academic year 2015–16. The format and structure of the examination may change in future years, and any such changes will be publicised on the virtual learning environment (VLE).

Information about the subject guide and the Essential reading references

Unless otherwise stated, all cross-references will be to the latest version of the subject guide (2016). You should always attempt to use the most recent edition of any Essential reading textbook, even if the commentary and/or online reading list and/or subject guide refer to an earlier edition. If different editions of Essential reading are listed, please check the VLE for reading supplements – if none are available, please use the contents list and index of the new edition to find the relevant section.

Comments on specific questions – Zone B

Candidates should answer **EIGHT** of the following **TEN** questions: **ALL** of the questions in Section A (8 marks each) and **THREE** questions from Section B (20 marks each). Candidates are strongly advised to divide their time accordingly.

Section A

Answer all questions from this section.

Question 1

What do you understand by an instrumental variable (IV)? How you would estimate a model $y_t = \beta x_t + u_t$ using an IV z_t; $t = 1, 2, \ldots, T$? Examine the consistency of the IV estimator.

(8 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fourth edition): Chapter 8.5 (Instrumental variables).

Dougherty, C. Subject guide: Chapter 8.5 (Stochastic regressor and measurement errors).

Approaching the question

Consider the model:

$$y_t = \beta x_t + u_t, \qquad t = 1, 2, \ldots, T.$$

If x_t is not independently distributed of u_t then the OLS estimator of β will be inconsistent. In this situation an IV should be used. An IV should be:

- correlated with x
- should not be correlated with the disturbance term
- should not be an explanatory variable in its own right.

Let z be an IV. An estimator of β based on z is known as an IV estimator. It is defined as:

$$\widehat{\beta}_{IV} = \frac{\sum z_t y_t}{\sum z_t x_t}.$$

It can be shown that $\widehat{\beta}_{IV}$ is a consistent estimator of β.

$$\widehat{\beta}_{IV} = \frac{\sum z_t y_t}{\sum z_t x_t} = \frac{\sum z_t(\beta x_t + u_t)}{\sum z_t x_t} = \beta + \frac{\sum z_t u_t}{\sum z_t x_t}$$

and:

$$\text{plim}(\widehat{\beta}_{IV}) = \beta + \frac{\text{plim}\left(\sum z_t u_t/T\right)}{\text{plim}\left(\sum z_t x_t/T\right)} = \beta$$

hence $\widehat{\beta}_{IV}$ is a consistent estimator of β.

Note that:

$$\text{plim}\left(\sum z_t u_t/T\right) = 0 \quad \text{and} \quad \text{plim}\left(\sum z_t x_t/T\right) \neq 0.$$

Question 2

Derive the order of integration of x_t in the following models. Assume in each case that u_t is stationary, x_0 is fixed and $E(u_s u_t) = 0$ if $s \neq t$.

(a) $\boldsymbol{x_t = \alpha_0 + u_t + u_{t-1};\ t = 1, 2, \ldots, T.}$

(3 marks)

(b) $\boldsymbol{x_t = \alpha_0 + \alpha_1 x_{t-1} + \alpha_2 t + u_t;\ |\alpha| < 1;\ t = 1, 2, \ldots, T.}$

(5 marks)

Reading for this question

Dougherty, C. *Introduction to Econometrics* (fourth edition): Chapter 13.1 (Stationarity and nonstationarity).

Dougherty, C. Subject guide: Chapter 13 (Introduction to nonstationary time series).

Gujarati, D.N. and D.C. Porter *Basic Econometrics* (fifth edition): (ISBN 9780071276252), Chapters 21.5 (Trend Stationary (TS) and Difference Stationary (DS) Stochastic Processes) and 21.6 (Integrated Stochastic Process).

Approaching the question

In both (a) and (b), it should be shown that the mean and the variance of the variable is constant and the covariances are independent of time but may depend on the length of the lag. The solution is as follows.

(a) $E(x_t) = \alpha_0$ and $\text{var}(x_t) = \text{var}(u_t - u_{t-1}) = \text{var}(u_t) + \text{var}(u_{t-1}) = 2\sigma^2$ as $E(u_s u_t) = 0$ if $s \neq t$ and $E(x_t x_{t-s}) = \sigma^2$ if $s = 1$ and 0 if $s > 1$. Therefore, x_t is stationary or I(0).

(b) Lag this equation by one period to get:

$$x_{t-1} = \alpha_0 + \alpha_1 x_{t-2} + \alpha_2(t-1) + u_{t-1}.$$

Subtracting this from the first equation gives:

$$\Delta x_t = \alpha_1 \Delta x_{t-1} + \alpha_2 + u_t - u_{t-1}.$$

Since $|\alpha_1| < 1$, Δx_t follows a stationary AR(1) process.

As $\mathrm{E}(x_t) = \alpha_0 + \alpha_1 \mathrm{E}(x_{t-1}) + \alpha_2 t$, $\mathrm{E}(x_t)$ is clearly a function of time and $\mathrm{E}(x_t) \neq \mathrm{E}(x_{t-1})$. Therefore, x_t is non-stationary. Hence x_t is I(1) and it is trend stationary.

Question 3

Consider a model:

$$\boldsymbol{Y_i = \alpha + \beta X_i + u_i; \qquad i = 1, 2, \ldots, 6}$$

where $\mathrm{E}(u_i) = 0$; $\mathrm{E}(u_i^2) = \sigma^2$ and $\mathrm{E}(u_i u_j) = 0$ if $i \neq j$.

The observations on X_i are

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

The OLS estimator of β is $\widehat{\beta}$ and $\mathrm{V}(\widehat{\beta}) = \frac{\sigma^2}{17.5}$.

An alternative estimator of β is $\tilde{\beta} = \frac{1}{8}[Y_6 + Y_5 - Y_2 - Y_1]$.

Compare the sampling variance of $\tilde{\beta}$ with that of $\widehat{\beta}$.

(8 marks)

Reading for this question

Dougherty, C. *Introduction to Econometrics* (third edition) Chapters R.5, R.6, 1.4, 2.5, 12.3 and 12.4.

Gujarati, D.N. and D.C. Porter *Basic Econometrics* (fifth edition): (ISBN 9780071276252), Chapters 3.4 (Properties of Least Squares Estimators: The Gauss–Markov Theorem) and 3A.3 (Variances and Standard Errors of Least Squares Estimators).

Approaching the question

Replace the values of the X_is in the equation given by $\tilde{\beta}$. The variance of $\tilde{\beta}$ has to be calculated and compared with the variance of $\widehat{\beta}$ (the variance of $\widehat{\beta}$ is given). While deriving the variance, be careful to take into account that $\mathrm{E}(u_i u_j) = 0$ if $i \neq j$. The answer is as follows.

Replace the value of the X_is:

$$\begin{aligned}
\tilde{\beta} &= \frac{1}{8}[Y_6 + Y_5 - Y_2 - Y_1] \\
&= \frac{1}{8}[(\alpha + 6\beta + u_6) + (\alpha + 5\beta + u_5) - (\alpha + 2\beta + u_2) - (\alpha + \beta + u_1)] \\
&= \frac{1}{8}[8\beta + u_6 + u_5 - u_2 - u_1] \qquad \text{(i).}
\end{aligned}$$

From (i), it is easy to see that:

$$\begin{aligned}
\text{var}(\tilde{\beta}) &= \text{var}\left[\frac{1}{8}(u_6 + u_5 - u_2 - u_1)\right] \qquad (\text{since var}(8\beta) = 0) \\
&= \frac{1}{8^2}[\text{var}(u_6) + \text{var}(u_5) + \text{var}(u_2) + \text{var}(u_1)] \qquad (\text{since E}(u_i u_j) = 0 \text{ if } i \neq j) \\
&= \frac{4\sigma^2}{64} \\
&= \frac{\sigma^2}{16} \qquad (\text{since var}(u_i) = \sigma^2 \text{ for all } i).
\end{aligned}$$

Hence $\text{var}(\tilde{\beta}) > \text{var}(\widehat{\beta})$ and so $\widehat{\beta}$ is more efficient in comparison to $\tilde{\beta}$.

Question 4

Consider a model:

$$\boldsymbol{y_t = \alpha y_{t-1} + u_t}$$

where

$$\boldsymbol{u_t = \theta\varepsilon_{t-1} + \varepsilon_t; \qquad t = 1, 2, \ldots, T}$$

and $\boldsymbol{\text{E}(\varepsilon_t) = 0}$**,** $\boldsymbol{\text{E}(\varepsilon_t^2) = \sigma^2}$ **and** $\boldsymbol{\text{E}(\varepsilon_s\varepsilon_t) = 0}$ **if** $\boldsymbol{s \neq t}$ **for all** $\boldsymbol{s, t = 1, 2, \ldots, T}$**.**

Show that the OLS estimator of $\boldsymbol{\alpha}$ **is inconsistent if** $\boldsymbol{\theta \neq 0}$**.**

(8 marks)

Reading for this question

Dougherty, C. *Introduction to Econometrics* (fourth edition): Chapter 12 (Definition and consequences of autocorrelation).

Dougherty, C. Subject guide: Chapter 12 (Properties of regression models with time series data).

Approaching the question

The model contains a lagged dependent variable as an explanatory variable and also has autocorrelation. The ordinary least squares estimator will be inconsistent. To prove inconsistency, plim should be used. The answer is as follows.

The OLS estimator of α is:

$$\widehat{\alpha} = \frac{\sum y_t y_{t-1}}{\sum y_{t-1}^2} = \alpha + \frac{\sum u_t y_{t-1}}{\sum y_{t-1}^2} = \alpha + \frac{\sum \varepsilon_t y_{t-1}}{\sum y_{t-1}^2} + \frac{\theta \sum \varepsilon_{t-1} y_{t-1}}{\sum y_{t-1}^2}$$

where y_t has been substituted as $y_t = \alpha y_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1}$.

Since $y_{t-1} = \alpha y_{t-2} + \varepsilon_{t-1} + \theta\varepsilon_{t-2}$ hence:

$$\text{plim}\,\frac{\sum \varepsilon_t y_{t-1}}{T} = 0.$$

It is easy to check that:

$$\text{plim}\,\frac{\sum \varepsilon_{t-1} y_{t-1}}{T} \neq 0.$$

It is also assumed:

$$\text{plim}\,\frac{\sum y_{t-1}^2}{T}$$

is greater than zero and finite.

Therefore, $\text{plim}(\widehat{\alpha}) \neq \alpha$, and so $\widehat{\alpha}$ is inconsistent.

Question 5

Consider an ADL(1, 1) model:

$$\boldsymbol{Y_t = \alpha_1 + \alpha_2 Y_{t-1} + \alpha_3 X_t + \alpha_4 X_{t-1} + u_t; \qquad t = 1, 2, \ldots, T.}$$

where both $\boldsymbol{Y_t}$ and $\boldsymbol{X_t}$ are I(1). $\boldsymbol{u_t}$ is the disturbance term where $\boldsymbol{\mathrm{E}(u_t) = 0}$; $\boldsymbol{\mathrm{E}(u_t^2) = \sigma^2}$ and $\boldsymbol{\mathrm{E}(u_s u_t) = 0}$ for all $\boldsymbol{s \neq t}$. Express the ADL(1, 1) model in an error correction form and interpret the coefficients of the error correction model.

(8 marks)

Reading for this question

Dougherty, C. *Introduction to Econometrics* (fourth edition): Chapter 13.6 (Fitting models with nonstationary time series).

Gujarati, D.N. and D.C. Porter, *Basic Econometrics* (fifth edition): (ISBN 9780071276252), Chapter 21.11 (Cointegration: Regression of a Unit Root Time Series on Another Unit Root Time Series).

Approaching the question

The ADL(1,1) model should be expressed in error correction form. It involves some simple algebraic manipulations. A detailed interpretation of the coefficients of the error correction model should be given. The answer is as follows.

Consider a simple ADL(1,1) model (this is also known as ARDL(1, 1)):

$$Y_t = \alpha_1 + \alpha_2 Y_{t-1} + \alpha_3 X_t + \alpha_4 X_{t-1} + u_t \qquad \text{(i)}$$

Rewrite (i) as:

$$\begin{aligned}
Y_t - Y_{t-1} &= \alpha_1 + \alpha_2 Y_{t-1} - Y_{t-1} + \alpha_3 X_t - \alpha_3 X_{t-1} + \alpha_3 X_{t-1} + \alpha_4 X_{t-1} + u_t \\
\Delta Y_t &= \alpha_1 - (1-\alpha_2) Y_{t-1} + \alpha_3 \Delta X_t + (\alpha_3 + \alpha_4) X_{t-1} + u_t \\
\Delta Y_t &= \alpha_3 \Delta X_t - (1-\alpha_2)\left[Y_{t-1} - \frac{\alpha_1}{1-\alpha_2} - \frac{\alpha_3+\alpha_4}{1-\alpha_2} X_{t-1}\right] + u_t \\
\Delta Y_t &= \alpha_3 \Delta X_t - (1-\alpha_2)[Y_{t-1} - \beta_1 - \beta_2 X_{t-1}] + u_t
\end{aligned}$$

or:

$$\Delta Y_t = \alpha_3 \Delta X_t - \pi[Y_{t-1} - \beta_1 - \beta_2 X_{t-1}] + u_t \qquad \text{(ii)}$$

where:

$$\pi = 1 - \alpha_2, \qquad \beta_1 = \frac{\alpha_1}{1-\alpha_2} \qquad \text{and} \qquad \frac{\alpha_3 + \alpha_4}{1-\alpha_2}.$$

Equation (ii) is the ECM.

When the two variables Y and X are cointegrated, the ECM incorporates not only the short-run effects but also the long-run effects. The long-run equilibrium $Y_t - \beta_1 - \beta_2 X_{t-1}$ is included in the model together with the short-run effect captured by the differenced term.

All the terms in the ECM, given by (ii), are stationary. As Y and X are I(1), then ΔX and ΔY are I(0). As Y and X are cointegrated their linear combination:

$$u_t = Y_t - \beta_1 - \beta_2 X_{t-1}$$

is I(0).

The coefficient π provides us with the information about the speed of adjustment in cases of disequilibrium.

- If $\pi = 1$ then 100% of the adjustment takes place within the period. In other words adjustment is instantaneous and full.
- If $\pi = 0.5$ then 50% adjustment takes place each period.
- If $\pi = 0$ then there is no adjustment.

Section B

Answer three questions from this section.

Question 6

(a) Describe an adaptive expectations model.

(6 marks)

(b) Koyck investigated the relationship between investment in railcars and the volume of freight carried on the U.S. railroads using annual data for the period 1884–1939. Assuming that the desired stock of railcars in year t depended on the volume of the freight in year $t-1$ and $t-2$ and a time trend, and assuming that investment in railcars was subject to a partial adjustment process, he fitted the following regression using ordinary least squares:

$$\widehat{I_t} = 0.077F_{t-1} + 0.017F_{t-2} - 0.003t - 0.110K_{t-1}; \qquad R^2 = 0.85; \quad t = 1, 2, \ldots, T$$

where $I_t = K_t - K_{t-1}$ is investment in railcars in year t (thousands), K_t is the stock of railcars at the end of year t (thousands), and F_t is the volume of freight handled in year t (ton-miles).

i. Explain how Koyck's model can be derived from an adjustment equation $K_t - K_{t-1} = \lambda(K_t^* - K_{t-1})$ and a behavioural equation with dependent variable K_t^*, using variables F (lagged) and t (time).

(6 marks)

ii. Using Koyck's estimated equation estimate the parameters of your behavioural equation. What are the implications for the behaviour in the long run?

(8 marks)

Reading for this question

Dougherty, C. *Introduction to Econometrics* (fourth edition) Chapter 11.4 (Models with a lagged dependent variable).

Dougherty, C. Subject guide: Chapter 11 (Models using time series data).

Gujarati, D.N. and D.C. Porter *Basic Econometrics* (fifth edition): (ISBN 9780071276252), Chapter 17.5 (Rationalization of the Koyck Model: The Adaptive Expectations Model).

Approaching the question

(a) An adaptive expectations model involves a learning process in which, in each time period, the actual value of the variable is compared with the value that has been expected. If the actual value is greater, the expected value is adjusted upwards for the next period. If it is lower, the expected value is adjusted downwards. The size of the adjustment is hypothesised to be proportional to the discrepancy between the actual and expected value.

If X is the variable in question, and X_t^e is the value expected in time period t given the information available at time period $t-1$, then:

$$X_{t+1}^e - X_t^e = \lambda(X_t - X_t^e) \qquad 0 \leq \lambda \leq 1$$

or:

$$X_{t+1}^{e} = \lambda X_t + (1-\lambda)X_t^{e}.$$

The model is derived to capture the changing nature of the formation of expectations, often in variables that are also changing with time. It is an attempt at a 'simple learning' solution to model building in order to forecast macroeconomic variables. Such variables include investment, savings and the demand for assets.

The model is estimated by repeated substitution for the expected variable, by its lagged variant which has known components of the previous period and the unobserved expectation lagged, until the term on the unobserved expectation $(1-\lambda)^s$ is so small as to be ignored resulting in a model where all the variables are observed, where s is the period lagged and λ is the speed of adjustment of expected and actual and λ is between 0 and 1. Technical details should be given.

(b) i. Given the information in the question, the model may be written as:

$$K_t^* = \beta_1 F_{t-1} + \beta_2 F_{t-2} + \beta_3 t + u_t.$$

It is given that:

$$K_t - K_{t-1} = I_t = \lambda(K_{t-1}^* - K_{t-1}).$$

Hence:

$$I_t = \lambda\beta_1 F_{t-1} + \lambda\beta_2 F_{t-2} + \lambda\beta_3 t - \lambda K_{t-1} + \lambda u_t.$$

ii. Estimates of the coefficients should be obtained and interpreted. The working and interpretations are given below.

From the fitted equation:

$$\widehat{\lambda} = 0.110, \quad \widehat{\beta}_1 = \frac{0.077}{0.110} = 0.70, \quad \widehat{\beta}_2 = \frac{0.017}{0.110} = 0.15 \quad \text{and} \quad \widehat{\beta}_3 = -\frac{0.0033}{0.110} = -0.030.$$

Hence the short-run effect of an increase of 1 million ton-miles of freight is to increase investment in rail-cars by 77 one year later and 17 two years later. It does not make much sense to talk of the short-run effect of a time trend.

In the long-run equilibrium, neglecting the effects of the disturbance term, K_t and K_t^* are both equal to the equilibrium value $\bar{K}$, and F_{t-1} and F_{t-2} are both equal to their equilibrium value $\bar{F}$. Hence:

$$\bar{K} = (\beta_1 + \beta_2)\bar{F} + \beta_3 t.$$

Therefore, an increase of one million ton-miles of freight will increase the stock of rail-cars by 850 and the time trend will be responsible for a secular decline of 33 rail-cars per year (obsolescence/damage).

Question 7

The following estimates were calculated from a sample of 7,634 women respondents from the General Household Survey 1995. The dependent variable takes the value 1 if the woman was in paid employment, and 0 otherwise.

	OLS	Logit	Probit
high	0.093	0.423	0.259
	(0.015)	(0.071)	(0.043)
noqual	−0.210	−0.898	−0.554
	(0.013)	(0.056)	(0.035)
age	0.038	0.173	0.107
	(0.003)	(0.124)	(0.008)
age2	−0.051	−0.230	−0.142
	(0.003)	(0.069)	(0.009)
mar	0.024	0.103	0.063
	(0.009)	(0.057)	(0.035)
Constant	−0.068	−2.587	−1.593
	(0.049)	(0.225)	(0.137)

Where high is one if the respondent has a higher educational qualification, zero otherwise; noqual is one if the respondent has no qualifications, zero otherwise; age is age in years; age2 is (age $\times$ age) 100; mar is one if married, zero otherwise. Conventionally calculated standard errors are given in brackets for the ordinary least squares (OLS) results and asymptotic standard errors in brackets elsewhere.

(a) Explain briefly how Probit estimates are calculated when the model has no intercept and only one explanatory variable.

(6 marks)

(b) For all three sets of estimates, test the null hypothesis that the coefficient of mar is zero. Which test statistics would you consider more reliable? Explain.

(6 marks)

(c) Using the OLS and Probit estimates, calculate the estimated probabilities of being in employment for a married woman aged 40 with a higher educational qualification. Comment on your results.

(5 marks)

(d) Test the null hypothesis that all the slope coefficients of the probit model are jointly equal to zero, given that:

$$\ln L_R = -416.01$$

$$\ln L_U = -321.25$$

where $\ln L_R$ and $\ln L_U$ are the logs of the likelihood from the restricted and the unrestricted probit models, respectively.

(3 marks)

Reading for this question

Dougherty, C. *Introduction to Econometrics* (fourth edition) Chapters 10.1 (The linear probability model), 10.3 (Probit analysis) and 10.6 (An introduction to maximum likelihood estimation).

Dougherty, C. Subject guide (2011) Chapter 10 (Binary choice and limited dependent variable models, and maximum likelihood estimation).

Gujarati, D.N. and D.C. Porter *Basic Econometrics* (fifth edition): (ISBN 9780071276252), Chapters 15.2 (The Linear Probability Model (LPM)), 15.5 (The Logit Model) and 15.9 (The Probit Model).

Approaching the question

(a) The probit model uses the cumulative standardised normal distribution. The maximum likelihood technique is used to obtain the estimates of the parameters. Estimates have the standard maximum likelihood properties, i.e. the estimators are consistent, asymptotically efficient and asymptotically normally distributed.

[For technical details see Dougherty (fourth edition) Section 10.3].

(b) To test the null hypothesis the t test should be used. t tables are attached with the examination paper. Candidates should know how to look at the critical values. The null and alternative hypotheses should be clearly stated:

$$\text{H}_0 : \text{coefficient of mar} = 0$$

$$\text{H}_1 : \text{coefficient of mar} \neq 0.$$

t test statistics are 2.67 (OLS), 1.81 (logit), and 1.8 (probit).

The test based on the OLS estimates gives outright rejection. However, the standard errors for this test are wrongly calculated because of the heteroskedasticity of the error term. Therefore, we prefer probit or logit estimates.

(c) A large number of candidates wrongly thought that to obtain the probability in the case of probit a sophisticated calculator is needed. The probability in the case of probit can be read directly from the supplied statistical tables. We have:

$$0.093 + 0.038 \times 40 - 0.051 \times 16 + 0.024 \times 0.068 = 0.753 \qquad \text{(OLS)}$$

$$0.259 + 0.107 \times 40 - 0.142 \times 16 + 0.063 - 1.593 = 0.737 \qquad \text{(probit).}$$

The statistical tables give a probability of 0.77, approximately.

They are all fairly close. The OLS estimates do not fall outside the probability bounds.

(d) To test the null hypothesis, the large sample likelihood-ratio test should be used. In large samples $-2(\ln L_R - \ln L_U)$ is distributed as a chi-squared distribution with degrees of freedom equal to the number of restrictions imposed by the null hypothesis. The null and alternative hypotheses should be clearly stated as:

$$\text{H}_0 : \text{all the slope coefficients are} = 0$$

$$\text{H}_1 : \text{at least one slope coefficient is} \neq 0.$$

We have:

$$-2(\ln L_R - \ln L_U) \sim \chi^2_5$$

and:

$$-2(-416.01 - (-321.25)) = 189.92.$$

The critical value of χ^2_5 at the 5% significance level is 11.07, hence we reject H_0.

Question 8

Economists have tried to examine the catch-up hypothesis where it is predicted that poorer countries will grow faster than richer countries to 'catch-up' to the richer countries. To assess this hypothesis an economist regresses the growth rate of GDP (gross domestic product), gr_i, on log of GDP per capita, $\ln(gdp_i)$, as well as a regression including other variables which might influence growth rates. The following regression results come from regressions using data from 96 countries:

$$\begin{aligned} gr_i &= \underset{(0.023)}{-0.058} - \underset{(0.002)}{0.009} \ln(gdp_i) + e_{1i} \qquad \text{(A)} \end{aligned}$$

$$n = 96, R^2 = 0.122 \text{ and } s = 0.025$$

where estimated standard errors are given in brackets, gr_i is the growth rate of the ith country over the period 1985–1990, $\ln(gdp_i)$ is the logarithm of GDP for country i, e_{1i} is the residual and s is the estimated standard error of the residuals.

A second regression yields

$$\underset{}{gr_i} = \underset{}{-0.019} + \underset{(0.023)}{0.001}\ln(gdp_i) + \underset{(0.002)}{0.12}inv_i + e_{2i} \qquad \text{(B)}$$

$$n = 96, R^2 = 0.196 \text{ and } s = 0.024$$

where e_{2i} is the estimated residual and inv_i is the ratio of investment to GDP in 1985.

(a) Test, for each regression, the hypothesis that the coefficient on $\ln(gdp_i)$ is zero. What does the result of your tests say about the 'catch-up' hypothesis?

(7 marks)

(b) The R^2 measure is higher for equation (B) than for (A). Is this what you would expect? Explain.

(5 marks)

(c) The regression results may be affected by heteroskedasticity. Explain what you understand about heteroskedasticity, why the estimation results might be affected by heteroskedasticity and what the effects are if heteroskedasticity is present.

(8 marks)

Reading for this question

Dougherty, C. *Introduction to Econometrics* (fourth edition) Chapters 3.3 (Properties of multiple regression coefficients), 3.5 (Goodness of fit: R^2) and 7.1 (Heteroscedasticity and its implications).

Dougherty, C. Subject guide: Chapters 3 (Multiple regression analysis) and 7 (heteroscedasticity).

Gujarati, D.N. and D.C. Porter *Basic Econometrics* (fifth edition): (ISBN 9780071276252), Chapters 3.5 (The Coefficient of Determination R^2: A Measure of 'Goodness of Fit'), 11.1 (The Nature of Heteroscedasticity), 11.2 (OLS Estimation in the Presence of Heteroscedasticity) and 13.9 (Model Selection Criteria).

Approaching the question

(a) It is required to conduct tests of significance and relate the results of the tests to the 'catch-up' hypothesis.

The t-values are (A) 4.5 and (B) 0.043. The degrees of freedom are (A) 94 and (B) 93 giving 95% two-tailed critical values of ± 1.99. We reject the null hypothesis for (A), but do not reject for (B). A better answer would be to use a one-tailed test since the alternative hypothesis is that as GDP increases, the growth rate will fall, hence the coefficient will be negative, but the answer will be the same although the critical value is now -1.66. The results show that (A) supports the 'catch-up' hypothesis, but (B) does not. If (B) is the true model then (A) will suffer from omitted variable bias so that the 'catch-up' hypothesis is in doubt.

(b) It should be explained that R^2 is a non-decreasing function of the number of explanatory variables in the model. Candidates should show that as the number of explanatory variables is increased, R^2 will never decrease.

R^2 is defined as:

$$R^2 = \frac{\text{explained sum of squares}}{\text{total sum of squares}}.$$

As extra explanatory variables are added to an equation the explained sum of squares increases or does not change. Total sum of squares does not change, hence R^2 will never decrease. The result is as expected.

(c) The definition of heteroskedasticity is required. Candidates should explain the consequences of heteroskedasticity on the properties of ordinary least squares estimators.

The data are cross-sectional and deal with countries which are probably at very different levels of GDP, hence heteroskedasticity is to be expected. Heteroskedasticity is the condition where the error term variance is not constant and ordinary least squares produces unbiased and consistent, but inefficient, parameter estimates. Also standard errors are not correct hence making t and F tests invalid.

Question 9

(a) Explain the concept of the likelihood function and the maximum likelihood estimator.

(6 marks)

(b) Suppose that an event occurs with probability p. In a simple random sample of size n the event occurs m times.

i. Show that the maximum likelihood estimator (MLE) of p is $\frac{m}{n}$. Verify that the MLE maximises the likelihood function.

(8 marks)

ii. Derive the likelihood ratio statistic for the null hypothesis $p = p_0$. If $m = 40$ and $n = 100$, test the null hypothesis $p = 0.5$.

(6 marks)

Reading for this question

Dougherty, C. *Introduction to Econometrics* (fourth edition) Chapter 10.6 (An introduction to maximum likelihood estimation).

Dougherty, C. Subject guide: Chapter 10 (Binary choice and limited dependent variable models, and maximum likelihood estimation).

Gujarati, D.N. and D.C. Porter *Basic Econometrics* (fifth edition): (ISBN 9780071276252), Chapters 4.4 (The Method of Maximum Likelihood), 4A.1 (Maximum Likelihood Estimation of Two-Variable Regression Model) and 8A2 (Likelihood Ratio (LR) Test).

Approaching the question

(a) It is required to explain the concept of the likelihood function and the maximum likelihood estimator (MLE).

Definition: Let the probability density function (pdf) of X be $f(x;\theta)$ where θ is a parameter, then the likelihood function of a simple random sample $\{X_1, X_2, \dots, X_n\}$ from X is the product of the individual densities of the X_is taken as a function of θ.

The joint pdf of $\{X_1, X_2, \dots, X_n\}$ is:

$$f(x_1, x_2, \dots, x_n; \theta) = f(x_1;\theta)\, f(x_2, \theta) \cdots f(x_n;\theta).$$

The likelihood function $L(\theta; x_1, x_2, \dots, x_n)$ has the same formulation as the joint pdf, but now it is a function of θ.

Maximum Likelihood Estimator (MLE)

Let $X \sim f(x;\theta)$. The MLE $\widehat{\theta}$ of the parameter θ is an estimator that maximises the likelihood function.

To obtain the MLE set:

$$\frac{\partial L}{\partial \theta} = 0$$

and solve for $\widehat{\theta}$.

It is easier to take logs and then maximise, as in both the situations we will arrive at the same maximum, because:

$$\frac{\partial \ln L}{\partial \theta} = \frac{1}{L}\frac{\partial L}{\partial \theta} = 0 \qquad \Rightarrow \qquad \frac{\partial L}{\partial \theta} = 0.$$

To verify the maximum, the second derivative evaluated at the value of the MLE should be checked. It should be negative.

(b) To test the null hypothesis a large sample likelihood ratio test should be used.

i. p is the probability of the event occurring, hence $(1-p)$ is the probability that the event does not occur. The event occurs m times and does not occur $(n-m)$ times (n is the total number of observations).

The joint probability of the event occurring and not occurring is $p^m\,(1-p)^{n-m}$. The log-likelihood function is:

$$\log L(p) = m\log(p) + (n-m)\log(1-p).$$

Differentiating with respect to p we get:

$$\frac{\mathrm{d}\log p}{\mathrm{d}p} = \frac{m}{p} - \frac{n-m}{1-p}.$$

Equating it to zero we obtain the maximum likelihood estimator $\widehat{p}$ of p as:

$$\widehat{p} = \frac{m}{n}.$$

We should check that the second differential is negative to verify the maximum:

$$\frac{\mathrm{d}^2\log p}{\mathrm{d}p^2} = -\frac{m}{p^2} - \frac{n-m}{(1-p)^2}.$$

Evaluating it at $p = m/n$ we obtain:

$$\frac{\mathrm{d}^2\log p}{\mathrm{d}p^2} = -\frac{n^2}{m} - \frac{n-m}{(1-m/n)^2} < 0.$$

So we have chosen a value of p which maximises the likelihood function.

ii. The log-likelihood function is:

$$\log L(p) = m\log(p) + (n-m)\log(1-p).$$

The likelihood-ratio statistic is:

$$\begin{aligned} LR &= 2\left[\left(m\log\frac{m}{n} + (n-m)\log\left(1-\frac{m}{n}\right)\right) - (m\log p_0 + (n-m)\log(1-p_0))\right] \\ &= 2\left[m\log\left(\frac{m/n}{p_0}\right) + (n-m)\log\left(\frac{1-m/n}{1-p_0}\right)\right]. \end{aligned}$$

If $m = 40$ and $n = 100$ the LR statistic for $\mathrm{H}_0 : p = 0.5$ is:

$$LR = 2\left[40\log\left(\frac{0.4}{0.5}\right) + 60\log\left(\frac{1-0.4}{1-0.5}\right)\right] = 4.03.$$

The critical value of χ^2_1 at the 5% significance level is 3.84, and at the 1% significance level is 6.64. We would reject H_0 at the 5% but not at the 1% significance level.

Question 10

(a) **There are six missing values (denoted by x's) in the given Stata output**

  i. **Model SS**

  ii. **Model df**

  iii. ***R*-squared**

  iv. ***avetemp t* value**

  v. ***harvrain* 95% Conf. Interval.**

Give the formulae for them and obtain their values based on the given output.

```
      Source |       SS       df       MS              Number of obs =      27
-------------+------------------------------           F(  3,    23) =   21.90
       Model |  xxxxxxxxxx    xx  2.58519261           Prob > F      =  0.0000
    Residual |  2.71468685    23  .118029863           R-squared     =  xxxxxx
-------------+------------------------------           Adj R-squared =  0.7069
       Total |  10.4702647    26  .402702488           Root MSE      =  .34355
------------------------------------------------------------------------------
     lnprice |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    wintrain |    .001282   .0005765     2.22   0.036      .0000894    .0024747
     avetemp |   .7123178   .1087674     xxxx   0.000      .4873154    .9373202
    harvrain |  -.0036242   .0009646    -3.76   0.001     xxxxxxxxx   xxxxxxxxx
       _cons |  -13.44433   1.969396    -6.83   0.000     -17.51834   -9.370326
------------------------------------------------------------------------------
```

where the explanatory variables are

***wintrain*: the level of winter rain (October – March) in millimetres**

***avetemp*: average temperature in the growing season (April – September)**

***harvrain*: level of harvest rain (August – September) in millimetres**

and the dependent variable (*lnprice*) is the price of mature red wine from the Bordeaux region of France at harvest time. The data relate to wines of different vintages (ages) in 1987.

(10 marks)

(b) **Interpret the regression results in (a).**

(3 marks)

(c) **If the age of the wine is included in the equation the estimates become:**

```
      Source |       SS       df       MS              Number of obs =      27
-------------+------------------------------           F(  4,    22) =   26.39
       Model |  8.66443586     4  2.16610897           Prob > F      =  0.0000
    Residual |  1.80582883    22  .082083129           R-squared     =  0.8275
-------------+------------------------------           Adj R-squared =  0.7962
       Total |  10.4702647    26  .402702488           Root MSE      =   .2865
------------------------------------------------------------------------------
     lnprice |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    wintrain |   .0011668    .000482     2.42   0.024      .0001671    .0021665
     avetemp |   .6163926   .0951755     6.48   0.000      .4190107    .8137745
    harvrain |  -.0038606   .0008075    -4.78   0.000     -.0055353   -.0021858
         age |   .0238474   .0071667     3.33   0.003     0.0089846    .0387103
       _cons |  -12.31227   1.677212    -7.34   0.000     -15.79059   -8.833945
------------------------------------------------------------------------------
```

  i. **What is the interpretation of the coefficient of the variable age?**

(3 marks)

ii. If the true model is given by the specification in (c) do you think that the model in (a) should show evidence of specification error? Does it? Carefully explain your observations.

(4 marks)

Reading for this question

Dougherty, C. *Introduction to Econometrics* (fourth edition) Chapters 4.2 (Logarithmic Transformations) and 6.2 (The effect of omitting a variable that ought to be included).

Dougherty, C. Subject guide: Chapters 4 (Transformation of variables) and 6 (Specification of regression variables).

Gujarati D.N. and D.C. Porter *Basic Econometrics* (fifth edition): (ISBN 9780071276252), Chapters 6.6 (Semilog Models: Log-Lin and Lin-Log Models) and 13.3 (Consequences of Model Specification Errors).

Approaching the question

(a) It is required to calculate the missing values. Formulae for each must be given.

  i. Model SS = 7.7556.

  ii. R-squared = 0.741.

  iii. Model df = 3.

  iv. *avetemp* t value = 6.55.

  v. *harvrain* 95% CI = $[-0.0056, -0.0016]$.

(b) The dependent variable is in log form. The answer must explain the implication of this.

The critical t value is $t_{27-4} = 2.069$ hence all coefficients are significantly different from 0. *wintrain* increases the log price by 0.001 per millimetre, *avetemp* increases log price by 0.712 per degree and *harvrain* reduces log price by 0.0036 per millimetre.

(c) In answering part (ii), it should be mentioned that *age* appears to be part of the true model. Hence we have an omitted variable. The sign and size of omitted variables bias should be discussed.

The model in (a) has a specification error in that *age* appears to be part of the true model. The omitted variable bias will be related to the true coefficient and the correlation between the omitted variable and the remaining variables – the true coefficient is presumably non-zero and positive, but *age* is unlikely to be correlated to any of the included variables, hence one would not expect any measurable bias – and this is what you observe. The difference between the two models is very small.

Examiners' commentaries 2017

EC2020 Elements of econometrics

Important note

This commentary reflects the examination and assessment arrangements for this course in the academic year 2016–17. The format and structure of the examination may change in future years, and any such changes will be publicised on the virtual learning environment (VLE).

Information about the subject guide and the Essential reading references

Unless otherwise stated, all cross-references will be to the latest version of the subject guide (2016). You should always attempt to use the most recent edition of any Essential reading textbook, even if the commentary and/or online reading list and/or subject guide refer to an earlier edition. If different editions of Essential reading are listed, please check the VLE for reading supplements – if none are available, please use the contents list and index of the new edition to find the relevant section.

General remarks

Learning outcomes

At the end of the course, and having completed the Essential reading and activities, you should be able to:

- describe and apply the classical regression model and its application to cross-section data
- describe and apply the:
  - Gauss–Markov conditions and other assumptions required in the application of the classical regression model
  - reasons for expecting violations of these assumptions in certain circumstances
  - tests for violations
  - potential remedial measures, including, where appropriate, the use of instrumental variables
- recognise and apply the advantages of logit, probit and similar models over regression analysis when fitting binary choice models
- competently use regression, logit and probit analysis to quantify economic relationships using standard regression programmes (Stata and EViews) in simple applications
- describe and explain the principles underlying the use of maximum likelihood estimation
- apply regression analysis to time-series models using stationary time series, with awareness of some of the econometric problems specific to time series applications (for example, autocorrelation) and remedial measures
- recognise the difficulties that arise in the application of regression analysis to nonstationary time series, know how to test for unit roots, and know what is meant by cointegration.

Common mistakes committed by candidates

A large number of candidates are not able to clearly distinguish between sample variance and covariance, and population variance and covariance (this is happening year after year).

The use of $\text{Cov}(X, Y)$ and $\text{Var}(X)$ should be restricted to describing the population covariance and variances, respectively, with definitions:

$$\text{Cov}(X, Y) = \text{E}\left((X - \text{E}(X))(Y - \text{E}(Y))\right) = \text{E}(XY) - \text{E}(X)\,\text{E}(Y)$$

and:

$$\text{Var}(X) = \text{E}((X - \text{E}(X))^2) = \text{E}(X^2) - (\text{E}(X))^2$$

(you also may denote $\text{Cov}(X, Y) = \sigma_{XY}$ and $\text{Var}(X) = \sigma_X^2$). They are typically unknown, but fixed, quantities.

The sample covariance and variance are estimators of the population covariance and variance, respectively. They are defined as:

$$\text{Sample Cov}(X, Y) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})$$

and:

$$\text{Sample Var}(X) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

(you also may use $\widehat{\sigma}_{XY}$ and $\widehat{\sigma}_X^2$). You can compute them given the data.

With a slight abuse of notation, we often divide by n instead, which is irrelevant if we let n be large. The division by $n - 1$ is a finite sample issue only (unbiasedness).

The sample covariance and variance show up in our definition of the OLS estimator of the slope in the simple linear regression model, not the population covariance and variance, as:

$$\widehat{\beta} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \frac{\text{Sample Cov}(X, Y)}{\text{Sample Var}(X)} \neq \frac{\text{Cov}(X, Y)}{\text{Var}(X)}.$$

Treating them as being the same results in incorrect analyses and candidates losing significant marks.

Candidates should realise that $\frac{1}{n}\sum_{i=1}^{n} u_i$ is not the same as $\text{E}(u_i)$. So, while we typically assume $\text{E}(u_i) = 0$, this does not guarantee that $\frac{1}{n}\sum_{i=1}^{n} u_i = 0$. Also, while we may be happy to assume $\text{E}(x_i u_i) = 0$ (uncorrelatedness between the errors and regressors), this does not guarantee that $\frac{1}{n}\sum_{i=1}^{n} x_i u_i = 0$. Note that:

- both $\frac{1}{n}\sum_{i=1}^{n} u_i$ and $\frac{1}{n}\sum_{i=1}^{n} x_i u_i$ are random variables, which take the value 0 with probability 0 (continuous random variables)!
- $\text{E}(u_i) = 0$ and $\text{E}(x_i u_i) = 0$ are fixed, not stochastic!

The differences between sample and population moments need to come across clearly when looking at unbiasedness and making consistency arguments. In both cases, we first simplify our estimator (plug in the true model) to obtain:

$$\widehat{\beta} = \beta + \frac{\sum_{i=1}^{n}(X_i - \bar{X})u_i}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \beta + \frac{\sum_{i=1}^{n} x_i u_i}{\sum_{i=1}^{n} x_i^2} \qquad \text{with } x_i = X_i - \bar{X}.$$

- For *unbiasedness*, clearly indicate that you want to show that $\mathrm{E}(\widehat{\beta}) = \beta$. Unbiasedness does not follow from $\sum_{i=1}^{n} x_i u_i = 0$, instead it follows from $\mathrm{E}\left(\frac{\sum_{i=1}^{n} x_i u_i}{\sum_{i=1}^{n} x_i^2}\right) = 0$.

  If we treat x_i as fixed, $\mathrm{E}\left(\frac{\sum_{i=1}^{n} x_i u_i}{\sum_{i=1}^{n} x_i^2}\right) \equiv \mathrm{E}\left(\sum_{i=1}^{n} d_i u_i\right) = \sum_{i=1}^{n} d_i \,\mathrm{E}(u_i)$ and then unbiasedness follows as $\mathrm{E}(u_i) = 0$.
- For *consistency*, clearly indicate that you want to show that $\mathrm{plim}(\widehat{\beta}) = \beta$. Using the plim properties, we show:

$$\begin{aligned}
\mathrm{plim}\ \widehat{\beta} = \beta + \mathrm{plim}\left(\frac{\sum_{i=1}^{n} x_i u_i}{\sum_{i=1}^{n} x_i^2}\right) &= \beta + \frac{\mathrm{plim}\left(\frac{1}{n}\sum_{i=1}^{n} x_i u_i\right)}{\mathrm{plim}\left(\frac{1}{n}\sum_{i=1}^{n} x_i^2\right)} \\
&\equiv \beta + \frac{\mathrm{plim}\,(\text{Sample Cov}\,(x,u))}{\mathrm{plim}\,(\text{Sample Var}\,(x))} \\
&= \beta + \frac{\mathrm{Cov}\,(x,u)}{\mathrm{Var}\,(x)} \qquad \text{using the law of large numbers}
\end{aligned}$$

  where $\mathrm{Cov}(x,u) = 0$ and $\mathrm{Var}(x) > 0$, ensuring we get consistency.
- Remember, the law of large numbers ensures that sample averages converge to their population analogues.

Candidates struggled to give competent answers to the interpretation of empirical results. When interpreting an empirical result you should discuss the significance of the coefficients, magnitude and sign of the coefficients.

When conducting hypothesis tests, you should make sure that the Gauss–Markov conditions hold. The Gauss–Markov conditions have to be explicitly specified. Only writing that the Gauss–Markov conditions hold is not sufficient. As good practice, begin your examination by explicitly providing the Gauss–Markov conditions. You can then refer back to them thereafter. Moreover, ensure when conducting hypothesis testing that you clearly indicate the null and alternative hypotheses (in terms of the true parameters, say β_1), the test statistic (in terms of the parameter estimates, here $\widehat{\beta}_1$), its distribution (with degrees of freedom), the rejection rule (one-sided or-two sided) for a given significance level (typically 5%) with suitable critical values, and provide an interpretation of your result.

Just as last year, many candidates do not answer all parts of the question. Make sure you read the questions properly and provide all details that are requested. Not answering a question will automatically earn you a zero mark for that question.

Key steps to improvement

Essential reading for this course includes the subject guide and the following:

- Dougherty, C. *Introduction to econometrics.* (Oxford: Oxford University Press, 2016) 5th edition [ISBN 9780199676828]; http://oxfordtextbooks.co.uk/orc/dougherty5e/

Apart from the Essential readings you should do some supplementary reading. One very good book at the same level is:

- Gujarati, D.N. and D.C. Porter *Basic econometrics.* (McGraw–Hill, 2009, International edition) 5th edition [ISBN 9780071276252].

To understand the subject clearly it is important to supplement Dougherty's *Introduction to econometrics* (fifth edition) with the subject guide **EC2020 Elements of econometrics** (2016), especially Chapter 10 which covers maximum likelihood estimation. It is very important to carefully go through the subject guide. The subject guide contains solutions to the questions given in the main textbook and also some additional questions and solutions. Working through these will improve your understanding of the subject.

The chapter in the subject guide on maximum likelihood (Chapter 10) includes some additional theory which has not been covered in the main textbook. It is important to read the additional theory given in the subject guide to have a better understanding of the principles of maximum likelihood and tests based on the likelihood function.

Please check the VLE course page for resources for this subject such as a downloadable copy of the subject guide **EC2020 Elements of econometrics** (2016), PowerPoint slideshows that provide a graphical treatment of the topics covered in the textbook, datasets and statistical tables. Candidates should utilise datasets using standard regression programmes (STATA or EViews). This will help in the understanding of the subject.

Examination revision strategy

Many candidates are disappointed to find that their examination performance is poorer than they expected. This may be due to a number of reasons, but one particular failing is '**question spotting**', that is, confining your examination preparation to a few questions and/or topics which have come up in past papers for the course. This can have serious consequences.

We recognise that candidates might not cover all topics in the syllabus in the same depth, but you need to be aware that examiners are free to set questions on **any aspect** of the syllabus. This means that you need to study enough of the syllabus to enable you to answer the required number of examination questions.

The syllabus can be found in the Course information sheet available on the VLE. You should read the syllabus carefully and ensure that you cover sufficient material in preparation for the examination. Examiners will vary the topics and questions from year to year and may well set questions that have not appeared in past papers. Examination papers may legitimately include questions on any topic in the syllabus. So, although past papers can be helpful during your revision, you cannot assume that topics or specific questions that have come up in past examinations will occur again.

If you rely on a question-spotting strategy, it is likely you will find yourself in difficulties when you sit the examination. We strongly advise you not to adopt this strategy.

Examiners' commentaries 2017

EC2020 Elements of econometrics

Important note

This commentary reflects the examination and assessment arrangements for this course in the academic year 2016–17. The format and structure of the examination may change in future years, and any such changes will be publicised on the virtual learning environment (VLE).

Information about the subject guide and the Essential reading references

Unless otherwise stated, all cross-references will be to the latest version of the subject guide (2016). You should always attempt to use the most recent edition of any Essential reading textbook, even if the commentary and/or online reading list and/or subject guide refer to an earlier edition. If different editions of Essential reading are listed, please check the VLE for reading supplements – if none are available, please use the contents list and index of the new edition to find the relevant section.

Comments on specific questions – Zone A

Candidates should answer **EIGHT** of the following **TEN** questions: **ALL** of the questions in Section A (8 marks each) and **THREE** questions from Section B (20 marks each). **Candidates are strongly advised to divide their time accordingly.**

Section A

Answer all questions from this section.

Question 1

Discuss the drawbacks and advantages of using the Linear Probability Model when trying to explain a binary decision. In your answer clearly indicate what the Linear Probability Model is.

(8 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Chapter 10.1 (The linear probability model).

Subject guide (2016), Chapter 10.

Approaching the question

Candidates should clearly indicate that a linear probability model (LPM) is used to denote a model in which the dependent variable is binary, which takes the value 1 if the event occurs and

0 if it does not. It is estimated by ordinary least squares (OLS). Answers should discuss both drawbacks and advantages of using the LPM. A discussion of Logit/Probit and maximum likelihood estimation (MLE) is inappropriate here. The answer is as follows.

The LPM specifies $P_i = \Pr(Y_i = 1 \mid X_i) = \beta_1 + \beta_2 X_i$ and applies OLS to the model:

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

where $\mathrm{E}(u_i \mid X_i) = 0$. Recall, for the binary (discrete) random variable Y_i, we have:

$$\mathrm{E}(Y_i \mid X_i) = 1 \times \Pr(Y_i = 1 \mid X_i) + 0 \times \Pr(Y_i = 0 \mid X_i) = \Pr(Y_i = 1 \mid X_i)$$

hence we should interpret $\mathrm{E}(Y_i \mid X_i) = \beta_1 + \beta_2 X_i$ as the probability that the event will occur, given X_i.

Advantages: The results of the LPM are *easy to interpret* as marginal effects and/or *easy to estimate*. If we denote $\widehat{\beta}_0$ and $\widehat{\beta}_1$ as estimates of β_0 and β_1, respectively, then:

$$\widehat{Y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 X_i = \widehat{P}_i$$

is the estimated probability that the event will occur, and $\widehat{\beta}_1$ provides the marginal effect the explanatory variable X has on the probability of $Y = 1$, *ceteris paribus*.

Drawbacks: Since $\mathrm{E}(u_i) = 0$, we have:

$$\mathrm{E}(u_i) = (1 - \beta_0 - \beta_1 X_i)\underbrace{\Pr(Y_i = 1)}_{\beta_0 + \beta_1 X_i} + (-\beta_0 - \beta_1 X_i)\underbrace{\Pr(Y_i = 0)}_{1 - (\beta_0 + \beta_1 X_i)} = 0$$

and $\mathrm{Var}(u_i) = \mathrm{E}(u_i^2)$ exhibits *heteroskedasticity* rendering OLS *inefficient*. We have:

$$\begin{aligned}
\mathrm{E}(u_i^2) &= (1 - \beta_0 - \beta_1 X_i)^2 \underbrace{\Pr(Y_i = 1)}_{\beta_0 + \beta_1 X_i} + (-\beta_0 - \beta_1 X_i)^2 \underbrace{\Pr(Y_i = 0)}_{1 - (\beta_0 + \beta_1 X_i)} \\
&= (1 - \beta_0 - \beta_1 X_i)(\beta_1 + \beta_2 X_i)\left[(1 - \beta_0 - \beta_1 X_i) + (\beta_0 + \beta_1 X_i)\right] \\
&= (\beta_1 + \beta_2 X_i)(1 - \beta_0 - \beta_1 X_i) \\
&= \mathrm{E}(Y_i)(1 - \mathrm{E}(Y_i)) \\
&= P_i(1 - P_i) \qquad \text{for all } i.
\end{aligned}$$

In addition, as Y_i can take only two values, 1 or 0, u_i can only take the values $1 - \beta_0 - \beta_1 X_i$ when $Y_i = 1$ and $-\beta_0 - \beta_1 X_i$ when $Y_i = 0$ for given X. Therefore, the *errors are highly non-normal*. Finally, a major drawback with the LPM is that the *estimated probability can be negative or greater than 1, which are unreasonable results*. The fact that the marginal effects in the LPM are constant gives rise to this problem as OLS.

Question 2

Discuss the consequences of measurement error.

In your answer consider the following empirical study attempting to estimate the relationship between advertising and magazine circulation rates. A simple linear relationship is postulated between A_t, the advertising rate in magazine t, and C_t, the circulation figure for the magazine in question:

$$A_t = \alpha + \beta C_t + \varepsilon_t, \qquad t = 1, \ldots, T.$$

The relation is estimated by least squares on data for $T = 75$ magazines. Unfortunately, there are considerable errors in measurement in the reported circulation figures. Critically discuss the following statement: 'The estimated coefficient $\widehat{\beta}$ will be too small.' Rigour of your answer will be rewarded.

(8 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Chapter 8.4 (The consequences of measurement error).

Subject guide (2016), Chapter 8.

Approaching the question

Candidates should clearly discuss, with technical details (rigour), the consequences of measurement error. Inappropriate comments, for example stating that measurement error is caused by omitted variables, are penalised. It is important to clearly point out that measurement error in the explanatory variable (as is the case in our setting) induces correlation between the error and regressors – failure to point this out is serious. The answer is as follows.

Assuming we have classical measurement error, the measurement error on C_t causes our estimate of β to be *biased towards zero* – we also call this problem 'attenuation bias'. This is not the same as saying that the estimated coefficient will be too small.

Measurement in the circulation figures induces the problem of correlation between error and regressors, $\mathrm{E}(x_i \varepsilon_i) \neq 0$, giving rise to such inconsistency (bias).

Candidates should set up the model as follows. We are given the *true model*:

$$A_t = \alpha + \beta C_t + \varepsilon_t \qquad \text{where } \mathrm{Cov}(C_t, \varepsilon_t) = 0.$$

Unfortunately, we are told that C_t is *unobservable.* Instead we *observe* C_t^*, the circulation figures measured with error:

$$C_t^* = C_t + v_t \qquad \text{where } v_t \text{ is the } \textit{measurement error}, \quad v_t \text{ is i.i.d. } (0,\, \sigma_v^2).$$

The classical measurement assumptions ensure that:

$$\mathrm{Cov}(v_t, \varepsilon_t) = 0 \qquad \text{and} \qquad \mathrm{Cov}(v_t, C_t) = 0$$

with v_t independent of anything else in the model.

Therefore, the *estimable model* we can use to estimate β becomes:

$$A_t = \alpha + \beta C_t^* + u_t \qquad \text{with } u_t = \varepsilon_t - \beta v_t.$$

This estimator will be *inconsistent* as $\mathrm{Cov}(C_t^*, u_t) \neq 0$. Specifically, under the above assumptions:

$$\mathrm{Cov}(C_t^*, u_t) = \mathrm{Cov}(C_t + v_t, \varepsilon_t - \beta v_t) = -\beta \, \mathrm{Var}(v_t).$$

Candidates should clearly indicate the estimator, whose properties we need to discuss, as:

$$\widehat{\beta} = \frac{\sum (C_t^* - \bar{C}^*)(A_t - \bar{A})}{\sum (C_t^* - \bar{C}^*)^2} = \beta + \frac{\sum (C_t^* - \bar{C}^*)(u_t - \bar{u})}{\sum (C_t^* - \bar{C}^*)^2} = \beta + \frac{\text{Sample Cov}(C_t^*, u_t)}{\text{Sample Var}(C_t^*)}.$$

Using the *plim operator* and the *law of large numbers*, we then obtain:

$$\begin{aligned}
\mathrm{plim}(\widehat{\beta}) &= \beta + \frac{\mathrm{plim} \frac{1}{T} \sum (C_t^* - \bar{C}^*)(u_t - \bar{u})}{\mathrm{plim} \frac{1}{T} \sum (C_t^* - \bar{C}^*)^2} \\
&= \beta + \frac{\mathrm{plim}(\text{Sample Cov}(C_t^*, u_t))}{\mathrm{plim}(\text{Sample Var}(C_t^*))} \\
&= \beta + \frac{-\beta \, \mathrm{Var}(v_t)}{\mathrm{Var}(C_t^*)} \\
&= \beta \left(1 - \frac{\sigma_v^2}{\sigma_C^2 + \sigma_v^2} \right)
\end{aligned}$$

as $\text{Var}(C_t^*) = \text{Var}(C_t + v_t) = \sigma_C^2 + \sigma_v^2$.

Since $0 < \left(1 - \dfrac{\sigma_v^2}{\sigma_c^2 + \sigma_v^2}\right) < 1$, we have completed the proof of the attenuation bias.

Question 3

Consider the following regression model:

$$\boldsymbol{Y_i = \beta_0 + \beta_1 \frac{1}{X_i} + u_i, \qquad i = 1, \ldots, n.}$$

We assume that the errors $\{u_i\}_{i=1}^n$ are independent normal random variables with zero mean and variance σ^2/X_i^2. The regressor, $1/X_i$, is nonstochastic with positive sample variability and $X_i \neq 0$ for all i. You are interested in testing the hypothesis $\text{H}_0 : \beta_0 = 0$ against $\text{H}_1 : \beta_0 \neq 0$. You are advised to use the BLUE estimator of β_0 for this purpose.

Discuss how you would obtain the BLUE estimator of β_0 (note, you are not asked to derive this estimator).

Give two reasons why you would prefer using the BLUE estimator for β_0 instead of the OLS estimator $\widehat{\beta}_{0,OLS} = \bar{Y} - \widehat{\beta}_{1,OLS}\overline{(1/X)}$ when testing this hypothesis, where $\bar{Y} = \dfrac{1}{n}\sum Y_i$ and $\overline{(1/X)} = \dfrac{1}{n}\sum \dfrac{1}{X_i}$.

(8 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): R.10 (Type II error and the power of a test), Chapter 2.5 (The Gauss–Markov theorem), Chapter 2.6 (Testing hypotheses relating to the regression coefficients), Chapter 7.1 (Heteroskedasticity and its implications) and 7.3 (Remedies for heteroskedasticity).

Subject guide (2016), Chapter 7.

Approaching the question

Candidates should show a clear understanding of what it means for an estimator to be BLUE. Failure to recognise the presence of heteroskedasticity in this question is serious. Arguments that OLS on the regression itself would not be linear because of the form of the regressor $1/X_i$ are wrong. OLS on the regression model given is linear (it is linear in the parameters!), but due to the heteroskedasticity it is not efficient (BLUE)! Many candidates lost points simply because they did not discuss two reasons why, *for testing purposes,* one would prefer to use this BLUE estimator. Answer all parts of the question. The answer is as follows.

Since the only Gauss–Markov violation here is the presence of heteroskedasticity, we should propose to use weighted least squares (WLS) to make the problem go away. Regress:

$$Y_i X_i = \beta_0 X_i + \beta_1 + u_i X_i \qquad \text{for } i = 1, \ldots, n$$

or:

$$Y_i^* = \beta_1 + \beta_0 X_i + u_i^*.$$

This transformed model satisfies all the Gauss–Markov assumptions, $\text{E}(u_i^*) = \text{E}(u_i)\, X_i = 0$ and $\text{Var}(u_i^*) = \text{Var}(u_i X_i) = \text{Var}(u_i)\, X_i^2 = \sigma^2$. Do make sure you mention that the *Gauss–Markov theorem* ensures that OLS on this regression will yield our BLUEs of β_0 and β_1.

Benefits to using the BLUE of β_0 are (i) it is an efficient (more precise) estimator and because of this efficiency the test will have higher power (easier to reject the null hypothesis when it is false), and (ii) using the BLUE estimator allows us to directly use its standard error for inference. Had we used the OLS estimator instead, we would have needed to obtain heteroskedasticity-robust standard errors.

Question 4

Consider the following non-stationary process:

$$y_t = \gamma_0 + \gamma_1 t + u_t, \qquad \text{with } u_t = \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2}$$

and ε_t i.i.d. $(0, \sigma^2)$. Indicate (with explanation) the source(s) of non-stationarity of y_t. Discuss how you would test whether y_t indeed is non-stationary. Clearly indicate the null and the alternative hypothesis, the test statistic and the rejection rule. What name do we give such a non-stationary process?

(8 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Chapter 12.1 (Definition and consequences of autocorrelation), 13.1 (Stationarity and non-stationarity) and 13.5 (Tests of deterministic trends).

Subject guide (2016), Chapters 12 and 13.

Approaching the question

Candidates should show a clear understanding of what non-stationarity means. A discussion of a unit root test here is inappropriate as the dependence structure is given by an MA(2) process, not an AR process. The only source of non-stationarity is the presence of the deterministic trend, $\gamma_1 \neq 0$. Many candidates failed to recognise this. Answer all parts of the question, for example do not forget to answer how we call such a non-stationary process! The answer is as follows.

A process $\{y_t\}$ is (covariance) stationary if its mean and variance exist and do not depend on time and its covariance is a function of distance in time only (not location). Violation of either of these requirements renders the process non-stationary.

In this case $\{y_t\}$ is non-stationary because $\mathrm{E}(y_t) = \gamma_0 + \gamma_1 t$ depends on time when $\gamma_1 \neq 0$. Indeed, the only problem of non-stationarity here is the *presence of the deterministic trend*, as the dependence of $\{u_t\}$, and therefore $\{y_t\}$, is MA(2) and all finite-order moving average processes are (covariance) stationary.

Therefore, candidates should propose to test $\mathrm{H}_0 : \gamma_1 = 0$ against $\mathrm{H}_1 : \gamma_1 > 0$ (when trending upwards). For this we will use the t test $\widehat{\gamma}/\mathrm{SE}(\widehat{\gamma})$. Critical values are given by the $N(0, 1)$ distribution (since T is large). Reject H_0 if $\widehat{\gamma}/\mathrm{SE}(\widehat{\gamma}) > 1.645$ at the 5% significance level. (Suggesting a two-sided test is fine as well, but has less power.) Due to the presence of the dependence in the errors, *robust standard errors* should be used when implementing the test (this point was made by only a few candidates).

If we reject the null hypothesis, we call such a process $\{y_t\}$ *trend-stationary.*

Question 5

Suppose you are given a random sample $X_1, \ldots, X_n$ from the exponential distribution:

$$f(x) = \lambda \exp(-\lambda x), \qquad x > 0,\ \lambda > 0.$$

According to this distribution $\mathbf{E}(X_i) = 1/\lambda$ and $\mathbf{Var}(X_i) = 1/\lambda^2$ for $i = 1, \ldots, n$.

Show that the maximum likelihood estimator for λ is $1/\bar{X}$ where $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Is the estimator unbiased and/or consistent? Prove your claims.

(8 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Chapter 4.2 (Logarithmic transformations) and Chapter 10.6 (An introduction to maximum likelihood estimation).

Subject guide (2016), Chapter 10.

Approaching the question

Candidates should clearly conduct the maximum likelihood procedure and ensure proper uses of the product and logarithm operators are displayed. Many candidates made the classic error of not recognising that $\mathrm{E}(1/\bar{X}) \neq 1/\mathrm{E}(\bar{X})$ and also failed to realise that the proof of consistency should have used the plim operator and the law of large numbers as $\operatorname{plim}(1/\bar{X}) = 1/(\operatorname{plim} \bar{X})$. Instead, many candidates attempted (unsuccessfully) to look at the sufficient conditions (what is $\mathrm{Var}(1/\bar{X})$?). Finally, candidates should make a clear distinction between the unknown (fixed) parameter λ and its MLE $\widehat{\lambda}$ (a random variable). The answer is as follows.

The likelihood function is given by the joint density of the data (which equals the product of the marginals given the independence of our observations):

$$L(\lambda) = \prod_{i=1}^{n} \lambda \exp(-\lambda X_i) = \lambda^n \exp\left(-\lambda \sum_{i=1}^{n} X_i\right).$$

The MLE is the value of λ which maximises this function, or by monotonicity, the log-likelihood function:

$$\ln L(\lambda) = \ln\left(\lambda^n \exp\left(-\lambda \sum_{i=1}^{n} X_i\right)\right) = n \ln \lambda - \lambda \sum_{i=1}^{n} X_i.$$

By the first-order condition, we require:

$$\left.\frac{\partial \ln L}{\partial \lambda}\right|_{\widehat{\lambda}_{MLE}} = \frac{n}{\widehat{\lambda}_{MLE}} - \sum_{i=1}^{n} X_i = 0 \quad \text{or} \quad \widehat{\lambda}_{MLE} = \frac{1}{\bar{X}} \quad \text{with} \quad \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

(It should clearly state $\widehat{\lambda}_{MLE} = 1/\bar{X}$, saying $\lambda = 1/\bar{X}$ is incorrect.)

The MLE is *not unbiased*:

$$\mathrm{E}(\widehat{\lambda}_{MLE}) = \mathrm{E}\left(\frac{1}{\bar{X}}\right) \overset{\text{Jensen's Ineq}}{\neq} \frac{1}{\mathrm{E}(\bar{X})} = \frac{1}{\frac{1}{n}\sum \mathrm{E}\,(X_i)} = \lambda.$$

The MLE is *consistent*:

$$\operatorname{plim} \widehat{\lambda}_{MLE} \overset{\text{plim rules}}{=} \frac{1}{\operatorname{plim}(\bar{X})} \overset{\text{LLN}}{=} \frac{1}{\mathrm{E}(X_i)} = \lambda.$$

Section B

Answer three questions from this section.

Question 6

The following equation was estimated by Ordinary Least Squares based on second semester candidates in the fall term

$$\begin{aligned}
\widehat{trmgpa} &= \underset{\substack{(.55)\\ [.55]}}{-2.12} + \underset{\substack{(.175)\\ [.166]}}{.900}\, crsgpa + \underset{\substack{(.064)\\ [.074]}}{.193}\, cumgpa + \underset{\substack{(.0012)\\ [.0012]}}{.0014}\, tothrs \\
&\quad + \underset{\substack{(.0002)\\ [.0002]}}{.0018}\, sat - \underset{\substack{(.0018)\\ [.0019]}}{.0039}\, hsperc + \underset{\substack{(.085)\\ [.079]}}{.351}\, female - \underset{\substack{(.098)\\ [.080]}}{.157}\, season \\
n &= 269;\ R^2 = .465 \qquad (6.1)
\end{aligned}$$

where $trmgpa_i$ is the term GPA (grade point average) of individual i. $crsgpa_i$ is a measure of difficulty of courses taken by individual i (weighted overall average of GPA in selected courses), $cumgpa_i$ is the GPA of individual i prior to the current semester, $tothrs_i$ is the total credit hours of individual i prior to the semester, sat_i is his/her SAT score (test taken for college admission in the USA), $hsperc_i$ is the graduating percentile of individual i in high school class, $female_i$ is a gender dummy, and $season_i$ is a dummy variable equal to unity if the student's sport is in season during the fall term. The usual standard errors are in parentheses and the White's heteroskedasticity-robust standard errors are in squared brackets.

(a) **Explain the concept of heteroskedasticity and discuss the properties of the OLS estimator in the presence of heteroskedasticity.**

(5 marks)

(b) **Do the variables *crsgpa*, *cumgpa* and *tothrs* have the expected estimated effects? Which of these variables are statistically significant at the 5% level? Does it matter which standard errors are used?**

(5 marks)

(c) **In view of your concern about the presence of heteroskedasticity, provide the 95% confidence interval for β_{crsgpa} and use it to test the hypothesis $H_0 : \beta_{crsgpa} = 1$ against $H_1 : \beta_{crsgpa} \neq 1$. Describe your conclusions. In view of your answer, indicate whether the p-value of the test is bigger or smaller than 5%.**

(5 marks)

(d) **Discuss how you would conduct a test for heteroskedasticity in this setting. Clearly indicate the assumptions that underlie the test you suggest. Detail of your answer will be rewarded.**

(5 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Chapter 2.6 (Testing hypotheses relating to the regression coefficients), R.12, and Chapter 7 (Heteroskedasticity).

Subject guide (2016), Chapter 7.

Approaching the question

In part (a), the concept of heteroskedasticity and the consequences on the OLS estimator need to be discussed. In part (b), the sign and significance of the coefficients should be explained. To conduct the test, a t test should be used where robust standard errors should be used because of the presence of heteroskedasticity. In part (c), candidates should provide the 95% confidence interval and its interpretation. In part (d), a suitable test for heteroskedasticity should be provided. This question was answered by 85% of the candidates in Zone A. The answer is as follows.

(a) Discussion of the concept is standard bookwork, see Chapter 7.1 of Dougherty.

The properties of OLS are *inefficiency* (as heteroskedasticity is a violation of the Gauss–Markov assumptions), *unbiasedness* (heteroskedasticity does not violate the assumption that $\mathrm{E}(u \mid X) = 0$ needed for unbiasedness) and *consistency* (heteroskedasticity does not violate the assumption of contemperaneous uncorrelatedness needed for consistency). Moreover, the *usual standard errors of the OLS estimator will be invalid* (we will need to use White's heteroskedasticity-robust standard errors).

(b) These coefficients have the anticipated sign. If a student takes courses where grades are, on average, higher – as reflected by higher $crsgpa$ – then his/her grades will be higher. The better the student has been in the past – as measured by $cumgpa$ – the better the student does (on average) in the current semester. Finally, $tothrs$ is a measure of experience, and its coefficient indicates an increasing return to experience.

We should use the t statistic to determine whether parameters are statistically significant: $\mathrm{H}_0 : \beta_i = 0$ against $\mathrm{H}_1 : \beta_i > 0$. Our test statistic is $\widehat{\beta}_i/\mathrm{SE}(\widehat{\beta}_i)$, which has a t_{n-8} distribution under H_0. (Exact distribution under Gauss–Markov assumptions + normality of the errors; approximate distribution when using robust standard errors in which case you may simply use $N(0, 1)$ critical values instead). Using a 5% significance level you should reject H_0 if $\widehat{\beta}_i/\mathrm{SE}(\widehat{\beta}_i) > 1.645$ with $i = \{crsgpa, cumgpa, tothrs\}$.

All parameters enter significantly. The t statistic for *crsgpa* is very large, over five using the usual standard error (which is the larger of the two). Using the robust standard error for *cumgpa*, its t statistic is about 2.61, which is also significant at the 5% significance level. The t statistic for *tothrs* is only about 1.17 using either standard error, so it is not significant. While the decision remains unchanged whichever standard errors are used, in the presence of heteroskedasticity we should use the robust standard errors as the other ones are invalid.

(c) The confidence interval, using the heteroskedasticity-robust standard errors, is given by:

$$\left[\widehat{\beta}_{crsgpa} - 1.96 \times \mathrm{SE}(\widehat{\beta}_{crsgpa}),\ \widehat{\beta}_{crsgpa} + 1.96 \times \mathrm{SE}(\widehat{\beta}_{crsgpa})\right] \qquad \text{or} \qquad [0.575,\ 1.255].$$

Since 1 lies in this 95% confidence interval, we do not reject the hypothesis that $\beta_{crgpga} = 1$. Conclusion: Everything else constant, candidates with a one-unit higher course GPA are expected to have a one-unit higher term GPA score. Many candidates were unable to provide this confidence interval and/or failed to interpret the conclusion from our test. Answers that conduct the test without reference to the confidence interval are incorrect.

The p-value is the lowest level of significance at which we want to reject the null hypothesis. Since we do not reject at the 5% significance level, the p-value will be bigger than 5%.

(d) Candidates can discuss here either the Goldfeld–Quandt test or White's test, see Chapter 7.2 of Dougherty.

Candidates should make it clear what the assumptions are that underlie either test. For the Goldfeldt–Quandt test they will have to specify the variable that enables them to divide the sample into large variance versus small variance observations (say *cumgpa*). When discussing White's test, on the other hand, no prior assumption is required about the form of heteroskedasticity.

Details of the test, test statistic, distribution and rejection rule should be provided (standard textbook answer).

Question 7

Consider the model:

$$y_t = \alpha_1 y_{t-1} + u_t, \qquad t = 1, \ldots, T$$

where $y_0 = 0$. $\mathrm{E}(u_t) = 0$, $\mathrm{E}(u_t^2) = \sigma^2$ and $\mathrm{E}(u_t u_s) = 0$ when $s \neq t$, for all $s, t = 1, \ldots, T$.

(a) **Discuss what we mean by the concept of stationarity (more precisely 'covariance stationarity') and indicate under what condition $\{y_t\}_{t=1}^T$ will be stationary.**

(3 marks)

(b) **Discuss the Dickey–Fuller procedure used to test for the presence of a unit root in the above model. Clearly indicate the null and alternative hypothesis, test statistic and rejection rule.**

(5 marks)

(c) **Consider a slight variation of the above model:**

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + v_t, \qquad t = 1, \ldots, T$$

where $y_0 = 0$. $\mathrm{E}(v_t) = 0$, $\mathrm{E}(v_t^2) = \sigma^2$ and $\mathrm{E}(v_t v_s) = 0$ when $s \neq t$, for all $s, t = 1, \ldots, T$. What do we call such a process? Discuss the problem you will have when conducting your test as in (b).

(3 marks)

(d) **Instead of conducting the Dickey–Fuller procedure, you are told to apply the Augmented DF test. Indicate how you would conduct the test for the presence of a unit root here. Derivation of the test equation will be required for full marks.**

(5 marks)

(e) **What are the potential problems associated with performing a regression with $I(1)$ variables? In your answer explain what it means for a variable to be $I(1)$.**

(4 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Chapter 13.1 (Stationarity and nonstationarity), Chapter 13.4–13.5 (Tests of nonstationarity), and Chapter 13.6 (Cointegration).

Subject guide (2016), Chapter 13.

Approaching the question

In part (a), the concept of (covariance) stationarity needs to be given. In part (b), you need to provide the Dickey–Fuller test for unit roots and in part (d) you need to provide the Augmented Dickey–Fuller test. The latter allows us to deal with the fact that when conducting the Dickey–Fuller test we cannot have any autocorrelation, which would be the case if the dependence is not AR(1), but as is the case here AR(2). In part (e), a discussion of the spurious regression result is expected. This question was answered by 60% of the candidates in Zone A. The answer is as follows.

(a) Candidates should provide a textbook definition of (covariance) stationarity.

The requirement for (covariance) stationarity here is that $|\alpha_1| < 1$. (Stating $\alpha_1 < 1$ is permitted.) Here $\{y_t\}$ is described by an AR(1) process, which is stationary provided its coefficient, α_1, is smaller than 1 (in magnitude). If α_1 is equal to one we have a unit root, whereas if $|\alpha_1| > 1$ we have an explosive process.

(b) To test for a unit root, we want to estimate the following regression:

$$\Delta y_t = \gamma y_{t-1} + u_t \qquad \text{where } \gamma = (\alpha_1 - 1).$$

We test $\mathrm{H}_0 : \gamma = 0$ (nonstationarity) against $\mathrm{H}_1 : \gamma < 0$ (stationarity). (Alternatively, we estimate the original model and test $\mathrm{H}_0 : \alpha_1 = 1$ (nonstationarity, unit root) against $\mathrm{H}_1 : \alpha_1 < 1$ (stationary AR process).)

We use the Dickey–Fuller t test $\widehat{\gamma}/\mathrm{SE}(\widehat{\gamma}) = (\widehat{\alpha}_1 - 1)/\mathrm{SE}(\widehat{\alpha}_1)$. Because we have non-stationarity under the null hypothesis, this test is not standard and we have to use the Dickey–Fuller critical values. Given the alternative hypothesis, we reject H_0 when $\widehat{\gamma}/\mathrm{SE}(\widehat{\gamma}) < \tau$, where τ is the Dickey–Fuller critical value (no trend or intercept in the model). Candidates can, alternatively, discuss the Dickey–Fuller scaled coefficient test which is $T(\widehat{\alpha}_1 - 1) = T\widehat{\gamma}$.

If we reject the null hypothesis we have found evidence that our process is stationary (weakly dependent).

In the presence of a unit root, a shock to u_t will have an everlasting effect on the process $\{y_t\}$. Unit roots are persistent and strongly dependent.

(c) Here $\{y_t\}$ is an AR(2) process. (ADL(2, 0) is also acceptable.)

The problem with (b) will be serious, as *dependence (autocorrelation) in the error* will cause the error term u_t (which would then be equal to $\alpha_2 y_{t-2} + v_t$) and regressor y_{t-1} to be correlated and hence OLS will be inconsistent. Hence our test $\widehat{\gamma}/\mathrm{SE}(\widehat{\gamma})$ (or $T(\widehat{\alpha}_1 - 1)$) will be *invalid*.

(d) The Augmented Dickey–Fuller test suggests that we add further lags in our testing equation to remove the autocorrelation. In particular, we should estimate the following model:

$$\Delta y_t = \gamma_1 y_{t-1} + \gamma_2 \Delta y_{t-1} + v_t$$

and test the hypothesis $\mathrm{H}_0 : \gamma_1 = 0$ (non-stationarity) against $\mathrm{H}_1 : \gamma_1 < 0$ (stationarity). We should use the test given by $ADF = \widehat{\gamma}_1/\mathrm{SE}(\widehat{\gamma}_1)$ and we should reject H_0 when ADF is smaller than the critical value given by the Dickey–Fuller tables for a given significance level.

To derive this result observe that:

$$\begin{aligned}
\Delta y_t &= \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + v_t - y_{t-1} \\
&= (\alpha_1 - 1)\, y_{t-1} + \alpha_2 y_{t-2} + v_t \\
&= (\alpha_1 - 1)\, y_{t-1} + \alpha_2 \left(\underbrace{y_{t-2} - y_{t-1}}_{-\Delta y_{t-1}} + y_{t-1} \right) + v_t \\
&= (\alpha_1 + \alpha_2 - 1)\, y_{t-1} - \alpha_2 \Delta y_{t-1} + v_t
\end{aligned}$$

so $\gamma_1 = (\alpha_1 + \alpha_2 - 1)$ and $\gamma_2 = -\alpha_2$. Therefore, to test for a unit root in the AR(2) model we test whether $\alpha_1 + \alpha_2 < 1$.

(e) To say that a variable is $I(1)$ indicates that the variable has a unit root. We say that the variable then is integrated of order 1, revealing that by differencing the variable once we can make it stationary.

The potential problem associated with performing such a regression is that we may get a *spurious relation*. This is the setting where, due to the fact that both variables are trending, there is an appearance of a relationship that does not exist at all (high t statistics and a large R^2). To ensure that we have a meaningful (long-run) relationship between variables, we want to verify that instead we are dealing with a *cointegrating relationship*.

Question 8

For the US economy, let *gprice* denote the monthly growth in the overall price level and let *gwage* be the monthly growth in hourly wages. We have estimated the following distributed lag model (ADL(0, 12)):

$$\begin{aligned}
\widehat{gprice}_t \;&=\; \underset{(.00057)}{-.00093} + \underset{(.052)}{.119}\, gwage_t + \underset{(.039)}{.097}\, gwage_{t-1} + \underset{(.039)}{.040}\, gwage_{t-2} \\
&\quad + \underset{(.039)}{.038} gwage_{t-3} + \underset{(.039)}{.081}\, gwage_{t-4} + \underset{(.039)}{.107}\, gwage_{t-5} + \underset{(.039)}{.095}\, gwage_{t-6} \\
&\quad + \underset{(.039)}{.104} gwage_{t-7} + \underset{(.039)}{.103}\, gwage_{t-8} + \underset{(.039)}{.159}\, gwage_{t-9} + \underset{(.039)}{.110}\, gwage_{t-10} \\
&\quad + \underset{(.039)}{.103}\, gwage_{t-11} + \underset{(.052)}{.016}\, gwage_{t-12}, \\
n \;&=\; 273,\ R^2 = .317,\ \bar{R}^2 = .283,\ DW = .99.
\end{aligned}$$

The usual standard errors are in parentheses.

(a) What is the estimated long-run effect (LRP)? Is it very different from one? Explain what the LRP tells us in this example. How does this differ from the short-run effect?

(4 marks)

(b) We want to test whether the LRP is significantly smaller than one. Clearly indicating the null and the alternative hypothesis, give the test statistic and the rejection rule. What regression would you run to obtain the standard error of the LRP directly?

(6 marks)

(c) Your result in (b) may be affected by the presence of autocorrelation. Briefly explain this and discuss how you could use the Durbin–Watson test to detect whether this is indeed the case. Indicate clearly the assumptions underlying the Durbin–Watson test, the test statistic and the rejection rule.

(5 marks)

***Hint:* For the critical values of the DW test, use the largest sample size and number of regressors available in the table.**

(d) Assuming the Durbin–Watson test finds evidence of autocorrelation, discuss how you could resolve the problem by means of the iterated Cochrane–Orcutt procedure. For notational purposes, you may simplify your model to an ADL(0, 1).

(5 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Chapter 6.5 (Testing a linear restriction), Chapter 11.3 (Models with lagged explanatory variables), and Chapter 12.1–12.3 (Definition, consequences and detection of autocorrelation; Fitting a model subject to AR(1) autocorrelation).

Subject guide (2016), Chapters 6 and 11.

Approaching the question

In part (a), an interpretation of parameters in distributed lag models (short-run and long-run effects) needs to be given. In part (b), a one-sided t test for the LRP is required and candidates are expected to discuss a reparameterisation of the model that enables one to get the standard error of the LRP parameter. In part (c), the validity of our test in (b) is put into question due to the presence of autocorrelation and a discussion of the Durbin–Watson test is required. In part (d), a discussion of the Cochrane–Orcutt procedure is required. Many candidates were unable to answer this question. (In view of the fact that the Cochrane–Orcutt procedure is not explicitly mentioned in Appendix A of the subject guide, with the approval of the external examiner, steps were undertaken to ensure candidates were not affected by this in their overall classification.) This question was answered by 25% of the candidates in Zone A. The answer is as follows.

(a) The estimated long-run effect is given by:

$$\widehat{\theta} = \widehat{\beta}_1 + \cdots + \widehat{\beta}_{13} = 1.172$$

where $\widehat{\beta}_i$ is the estimated coefficient on $gwage_{t-i+12}$ (numerical value irrelevant).

The LRP tells us what the long-run effect of *gwage* is on *gprice*. It shows the effect a permanent change in *gwage* with one unit has on *gprice* after all 12 periods (the last lagged response) have passed.

The short-run effect is $\widehat{\beta}_1$ and it tells us what the impact of current *gwage* on current *gprice* is.

(b) We want to test $\mathrm{H}_0 : \theta = 1$ against $\mathrm{H}_1 : \theta < 1$. By rejecting the null hypothesis against this one-sided alternative hypothesis, we may find evidence that the LRP is significantly smaller than 1!

The test statistic we should use is:

$$t = \frac{\widehat{\theta} - 1}{\mathrm{SE}(\widehat{\theta})} \sim t_{n-14} \qquad \text{under } \mathrm{H}_0 \text{ (Gauss–Markov assumptions + normality).}$$

Reject H_0 if $t < -1.645$ at the 5% significance level. (Important: candidates should recognise that $\mathrm{SE}(\widehat{\theta}) \neq \mathrm{SE}(\widehat{\beta}_1) + \cdots + \mathrm{SE}(\widehat{\beta}_{13})$.) If we reject H_0, we find evidence that the LRP is significantly smaller than one.

A candidate should discuss that one can rewrite (reparameterise) the model so that $\mathrm{SE}(\widehat{\theta})$ needed for this test can be obtained directly by our regression output (see Chapter 11.3 – Estimating long-run effects). Specifically:

$$gprice_t = \beta_0 + \theta gwage_t + \beta_2 \left(gwage_{t-1} - gwage_t\right) + \cdots + \beta_{13} \left(gwage_{t-12} - gwage_t\right) + u_t.$$

The model is exactly the same, observe $\theta - \beta_2 - \cdots - \beta_{13} = \beta_1$. Because both models are the same their R^2s are the same – it is only the interpretation of the parameters (here the parameter of $gwage_t$) that is different.

(c) In the presence of autocorrelation, we cannot trust the usual standard errors and, therefore, our test statistics will be invalid and our conclusions may be wrong.

The Durbin–Watson test allows us to test for the absence of autocorrelation against the alternative of an AR(1). The Durbin–Watson test requires the regressors to be deterministic (for example, cannot contain a lagged dependent variable). (An alternative test that does not suffer from this problem is the Breusch–Pagan test.)

Discussion of the Durbin–Watson test:

- $\mathrm{H}_0 : \rho = 0$ against $\mathrm{H}_1 : u_t = \rho u_{t-1} + \varepsilon_t$, for $|\rho| < 1$.
- $DW = \frac{\sum_{t=2}^{T} (\widehat{u}_t - \widehat{u}_{t-1})^2}{\sum_{t=2}^{T} \widehat{u}_t{}^2} \simeq 2(1 - \widehat{\rho})$, where $\widehat{u}_t$ are the OLS residuals from the regression in the question.
- Choose a significance level, α, (willingness to commit a Type I error) and find the critical values d_L and d_U.
- One-sided test: Reject H_0 in favour of positive autocorrelation if $DW < d_L$; test is inconclusive if it lies between d_L and d_U; do not reject H_0 when $DW > d_U$.
- k' is taken to be 5 (number of explanatory variables, here 13) and n is taken to be 100 (number of observations, here 273). Therefore, $d_L = 1.57$ and $d_U = 1.78$ with a 5% significance level. With $DW = 0.99$ we reject H_0 in favour of positive autocorrelation.

(d) Simplify the model to:

$$gprice_t = \beta_0 + \beta_1 gwage_t + \beta_2 gwage_{t-1} + u_t, \qquad t = 2, \ldots, T$$

and note that Durbin–Watson suggests $u_t = \rho u_{t-1} + \varepsilon_t$.

Discussion of the Cochrane–Orcutt procedure (see also Box 12.1 of Dougherty):

- Step 1: Obtain $\widehat{\beta}$ by running the original model by OLS, and compute the residuals $\widehat{u}_t = gprice_t - \widehat{\beta}_0 - \widehat{\beta}_1 gwage_t - \widehat{\beta}_2 gwage_{t-1}$.
- Step 2: Obtain an estimate of ρ using the residuals $\widehat{\rho} = \dfrac{\sum_{t=2}^{T} \widehat{u}_t \widehat{u}_{t-1}}{\sum_{t=2}^{T} \widehat{u}_t^{\,2}}$.
- Step 3: Estimate the transformed model:

$$\begin{aligned} & gprice_t - \widehat{\rho} gprice_{t-1} \\ &= \beta_0 (1 - \widehat{\rho}) + \beta_1 (gwage_t - \widehat{\rho} gwage_{t-1}) + \beta_2 (gwage_{t-1} - \widehat{\rho} gwage_{t-2}) + v_t \\ &\qquad \text{for } t = 3, \ldots, T \qquad \text{(lose one observation).} \end{aligned}$$

  Use the new parameter estimates of β_0, β_1 and β_2 and recompute the residuals $\widehat{u}_t$.
- Step 4: Return to Step 2, until convergence.

If the sample is small, losing one observation is not a good idea, and it is preferable to use the Prais–Winsten estimator (Chapter 12.3 – Fitting a model subject to AR(1) correlation).

Alternative non-linear regressions to deal with the AR(1) correlation are available as well.

Question 9

Let us consider the demand for fish. Using 97 daily price ($avgprc$) and quantity ($totqty$) observations on fish prices at the Fulton Fish Market in Manhattan, the following results were obtained by OLS:

$$\begin{aligned} \log(\widehat{totqty_t}) &= \underset{(.163)}{8.244} - \underset{(.176)}{.425} \log(avgprc_t) - \underset{(.226)}{.311}\, mon_t - \underset{(.223)}{.683}\, tues_t \\ &\quad - \underset{(.220)}{.533}\, wed_t + \underset{(.220)}{.067}\, thurs_t \end{aligned} \tag{9.1}$$

The equation allows demand to differ across the days of the week, and Friday is the excluded dummy variable. The standard errors are in parentheses.

(a) Interpret the coefficient of log($avgprc$) and discuss whether it is significant.

(3 marks)

(b) It is commonly thought that prices are jointly determined with quantity in equilibrium where demand equals supply. What are the consequences of this simultaneity for the properties of the OLS estimator?

(3 marks)

(c) The variables $wave2_t$ and $wave3_t$ are measures of ocean wave heights over the past several days. In view of your answer in (b), what two assumptions do we need to make in order to use $wave2_t$ and $wave3_t$ as instruments for log($avgprc_t$) in estimating the demand equation? Discuss whether these assumptions are reasonable.

(4 marks)

(d) Below we report two sets of regression results, where the dependent variable is log($avgprc_t$). Are $wave2_t$ and $wave3_t$ jointly significant? State the test statistic and rejection rule. How is your finding related to your answer in (c)?

(4 marks)

	Regressors						
Dependent Variable $\log(avgprc_t)$	*constant*	*wave2*	*wave3*	*day-of-the-week dummies*	R^2	RSS	n
Regression (9.2)	-1.022 (.144)	.094 (.021)	.053 (.020)	*yes*	.3041	10.934	97
Regression (9.3)	$-.276$ (.092)	—	—	*yes*	.0088	15.576	97

(e) **The following IV results were obtained in Stata:**

$$\begin{aligned}\log(\widehat{totqty_t}) \;=\;& \underset{(.182)}{8.164} - \underset{(.327)}{.815}\log(avgprc_t) - \underset{(.229)}{.307}\,mon_t - \underset{(.226)}{.685}\,tues_t \\ & - \underset{(.223)}{.521}\,wed_t + \underset{(.225)}{.095}\,thurs_t \end{aligned} \qquad (9.4)$$

Discuss how these results can be obtained using Two Stage Least Squares (2SLS) and briefly discuss how you would test whether the results in (9.1) and (9.4) are significantly different from each other.

(6 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Chapter 6.5 (Testing a linear restriction), Chapter 8.3 (Instrumental variables), and Chapter 9 (Simultaneous equations estimation).

Subject guide (2016), Chapter 9.

Approaching the question

In part (a), a parameter interpretation is required (be specific as it is an elasticity here!) and a t test should be used to test for its significance. In part (b), a discussion of simultaneous equation bias is required. The joint determinacy in simultaneous equations induces an endogeneity problem: a problem where we have a correlation between the errors and regressors. This correlation renders the OLS parameter estimates inconsistent. Recognising that the IV estimator avoids this problem, in part (c) the requirements on the instruments need to be discussed as they relate to the particular example. In part (d), an F test needs to be proposed to test for the relevance of the instruments, and in part (e) the 2SLS estimator and the Durbin–Wu–Hausman test need to be discussed. This question was answered by 80% of the candidates in Zone A. The answer is as follows.

(a) We want to interpret the parameter as the *price elasticity of demand* after controlling for days-of-the-week differences. If the price increases by 1%, then quantity demanded decreases by 0.425%.

Use the t statistic $\widehat{\beta}/\mathrm{SE}(\widehat{\beta}) = -0.425/0.176 = -2.415$. It is significant as the 5% significance level critical value of t_{n-6} is 1.96 (Gauss–Markov assumptions + normality assumed).

(b) This is the *endogeneity problem*, whereby $\log(avgprc)$ will be correlated with the error term. Candidates making incorrect claims, for example it is a problem of multicollinearity, are penalised. If we were given the structural form equations for both supply and demand, we could explicitly derive this correlation by deriving the reduced form equation for $\log(avgprc)$. The error term in that reduced form would contain the error term of the demand equation, which induces this correlation.

OLS will result in inconsistent parameter estimates, or simply stated will cause simultaneous equation bias.

(c) To estimate the demand equation, we need at least one exogenous variable that appears in the supply equation (because we have one 'bad' variable) that does not also appear in the demand equation (relevance) and that variable needs to be uncorrelated with the error term in the demand equation (validity). We call such variables *instruments*.

Exclusion and validity: we need to assume that the instruments can properly be excluded from the demand equation and are uncorrelated with the demand error term. This may not be entirely reasonable – wave heights are determined partly by weather and demand at a local fish market could depend on weather.

Relevance: we need to make sure that the instruments are correlated with the 'bad' variable $\log(avgrpc)$. By ensuring that at least one of the instruments ($wave2_t$ and $wave3_t$) appears in the supply equation, we will be able to show that the reduced form for $\log(avgprc_t)$ will have these variables on the right-hand side, hence that there is a correlation between the instruments and the 'bad' regressor. Seems to be quite reasonable, and indeed there is indirect evidence of this in part (d), as the two variables are jointly significant in the reduced form for $\log(avgprc_t)$.

(d) We test the joint hypothesis $\mathrm{H}_0 : \beta_{wave2} = \beta_{wave3} = 0$ against $\mathrm{H}_1 : \beta_{wave2} \neq 0$ and/or $\beta_{wave3} \neq 0$. The F test is obtained as:

$$F = \frac{(RRSS - URSS)/2}{URSS/(97-7)} = \frac{(15.576 - 10.934)/2}{10.934/90} = 19.0.$$

Under the Gauss–Markov assumptions + normality, this gives us an $F_{2,\,90}$ random variable under the null hypothesis. For any reasonable significance level we will want to reject H_0 (at the 5% significance level the critical value is 3.10). An alternative formula that could have been used is:

$$\frac{(R_u^2 - R_r^2)/2}{(1 - R_u^2)/90}.$$

Conclusion: the test result ensures that our instrumental variables are indeed relevant. (Many candidates failed to mention this: answer all parts of the question.)

(e) Following a discussion of 2SLS (given below), a brief, non-technical, discussion of the Durbin–Wu–Hausman test should be provided.

The Durbin–Wu–Hausman test is a chi-squared test where the two sets of parameter estimates (OLS and 2SLS) are compared. Under the null hypothesis (there is no endogeneity problem and both 2SLS and OLS will be consistent) we should expect $\widehat{\beta}_{OLS} - \widehat{\beta}_{2SLS} \approx 0$, whereas under the alternative hypothesis (there is an endogeneity problem and only 2SLS is consistent) the parameter estimates can be quite different. The fact that under the null hypothesis OLS is efficient (the best instrument for something which is 'good' is always itself) makes it easy to work out the precision (variance) of $\widehat{\beta}_{OLS} - \widehat{\beta}_{2SLS}$ required for the test.

The 2SLS approach is as follows.

- Step 1: Requires us to make the 'bad' variable 'good' by regressing it (OLS) on all exogenous variables in our model:

$$\log(avgprc_t) = \delta_0 + \delta_1 mon_t + \cdots + \delta_4 thurs_t + \delta_5 wave2_t + \delta_6 wave3_t + e_t.$$

We obtain the fitted values of this regression: $\log(\widehat{avgprc_t})$. All exogenous variables have to be included here!

- Step 2: Requires us to estimate the equation:

$$\log(totqty_t) = \beta_0 + \beta_1 \log(\widehat{avgprc_t}) + \beta_2 mon_t + \beta_3 tues_t + \beta_4 wed_t + \beta_5 thurs_t + u_t$$

by OLS (hence *two* stage LS). The parameter estimates of this regression are our 2SLS estimates. Candidates may alternatively indicate that the second stage is an IV estimator, where $\log(\widehat{avgprc_t})$ (super instrument) is used as an instrument for $\log(avgprc_t)$.

Question 10

Let us consider an analysis on recidivism (probability of re-arrest) among a group of young men in California who have at least one arrest prior to 1986. The dependent variable, $arr86$, is equal to unity if the man was arrested at least once during 1986, and zero otherwise.

	OLS A	OLS B	Logit A	Logit B	Logit B Marginal Effect
$pcnv$	$-.152$ (.021)	$-.162$ (.021)	$-.880$ (.122)	$-.901$ (.120)	$-.176$ (.023)
$avgsen$	.005 (.006)	.006 (.006)	.027 (.035)	.031 (.034)	.006 (.007)
$tottime$	$-.003$ (.005)	$-.002$ (.005)	$-.014$ (.028)	$-.010$ (.027)	$-.002$ (.005)
$ptime86$	$-.023$ (.005)	$-.022$ (.005)	$-.140$ (.031)	$-.127$ (.031)	$-.025$ (.006)
$qemp86$	$-.038$ (.005)	$-.043$ (.005)	$-.199$ (.028)	$-.216$ (.028)	$-.042$ (.005)
$black$	.170 (.024)	—	.823 (.117)	—	—
$hispan$	.096 (.021)	—	.522 (.109)	—	—
$constant$	.380 (.019)	.380 (.019)	$-.464$ (.095)	$-.169$ (.084)	
R^2	.068	.047			
$\log L$			-1512.35	-1541.24	

$pcnv$ is the proportion of prior arrests that led to a conviction, $avgsen$ is the average sentence served from prior convictions, $tottime$ is the months spent in prison since age 18 prior to 1986, $ptime86$ is months spent in prison in 1986, $qemp86$ is the number of quarters the man was legally employed in 1986, while $black$ and $hispan$ are two race dummies ($white$ the excluded dummy). The standard errors are reported in parentheses.

(a) When estimating the parameters by OLS, we are using the Linear Probability Model. Why might you then report heteroskedasticity-robust standard errors?

(2 marks)

(b) Using the OLS B results, what is the estimated effect on the probability of arrest if $pcnv$ goes from 0.25 to 0.75 holding everything else constant?

(4 marks)

(c) It is argued that the linear probability model is not appropriate for explaining the binary variable $arr86$ and a logit regression model has been estimated. Explain how the Logit estimates are obtained.

(5 marks)

Hint: You may recall, that for the Logit model A, we will specify

$$\Pr(arr86_i = 1) = \Lambda\left(\beta_0 + \beta_1 pcnv + \beta_2 avgsen + \cdots + \beta_6 black + \beta_7 hispan\right)$$

where $\Lambda(z) = \frac{1}{1+\exp(-z)}$.

(d) Using the Logit model results, discuss whether $black$ and $hispan$ are jointly significant. Clearly indicate the null and alternative hypothesis, the test statistic and the rejection rule.

(3 marks)

(e) An important distinction between the two approaches is that the marginal effect of $pcnv$ on the probability of re-arrest is constant for the LPM unlike the marginal effect using the logit analysis. What this means for instance is that the estimated effect on the probability of arrest if $pcnv$ goes from 0.25 to 0.75 will depend on the other characteristics.

i. **Explain how the marginal effects evaluated at the mean values of the explanatory variables (reported in the last column) were obtained. Give a brief comment as to how they compare to the marginal effect of the associated LPM.**

(3 marks)

ii. **Using the Logit B results, how would you obtain the estimated effect on the probability of arrest if $pcnv$ goes from 0.25 to 0.75 for a $white$ man, with characteristics $avgsen = 1$, $tottime = 1$, $ptime86 = 0$ and $qemp86 = 2$. A clear explanation of what calculations are required is sufficient.**

(3 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Chapter 10.1 (The linear probability model), Chapter 10.2–10.3 (Logit and Probit analysis) and Chapter 10.6 (An introduction to maximum likelihood estimation).

Subject guide (2016), Chapter 10.

Approaching the question

In parts (a) and (b), a discussion of the LPM and an interpretation of its parameters – constant marginal effects – need to be given. In parts (c) and (e), a related discussion needs to be given for the logit model. Unlike in the LPM where we imposed linearity, the marginal effects in the logit model are no longer constant due to the fact that in the logit model we use a non-linear, $\Lambda(z)$, specification of the probabilities. In part (c), a discussion of the maximum likelihood estimator used for our logit parameter estimates is required, and in part (d) a test for joint significance using the likelihood ratio test is needed. This question was answered by 40% of the candidates in Zone A. The answer is as follows.

(a) The linear probability model suffers from heteroskedasticity. Specifically, the heteroskedasticity takes the form $\mathrm{Var}(arr86_i \mid X_i) = p_i\,(1-p_i)$ with:

$$p_i = \mathrm{E}(arr86_i \mid X_i) = \beta_0 + \beta_1 pcnv_i + \beta_2 avgsen_i + \cdots + \beta_6 black_i + \beta_7 hispan_i.$$

Since this invalidates the standard errors, we will want to report heteroskedasticity-robust standard errors instead.

(b) Candidates should clearly indicate that $\widehat{P}_i = \widehat{\beta}_0 + \widehat{\beta}_1 pcnv_i + \widehat{\beta}_2 avgsen_i + \cdots + \widehat{\beta}_5 qemp86_i$.

The estimated effect is:

$$\widehat{\beta}_1(0.75 - 0.25) = -0.162 \times 0.5 = -0.081.$$

Therefore, the probability of re-arrest decreases by 0.081 or by 8.1 percentage points. (An 8.1% decrease is different from an 8.1 percentage point decrease!)

(c) Candidates are expected to indicate that we use the maximum likelihood estimator (MLE) to estimate the β parameters. Make sure you answer the question, you were asked how the estimates are obtained! Few candidates were able to give relevant details.

The log-likelihood function that will be maximised is given by:

$$\begin{aligned}
\log L(\beta) &= \sum_{i=1}^{n} arr86_i \times \log(\Pr(arr86_i = 1)) + (1 - arr86_i) \times \log(\Pr(arr86_i = 0)) \\
&= \sum_{i=1}^{n} arr86_i \times \log(\Lambda(z_i)) + (1 - arr86_i) \times \log(1 - \Lambda(z_i)) \\
\text{where } z_i &= \beta_0 + \beta_1 pcnv_i + \beta_2 avgsen_i + \cdots + \beta_6 black_i + \beta_7 hispan_i.
\end{aligned}$$

The MLE will set the partial derivatives equal to zero:

$$\left.\frac{\partial \log L(\beta)}{\partial \beta_i}\right|_{\widehat{\beta}_{MLE}} = 0 \qquad \text{for } i = 0, \ldots, 7$$

i.e. it chooses those values of $\widehat{\beta}$ that will set the derivatives equal to zero. Heuristically, these parameter estimates ensure that the predicted $\Pr(arr86_i = 1)$ is large for individuals for whom $arr86_i = 1$, and the predicted $\Pr(arr86_i = 0)$ is large for individuals for whom $arr86_i = 0$. This ensures that the 'likelihood' of observing the data is largest for these values of β.

(d) Candidates should indicate they want to test $\mathrm{H}_0 : \beta_{black} = \beta_{hispan} = 0$ against $\mathrm{H}_1 : \beta_{black} \neq 0$ and/or $\beta_{hispan} \neq 0$.

The likelihood ratio (LR) test statistic equals $-2(\ln L^R - \ln L^U) \overset{a}{\sim} \chi^2_2$ under H_0, where $\ln L^R$ is the log-likelihood function of the restricted model (Logit B) and $\ln L^U$ is the log-likelihood function of the unrestricted model (Logit A). $LR = -2 \times (-1541.24 - -1512.35) = 57.78$.

With a 5% significance level our critical value is given by 5.99, so we clearly reject the null hypothesis, rendering ethnicity an important factor in explaining re-offending rates.

(e) i. The marginal effects of interest describe how $\Pr(arr86_i)$ changes as a result of the explanatory variables.

For continuous variables that means, for example:

$$\begin{aligned} \frac{\partial \Pr(arr86_i)}{\partial pcnv} &= f(z_i) \times \beta_1 \qquad \text{with } f(z) = \frac{\mathrm{d}\Lambda(z)}{\mathrm{d}z} = \frac{\mathrm{e}^{-z}}{(1+\mathrm{e}^{-z})^2} \\ \text{with } z_i &= \beta_0 + \beta_1 pcnv_i + \beta_2 avgsen_i + \cdots + \beta_6 black_i + \beta_7 hispan_i \end{aligned}$$

(no need to give the derivative itself). Therefore, the marginal effects depend on the characteristics of the individual. The marginal effects reported are for an individual with average characteristics:

$$\bar{z} = \beta_0 + \beta_1 \overline{pcnv} + \beta_2 \overline{avgsen} + \cdots + \beta_6 \overline{black} + \beta_7 \overline{hispan}.$$

(Whether this individual exists is another matter. Indeed, it may be more interesting to report the average of these marginal effects over all individuals.)

The marginal effects for the average person are quite comparable to those obtained by the LPM (an easy mark, so do not forget to mention this – answer all parts of the question!). Candidates making these points clearly in (c) were rewarded here.

ii. Candidates here will need to observe that we need to compare predicted probabilities using the logit specification of the probabilities: $\Lambda(z) = \dfrac{1}{1+\exp(-z)}$.

$\widehat{\Pr}(arr86 = 1 \,|\, pcnv = 0.25, avgsen = 1, tottime = 1, ptime86 = 0, qemp86 = 2) = \Lambda(z_1)$, where:

$$z_1 = 0.25 \times (-0.901) + 1 \times (0.031) + 1 \times (-0.010) + 0 \times (-0.127) + 2 \times (-0.216) = -0.636$$

and $\Lambda(z_1) = \dfrac{1}{1+\mathrm{e}^{0.636}} = 0.346$.

$\widehat{\Pr}(arr86 = 1 \,|\, pcnv = 0.75, avgsen = 1, tottime = 1, ptime86 = 0, qemp86 = 2) = \Lambda(z_2)$, where:

$$z_2 = 0.75 \times (-0.901) + 1 \times (0.031) + 1 \times (-0.010) + 0 \times (-0.127) + 2 \times (-0.216) = -1.087$$

and $\Lambda(z_2) = \dfrac{1}{1+\mathrm{e}^{1.087}} = 0.252$.

So we see a drop in the probability of re-arrest equalling $0.346 - 0.252 = 0.094$.

Candidates are not expected to provide explicit numbers. Full marks can be obtained for clarity of approach. Failure to recognise the importance of Λ is serious.

Examiners' commentaries 2017

EC2020 Elements of econometrics

Important note

This commentary reflects the examination and assessment arrangements for this course in the academic year 2016–17. The format and structure of the examination may change in future years, and any such changes will be publicised on the virtual learning environment (VLE).

Information about the subject guide and the Essential reading references

Unless otherwise stated, all cross-references will be to the latest version of the subject guide (2016). You should always attempt to use the most recent edition of any Essential reading textbook, even if the commentary and/or online reading list and/or subject guide refer to an earlier edition. If different editions of Essential reading are listed, please check the VLE for reading supplements – if none are available, please use the contents list and index of the new edition to find the relevant section.

Comments on specific questions – Zone B

Candidates should answer **EIGHT** of the following **TEN** questions: **ALL** of the questions in Section A (8 marks each) and **THREE** questions from Section B (20 marks each). **Candidates are strongly advised to divide their time accordingly.**

Section A

Answer all questions from this section.

Question 1

Discuss the consequences of omitting relevant variables.

In your answer consider the following empirical study attempting to estimate the social benefits of an increase in public spending on education. Using data from a cross-section survey of employees, the following regression is estimated:

$$\log W_i = \alpha_1 + \alpha_2 S_i + \varepsilon_i,$$

where W_i is the hourly wage rate and S_i is the number of years of schooling completed of employee i. The coefficient of S_i is found to be positive and strongly significant. What can be concluded from this? In your answer discuss the interpretation of α_2, and discuss whether we are likely to obtain an unbiased (consistent) estimator of it when using OLS. Rigour of your answer will be rewarded.

(8 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Chapter 6.2 (The effect of omitting a variable that ought to be included).

Subject guide (2016), Chapter 6.

Approaching the question

Candidates should clearly discuss, with technical details (rigour), the consequences of omitting relevant variables. Inappropriate comments, for example relating this to multicollinearity are penalised. It is important to clearly point out that omitting relevant variables only causes serious problems if the omitted variable is correlated with included regressors. Only then will omitting relevant variables induce a correlation between the error (which then contains this omitted variable) and the included regressors (which cause inconsistency). Failure to point this out is serious. Candidates should answer all parts of the question and explain why the model in question is likely to suffer from this problem. The answer is as follows.

The interpretation of α_2 is as the returns to education – it provides the proportional change in wage per unit change in schooling, in other words $100 \times \alpha_2$ provides the percentage change in wage per unit change in schooling. If $\alpha_2 = 0.06$, this represents a 6% return. Many candidates stated $\alpha_2\%$, which is wrong!

It is indeed likely in this case that we have omitted relevant variables, such as ability, which in general generates *omitted variable bias* (OVB) for our incorrectly-specified model when using OLS as ability will be correlated with education.

Derivation of OVB:

Provide estimator: The model we consider here is:

$$\log W_i = \alpha_1 + \alpha_2 S_i + \varepsilon_i$$

and our interest is in the OLS estimator of α_2, which is given by:

$$\widehat{\alpha}_2 = \frac{\sum (S_i - \bar{S})(\log W_i - \overline{\log W})}{\sum (S_i - \bar{S})^2} = \frac{\text{Sample Cov}(S_i, \log W_i)}{\text{Sample Var}(S_i)}.$$

Proof of bias/inconsistency: Plugging in the true model:

$$\log W_i = \alpha_1 + \alpha_2 S_i + \gamma \mathit{Ability}_i + \varepsilon_i$$

we obtain:

$$\widehat{\alpha}_2 = \alpha_2 + \frac{\text{Sample Cov}(S_i, \mathit{Abilty}_i)}{\text{Sample Var}(S_i)}\gamma + \frac{\text{Sample Cov}(S_i, \varepsilon_i)}{\text{Sample Var}(S_i)}.$$

To consider the bias, we take expectations. *Assuming that S_i and Abilty_i are non-stochastic:*

$$\mathrm{E}(\widehat{\alpha}_2) = \alpha_2 + \underbrace{\frac{\text{Sample Cov}(S_i, \mathit{Abilty}_i)}{\text{Sample Var}(S_i)}\gamma}_{\text{Omitted Variable Bias}} \neq \alpha_2$$

that is, $\widehat{\alpha}_2$ is a biased estimator of α_2. This bias does not go away when the sample size tends to infinity $\rightarrow$ the estimator is inconsistent as well.

Candidates may directly establish the inconsistency as well, recognising that most likely S_i and *Abilty_i are stochastic*:

$$\begin{aligned}
\text{plim } \widehat{\alpha}_2 &= \alpha_2 + \frac{\text{plim Sample Cov}(S_i, \mathit{Abilty}_i)}{\text{plim Sample Var}(S_i)}\gamma + \frac{\text{plim Sample Cov}(S_i, \varepsilon_i)}{\text{plim Sample Var}(S_i)} \\
&= \alpha_2 + \frac{\text{Cov}(S_i, \mathit{Abilty}_i)}{\text{Var}(S_i)}\gamma + \frac{\text{Cov}(S_i, \varepsilon_i)}{\text{Var}(S_i)} \\
&= \alpha_2 + \frac{\text{Cov}(S_i, \mathit{Abilty}_i)}{\text{Var}(S_i)}\gamma
\end{aligned}$$

and:

$$\frac{\text{Cov}(S_i, Abilty_i)}{\text{Var}(S_i)}\gamma > 0.$$

Candidates need to clearly indicate (in proof or verbally) that OVB necessitates that our omitted variable is correlated with variables included in our incorrectly-specified regression, and here we expect that to be the case.

(Intuitively, we note that $\widehat{\alpha}_2$ incorporates the indirect effect $Ability_i$ has on $\log W_i$ through its correlation with S_i in addition to the direct effect S_i has on $\log W_i$. That is, $\widehat{\alpha}_2$ will take on some of the explanatory effect of $Ability_i$ due to the correlation between S_i and $Ability_i$.

The t test used for signalling significance of α_2 in fact is also invalid due to the fact that another serious consequence of omitting a relevant variable is that the standard errors of the coefficients and the test statistics in general are invalid.

Note: The severe econometric consequences of omitting relevant variables contrasts sharply with the less severe consequences of including irrelevant variables where unbiasedness remains and the test statistics also remain valid. The efficiency of our estimator is affected though. General result: Imposing valid linear restrictions allows us to obtain unbiased, less variable (more efficient) parameter estimates (lower standard errors) which enhance the power of our tests.)

Question 2

Suppose you are given a random sample $X_1, \ldots, X_n$ from the exponential distribution:

$$f(x) = \lambda \exp(-\lambda x), \qquad x > 0,\ \lambda > 0.$$

According to this distribution $\text{E}(X_i) = 1/\lambda$ and $\text{Var}(X_i) = 1/\lambda^2$ for $i = 1, \ldots, n$.

Show that the maximum likelihood estimator for λ is $1/\bar{X}$ where $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Is the estimator unbiased and/or consistent? Prove your claims.

(8 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Chapter 4.2 (Logarithmic transformations) and Chapter 10.6 (An introduction to maximum likelihood estimation).

Subject guide (2016), Chapter 10.

Approaching the question

Candidates should clearly conduct the maximum likelihood procedure and ensure proper uses of the product and logarithm operators are displayed. Many candidates made the classic error of not recognising that $\text{E}(1/\bar{X}) \neq 1/\text{E}(\bar{X})$ and also failed to realise that the proof of consistency should have used the plim operator and the law of large numbers as $\text{plim}(1/\bar{X}) = 1/(\text{plim}\ \bar{X})$. Instead, many candidates attempted (unsuccessfully) to look at the sufficient conditions (what is $\text{Var}(1/\bar{X})$?). Finally, candidates should make a clear distinction between the unknown (fixed) parameter λ and its MLE $\widehat{\lambda}$ (a random variable). The answer is as follows.

The likelihood function is given by the joint density of the data (which equals the product of the marginals given the independence of our observations):

$$L(\lambda) = \prod_{i=1}^{n} \lambda \exp(-\lambda X_i) = \lambda^n \exp\left(-\lambda \sum_{i=1}^{n} X_i\right).$$

The MLE is the value of λ which maximises this function, or by monotonicity, the log-likelihood function:

$$\ln L(\lambda) = \ln\left(\lambda^n \exp\left(-\lambda\sum_{i=1}^n X_i\right)\right) = n\ln\lambda - \lambda\sum_{i=1}^n X_i.$$

By the first-order condition, we require:

$$\left.\frac{\partial \ln L}{\partial \lambda}\right|_{\widehat{\lambda}_{MLE}} = \frac{n}{\widehat{\lambda}_{MLE}} - \sum_{i=1}^n X_i = 0 \qquad \text{or} \qquad \widehat{\lambda}_{MLE} = \frac{1}{\bar{X}} \qquad \text{with} \qquad \bar{X} = \frac{1}{n}\sum_{i=1}^n X_i.$$

(It should clearly state $\widehat{\lambda}_{MLE} = 1/\bar{X}$, saying $\lambda = 1/\bar{X}$ is incorrect.)

The MLE is *not unbiased*:

$$\mathrm{E}(\widehat{\lambda}_{MLE}) = \mathrm{E}\left(\frac{1}{\bar{X}}\right) \overset{\text{Jensen's Ineq}}{\neq} \frac{1}{\mathrm{E}(\bar{X})} = \frac{1}{\frac{1}{n}\sum \mathrm{E}\left(X_i\right)} = \lambda.$$

The MLE is *consistent*:

$$\operatorname{plim}\widehat{\lambda}_{MLE} \overset{\text{plim rules}}{=} \frac{1}{\operatorname{plim}(\bar{X})} \overset{\text{LLN}}{=} \frac{1}{\mathrm{E}(X_i)} = \lambda.$$

Question 3

Consider the following non-stationary process:

$$\boldsymbol{y_t = \gamma_0 + \gamma_1 t + u_t, \qquad \text{with } u_t = \rho u_{t-1} + \varepsilon_t}$$

and ε_t i.i.d. (0, σ^2). Indicate (with explanation) the source(s) of non-stationarity of y_t. Show that you can rewrite the model as:

$$\boldsymbol{\Delta y_t = \beta_0 + \beta_1 t + \beta_2 y_{t-1} + \varepsilon_t.}$$

Clearly indicate the one-to-one relation between $(\gamma_0, \gamma_1, \rho)$ and $(\beta_0, \beta_1, \beta_2)$. What name do we give the non-stationary process y_t when $\rho = 1$? Briefly indicate how to test whether this is the case.

(8 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Chapter 12.1 (Definition and consequences of autocorrelation), Chapter 13.1 (Stationarity and non-stationarity), and Chapter 13.4–13.5 (Tests of non-stationarity).

Subject guide (2016), Chapter 13.

Approaching the question

Candidates should show a clear understanding of what non-stationarity means. Here there are two possible sources of non-stationarity: the presence of the deterministic trend, $\gamma_1 \neq 0$, and the presence of a unit root, $\rho = 1$. The dependence structure is given by an AR(1) which may exhibit unit roots giving rise to long dependence. A discussion of the Dickey–Fuller test for the presence of a unit root is required. Answer all parts of the question: many candidates failed to answer how we call such a non-stationary process when $\rho = 1$! The answer is as follows.

A process $\{y_t\}$ is (covariance) stationary if its mean and variance exist and do not depend on time and its covariance is a function of distance in time only (not location). Violation of either of these requirements renders the process non-stationary.

In this case $\{y_t\}$ is non-stationary, because $\mathrm{E}(y_t) = \gamma_0 + \gamma_1 t$ depends on time when $\gamma_1 \neq 0$. If $\rho = 1$, there is a more serious non-stationarity problem as that indicates the presence of a unit root: in that case the variance will depend on time as well, and there will be strong dependence.

$\{u_t\}$ and, therefore, $\{y_t\}$ is an AR(1) process, which is (trend) stationary only if $|\rho| < 1$.

To obtain the testing equation, observe that we need to subtract from:

$$y_t = \gamma_0 + \gamma_1 t + u_t$$

the equation:

$$\rho y_{t-1} = \rho\gamma_0 + \rho\gamma_1 (t-1) + \rho u_{t-1}$$

to yield:

$$y_t - \rho y_{t-1} = \gamma_0 (1-\rho) + \rho\gamma_1 + \gamma_1 (1-\rho) t + \underbrace{u_t - \rho u_{t-1}}_{\varepsilon_t}.$$

By rewriting we get:

$$\Delta y_t = \underbrace{\gamma_0 (1-\rho) + \rho\gamma_1}_{\beta_0} + \underbrace{\gamma_1 (1-\rho)}_{\beta_1} t + \underbrace{(\rho - 1)}_{\beta_2} y_{t-1} + \varepsilon_t.$$

A *common mistake* is not realising that when lagging the original equation by one period that t becomes $t-1$.

If $\rho = 1$ then we call the process *difference stationary*, or integrated of order 1, $I(1)$. Observe, we would be left with:

$$\Delta y_t = \gamma_1 + \varepsilon_t \qquad \text{or} \qquad y_t = \gamma_1 + y_{t-1} + \varepsilon_t$$

i.e. a random walk with drift. Realise that if $\rho = 1$ we get $\beta_2 = 0$ and $\beta_1 = 0$.

If $\rho < 1$ while $\gamma_1 \neq 0$ then we call the process *trend stationary.* Observe, we would be left with:

$$y_t = \beta_0 + \beta_1 t + \rho y_{t-1} + \varepsilon_t.$$

Here y_t has a deterministic trend with short-run fluctuations represented by a stationary AR(1).

We can use a Dickey–Fuller test to test $\mathrm{H}_0 : \rho = 1$ (difference stationary) against $\mathrm{H}_1 : \rho < 1$ (trend stationary). Using the test statistic $\widehat{\beta}_2/\mathrm{SE}(\widehat{\beta}_2)$ and the Dickey–Fuller critical values (with trend and constant), we reject H_0 if $\widehat{\beta}_2/\mathrm{SE}(\widehat{\beta}_2) < \tau$.

Question 4

Explain the RESET test as a general test for functional form misspecification and discuss the drawbacks and advantages of this test.

In your answer consider the following multiple linear regression model:

$$\boldsymbol{y_i = \gamma_1 + \gamma_2 x_{2i} + \gamma_3 x_{3i} + u_i \qquad i = 1, \ldots, n,}$$

where $\boldsymbol{x_{2i}}$ and $\boldsymbol{x_{3i}}$ are exogenous variables known to affect $\mathbf{E}(\boldsymbol{y_i})$.

(8 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Chapter 4.3 (Models with quadratic and interactive variables).

Subject guide (2016), Chapter 4.

Approaching the question

Candidates should show a clear understanding of what the RESET test is and what the drawbacks and advantages of the test are. Many candidates were unable to answer this question. (In view of the fact that the RESET test is not explicitly mentioned in Appendix A of the subject guide, with the approval of the external examiner, steps were undertaken to ensure candidates were not affected by this in their overall classification.)

To perform the RESET test, we first perform OLS on our model and obtain the fitted values:

$$\widehat{y}_i = \widehat{\gamma}_1 + \widehat{\gamma}_2 x_{2i} + \widehat{\gamma}_3 x_{3i} \qquad \text{for } i = 1, \ldots, n.$$

Next we obtain $\widehat{y}_i^2 = (\widehat{\gamma}_1 + \widehat{\gamma}_2 x_{2i} + \widehat{\gamma}_3 x_{3i})^2$. This variable will pick up quadratic and interactive non-linearities, if present, without necessarily being highly correlated with any of the x variables and consuming only 1 degrees of freedom.

Next, we perform the following regression:

$$y_i = \gamma_1 + \gamma_2 x_{2i} + \gamma_3 x_{3i} + \gamma_4 \widehat{y}_i^2 + e_i \qquad \text{for } i = 1, \ldots, n$$

and test for the significance of γ_4 with $\text{H}_0 : \gamma_4 = 0$ against $\text{H}_1 : \gamma_4 \neq 0$. For this we use the t test where $\widehat{\gamma}_4/\text{SE}(\widehat{\gamma}_4) \sim t_{n-4}$ under H_0. If we reject H_0 then we find evidence of some type of non-linearity. (The RESET test may include $\widehat{y}_i^3$ as well, in which case it becomes an F test.)

Drawbacks: The test does not indicate the actual form of the non-linearity and it may fail to detect other types of non-linearity.

Advantages: Easy to implement, without incurring a serious loss in degrees of freedom or inducing problems of near multicollinearity.

Question 5

Consider the following regression model:

$$\boldsymbol{Y_i = \beta_0 + \beta_1 \frac{1}{X_i} + u_i, \qquad i = 1, \ldots, n.}$$

We assume that the errors $\{u_i\}_{i=1}^n$ are independent normal random variables with zero mean and variance σ^2/X_i^2. The regressor, $1/X_i$, is non-stochastic with positive sample variability. You are interested in testing the hypothesis $\text{H}_0 : \beta_0 = 0$ against $\text{H}_1 : \beta_0 \neq 0$. You are advised to use the BLUE estimator of β_0 for this purpose.

Discuss how you would obtain the BLUE estimator of β_0 (note, you are not asked to derive this estimator).

Give two reasons why you would prefer using the BLUE estimator for β_0 instead of the OLS estimator $\widehat{\beta}_{0,OLS} = \bar{Y} - \widehat{\beta}_{1,OLS}\overline{(1/X)}$ when testing this hypothesis, where $\bar{Y} = \dfrac{1}{n}\sum Y_i$ and $\overline{(1/X)} = \dfrac{1}{n}\sum \dfrac{1}{X_i}$.

(8 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): R.10 (Type II error and the power of a test), Chapter 2.5 (The Gauss–Markov theorem), Chapter 2.6 (Testing hypotheses relating to the regression coefficients), Chapter 7.1 (Heteroskedasticity and its implications) and 7.3 (Remedies for heteroskedasticity).

Subject guide (2016), Chapter 7.

Approaching the question

Candidates should show a clear understanding of what it means for an estimator to be BLUE. Failure to recognise the presence of heteroskedasticity in this question is serious. Arguments that OLS on the regression itself would not be linear because of the form of the regressor $1/X_i$ are wrong. OLS on the regression model given is linear (it is linear in the parameters!), but due to the heteroskedasticity it is not efficient (BLUE)! Many candidates lost points simply because they did not discuss two reasons why, *for testing purposes,* one would prefer to use this BLUE estimator. Answer all parts of the question. The answer is as follows.

Since the only Gauss–Markov violation here is the presence of heteroskedasticity, we should propose to use weighted least squares (WLS) to make the problem go away. Regress:

$$Y_i X_i = \beta_0 X_i + \beta_1 + u_i X_i \qquad \text{for } i = 1, \ldots, n$$

or:

$$Y_i^* = \beta_1 + \beta_0 X_i + u_i^*.$$

This transformed model satisfies all the Gauss–Markov assumptions, $\mathrm{E}(u_i^*) = \mathrm{E}(u_i)\, X_i = 0$ and $\mathrm{Var}(u_i^*) = \mathrm{Var}(u_i X_i) = \mathrm{Var}(u_i)\, X_i^2 = \sigma^2$. Do make sure you mention that the *Gauss–Markov theorem* ensures that OLS on this regression will yield our BLUEs of β_0 and β_1.

Benefits to using the BLUE of β_0 are (i) it is an efficient (more precise) estimator and because of this efficiency the test will have higher power (easier to reject the null hypothesis when it is false), and (ii) using the BLUE estimator allows us to directly use its standard error for inference. Had we used the OLS estimator instead, we would have needed to obtain heteroskedasticity-robust standard errors.

Section B

Answer three questions from this section.

Question 6

Let us consider a model for the sale price of Monet paintings. The data we have contains the sale prices, widths, and heights of 430 Monet paintings, which sold at auction for prices ranging from \$10,000 to \$33 million. A linear regression provided the following results:

$$\widehat{\ln \text{Price}}_i = \underset{(0.612)}{-8.427} + \underset{(0.091)}{1.334} \ln \text{Area}_i - \underset{(0.128)}{0.165} \text{Aspect Ratio}_i$$

$$N = 430, R^2 = 0.336$$

where Area = Width × Height and Aspect Ratio = Height/Width. The standard errors are given in parentheses.

(a) Test the joint significance of the regression. Discuss its relation to the goodness of fit measure: R^2.

(5 marks)

(b) You want to test the hypothesis that auction prices are inelastic with respect to area. Specifically, you are asked to test $H_0 : \beta_{\ln \text{Area}} = 1$ against $H_1 : \beta_{\ln \text{Area}} > 1$. Perform this test, clearly indicating the test statistic and the rejection rule. In view of your answer, indicate whether the p-value of the test is bigger or smaller than 5%.

(5 marks)

(c) A friend points out that you should be worried about the presence of heteroskedasticity. Explain the concept of heteroskedasticity and discuss the properties of the OLS estimator in the presence of heteroskedasticity.

(5 marks)

(d) **Discuss how you should modify the test in (b) if it is known that the variance of the error term is given by:**

$$\sigma_i^2 = \exp(\gamma_0 + \gamma_1 \ln \text{Area}_i + \gamma_2 \text{Aspect Ratio}_i),$$

where $(\gamma_0, \gamma_1, \gamma_2)$ are unknown parameters. You are told to use the FGLS estimator. Explain this estimator clearly.

(5 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Chapter 2.6 (Testing hypothesis relating to the regression coefficients), Chapter 3.5 (Goodness of fit: R^2), Chapter 7.1 (Heteroskedasticity and its implications) and Chapter 7.3 (Remedies for heteroskedasticity).

Subject guide (2016), Chapter 7.

Approaching the question

In part (a), an F test should be used to test the significance of the regression and its relation with the goodness of fit measure, R^2, should be discussed. In part (b), a one-sided t test should be used. In part (c), the concept of heteroskedasticity and the consequences on the OLS estimator need to be discussed. In part (d), we are told that there is evidence of heteroskedasticity. Failure to recognise this is serious. Candidates are asked to discuss the (feasible) weighted least squares estimator. Unfortunately, many candidates were confused because of the request to use the FGLS estimator, which is not discussed in the textbook or subject guide. The issue was raised with the external examiner, and steps were undertaken to ensure candidates were not affected by this in their overall classification. This question was answered by 85% of the candidates in Zone B. The answer is as follows.

(a) We need to test $\text{H}_0 : \beta_{\ln area} = \beta_{aspect} = 0$ against H_1: $\beta_{\ln area} \neq 0$ and/or $\beta_{aspect} \neq 0$, assuming all Gauss–Markov assumptions + normality, we use the F test:

$$\frac{(RRSS - URSS)/2}{URSS/(430-3)} = \frac{R^2}{1-R^2} \times \frac{427}{2} = \frac{0.336}{1-0.336} \times \frac{427}{2} = 108.04 \sim F_{2,427}$$

under H_0. At the 5% significance level the critical value is 3.00, so we strongly reject finding evidence that the regressors are important.

If the goodness of fit $R^2 = \dfrac{ESS}{TSS} = 1 - \dfrac{RSS}{TSS}$ is large, then our explanatory variables are important (help to explain the variation in the dependent variable), hence we should reject the null hypothesis that neither variable is important.

(b) Given the Gauss–Markov assumptions + normality, we will use the t statistic, here:

$$\frac{\widehat{\beta}_{\ln area} - 1}{\text{SE}(\widehat{\beta}_{\ln area})} = \frac{1.334 - 1}{0.091} = 3.6703$$

which is t_{427} under H_0. At the 5% significance level we need to reject H_0 when it exceeds 1.645 (one-sided test), which it does. Therefore, we find evidence that auction prices are elastic.

The p-value is the lowest significance level at which we want to reject H_0. Since we reject H_0 at the 5% significance level, the p-value will be smaller than 5%.

(c) Discussion of the concept is standard bookwork, see Chapter 7.1 of Dougherty.

The properties of OLS are *inefficiency* (as heteroskedasticity is a violation of the Gauss–Markov assumptions), *unbiasedness* (heteroskedasticity does not violate the assumption that $\text{E}(u \mid X) = 0$ needed for unbiasedness) and *consistency* (heteroskedasticity does not violate the assumption of contemporaneous uncorrelatedness needed for consistency). Moreover, the *usual standard errors of the OLS estimator will be invalid* (we will need to use White's heteroskedasticity-robust standard errors).

(d) Candidates should recognise that we are now told that there is evidence of *heteroskedasticity*, $\sigma_i^2 \neq \sigma^2$, which means that we should use weighted least squares instead for efficiency purposes. That is, we should perform OLS on:

$$\frac{\ln \text{Price}_i}{\sigma_i} = \beta_0 \frac{1}{\sigma_i} + \beta_1 \frac{\ln \text{Area}_i}{\sigma_i} + \beta_2 \frac{\text{Aspect Ratio}_i}{\sigma_i} + e_i.$$

This will give us the BLUE of our parameters, where:

$$\sigma_i = \exp(0.5(\gamma_0 + \gamma_1 \ln \text{Area}_i + \gamma_2 \text{Aspect Ratio}_i)).$$

A problem with its implementation is that we cannot run the above regression if we *do not know* γ_0, γ_1 and γ_2. This is why we want to propose FGLS (or *feasible weighted least squares*) – not in the syllabus. Once (consistent) estimates of γ_0, γ_1 and γ_2 are obtained, we can use them to estimate the weights:

$$\widehat{\sigma}_i^2 = \exp(\widehat{\gamma}_0 + \widehat{\gamma}_1 \ln \text{Area}_i + \widehat{\gamma}_2 \text{Aspect Ratio}_i)$$

and perform OLS on:

$$\frac{\ln \text{Price}_i}{\widehat{\sigma}_i} = \beta_0 \frac{1}{\widehat{\sigma}_i} + \beta_1 \frac{\ln \text{Area}_i}{\widehat{\sigma}_i} + \beta_2 \frac{\text{Aspect Ratio}_i}{\widehat{\sigma}_i} + \text{error}_i.$$

These estimates are the FGLS estimates which, if the sample size is large, are efficient.

Candidates were asked to indicate that these estimates, and their associated standard errors, should be used instead when we apply the test in (b). Alternatively, the robust standard errors should have been used in (b).

To estimate the γ parameters, the following regression can be run:

$$\ln \widehat{\varepsilon}_i^2 = \gamma_0 + \gamma_1 \ln \text{Area}_i + \gamma_2 \text{Aspect Ratio}_i + v_i$$

where $\widehat{\varepsilon}_i$ are the OLS residuals! Note: if there is heteroskedasticity, the OLS residuals should display this problem. As $\ln \sigma_i^2 = \gamma_0 + \gamma_1 \ln\text{Area}_i + \gamma_2 \text{Aspect Ratio}_i$, the above equation enables us to get these estimates needed to make our WLS feasible.

Question 7

In this question we look at a large data set on weekly hours worked by women having at least two children. Consider the following specification for labour supply, estimated by OLS

$$\begin{aligned} \widehat{hours} &= \underset{(6.503)}{-17.347} - \underset{(.116)}{2.242} kids + \underset{(.035)}{.938}\, educ + \underset{(.453)}{2.089} age \underset{(.008)}{-.028}\, age^2 - \underset{(.005)}{.075}\, nonmomi \\ n &= 31,857 \text{ and } R^2 = .052 \qquad (7.1) \end{aligned}$$

where *kids* is the total number of children, *educ* is the years of schooling, *age* is the woman's age in years, *nonmomi* is income from sources other than the mother's wage income, and *hours* is the hours worked per week. The standard errors are given in parentheses.

(a) **Interpret the coefficient of *kids* and discuss its statistical significance.**

(3 marks)

(b) **It is commonly thought that the decision to have more children is correlated with unobserved factors that affect labour supply. What name do we give such a problem in econometrics, and what consequences will this have for the properties of the OLS estimator? Support your answer using a simple model.**

(5 marks)

(c) **Consider the variables *multi2nd* and *samesex*, which are binary variables indicating whether the second birth was for multiple babies and whether the first two children are of the same gender. What properties do we need to assume for us to be able to use *multi2nd* and *samesex* as instruments for *kids*? Are these assumptions reasonable and can we test (if so how) these requirements?**

(6 marks)

You may assume that *educ*, *age* and *nonmomi* can be treated as being exogenous.

(d) **The following IV results were obtained in Stata:**

$$\begin{aligned}\widehat{hours} &= \underset{(7.020)}{-16.828} - \underset{(.134)}{2.504}kids + \underset{(.115)}{.916}\,educ + \underset{(.460)}{2.105}age \underset{(.008)}{-.028}\,age^2 - \underset{(.007)}{.076}\,nonmomi \\ n &= 31,857 \text{ and } R^2 = .052 \end{aligned} \quad (7.2)$$

Discuss how these results can be obtained using Two Stage Least Squares (2SLS) and briefly discuss how you would test whether the results in (7.1) and (7.2) are significantly different from each other.

(6 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Chapter 6.5 (Testing a linear restriction), Chapter 8.3 (Instrumental variables), and Chapter 9 (Simultaneous equations estimation).

Subject guide (2016), Chapter 9.

Approaching the question

In part (a), a parameter interpretation is required and a t test should be used to test for its significance. In part (b), a discussion of simultaneous equation bias is required. The joint determinacy in simultaneous equations induces an endogeneity problem: a problem where we have a correlation between the errors and regressors. This correlation renders the OLS parameter estimates inconsistent. Technical details need to be provided. Recognising that the IV estimator avoids this problem, in part (c) the requirements on the instruments need to be discussed as it relates to the particular example. In part (d), the 2SLS estimator and the Durbin–Wu–Hausman test need to be discussed. This question was answered by 80% of the candidates in Zone B. The answer is as follows.

(a) For the population of women who have at least two children, having an additional child lowers expected hours worked by about 2.2 hours per week on average, holding other factors fixed.

The resulting t statistic for the hypothesis $H_0 : \beta_{kids} = 0$ against $H_1 : \beta_{kids} \neq 0$ is $\widehat{\beta}_{kids}/\text{SE}(\widehat{\beta}_{kids})$, which is very large in magnitude (thanks to the large sample size), and the parameter is highly significant.

(b) This is the problem called *endogeneity.*

The consequences for our OLS estimator are severe, as we will not be able to establish consistency (neither unbiasedness for that matter).

Candidates are expected to prove in a simple model that correlation between the error and regressors yields inconsistency. Need to make use of plim rules and the law of large numbers.

(c) Instruments need to satisfy the requirements of validity (they need to be uncorrelated with the error term ($\text{E}(samesex_i u_i) = \text{E}(multi2nd_i u_i) = 0$)) – typically not easy to test (can consider an overidentification test, but do not discuss). Reasonable: Having a multiple birth with the second pregnancy, and having the first two children of the same gender, are random events in the sense that they cannot be influenced (or, at least currently, are not). This suggests that the dummy variables *samesex* and *multi2nd* might be exogenous to the labour supply decision.

Instruments also need to be relevant – related to *kids*. The latter can be tested easily by looking at the joint significance of *multi2nd* and *samesex* in the (reduced form) equation:

$$kids_i = \delta_0 + \delta_1 educ_i + \delta_2 age_i + \delta_3 age_i^3 + \delta_4 nonmom_i + \delta_5 multi2nd_i + \delta_6 samesex_i + e_i$$

(that is testing $\mathrm{H}_0 : \delta_5 = \delta_6 = 0$) for which you can use the F test. Reasonable: Clearly having multiple births for the second pregnancy will increase the number of children and the interest of having children of different genders might indicate that when the first two births have the same gender that the parents might have tried again (i.e. increase total number of children).

(d) Discuss the 2SLS approach:

- Step 1: Requires us to make the 'bad' variable 'good' by regressing it on all exogenous variables in our model:

$$kids_i = \delta_0 + \delta_1 educ_i + \delta_2 age_i + \delta_3 age_i^3 + \delta_4 nonmom_i + \delta_5 multi2nd_i + \delta_6 samesex_i + e_i.$$

  We obtain the fitted values of this regression: $\widehat{kids}$. All exogenous variables have to be included here!

- Step 2: Requires us to estimate the equation:

$$hours_i = \beta_0 + \beta_1 \widehat{kids}_i + \beta_2 educ_i + \beta_3 age_i + \beta_4 age_i^2 + \beta_5 nonmom_i + u_i.$$

  The parameter estimates of this regression are our 2SLS estimates.

Candidates are expected to give a (not too technical) discussion of the Durbin–Wu–Hausman test. Requires to compare the IV and OLS estimates: we want to reject the null hypothesis of absence of endogeneity if they are quite different (relative to their precision). Since both estimators are consistent under the null hypothesis, both estimates should be similar if there is no endogeneity. Explicit form of the test need not be given for full marks.

Question 8

Let us consider an analysis on recidivism (probability of re-arrest) among a group of young men in California who have at least one arrest prior to 1986. The dependent variable, $arr86$, is equal to unity if the man was arrested at least once during 1986, and zero otherwise.

	OLS A	OLS B	Logit A	Logit B	Logit B Marginal Effect
pcnv	−.152 (.021)	−.162 (.021)	−.880 (.122)	−.901 (.120)	−.176 (.023)
avgsen	.005 (.006)	.006 (.006)	.027 (.035)	.031 (.034)	.006 (.007)
tottime	−.003 (.005)	−.002 (.005)	−.014 (.028)	−.010 (.027)	−.002 (.005)
ptime86	−.023 (.005)	−.022 (.005)	−.140 (.031)	−.127 (.031)	−.025 (.006)
qemp86	−.038 (.005)	−.043 (.005)	−.199 (.028)	−.216 (.028)	−.042 (.005)
black	.170 (.024)	—	.823 (.117)	—	—
hispan	.096 (.021)	—	.522 (.109)	—	—
constant	.380 (.019)	.380 (.019)	−.464 (.095)	−.169 (.084)	
R^2	.068	.047			
log L			−1512.35	−1541.24	

$pcnv$ is the proportion of prior arrests that led to a conviction, $avgsen$ is the average sentence served from prior convictions, $tottime$ is the months spent in prison since age 18 prior to 1986, $ptime86$ is months spent in prison in 1986, $qemp86$ is the number of quarters the man was legally employed in 1986, while $black$ and $hispan$ are two race dummies ($white$ the excluded dummy). The standard errors are reported in parentheses.

(a) When estimating the parameters by OLS, we are using the Linear Probability Model. Why might you then report heteroskedasticity-robust standard errors?

(2 marks)

(b) Using the OLS B results, what is the estimated effect on the probability of arrest if $pcnv$ goes from 0.25 to 0.75 holding everything else constant?

(4 marks)

(c) It is argued that the linear probability model is not appropriate for explaining the binary variable $arr86$ and a logit regression model has been estimated. Explain how the Logit estimates are obtained.

(5 marks)

***Hint:* You may recall, that for the Logit model A, we will specify**

$$\Pr(arr86_i = 1) = \Lambda\left(\beta_0 + \beta_1 pcnv + \beta_2 avgsen + \cdots + \beta_6 black + \beta_7 hispan\right)$$

where $\Lambda(z) = \frac{1}{1+\exp(-z)}$.

(d) Using the Logit model results, discuss whether $black$ and $hispan$ are jointly significant. Clearly indicate the null and alternative hypothesis, the test statistic and the rejection rule.

(3 marks)

(e) An important distinction between the two approaches is that the marginal effect of $pcnv$ on the probability of re-arrest is constant for the LPM unlike the marginal effect using the logit analysis. What this means for instance is that the estimated effect on the probability of arrest if $pcnv$ goes from 0.25 to 0.75 will depend on the other characteristics.

i. Explain how the marginal effects evaluated at the mean values of the explanatory variables (reported in the last column) were obtained. Give a brief comment as to how they compare to the marginal effect of the associated LPM.

(3 marks)

ii. Using the Logit B results, how would you obtain the estimated effect on the probability of arrest if $pcnv$ goes from 0.25 to 0.75 for a $white$ man, with characteristics $avgsen = 1$, $tottime = 1$, $ptime86 = 0$ and $qemp86 = 2$. A clear explanation of what calculations are required is sufficient.

(3 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Chapter 10.1 (The linear probability model), Chapter 10.2–10.3 (Logit and Probit analysis) and Chapter 10.6 (An introduction to maximum likelihood estimation).

Subject guide (2016), Chapter 10.

Approaching the question

In parts (a) and (b), a discussion of the LPM and an interpretation of its parameters – constant marginal effects – need to be given. In parts (c) and (e), a related discussion needs to be given for the logit model. Unlike in the LPM where we imposed linearity, the marginal effects in the logit model are no longer constant due to the fact that in the logit model we use a non-linear, $\Lambda(z)$, specification of the probabilities. In part (c), a discussion of the maximum likelihood estimator used for our logit parameter estimates is required, and in part (d) a test for joint significance using the likelihood ratio test is needed. This question was answered by 39% of the candidates in Zone B. The answer is as follows.

(a) The linear probability model suffers from heteroskedasticity. Specifically, the heteroskedasticity takes the form $\text{Var}(arr86_i \mid X_i) = p_i\,(1 - p_i)$ with:

$$p_i = \text{E}(arr86_i \mid X_i) = \beta_0 + \beta_1 pcnv_i + \beta_2 avgsen_i + \cdots + \beta_6 black_i + \beta_7 hispan_i.$$

Since this invalidates the standard errors, we will want to report heteroskedasticity-robust standard errors instead.

(b) Candidates should clearly indicate that $\widehat{P}_i = \widehat{\beta}_0 + \widehat{\beta}_1 pcnv_i + \widehat{\beta}_2 avgsen_i + \cdots + \widehat{\beta}_5 qemp86_i$.

The estimated effect is:

$$\widehat{\beta}_1(0.75 - 0.25) = -0.162 \times 0.5 = -0.081.$$

Therefore, the probability of re-arrest decreases by 0.081 or by 8.1 percentage points. (An 8.1% decrease is different from an 8.1 percentage point decrease!)

(c) Candidates are expected to indicate that we use the maximum likelihood estimator (MLE) to estimate the β parameters. Make sure you answer the question, you were asked how the estimates are obtained! Few candidates were able to give relevant details.

The log-likelihood function that will be maximised is given by:

$$\begin{aligned}
\log L(\beta) &= \sum_{i=1}^{n} arr86_i \times \log(\Pr(arr86_i = 1)) + (1 - arr86_i) \times \log(\Pr(arr86_i = 0)) \\
&= \sum_{i=1}^{n} arr86_i \times \log(\Lambda(z_i)) + (1 - arr86_i) \times \log(1 - \Lambda(z_i)) \\
\text{where } z_i &= \beta_0 + \beta_1 pcnv_i + \beta_2 avgsen_i + \cdots + \beta_6 black_i + \beta_7 hispan_i.
\end{aligned}$$

The MLE will set the partial derivatives equal to zero:

$$\left.\frac{\partial \log L(\beta)}{\partial \beta_i}\right|_{\widehat{\beta}_{MLE}} = 0 \qquad \text{for } i = 0, \ldots, 7$$

i.e. it chooses those values of $\widehat{\beta}$ that will set the derivatives equal to zero. Heuristically, these parameter estimates ensure that the predicted $\Pr(arr86_i = 1)$ is large for individuals for whom $arr86_i = 1$, and the predicted $\Pr(arr86_i = 0)$ is large for individuals for whom $arr86_i = 0$. This ensures that the 'likelihood' of observing the data is largest for these values of β.

(d) Candidates should indicate they want to test $\text{H}_0 : \beta_{black} = \beta_{hispan} = 0$ against $\text{H}_1 : \beta_{black} \neq 0$ and/or $\beta_{hispan} \neq 0$.

The likelihood ratio (LR) test statistic equals $-2(\ln L^R - \ln L^U) \overset{a}{\sim} \chi^2_2$ under H_0, where $\ln L^R$ is the log-likelihood function of the restricted model (Logit B) and $\ln L^U$ is the log-likelihood function of the unrestricted model (Logit A). $LR = -2 \times (-1541.24 - -1512.35) = 57.78$.

With a 5% significance level our critical value is given by 5.99, so we clearly reject the null hypothesis, rendering ethnicity an important factor in explaining re-offending rates.

(e) i. The marginal effects of interest describe how $\Pr(arr86_i)$ changes as a result of the explanatory variables.

For continuous variables that means, for example:

$$\begin{aligned}
\frac{\partial \Pr(arr86_i)}{\partial pcnv} &= f(z_i) \times \beta_1 \qquad \text{with } f(z) = \frac{\text{d}\Lambda(z)}{\text{d}z} = \frac{\text{e}^{-z}}{(1 + \text{e}^{-z})^2} \\
\text{with } z_i &= \beta_0 + \beta_1 pcnv_i + \beta_2 avgsen_i + \cdots + \beta_6 black_i + \beta_7 hispan_i
\end{aligned}$$

(no need to give the derivative itself). Therefore, the marginal effects depend on the characteristics of the individual. The marginal effects reported are for an individual with average characteristics:

$$\bar{z} = \beta_0 + \beta_1 \overline{pcnv} + \beta_2 \overline{avgsen} + \cdots + \beta_6 \overline{black} + \beta_7 \overline{hispan}.$$

(Whether this individual exists is another matter. Indeed, it may be more interesting to report the average of these marginal effects over all individuals.)

The marginal effects for the average person are quite comparable to those obtained by the LPM (an easy mark, so do not forget to mention this – answer all parts of the question!).

Candidates making these points clearly in (c) were rewarded here.

ii. Candidates here will need to observe that we need to compare predicted probabilities using the logit specification of the probabilities: $\Lambda(z) = \dfrac{1}{1+\exp(-z)}$.

$\widehat{\Pr}(arr86 = 1 \mid pcnv = 0.25, avgsen = 1, tottime = 1, ptime86 = 0, qemp86 = 2) = \Lambda(z_1)$, where:

$$z_1 = 0.25 \times (-0.901) + 1 \times (0.031) + 1 \times (-0.010) + 0 \times (-0.127) + 2 \times (-0.216) = -0.636$$

and $\Lambda(z_1) = \dfrac{1}{1+\mathrm{e}^{0.636}} = 0.346$.

$\widehat{\Pr}(arr86 = 1 \mid pcnv = 0.75, avgsen = 1, tottime = 1, ptime86 = 0, qemp86 = 2) = \Lambda(z_2)$, where:

$$z_2 = 0.75 \times (-0.901) + 1 \times (0.031) + 1 \times (-0.010) + 0 \times (-0.127) + 2 \times (-0.216) = -1.087$$

and $\Lambda(z_2) = \dfrac{1}{1+\mathrm{e}^{1.087}} = 0.252$.

So we see a drop in the probability of re-arrest equalling $0.346 - 0.252 = 0.094$.

Candidates are not expected to provide explicit numbers. Full marks can be obtained for clarity of approach. Failure to recognise the importance of Λ is serious.

Question 9

Let us consider a distributed lag model:

$$y_t = \beta_0 + \delta_0 z_t + \delta_1 z_{t-1} + \cdots + \delta_q z_{t-q} + \beta_1 x_t + u_t\,, \qquad t = 1, \ldots, T$$

where u_t is independent of $x_t, z_t, z_{t-1}, \ldots,$ and z_{t-q} with zero mean and constant variance. For simplicity, we will argue that there is no autocorrelation in the errors.

1. **Explain the concept of autocorrelation and indicate for the above model what the consequence of autocorrelation for our OLS estimator would be.**

(4 marks)

Let us consider the example of the effects of tax policy on the U.S. fertility rates. Let gfr denote the number of children born per 1,000 women aged 15–44, pe denotes the real value of the personal tax exemption, and $ww2$ and $pill$ are dummy variables (WW II, availability of the birth control pill). Using annual data, the following OLS results were obtained:

$$\begin{aligned}\widehat{gfr}_t \;=\;& \underset{(3.30)}{92.50} + \underset{(.126)}{.089}\,pe_t - \underset{(.153)}{.004}\,pe_{t-1} + \underset{(.165)}{.007}\,pe_{t-2} + \underset{(.154)}{.018}\,pe_{t-3} + \underset{(.105)}{.014}\,pe_{t-4} \\ & - \underset{(11.03)}{21.48}\,ww2_t - \underset{(3.94)}{31.25}\,pill_t, \qquad R^2 = .537, T = 68.\end{aligned} \tag{9.1}$$

The usual standard errors are reported in parentheses.

(b) **A novice in econometrics argues that the result indicates that the effect of tax policy on the US fertility rate is ineffective because of the insignificance of the distributed lag coefficients (the coefficients on pe_t and its lags). Explain why he/she could be wrong. In your answer you may discuss what the short and long run effect of the tax policy are according to these results.**

(5 marks)

(c) **Your friend argues that only two lags should have been included. Using the same sample, he obtains the following result:**

$$\begin{aligned}\widehat{gfr}_t &= \underset{(3.27)}{92.52} + \underset{(.119)}{.101}pe_t - \underset{(.147)}{.011}pe_{t-1} + \underset{(.119)}{.033}pe_{t-2} \\ &\quad - \underset{(10.13)}{22.95}ww2_t - \underset{(3.76)}{30.83}pill_t, \qquad R^2 = .536,\ T = 68.\end{aligned} \tag{9.2}$$

Test for the joint significance of the third and fourth lag. Clearly indicate H_0 and H_1, the test statistic, the rejection rule and interpret your results.

(4 marks)

(d) **Your result in (c) may be affected by the presence of autocorrelation. Discuss how you would conduct the Breusch–Godfrey test for the presence of autocorrelation in (9.2). Clearly indicate H_0 and H_1, the test statistic, the rejection rule and interpret your results. Briefly indicate what you might want to do to try and remove the autocorrelation.**

(7 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Chapter 3.4 (Multicollinearity), Chapter 6.5 (Testing a linear restriction), Chapter 11.5 (Assumption C.7 and the properties of estimators), Chapter 11.3 (Models with lagged explanatory variables) and Chapter 12.1–12.3 (Definition and consequences of autocorrelation; Fitting a model subject to AR(1) autocorrelation).

Subject guide (2016), Chapters 11 and 12.

Approaching the question

In part (a), the concept of autocorrelation and the consequences on the OLS estimator need to be discussed. In part (b), candidates are expected to discuss the problem associated with multicollinearity. An interpretation of the parameters in distributed lag models (short- and long-run effects) can be given for partial credit. In part (b), a one sided t test for the LRP is required and candidates are expected to discuss a reparameterisation of the model that enables one to get the standard error of the LRP parameter. In part (c), an F test should be used to test for the joint significance of the third and fourth lag. Implementing a test for significance of the regression here is wrong – answer the question! In part (d), the Breusch–Godfrey test for the presence of autocorrelation should be discussed, and candidates should discuss ways of removing the problem of autocorrelation. This question was answered by 36% of the candidates in Zone B. The answer is as follows.

(a) Discussion of the concept is standard bookwork, see Chapter 12.1 of Dougherty.

The properties of OLS include *inefficiency* (autocorrelation is a violation of the Gauss–Markov assumptions). The OLS parameters will be *biased* (we do not have independence of the errors with all (future) values of the regressors – see Chapter 11.5 of Dougherty), but *consistent* (because of the presence of autocorrelation, it is important to point out that there is no lagged dependent variable here; consistency requires the uncorrelatedness between the error and regressors which are satisfied) and *invalid standard errors* and we will need to use HAC-robust standard errors.

(b) The distributed lag coefficients are $\delta_0, \delta_1, \ldots, \delta_4$ in this case. They tell us how if pe (z) increases by one unit today, but then falls back to its original level, y will change in each future period. Observe:

$$\begin{aligned}
\mathbf{y_t} &= \beta_0 + \boxed{\boldsymbol{\delta_0 z_t}} + \delta_1 z_{t-1} + \cdots + \delta_4 z_{t-4} + \beta_1 x_t + u_t \text{ (impact on } y_t\text{; contemporaneous)} \\
\mathbf{y_{t+1}} &= \beta_0 + \delta_0 z_{t+1} + \boxed{\boldsymbol{\delta_1 z_t}} + \cdots + \delta_4 z_{t-3} + \beta_1 x_{t+1} + u_{t+1} \text{ (impact on } y_{t+1}\text{)} \\
&\vdots \\
\mathbf{y_{t+4}} &= \beta_0 + \delta_0 z_{t+4} + \delta_1 z_{t+3} + \cdots + \boxed{\boldsymbol{\delta_4 z_t}} + \beta_1 x_{t+1} + u_{t+1} \text{ (impact on } y_{t+4}\text{).}
\end{aligned}$$

The novice is wrong, because the individual insignificance can be caused by the *near multicollinearity* of $pe_t, pe_{t-1}, \ldots, pe_{t-4}$. (Jointly, they are likely to be significant.)

By removing lags, more significant coefficients may be found.

The short-run impact is given by δ_0, while the long-run impact is given by $\delta_0 + \delta_1 + \cdots + \delta_4$. The LRP typically is significant even if the distributed lag coefficients are not.

(c) Candidates should test $\mathrm{H}_0 : \beta_{pe_{-3}} = 0$ and $\beta_{pe_{-4}} = 0$ against $\mathrm{H}_1 : \beta_{pe_{-3}} \neq 0$ and/or $\beta_{pe_{-4}} \neq 0$.

We want to use the F test here, which is distributed (asymptotically) as $F_{2,60}$, and at the 5% significance level we would reject H_0 if it exceeds 3.15.

With R_u^2 denoting the R^2 of the unrestricted model and R_r^2 denoting the R^2 of the restricted model, the F test is given by:

$$F = \frac{(R_u^2 - R_r^2)/2}{(1 - R_u^2)/60} = \frac{(0.537 - 0.536)/2}{(1 - 0.537)/60} = 0.065.$$

So this suggests that two lags would suffice, which would reduce the evidence of multicollinearity.

(d) When testing for autocorrelation, let us assume that:

$$u_t = \rho u_{t-1} + e_t \qquad \text{with } |\rho| < 1 \text{ and } e_t \text{ white noise (i.i.d. } (0, \sigma^2)).$$

A test for autocorrelation then becomes:

$$\mathrm{H}_0 : \rho = 0 \text{ (no autocorrelation)} \qquad \text{vs.} \qquad \mathrm{H}_1 : \rho \neq 0 \text{ (autocorrelation).}$$

As with a test of heteroskedasticity, a test of autocorrelation makes use of the OLS residuals – if there is autocorrelation, our OLS residuals will display this relationship.

The testing equation we should use is:

$$\widehat{u}_t = \gamma_0 + \rho \widehat{u}_{t-1} + \gamma_1 pe_t + \gamma_2 pe_{t-1} + \gamma_3 pe_{t-2} + \gamma_4 ww2_t + \gamma_1 pill_t + v_t.$$

We need to use the t test (and F test if general AR(p) or MA(p) is assumed) and we should reject the null hypothesis of zero autocorrelation if the test statistic $\widehat{\rho}/\mathrm{SE}(\widehat{\rho})$ in absolute value is larger than 1.96 (the asymptotic critical value).

To remove autocorrelation you may want to *introduce more dynamics in the model*, for example introduce lagged dependent variables. The Cochrane–Orcutt procedure (or other non-linear approach) can be suggested as well, but that requires us to know the form of autocorrelation which we may not (some details need to be provided).

Question 10

Consider the model:

$$\boldsymbol{y_t = \alpha_1 y_{t-1} + u_t, \qquad t = 1, \ldots, T}$$

where $\boldsymbol{y_0 = 0}$. $\boldsymbol{\mathrm{E}(u_t) = 0}$, $\boldsymbol{\mathrm{E}(u_t^2) = \sigma^2}$ and $\boldsymbol{\mathrm{E}(u_t u_s) = 0}$ when $\boldsymbol{s \neq t}$, for all $\boldsymbol{s, t = 1, \ldots, T}$.

(a) Discuss what we mean by the concept of stationarity (more precisely 'covariance stationarity') and indicate under what condition $\{y_t\}_{t=1}^T$ will be stationary.

(3 marks)

(b) Discuss the Dickey–Fuller procedure used to test for the presence of a unit root in the above model. Clearly indicate the null and alternative hypothesis, test statistic and rejection rule.

(5 marks)

(c) Consider a slight variation of the above model:

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + v_t, \qquad t = 1, \ldots, T$$

where $y_0 = 0$. $\mathrm{E}(v_t) = 0$, $\mathrm{E}(v_t^2) = \sigma^2$ and $\mathrm{E}(v_t v_s) = 0$ when $s \neq t$, for all $s, t = 1, \ldots, T$. What do we call such a process? Discuss what problem you will have when conducting your test as in (b).

(3 marks)

(d) Instead of conducting the Dickey–Fuller procedure, you are told to apply the Augmented DF test. Indicate how you would conduct the test for the presence of a unit root here. Derivation of the test equation will be required for full marks.

(5 marks)

(e) What are the potential problems associated with performing a regression with $I(1)$ variables? In your answer explain what it means for a variable to be $I(1)$.

(4 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Chapter 13.1 (Stationarity and nonstationarity), Chapter 13.4–13.5 (Tests of nonstationarity), and Chapter 13.6 (Cointegration).

Subject guide (2016), Chapter 13.

Approaching the question

In part (a), the concept of (covariance) stationarity needs to be given. In part (b), you need to provide the Dickey–Fuller test for unit roots and in (d) you need to provide the Augmented Dickey–Fuller test. The latter allows us to deal with the fact that when conducting the Dickey–Fuller test we cannot have any autocorrelation, which would be the case if the dependence is not AR(1), but as is the case here AR(2). In (e), a discussion of the spurious regression result is expected. This question was answered by 45% of the candidates in Zone B. The answer is as follows.

(a) Candidates should provide a textbook definition of (covariance) stationarity.

The requirement for (covariance) stationarity here is that $|\alpha_1| < 1$. (Stating $\alpha_1 < 1$ is permitted.) Here $\{y_t\}$ is described by an AR(1) process, which is stationary provided its coefficient, α_1, is smaller than 1 (in magnitude). If α_1 is equal to one we have a unit root, whereas if $|\alpha_1| > 1$ we have an explosive process.

(b) To test for a unit root, we want to estimate the following regression:

$$\Delta y_t = \gamma y_{t-1} + u_t \qquad \text{where } \gamma = (\alpha_1 - 1).$$

We test $\mathrm{H}_0 : \gamma = 0$ (nonstationarity) against $\mathrm{H}_1 : \gamma < 0$ (stationarity). (Alternatively, we estimate the original model and test $\mathrm{H}_0 : \alpha_1 = 1$ (nonstationarity, unit root) against $\mathrm{H}_1 : \alpha_1 < 1$ (stationary AR process).)

We use the Dickey–Fuller t test $\widehat{\gamma}/\mathrm{SE}(\widehat{\gamma}) = (\widehat{\alpha}_1 - 1)/\mathrm{SE}(\widehat{\alpha}_1)$. Because we have non-stationarity under the null hypothesis, this test is not standard and we have to use the

Dickey–Fuller critical values. Given the alternative hypothesis, we reject H_0 when $\widehat{\gamma}/\text{SE}(\widehat{\gamma}) < \tau$, where τ is the Dickey–Fuller critical value (no trend or intercept in the model). Candidates can, alternatively, discuss the Dickey–Fuller scaled coefficient test which is $T(\widehat{\alpha}_1 - 1) = T\widehat{\gamma}$.

If we reject the null hypothesis we have found evidence that our process is stationary (weakly dependent).

In the presence of a unit root, a shock to u_t will have an everlasting effect on the process $\{y_t\}$. Unit roots are persistent and strongly dependent.

(c) Here $\{y_t\}$ is an AR(2) process. (ADL(2, 0) is also acceptable.)

The problem with (b) will be serious, as *dependence (autocorrelation) in the error* will cause the error term u_t (which would then be equal to $\alpha_2 y_{t-2} + v_t$) and regressor y_{t-1} to be correlated and hence OLS will be inconsistent. Hence our test $\widehat{\gamma}/\text{SE}(\widehat{\gamma})$ (or $T(\widehat{\alpha}_1 - 1)$) will be *invalid*.

(d) The Augmented Dickey–Fuller test suggests that we add further lags in our testing equation to remove the autocorrelation. In particular, we should estimate the following model:

$$\Delta y_t = \gamma_1 y_{t-1} + \gamma_2 \Delta y_{t-1} + v_t$$

and test the hypothesis $H_0 : \gamma_1 = 0$ (non-stationarity) against $H_1 : \gamma_1 < 0$ (stationarity). We should use the test given by $ADF = \widehat{\gamma}_1/\text{SE}(\widehat{\gamma}_1)$ and we should reject H_0 when ADF is smaller than the critical value given by the Dickey–Fuller tables for a given significance level.

To derive this result observe that:

$$\begin{aligned}
\Delta y_t &= \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + v_t - y_{t-1} \\
&= (\alpha_1 - 1)\, y_{t-1} + \alpha_2 y_{t-2} + v_t \\
&= (\alpha_1 - 1)\, y_{t-1} + \alpha_2 \left(\underbrace{y_{t-2} - y_{t-1}}_{-\Delta y_{t-1}} + y_{t-1} \right) + v_t \\
&= (\alpha_1 + \alpha_2 - 1)\, y_{t-1} - \alpha_2 \Delta y_{t-1} + v_t
\end{aligned}$$

so $\gamma_1 = (\alpha_1 + \alpha_2 - 1)$ and $\gamma_2 = -\alpha_2$. Therefore, to test for a unit root in the AR(2) model we test whether $\alpha_1 + \alpha_2 < 1$.

(e) To say that a variable is $I(1)$ indicates that the variable has a unit root. We say that the variable then is integrated of order 1, revealing that by differencing the variable once we can make it stationary.

The potential problem associated with performing such a regression is that we may get a *spurious relation*. This is the setting where, due to the fact that both variables are trending, there is an appearance of a relationship that does not exist at all (high t statistics and a large R^2). To ensure that we have a meaningful (long-run) relationship between variables, we want to verify that instead we are dealing with a *cointegrating relationship*.

Examiners' commentaries 2018

EC2020 Elements of econometrics

Important note

This commentary reflects the examination and assessment arrangements for this course in the academic year 2017–18. The format and structure of the examination may change in future years, and any such changes will be publicised on the virtual learning environment (VLE).

Information about the subject guide and the Essential reading references

Unless otherwise stated, all cross-references will be to the latest version of the subject guide (2016). You should always attempt to use the most recent edition of any Essential reading textbook, even if the commentary and/or online reading list and/or subject guide refer to an earlier edition. If different editions of Essential reading are listed, please check the VLE for reading supplements – if none are available, please use the contents list and index of the new edition to find the relevant section.

General remarks

Learning outcomes

At the end of the course, and having completed the Essential reading and activities, you should be able to:

- describe and apply the classical regression model and its application to cross-section data
- describe and apply the:
  - Gauss–Markov conditions and other assumptions required in the application of the classical regression model
  - reasons for expecting violations of these assumptions in certain circumstances
  - tests for violations
  - potential remedial measures, including, where appropriate, the use of instrumental variables
- recognise and apply the advantages of logit, probit and similar models over regression analysis when fitting binary choice models
- competently use regression, logit and probit analysis to quantify economic relationships using standard regression programmes (Stata and EViews) in simple applications
- describe and explain the principles underlying the use of maximum likelihood estimation
- apply regression analysis to time-series models using stationary time series, with awareness of some of the econometric problems specific to time series applications (for example, autocorrelation) and remedial measures
- recognise the difficulties that arise in the application of regression analysis to nonstationary time series, know how to test for unit roots, and know what is meant by cointegration.

Common mistakes committed by candidates

A large number of candidates are not able to clearly distinguish between sample variance and covariance, and population variance and covariance (this is happening year after year).

The use of $\text{Cov}(X, Y)$ and $\text{Var}(X)$ should be restricted to describing the population covariance and variances, respectively, with definitions:

$$\text{Cov}(X, Y) = \text{E}\left((X - \text{E}(X))(Y - \text{E}(Y))\right) = \text{E}(XY) - \text{E}(X)\,\text{E}(Y)$$

and:

$$\text{Var}(X) = \text{E}((X - \text{E}(X))^2) = \text{E}(X^2) - (\text{E}(X))^2$$

(you also may denote $\text{Cov}(X, Y) = \sigma_{XY}$ and $\text{Var}(X) = \sigma_X^2$). They are typically unknown, but fixed, quantities.

The sample covariance and variance are estimators of the population covariance and variance, respectively. They are defined as:

$$\text{Sample Cov}(X, Y) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})$$

and:

$$\text{Sample Var}(X) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

(you also may use $\widehat{\sigma}_{XY}$ and $\widehat{\sigma}_X^2$). You can compute them given the data.

With a slight abuse of notation, we often divide by n instead, which is irrelevant if we let n be large. The division by $n - 1$ is a finite sample issue only (unbiasedness).

The sample covariance and variance show up in our definition of the OLS estimator of the slope in the simple linear regression model, not the population covariance and variance, as:

$$\widehat{\beta} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \frac{\text{Sample Cov}(X, Y)}{\text{Sample Var}(X)} \neq \frac{\text{Cov}(X, Y)}{\text{Var}(X)}.$$

Treating them as being the same results in incorrect analyses and candidates losing significant marks.

Candidates should realise that $\frac{1}{n}\sum_{i=1}^{n} u_i$ is not the same as $\text{E}(u_i)$. So, while we typically assume $\text{E}(u_i) = 0$, this does not guarantee that $\frac{1}{n}\sum_{i=1}^{n} u_i = 0$. Also, while we may be happy to assume $\text{E}(x_i u_i) = 0$ (uncorrelatedness between the errors and regressors), this does not guarantee that $\frac{1}{n}\sum_{i=1}^{n} x_i u_i = 0$. Note that:

- both $\frac{1}{n}\sum_{i=1}^{n} u_i$ and $\frac{1}{n}\sum_{i=1}^{n} x_i u_i$ are random variables, which take the value 0 with probability 0 (continuous random variables)!
- $\text{E}(u_i) = 0$ and $\text{E}(x_i u_i) = 0$ are fixed, not stochastic!

The differences between sample and population moments need to come across clearly when looking at unbiasedness and making consistency arguments. In both cases, we first simplify our estimator (plug in the true model) to obtain:

$$\widehat{\beta} = \beta + \frac{\sum_{i=1}^{n}(X_i - \bar{X})u_i}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \beta + \frac{\sum_{i=1}^{n} x_i u_i}{\sum_{i=1}^{n} x_i^2} \qquad \text{with } x_i = X_i - \bar{X}.$$

- For *unbiasedness*, clearly indicate that you want to show that $\mathrm{E}(\widehat{\beta}) = \beta$. Unbiasedness does not follow from $\sum_{i=1}^{n} x_i u_i = 0$, instead it follows from $\mathrm{E}\left(\frac{\sum_{i=1}^{n} x_i u_i}{\sum_{i=1}^{n} x_i^2}\right) = 0$.

  If we treat x_i as fixed, $\mathrm{E}\left(\frac{\sum_{i=1}^{n} x_i u_i}{\sum_{i=1}^{n} x_i^2}\right) \equiv \mathrm{E}\left(\sum_{i=1}^{n} d_i u_i\right) = \sum_{i=1}^{n} d_i \,\mathrm{E}(u_i)$ and then unbiasedness follows as $\mathrm{E}(u_i) = 0$.
- For *consistency*, clearly indicate that you want to show that $\mathrm{plim}(\widehat{\beta}) = \beta$. Using the plim properties, we show:

$$
\begin{aligned}
\mathrm{plim}\ \widehat{\beta} = \beta + \mathrm{plim}\left(\frac{\sum_{i=1}^{n} x_i u_i}{\sum_{i=1}^{n} x_i^2}\right) &= \beta + \frac{\mathrm{plim}\left(\frac{1}{n}\sum_{i=1}^{n} x_i u_i\right)}{\mathrm{plim}\left(\frac{1}{n}\sum_{i=1}^{n} x_i^2\right)} \\
&\equiv \beta + \frac{\mathrm{plim}\,(\text{Sample Cov}\,(x,u))}{\mathrm{plim}\,(\text{Sample Var}\,(x))} \\
&= \beta + \frac{\mathrm{Cov}\,(x,u)}{\mathrm{Var}\,(x)} \qquad \text{using the law of large numbers}
\end{aligned}
$$

  where $\mathrm{Cov}(x, u) = 0$ and $\mathrm{Var}(x) > 0$, ensuring we get consistency.
- Remember, the law of large numbers ensures that sample averages converge to their population analogues.

Candidates struggled to give competent answers to the interpretation of empirical results. When interpreting an empirical result you should discuss the significance of the coefficients, magnitude and sign of the coefficients.

When conducting hypothesis tests, you should make sure that the Gauss–Markov conditions hold. The Gauss–Markov conditions have to be explicitly specified. Only writing that the Gauss–Markov conditions hold is not sufficient. As good practice, begin your examination by explicitly providing the Gauss–Markov conditions. You can then refer back to them thereafter. Moreover, ensure when conducting hypothesis testing that you clearly indicate the null and alternative hypotheses (in terms of the true parameters, say β_1), the test statistic (in terms of the parameter estimates, here $\widehat{\beta}_1$), its distribution (with degrees of freedom), the rejection rule (one-sided or-two sided) for a given significance level (typically 5%) with suitable critical values, and provide an interpretation of your result.

Just as last year, many candidates do not answer all parts of the question. Make sure you read the questions properly and provide all details that are requested. Not answering a question will automatically earn you a zero mark for that question.

Key steps to improvement

Essential reading for this course includes the subject guide and the following:

- Dougherty, C. *Introduction to econometrics.* (Oxford: Oxford University Press, 2016) 5th edition [ISBN 9780199676828]; http://oxfordtextbooks.co.uk/orc/dougherty5e/

Apart from the Essential readings you should do some supplementary reading. One very good book at the same level is:

- Gujarati, D.N. and D.C. Porter *Basic econometrics.* (McGraw–Hill, 2009, International edition) 5th edition [ISBN 9780071276252].

To understand the subject clearly it is important to supplement Dougherty's *Introduction to econometrics* (fifth edition) with the subject guide **EC2020 Elements of econometrics** (2016), especially Chapter 10 which covers maximum likelihood estimation. It is very important to carefully go through the subject guide. The subject guide contains solutions to the questions given in the main textbook and also some additional questions and solutions. Working through these will improve your understanding of the subject.

The chapter in the subject guide on maximum likelihood (Chapter 10) includes some additional theory which has not been covered in the main textbook. It is important to read the additional theory given in the subject guide to have a better understanding of the principles of maximum likelihood and tests based on the likelihood function.

Please check the VLE course page for resources for this subject such as a downloadable copy of the subject guide **EC2020 Elements of econometrics** (2016), PowerPoint slideshows that provide a graphical treatment of the topics covered in the textbook, datasets and statistical tables. Candidates should utilise datasets using standard regression programmes (STATA or EViews). This will help in the understanding of the subject.

Examination revision strategy

Many candidates are disappointed to find that their examination performance is poorer than they expected. This may be due to a number of reasons, but one particular failing is '**question spotting**', that is, confining your examination preparation to a few questions and/or topics which have come up in past papers for the course. This can have serious consequences.

We recognise that candidates might not cover all topics in the syllabus in the same depth, but you need to be aware that examiners are free to set questions on **any aspect** of the syllabus. This means that you need to study enough of the syllabus to enable you to answer the required number of examination questions.

The syllabus can be found in the Course information sheet available on the VLE. You should read the syllabus carefully and ensure that you cover sufficient material in preparation for the examination. Examiners will vary the topics and questions from year to year and may well set questions that have not appeared in past papers. Examination papers may legitimately include questions on any topic in the syllabus. So, although past papers can be helpful during your revision, you cannot assume that topics or specific questions that have come up in past examinations will occur again.

If you rely on a question-spotting strategy, it is likely you will find yourself in difficulties when you sit the examination. We strongly advise you not to adopt this strategy.

Examiners' commentaries 2018

EC2020 Elements of econometrics

Important note

This commentary reflects the examination and assessment arrangements for this course in the academic year 2016–17. The format and structure of the examination may change in future years, and any such changes will be publicised on the virtual learning environment (VLE).

Information about the subject guide and the Essential reading references

Unless otherwise stated, all cross-references will be to the latest version of the subject guide (2016). You should always attempt to use the most recent edition of any Essential reading textbook, even if the commentary and/or online reading list and/or subject guide refer to an earlier edition. If different editions of Essential reading are listed, please check the VLE for reading supplements – if none are available, please use the contents list and index of the new edition to find the relevant section.

Comments on specific questions – Zone A

Candidates should answer **EIGHT** of the following **TEN** questions: **ALL** of the questions in Section A (8 marks each) and **THREE** questions from Section B (20 marks each). **Candidates are strongly advised to divide their time accordingly.**

Section A

Answer all questions from this section.

Question 1

We are interested in investigating the factors governing the precision of regression coefficients. Consider the model

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \varepsilon_i$$

with OLS parameter estimates $\widehat{\beta}_1$, $\widehat{\beta}_2$ and $\widehat{\beta}_3$. Under the Gauss–Markov assumptions, we have

$$\text{Var}\left(\widehat{\beta}_2\right) = \frac{\sigma_\varepsilon^2}{\sum_{i=1}^{n}(X_{2i} - \bar{X}_2)^2} \times \frac{1}{1 - r_{X_2X_3}^2},$$

where σ_ε^2 is the variance of ε and $r_{X_2X_3}$ is the sample correlation between X_2 and X_3.

(a) **Provide four factors that help with obtaining more precise parameter estimates for, say, $\widehat{\beta}_2$.**

(4 marks)

(b) **In light of your answer to (a), discuss the concept of near multicollinearity. What consequences does its presence have when considering single and joint significance testing of our slope parameters?**

(4 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Chapter 3.3 (Properties of the multiple linear regression coefficients), 3.4 (Multicollinearity) and Chapter 3.5, Relationship between F statistic and t statistic).

Dougherty, C. Subject guide (2016): Chapter 3.

Approaching the question

(a) Candidates should note that the expression of the variance can be rewritten as:

$$\mathrm{Var}(\widehat{\beta}_2) = \frac{\sigma_\varepsilon^2}{n \times MSD(X_2) \times (1 - r^2_{X_2X_3})}$$

where $MSD(X_2) \equiv n^{-1}\sum(X_{2i} - \bar{X}_2)^2$. The four factors that affect the precision then are: n, $MSD(X_2)$, $r^2_{X_2X_3}$ and σ_ε^2. Therefore, to obtain more precise parameter estimates of β_2 it is desirable to have: (i) small error variance, (ii) large sample size, (iii) large sample variability of the regressor X_2, and (iv) small correlation among the regressors X_2 and X_3.

Some candidates gave a discussion of the Gauss–Markov assumptions which is not the answer.

(b) The issue of near multicollinearity is associated with the setting where $r^2_{X_2X_3}$ is close to 1, yielding very imprecise parameter estimates (large variance). Concept and consequences of multicollinearity are standard bookwork.

Question 2

Consider the linear regression model

$$\boldsymbol{Y_t = \beta_0 + \beta_1 X_t + \beta_2 Y_{t-1} + u_t, \quad t = 1, \ldots, T}$$

where the errors u_t are distributed independently of the regressors X_t and $|\beta_2| < 1$. You suspect that the, mean zero, errors exhibit autocorrelation.

(a) **Explain what we mean by the concept of autocorrelation.**

(2 marks)

(b) **Assume that u_t follows an AR(1) process.**

i. **Discuss, for the given model, the consequences for the ordinary least squares estimator. Support your answers with suitable arguments.**

(3 marks)

ii. **Discuss how you would detect the presence of autocorrelation in the errors in this model. Clearly indicate the null and alternative hypothesis, the test statistic, and rejection rule.**

(3 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Chapter 12.3 (Fitting a model subject to AR(1) autocorrelation, and Chapter 12.2 (Detection of autocorrelation).

Dougherty, C. Subject guide (2016): Chapter 12.

Approaching the question

(a) Candidates should clearly indicate what autocorrelation is (standard bookwork) and indicate that in the presence of the lagged endogenous variable Y_{t-1} this yields inconsistency as $\text{Cov}(Y_{t-1}, u_t) \neq 0$.

(b) i. Specifically, since ε_t is uncorrelated with X_{t-1}, Y_{t-2} and u_{t-1} and since the errors u_t are distributed independently of the regressors.

$$\begin{aligned}\text{Cov}(Y_{t-1}, u_t) &= \text{Cov}(\beta_0 + \beta_1 X_{t-1} + \beta_2 Y_{t-2} + u_{t-1}, \rho u_{t-1} + \varepsilon_t) \\ &= \beta_2 \rho \text{Cov}(Y_{t-2}, u_{t-1}) + \rho \text{Cov}(u_{t-1}, u_{t-1}).\end{aligned}$$

Using the fact that $\text{Cov}(Y_{t-2}, u_{t-1}) = \text{Cov}(Y_{t-1}, u_t)$ by covariance stationarity we get:

$$\text{Cov}(Y_{t-1}, u_t) = (1 - \beta_2 \rho)^{-1} \rho \sigma_u^2 \neq 0.$$

ii. To detect autocorrelation the Breusch–Godfrey test should be proposed, which requires us to run the following auxiliary regression:

$$\widehat{u}_t = \gamma_0 + \gamma_1 X_t + \gamma_2 Y_{t-1} + \rho \widehat{u}_{t-1} + v_t.$$

The use of the Durbin–Watson test is incorrect, a Durbin h-test may be proposed as well. Candidates should clearly indicate H_0 and H_1, the test statistic and the rejection rule.

Question 3

For the population of men who grew up with disadvantaged backgrounds, let *poverty* be a dummy variable equal to one if a man is currently living below the poverty line, and zero otherwise. The variable *age* is age and *educ* is total years of schooling. Let *vocat* be an indicator equal to unity if a man's high school offered vocational training. Using a random sample of 850 men, you obtain

$$\widehat{\mathbf{Pr}(poverty = 1 \mid educ}, age, vocat) = \Lambda(0.453 - 0.016 age - 0.087 educ - 0.049 vocat)$$

where $\Lambda(z) = \exp(z) = (1 + \exp(z))$ is the logit function.

(a) It is argued that using the logit regression model is better than using the linear probability model when explaining the binary variable *poverty.* Discuss the benefits/drawback of using the logit regression model when trying to explain a binary variable.

(5 marks)

(b) For a 40-year old man, with 12 years of education, what is the estimated effect of having vocational training available in high school on the probability of currently living in poverty?

***Hint*: Clarity of computations required is enough, no need to give an exact number.**

(3 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Chapter 10.1 (the linear probability model), Chapter 10.2 (logit analysis), and Chapter 10.6 (introduction to maximum likelihood estimation.

Dougherty, C. Subject guide (2016): Chapter 10.

Approaching the question

(a) The logit model has two main advantages over the linear probability model (LPM): predicted probabilities are restricted to lie in $[0, 1]$ and MLE is (asymptotically) efficient whereas OLS (LPM) will be inefficient given the inherent presence of heteroskedasticity.

The main drawbacks of the logit model relative to the linear probability model are that the coefficients cannot be directly interpreted as the marginal effects of the regressor(s) of interest and it is also computationally more complicated.

(b) Candidates will here need to observe that we need to compare predicted probabilities using the logit specification of the probabilities:

$$\Lambda(z) = \frac{1}{1+\exp(-z)}.$$

For a 40-year old man with 12 years of education with vocational training the estimated probability of living in poverty is given by:

$$\widehat{\Pr(y_i = 1 \mid age_i = 40, educ_i} = 12, vocat_i = 1) = \frac{\exp(z_1)}{1+\exp(z_1)} \approx 0.218$$

where $z_1 = 0.453 - 0.016 \times 40 - 0.087 \times 12 - 0.049 \times 1 \approx -1.28$.

The estimated probability of living in poverty for the same man without the vocational training is given by:

$$\widehat{\Pr(y_i = 1 \mid age_i = 40; educ_i} = 12; vocat_i = 0) = \frac{\exp(z_2)}{1+\exp(z_2)} \approx 0.226$$

where $z_2 = z_1 + 0.049 \times 1 \approx -1.23$.

Therefore, for a 40-year old man with 12 years of education having vocational training in high school decreases the probability of living in poverty by 0.8 percentage points.

Question 4

The following model jointly determines monthly child support payments and monthly visitation rights for divorced couples with children:

$$\boldsymbol{support = \alpha_1 + \alpha_2 visits + \alpha_3 finc + \alpha_4 fremarr + \alpha_5 dist + \varepsilon_1}$$

$$\boldsymbol{visits = \beta_1 + \beta_2 support + \beta_3 mremarr + \beta_4 dist + \varepsilon_2}$$

We assume that children live with their mothers, so that fathers pay child support. Thus, the first equation is the father's 'reaction function': it describes the amount of child support paid for any given level of visitation rights and the other exogenous variables *finc* (father's income), *fremarr* (binary indicator if father remarried), and *dist* (miles currently between the mother and father's residence). Similarly the second equation is the mother's reaction function: it describes visitation rights for a given amount of child support; *mremarr* is a binary indicator for whether the woman is remarried.

(a) **Examine the identification of each structural equation.**

(3 marks)

(b) **Your friend suggests you should implement the IV estimator to estimate the β parameters consistently. He tells you to use *finc* as instrument for *support*. Provide a critical discussion of this suggestion.**

(5 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Chapter 8.3 (Instrumental variables), and Chapter 9 (Simultaneous equations estimation).

Dougherty, C. Subject guide (2016): Chapter 9.

Approaching the question

(a) The first equation is *exactly identified* since there is one exogenous variable ($mremarr_i$) available as instrument to the endogenous regressor ($visits_i$).

The second equation is *overidentified* since there are two exogenous variables ($finc_i$ and $fremarr_i$) available as instruments to the endogenous regressor ($support_i$).

Let G denote the number of equations in the system of simultaneous equations. It is true that in both equations there are $G-1$ endogenous variables (but you may be given a setting where there are fewer endogenous variables), and in each case you have excluded enough exogenous variables (k) you can use as instruments for these bad variables. Answers tend to be very vague.

(b) The IV approach suggested by the friend would yield consistent estimates for β parameters since $finc_i$ is exogenous and correlated with support (assuming $\alpha_3 \neq 0$). However, given this equation is overidentified, we could obtain more efficient estimates by using a two-stage least squares approach in which both $fremarr_i$ and $finc_i$ are included in the vector of instruments.

Question 5

Consider the simple linear regression model

$$\boldsymbol{Y_i = \beta_0 + \beta_1 X_i + u_i}$$

under the classical linear regression model assumptions, where $\boldsymbol{X_i}$ is fixed under repeated sampling. The usual OLS estimators $\boldsymbol{\widehat{\beta}_0}$ and $\boldsymbol{\widehat{\beta}_1}$ are unbiased for their respective population parameters. Let $\boldsymbol{\tilde{\beta}_1}$ be the estimator of $\boldsymbol{\beta_1}$ obtained by assuming the intercept is zero.

(a) Show that the restricted least squares estimator of $\boldsymbol{\beta_1}$ is given by

$$\tilde{\beta}_1 = \frac{\sum_{i=1}^{n} X_i Y_i}{\sum_{i=1}^{n} X_i^2}.$$

(4 marks)

(b) Find $\mathrm{E}(\boldsymbol{\tilde{\beta}_1})$ in terms of the $\boldsymbol{X_i}$, $\boldsymbol{\beta_0}$ and $\boldsymbol{\beta_1}$. Verify that $\boldsymbol{\tilde{\beta}_1}$ is unbiased for $\boldsymbol{\beta_1}$ when the population intercept is zero. Are there other cases where $\boldsymbol{\tilde{\beta}_1}$ is unbiased?

(4 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Chapter 1.3 (Derivation of the regression coefficients) and Chapter 2.3 (The random components and unbiasedness of the OLS regression coefficients).

Dougherty, C. Subject guide (2016): Chapters 1 and 2.

Approaching the question

(a) Formally, restricted least squares estimates of β_0 and β_1 solve the following problem:

$$(\tilde{\beta}_0, \tilde{\beta}_1) = \min_{b_0, b_1} \sum_{i=1}^{n} (Y_i - b_0 - b_1 X_i)^2, \quad \text{subject to } b_0 = 0.$$

This is the same as performing OLS on the model where $\beta_0 = 0$, that is performing OLS while leaving out the intercept.

Therefore, $\tilde{\beta}_0 = 0$ and $\tilde{\beta}_1 : \min_{b_1} \sum_{i=1}^{n} (Y_i - b_1 X_i)^2$. The first-order condition is given by:

$$-2 \sum_{i=1}^{n} (Y_i - \tilde{\beta}_1 X_i) X_i = 0 \qquad \Leftrightarrow \qquad \tilde{\beta}_1 = \frac{\sum_{i=1}^{n} X_i Y_i}{\sum_{i=1}^{n} X_i^2}.$$

(b) While candidates found the discussion of restricted least squares difficult, there was no reason not to answer the second part which was standard. Plug in the true model and use properties of sums to obtain:

$$\begin{aligned} \tilde{\beta}_1 &= \left(\sum_{i=1}^{n} X_i^2 \right)^{-1} \left(\sum_{i=1}^{n} X_i (\beta_0 + \beta_1 X_i + u_i) \right) \\ &= \left(\sum_{i=1}^{n} X_i^2 \right)^{-1} \left(\sum_{i=1}^{n} X_i \right) \beta_0 + \beta_1 + \left(\sum_{i=1}^{n} X_i^2 \right)^{-1} \left(\sum_{i=1}^{n} X_i u_i \right). \end{aligned}$$

Taking expectations using the fact that X_is are fixed and $\mathrm{E}(u_i) = 0$, we have:

$$\begin{aligned} \mathrm{E}(\tilde{\beta}_1) &= \left(\sum_{i=1}^{n} X_i^2 \right)^{-1} \left(\sum_{i=1}^{n} X_i \right) \beta_0 + \beta_1 + \left(\sum_{i=1}^{n} X_i^2 \right)^{-1} \left(\sum_{i=1}^{n} X_i \mathrm{E}(u_i) \right) \\ &= \left(\sum_{i=1}^{n} X_i^2 \right)^{-1} \left(\sum_{i=1}^{n} X_i \right) \beta_0 + \beta_1. \end{aligned}$$

Therefore, $\tilde{\beta}_1$ is unbiased if either: (i) $\beta_0 = 0$ or (ii) $\sum_{i=1}^{n} X_i = 0$.

Section B

Answer three questions from this section.

Question 6

Let us consider the estimation of a hedonic price function for houses. The hedonic price refers to the implicit price of a house given certain attributes (e.g., the number of bedrooms). The data contains the sale price of 546 houses sold in the summer of 1987 in Canada along with their important features. The following characteristics are available: the lot size of the property in square feet (*lotsize*), the numbers of bedrooms (*bedrooms*), the number of full bathrooms (*bathrooms*), and a dummy indicating the presence of airconditioning (*airco*).

Consider the following ordinary least squares results

$$\begin{aligned} \widehat{\log(price)}_i = \underset{\substack{(.232) \\ [.233]}}{7.094} + \underset{\substack{(.028) \\ [.028]}}{0.400} \log(lotsize)_i + \underset{\substack{(.015) \\ [.017]}}{0.078} bedrooms_i + \\ \underset{\substack{(.023) \\ [.024]}}{0.216} bathrooms_i + \underset{\substack{(.024) \\ [.023]}}{0.212} airco_i \quad n = 546, \; RSS = 32.622 \end{aligned} \tag{6.1}$$

The usual standard errors are in parentheses, the heteroskedasticity robust standard errors are in square brackets, and RSS measures the residual sum of squares.

(a) **Interpret the parameter estimates on log($lotsize$), $bedrooms$, and $airco$. Briefly discuss the statistical significance of the results.**

(5 marks)

(b) **Suppose that lot size was measured in square metres rather than square feet. How would this affect the parameter estimates of the slopes and intercept? How would this affect the fitted values? *Note*: the conversion (approximate) $1m^2 = 10ft^2$.**

(5 marks)

(c) **We are interested in testing the hypothesis H$_0$: $\beta_{bedrooms} = \beta_{bathrooms}$ against the alternative $H_1 : \beta_{bedrooms} \neq \beta_{bathrooms}$. Discuss a test for this hypothesis that makes use of the following restricted regression result**

$$\widehat{\log(price)}_i = \underset{(.234)}{6.994} + \underset{(.282)}{0.408}\log(lotsize)_i + \underset{(.011)}{0.127}bbrooms_i + \underset{(.024)}{0.215}airco_i \qquad (6.2)$$

$$n = 546, \;\; RSS = 33.758$$

where $bbrooms = bedrooms + bathrooms$. Clearly indicate the assumptions you are making for this test to be valid.

(5 marks)

(d) **You are interested in testing for the presence of heteroskedasticity. Say you are told that the variance is increasing with log($lotsize$). Discuss how you would test for the presence of heteroskedasticity. What is the name of the test you are proposing?**

(5 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Chapter 1.4 (Interpretation of a regression equation – units of measurement), Chapter 4.2 (Logarithmic transformations), Chapter 2.6 (Testing hypotheses relating to the regression coefficients), Chapter 7.1 (Heteroskedasticity and its implications), and Chapter 7.2 (Detection of heteroskedasticity).

Dougherty, C. Subject guide (2016): Chapter 7.

Approaching the question

(a) Clear discussion of interpretation required (units not always clear): On average, holding the remaining variables in the regression constant, (i) a 1% increase in lot size is associated with a 0.4% increase in house price, (ii) each extra bedroom is associated with a 7.8% increase in house price, and (iii) houses with air conditioning are 21.2% more expensive than those without. All estimates are statistically significant at 5% significance levels. Candidates should clearly indicate H$_0$ and H$_1$, the test statistic and the rejection rule.

(b) Let $lotsize_i$ be the lot size in square feet and $\widetilde{lotsize}_i$ be the lot size in square metres. We have that $\widetilde{lotsize}_i = (10)^{-1}lotsize_i$ and $\log(\widetilde{lotsize}_i) = \log((10)^{-1}) + \log(lotsize_i)$. Since this is an additive transformation of one of the explanatory variables we have that (i) the regression slopes will not be affected, (ii) the intercept will change to $7.094 - \log((10)^{-1}) \times 0.4$, and (iii) the fitted values will also not be affected. Many candidates made an error here, ignoring the fact that the variable whose measurement was changed entered in log form.

(c) Candidates would have to recognise that (6.2) is a restricted version of (6.1) where $\beta_{bedrooms} = \beta_{bathrooms}$ is imposed. We therefore need to use the F test. Test statistic:

$$F = \frac{RRSS - URSS}{URSS} \times \frac{n-K}{J} = \frac{33.758 - 32.622}{32.622} \times \frac{541}{1} \approx 18.84.$$

Assuming the Gauss–Markov assumptions plus normality of the error term hold: under H_0, $F \sim F_{1,\,541}$. At the 5% signicance level we reject H_0 since $F > 3.86$. Conclusion: The effect of one extra bathroom is different from the effect of one extra bedroom. Some candidates did not recognise this and were proposing a test on the coefficient of *bbrooms* equalling zero which is wrong.

(d) Assuming Gauss–Markov assumptions plus the normality of the error hold, he can use the Goldfeld–Quandt test for heteroskedasticity. For that purpose, we should first order the 546 observations by the magnitude of $\log(lotsize_i)$. Fit one regression for the first n^* observations and another for the last n^* observations (usually n^* equals one-third of the sample). Let RSS_1 and RSS_2 denote the sum of squared residuals in each of these regressions, respectively.

- $H_0 : \sigma_2^2 = \sigma_1^2$ vs. $H_1 : \sigma_2^2 > \sigma_1^2$.
- Test statistic: $GQ = RSS_2/RSS_1$.
- Under H_0, $GQ \sim F_{n^*-k,\, n^*-k}$.
- Reject if GQ is greater than the 95th percentile of the F distribution above.

Candidates need to be careful not to state $H_0 : RSS_2 = RSS_1$ vs. $H_1 : RSS_2 > RSS_1$. Both RSS_1 and RSS_2 are random variables. Because the sample sizes are identical it is also correct to state H_0 : RSS_1 and RSS_2 are not statistically different. Note that simply by having one sample larger than the other, you could also have a larger residual sum of squares.

Question 7

The following question concerns the effects of background characteristics and admission assessment scores on the performance of students in the final university examinations in a UK university. The following equation was estimated by Ordinary Least Squares:

$$\begin{aligned}\widehat{finalavg} &= \underset{(2.78)}{53.89} + \underset{(.04)}{0.03}tst\_reas + \underset{(.02)}{0.05}tst\_quan + \underset{(.02)}{0.06}interview \\ &\quad -\underset{(.78)}{0.04}indep + \underset{(.66)}{0.67}male + \underset{(.97)}{0.06}indep{*}male \\ n &= 325,\ R^2 = .06, \qquad (7.1)\end{aligned}$$

where *finalavg* is the average finals score (the outcome), *tst_reas* and *tst_quan* are the pre-admission reasoning and quantitative test scores respectively, *interview* is the pre-admission interview score, *indep* indicates whether the student attended an independent school (1 = yes, 0 = no), and *male* indicates whether the student is male (1 = yes, 0 = no). The usual standard errors are in parentheses.

(a) We want to test whether gender has a significant impact on students' finals performance. Clearly indicating the null and the alternative hypothesis, provide the test statistic and the rejection rule. Discuss what information you would need to enable you to implement this test. You are expected to provide the assumptions which underlie your test.

(5 marks)

(b) If we do not include the interaction term *indep*male* in our regression model, what are we implicitly assuming about the effect of gender and school background on finals performance?

(5 marks)

(c) Suppose students who did not attend an independent school, attended a state school. Using the results in (7.1), provide the parameter estimates you would obtain if you had applied Ordinary Least Squares to the equation

$$\begin{aligned}finalavg &= \beta_0 + \beta_1 tst\_reas + \beta_2 tst\_quan + \beta_3 interview \qquad (7.2)\\ &\quad \beta_4 state + \beta_5 male + \beta_6 state{*}male + \varepsilon,\end{aligned}$$

where *state* indicates whether the student attended a state school (1 = yes, 0 = no).

(5 marks)

(d) **Discuss any problem you may have in estimating the model if all males in your sample have attended an independent school prior to attending university. What name does this problem have and what can you do to mitigate this problem?**

(5 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Chapter 5.1–5.3 (Dummy variables).

Dougherty, C. Subject guide (2016): Chapter 5.

Approaching the question

(a) For given values of the remaining explanatory variables, the gender performance gap measured as the difference between male and female average final scores is given by β_5 if $indep_i = 0$ or $\beta_5 + \beta_6$ if $indep_i = 1$. Therefore, we want to perform the following test:

- $\text{H}_0 : \beta_5 = 0$ and $\beta_6 = 0$ vs. $\text{H}_1 : \beta_5 \neq 0$ or $\beta_6 \neq 0$.
- Need the R^2 of a regression of *finalavg* on a constant, *tst_reas*, *tst_quant*, *interview* and *indep* (the restricted model).
- Compute the test statistic:

$$F = \frac{R^2_{ur} - R^2_r}{1 - R^2_{ur}} \times \frac{n - K}{J} = \frac{0.06 - R^2_r}{1 - 0.06} \times \frac{325 - 7}{2}.$$

- Under Gauss–Markov assumptions plus normality, $F \sim F_{2,\,318}$.
- At the 5% signicance level, reject H_0 if $F > 3.07$.

(b) If we do not include the interaction term we are implicitly assuming that the effect of gender on performance is the same in independent and non-independent schools. We are also assuming that the effect of attending an independent school on final performance is the same for both male and female students.

(c) Let $\widehat{\theta}$s indicate the estimated parameters from the initial regression. From the original regression we have:

$$\begin{aligned}
\widehat{y}_i &= \widehat{\theta}_0 + \cdots + \widehat{\theta}_4 indep_i + \widehat{\theta}_5 male_i + \widehat{\theta}_6 indep_i \times male_i \\
&= \widehat{\theta}_0 + \cdots + \widehat{\theta}_4 (1 - state_i) + \widehat{\theta}_5 male_i + \widehat{\theta}_6 (1 - state_i) \times male_i \\
&= (\widehat{\theta}_0 + \widehat{\theta}_4) + \cdots - \widehat{\theta}_4 state_i + (\widehat{\theta}_5 + \widehat{\theta}_6) male_i - \widehat{\theta}_6 state_i \times male_i
\end{aligned}$$

where $\cdots = \widehat{\theta}_1 tst\_reas_i + \widehat{\theta}_2 tst\_quant_i + \widehat{\theta}_3 interview_i$. Mapping original estimates into estimated parameters in (7.2) yields:

- $\widehat{\beta}_0 = \widehat{\theta}_0 + \widehat{\theta}_4 = 53.89 - 0.04 = 53.85$
- $\widehat{\beta}_4 = -\widehat{\theta}_4 = -(-0.04) = 0.04$
- $\widehat{\beta}_6 = -\widehat{\theta}_6 = -0.06$
- $\widehat{\beta}_5 = \widehat{\theta}_5 + \widehat{\theta}_6 = 0.67 + 0.06 = 0.73$
- $\widehat{\beta}_2$, $\widehat{\beta}_3$ and $\widehat{\beta}_4$ remain the same from (7.1).

(d) If all males attended an independent school, then $male_i = male_i \times indep_i \; \forall \; i$, so the variables will be perfectly collinear and he would not be able to obtain the OLS estimates from (7.1). In this case, we could either drop the interaction term from the regression or obtain more data such that some men that have not attended independent school will also be in the sample.

Question 8

An OLS regression of y_t on x_t and x_{t-1} gives the following results (with the standard errors given in parentheses)

$$\widehat{y}_t = \underset{(2.30)}{8.88} + \underset{(3.01)}{5.07}\,x_t - \underset{(3.01)}{3.18}\,x_{t-1}; \quad R^2 = .095,\ T = 209 \tag{8.1}$$

(a) **What are the estimates of the short-run and long-run effect of x_t on y_t? Interpret these estimates.**

(4 marks)

(b) **Test the hypothesis that a one unit increase in x results in a ten unit increase in y in the same year. Under what assumptions is this test valid?**

(4 marks)

Let e_t be the OLS residuals from the above regression. An OLS regression of e_t on e_{t-1} yields

$$e_t = \underset{(2.12)}{0.55} + \underset{(0.18)}{0.44}\,e_{t-1} + \underset{(2.18)}{2.16}\,x_t - \underset{(.99)}{1.09}x_{t-1}; \quad R^2 = .175,\ T = 208 \tag{8.2}$$

(c) **Using this result, test for evidence of autocorrelation, clearly indicating the null and alternative hypotheses, the test statistic, rejection rule and assumptions underlying the test. What name do we give this test?**

(5 marks)

(d) **You are interested in testing whether the long-run effect of x_t on y_t is statistically significant.**

i. **Discuss how to reparameterise (8.1) to ensure that your regression output will provide you with a standard error for the long-run effect.**

(4 marks)

ii. **Discuss the problem of implementing your test using the standard error obtained in (d)i. when you do find evidence of autocorrelation in (8.1). Briefly indicate how you proceed with your test.**

(3 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Chapter 6.5 (Testing a linear restriction), Chapter 11.3 (Models with lagged explanatory variables), and Chapter 12.1–12.3 (Definition, consequences and detection of autocorrelation; Fitting a model subject to AR(1) autocorrelation).

Dougherty, C. Subject guide (2016): Chapters 6 and 12.

Approaching the question

(a) Short-run effect: It shows the immediate effect a change in x with 1 unit has on y. Increases with 5.07 units.

Long-run effect: It shows the effect a permanent change in x with one unit has on y after 1 period (the last lagged response) has passed. Increases with $5.07 - 3.18 = 1.89$ units.

(b) Asked to perform the t test for $\mathrm{H}_0 : \beta_1 = 10$ vs. $\mathrm{H}_1 : \beta_1 \neq 10$ (standard bookwork). Validity of Gauss–Markov + normality.

(c) The test equation tells us that we have to perform the Breusch–Godfrey test for first-order autocorrelation. Let ρ denote the coefficient associated with e_{t-1}.

- $\mathrm{H}_0 : \rho = 0$ (No autocorrelation) vs. $\mathrm{H}_1 : \rho \neq 0$ (Autocorrelation).
- Test statistic: $LM = nR^2 = 28 \times 0.175 \approx 4.9$.

- Under H_0, $LM \overset{a}{\sim} \chi^2_1$.
- Reject H_0 in favour of autocorrelation since $LM > 3.84$.

(d) i. We need to reparameterise (8.1) to obtain the standard error of $\widehat{\beta}_1 + \widehat{\beta}_2$ directly:

$$\begin{aligned} y_t &= \beta_0 + \beta_1 x_t + \beta_2 x_{t-1} + u_t \\ y_t &= \beta_0 + \beta_1 x_t + \beta_2 x_t - \beta_2 x_t + \beta_2 x_{t-1} + u_t \\ y_t &= \beta_0 + (\beta_1 + \beta_2) x_t - \beta_2 \Delta x_t + u_t. \end{aligned}$$

Regress y_t on a constant, x_t and Δx_t and the standard error of the estimated coefficient on x_t is the standard error of the long-run effect.

ii. The usual standard errors will be invalidated so we need to use HAC standard errors instead. Given the robust standard errors, one could compute the test statistic robust to autocorrelation as $t^{robust} = \widehat{\theta}/se(\widehat{\theta})^{HAC}$ and conclude the long-run effect is statistically significant at the 5% significance level if $|t^{robust}| > 1.96$.

Question 9

Consider the model

$$\begin{aligned} y_t &= \alpha + \beta t + \varepsilon_t, \quad t = 1, \ldots, T \qquad (9.1) \\ \varepsilon_t &= \rho \varepsilon_{t-1} + v_t, \quad \text{and} \end{aligned}$$

v_t is an i.i.d. (0, σ^2) innovation which is independent of the past. Let $|\rho| \leq 1$.

(a) What name do we give the ε_t process given above? Provide the condition(s) that ensures that ε_t is stationary. In your answer discuss what we mean by the concept of stationarity (more precisely 'covariance stationarity').

(4 marks)

(b) It will be important to distinguish between the above process for y_t being 'trend stationary' as opposed to 'difference stationary'.

i. Explain these concepts clearly. Why is it important to distinguish between these two types of non-stationarity?

(4 marks)

ii. Show that under the condition you provided in (a) that y_t is trend stationary.

(2 marks)

iii. Show that if ε_t is difference stationary then y_t is difference stationary.

(2 marks)

(c) Show that you can rewrite the above model in the following form

$$\Delta y_t = \gamma_1 + \gamma_2 t + \gamma_3 y_{t-1} + v_t. \qquad (9.2)$$

Clearly indicate the relation between $(\gamma_1, \gamma_2, \gamma_3)$ and (α, β, ρ).

(4 marks)

(d) What problem do you see here with using (9.2) to conducting the Dickey–Fuller Test to distinguish between trend and difference stationarity when v_t exhibits autocorrelation? What solution do you suggest we adopt?

(4 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Chapter 12.1 (Definition and consequences of autocorrelation), Chapter 13.1 (Stationarity and nonstationarity), and Chapter 13.4–13.5 (Tests of nonstationarity).

Dougherty, C. Subject guide (2016): Chapter 13.

Approaching the question

(a) Standard bookwork.

(b) i. A process $\{y_t\}_{t=-\infty}^{\infty}$ is said to be difference stationary if its first-difference is stationary ($\Delta y_t \sim I(0)$). It is said to be trend stationary if the process $y_t - \beta t$ is stationary. It is important to distinguish between the two because the source of non-stationarity has different implications on how we proceed to obtain a stationary time-series to use for regression analysis and for statistical inference.

ii. If $|\rho| < 1$ then ε_t is stationary, since $y_t - \beta t = \alpha + \varepsilon_t$ is also stationary. Therefore, y_t is trend stationary.

iii. Suppose that ε_t is difference stationary ($\Delta\varepsilon_t \sim I(0)$). Taking first differences of (9.1) we obtain $\Delta y_t = \beta + \Delta\varepsilon_t$, since β is just a constant we have that $\Delta y_t \sim I(0)$, that is, y_t is difference stationary.

(c) Lagging (9.1) by one period and multiplying both sides by ρ we obtain:

$$\rho y_{t-1} = \rho\alpha + \rho\beta(t-1) + \rho\varepsilon_{t-1}.$$

Subtracting the above from (9.1) yields:

$$y_t - \rho y_{t-1} = (1-\rho)\alpha + \beta t - \rho\beta t + \rho\beta + \varepsilon_t - \rho\varepsilon_{t-1}.$$

Let $v_t \equiv \varepsilon_t - \rho\varepsilon_{t-1}$ and rearrange:

$$y_t = ((1-\rho)\alpha + \rho\beta) + \beta(1-\rho)t + \rho y_{t-1} + v_t.$$

Finally, subtract y_{t-1} on both sides to obtain:

$$\Delta y_t = \underbrace{(1-\rho)\alpha + \rho\beta}_{\gamma_1} + \underbrace{\beta(1-\rho)t}_{\gamma_2} + \underbrace{\rho}_{\gamma_3} y_{t-1} + v_t.$$

(d) If v_t exhibits autocorrelation it must be eliminated before running the test regression otherwise the Dickey–Fuller test will not be valid. To eliminate it we should include the lags of Δy_t in the test equation (9.2). This test is then known as the augmented Dickey–Fuller test.

Question 10

Let $math10$ denote the percentage of students at a high school receiving a passing score on a standardised math test. We are interested in estimating the effect of per student spending on math performance. A simple model is

$$math10_i = \beta_0 + \beta_1 \log(expend_i) + \beta_2 \log(enroll_i) + \beta_3 poverty_i + u_i \qquad (10.1)$$

where, for each high school i; $poverty_i$ is the percentage of students living in poverty, $expend_i$ is the spending per student and $enroll_i$ the number of registered students. You may assume that this model satisfies all Gauss–Markov assumptions.

You are faced with the fact that data is unavailable on a key variable: $poverty$.

(a) Discuss the properties (unbiasedness and consistency) of the estimators when you drop the variable poverty. Explain your answers.

(5 marks)

You do have information available on a closely related variable: the percentage of students eligible for the federally funded school lunch program, $lnchprg_i$. Let us consider using $lnchprg_i$ as a proxy for $poverty_i$.

(b) **Briefly discuss why $lnchprg_i$ is a sensible proxy variable for the unobserved variable $poverty_i$.**

(2 marks)

(c) **It is unlikely that $lnchprg_i$ is an ideal proxy, in the sense that there is an exact linear relationship between them, instead, we will assume that**

$$poverty_i = \alpha_0 + \alpha_1 lnchprg_i + v_i, \quad \alpha_1 \neq 0 \qquad (10.2)$$

Discuss the assumptions you need to make to enable consistent parameter estimators of β_1 and β_2 using your estimable equation

$$math10_i = \gamma_0 + \gamma_1 \log(expend_i) + \gamma_2 \log(enroll_i) + \gamma_3 lnchprg_i + e_i,$$

***Hint*: Consider the relation between the γ and the β parameters and express e_i in terms of u_i and v_i.**

(5 marks)

(d) **The OLS results with and without $lnchprg_i$ as an explanatory variable are given by (standard errors in parentheses):**

$$\widehat{math10}_i = -\underset{(26.72)}{69.24} + \underset{(3.30)}{11.13} \log expend_i + \underset{(0.615)}{0.022} \log emroll_i$$
$$N = 428, \; R^2 = 0.0297$$

$$\widehat{math10}_i = -\underset{(24.99)}{23.14} + \underset{(3.04)}{7.75} \log expend_i - \underset{(0.58)}{1.26} \log emroll_i - \underset{(0.036)}{0.324} lnchprg_i$$
$$N = 428, \; R^2 = 0.1893$$

i. **Interpret the coefficient on $lnchprg$. What does this parameter tell us regarding the parameter of interest β_3?**

(4 marks)

ii. **Give an intuitive discussion explaining why the effect of expenditures on $math10_i$ is lower in the regression where $lnchprg_i$ is included than where it is excluded.**

(4 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Chapter 6.2 (The effect of omitting a variable that ought to be included), Chapter 6.4 (Proxy variables).

Dougherty, C. Subject guide (2016): Chapter 6.

Approaching the question

(a) Consider rewriting (10.1) as:

$$math10_i = \beta_0 + \beta_1 \log(expend_i) + \beta_2 \log(enroll_i) + \varepsilon_i$$

where $\varepsilon_i = \beta_3 poverty_i + u_i$. Assuming $\beta_3 \neq 0$, if poverty is correlated with either (log) expenditures and/or (log) enrollment the model will suffer from endogeneity due to omitted variables and OLS will be biased and inconsistent. That is likely to be the case since schools that have smaller expenditures tend to be located in poorer neighbourhoods and hence to have more students living in poverty conditions.

(b) It is a sensible proxy because it is likely to be correlated with poverty and to capture some of the effect of poverty since usually students eligible for the lunch program tend to be those with low levels of family income.

(c) Plug in (10.2) for poverty in (10.1) to obtain:

$$math10_i = \underbrace{(\beta_0 + \beta_3\alpha_0)}_{\gamma_0} + \underbrace{\beta_1}_{\gamma_1} \log(expend_i) + \underbrace{\beta_2}_{\gamma_2} \log(enroll_i) + \underbrace{\beta_3\alpha_1}_{\gamma_3} lnchprg_i + \underbrace{(\beta_3 v_i + u_i)}_{\varepsilon_i}.$$

Given that the model (10.1) satisfies all the Gauss–Markov assumptions, to obtain consistent estimates of β_1 and β_2 we need that: (i) v_i is uncorrelated with $\log(expend_i)$, $\log(enroll_i)$ and $lnchprg_i$, and (ii) $lnchprg_i$ is uncorrelated with u_i.

(d) i. On average, holding expenditure and enrollment constant, a 1 percentage point increase in the number of students eligible for the lunch program is associated with a 0.324 percentage point fall in the percentage of students receiving a passing score in the standardised math test. Since $\gamma_3 = \alpha_1\beta_3$ and assuming $\alpha_1 > 0$, the direction of the effect (sign) of poverty of $math10$ is the same as the effect of $lnchprg$ on $math10$, in this case, with a negative coefficient on $lnchprg$ we can infer that poverty has a negative effect on $math10$.

ii. Omitting relevant variables will result in the remaining parameters attempting to pick up its effect through the correlation these omitted variables have with the included regressors. We, therefore, expect the effect to be smaller, as part of the effect we attribute to expenditure in the short regression is actually coming from the fact that high schools that have larger expenditures tend to have fewer students eligible for the lunch program and those students tend to perform worse in the standardised math test.

Examiners' commentaries 2018

EC2020 Elements of econometrics

Important note

This commentary reflects the examination and assessment arrangements for this course in the academic year 2017–18. The format and structure of the examination may change in future years, and any such changes will be publicised on the virtual learning environment (VLE).

Information about the subject guide and the Essential reading references

Unless otherwise stated, all cross-references will be to the latest version of the subject guide (2016). You should always attempt to use the most recent edition of any Essential reading textbook, even if the commentary and/or online reading list and/or subject guide refer to an earlier edition. If different editions of Essential reading are listed, please check the VLE for reading supplements – if none are available, please use the contents list and index of the new edition to find the relevant section.

Comments on specific questions – Zone B

Candidates should answer **EIGHT** of the following **TEN** questions: **ALL** of the questions in Section A (8 marks each) and **THREE** questions from Section B (20 marks each). **Candidates are strongly advised to divide their time accordingly.**

Section A

Answer all questions from this section.

Question 1

Consider the consumption function

$$C_t = \alpha + \lambda Y_t + \varepsilon_t \tag{1.1}$$

where C_t is aggregate consumption at t, λ is marginal propensity to consume ($0 < \lambda < 1$) and Y_t is aggregate income at t defined as

$$Y_t = C_t + A_t,$$

where A_t is the sum of investment and government consumption at t. Assume that A_t is uncorrelated with ε_t and that the shock ε_t is mean zero i.i.d. across t. A random sample of size n containing Y_t, C_t and A_t is available.

(a) **Provide the reduced form equation for Y_t.**

(2 marks)

(b) **Show that the OLS estimator of λ in (1.1) is inconsistent. You are asked to indicate the direction of this inconsistency.**

***Note*: you are not expected to derive the OLS estimator.**

(6 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Chapter 9.1 (Structural and reduced form equations) and Chapter 9.2 (Simultaneous equation bias).

Dougherty, C. Subject guide (2016): Chapter 9.

Approaching the question

(a) Plugging in the consumption function into the aggregate income identity:

$$Y_t = (\alpha + \lambda Y_t + \varepsilon_t) + A_t.$$

Rearranging yields, $Y_t = (1-\lambda)^{-1}\alpha + (1-\lambda)^{-1}A_t + (1-\lambda)^{-1}\varepsilon_t$.

(b) The OLS estimator of λ is given by:

$$\widehat{\lambda} = \frac{\sum_{t=1}^{T}(C_t - \bar{C})(Y_t - \bar{Y})}{\sum_{t=1}^{T}(Y_t - \bar{Y})^2} = \frac{T^{-1}\sum_{t=1}^{T}(C_t - \bar{C})(Y_t - \bar{Y})}{T^{-1}\sum_{t=1}^{T}(Y_t - \bar{Y})^2}.$$

Taking probability limits and using a suitable LLN:

$$\text{plim}(\widehat{\lambda}) = \frac{\text{Cov}(C_t, Y_t)}{\text{Var}(Y_t)} = \frac{\text{Cov}(\alpha + \lambda Y_t + \varepsilon_t, Y_t)}{\text{Var}(Y_t)} = \lambda + \frac{\text{Cov}(\varepsilon_t, Y_t)}{\text{Var}(Y_t)}.$$

Using the reduced form from part (a) we can show OLS is inconsistent since:

$$\text{Cov}(\varepsilon_t, Y_t) = (1-\lambda)^{-1}\text{Cov}(\varepsilon_t, \alpha + A_t + \varepsilon_t) = (1-\lambda)^{-1}\text{Var}(\varepsilon_t) \neq 0$$

where the second equality uses that A_t is uncorrelated with ε_t. Finally, since $\lambda \in (0,1)$ and both $\text{Var}(Y_t)$ and $\text{Var}(\varepsilon_t)$ are positive, there is a positive inconsistency in OLS estimator, in other words, propensity to consume is overestimated in (1.1).

Question 2

Consider the simple linear regression model

$$\boldsymbol{Y_t = \beta X_t + u_t, \qquad t = 1, \ldots, T}$$

where the errors u_t are distributed independently of the regressors X_t. You suspect that the, mean zero, errors exhibit autocorrelation.

(a) **Explain what we mean by the concept of autocorrelation.**

(2 marks)

(b) **Assume you are told that u_t follows an MA(1) process.**

i. **Discuss whether the OLS estimator $\widehat{\beta}$ is a consistent estimator for β. Justify your answers with suitable technical derivations.**

***Note*: you are not expected to derive the OLS estimator.**

(3 marks)

ii. **Suppose you want to test $\mathbf{H_0}: \beta = 1$ against $\mathbf{H_1}: \beta < 1$. Discuss how you would conduct this test based on the OLS estimator, recognising the presence of autocorrelation in the error.**

(3 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Chapter 6.5 (Testing a linear restriction), Chapter 12.1–12.3 (Definition, consequences and detection of autocorrelation; Fitting a model subject to AR(1) autocorrelation).

Dougherty, C. Subject guide (2016): Chapters 6 and 12.

Approaching the question

(a) Candidates should indicate clearly what autocorrelation is (standard bookwork).

(b) i. The OLS estimator is given by:

$$\widehat{\beta} = \frac{\sum_{t=1}^{T} X_t Y_t}{\sum_{t=1}^{T} X_t^2} = \frac{T^{-1}\sum_{t=1}^{T} X_t Y_t}{T^{-1}\sum_{t=1}^{T} X_t^2}.$$

To analyse consistency we take probability limits and apply a LLN to obtain:

$$\text{plim}(\widehat{\beta}) = \frac{\text{E}(X_t Y_t)}{\text{E}(X_t^2)} = \frac{\text{E}(X_t(\beta X_t + u_t))}{\text{E}(X_t^2)} = \beta + \frac{\text{E}(X_t u_t)}{\text{E}(X_t^2)}.$$

The OLS estimator is consistent as long as $\text{E}(X_t u_t) = 0$ (and $\text{E}(X_t^2) \neq 0$). This is satisfied since the zero mean errors u_t are assumed to be independent (and hence uncorrelated) of the regressors X_t.

ii. Recognising the presence of autocorrelation in the error when testing the single linear restriction we need to make use of HAC standard errors. This fact was ignored by many. The test statistic we use is $t = (\widehat{\beta} - 1)/se(\widehat{\beta})^{HAC}$ and we should reject H_0 when t is too small. Under H_0, $t \overset{a}{\sim} N(0, 1)$ (also accepted would be $t \sim t_{n-k}$). At the 5% significance level we reject H_0 if $t < -1.645$.

Question 3

Consider the simple linear regression model

$$\boldsymbol{Y_i = \beta_0 + \beta_1 X_i + u_i}$$

under the classical linear regression model assumptions, where X_i is fixed under repeated sampling. The usual OLS estimators $\widehat{\beta}_0$ and $\widehat{\beta}_1$ are unbiased for their respective population parameters. Let $\tilde{\beta}_0$ be the estimator of β_0 when β_1 equals 1.

(a) Show that the restricted least squares estimator of β_0 is given by

$$\boldsymbol{\tilde{\beta}_0 = \bar{Y} - \bar{X}}$$

where $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$ and $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

(4 marks)

(b) Find $\text{E}(\tilde{\beta}_0)$ in terms of the X_i, β_0 and β_1. Verify that $\tilde{\beta}_0$ is unbiased for β_0 when $\beta_1 = 1$. Are there other cases where $\tilde{\beta}_0$ is unbiased?

(4 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Chapter 1.3 (Derivation of the regression coefficients) and Chapter 2.3 (The random components and unbiasedness of the OLS regression coefficients).

Dougherty, C. Subject guide (2016): Chapters 1 and 2.

Approaching the question

(a) Formally, restricted least squares estimates of β_0 and β_1 solve the following problem:

$$(\tilde{\beta}_0, \tilde{\beta}_1) = \min_{b_0, b_1} \sum_{i=1}^{n} (Y_i - b_0 - b_1 X_i)^2, \quad \text{subject to } b_1 = 1.$$

This is the same as performing OLS on the model where $\beta_1 = 1$.

Therefore, $\tilde{\beta}_1 = 1$ and $\tilde{\beta}_0 = \min_{b_0} \sum_{i=1}^{n} (Y_i - b_1 X_i)^2$. The first-order condition is given by:

$$-2 \sum_{i=1}^{n} (Y_i - \tilde{\beta}_0 - X_i) = 0 \qquad \Leftrightarrow \qquad \tilde{\beta}_0 = \bar{Y} - \bar{X}.$$

(b) While candidates found the discussion of restricted least squares difficult, there was no reason not to answer the second part which was standard. Using our estimator, notice that averaging the true model yields $\bar{Y} = \beta_0 + \beta_1 \bar{X} + \bar{u}$. Plugging in our estimator $\tilde{\beta}_0$, we have:

$$\tilde{\beta}_0 = (\beta_0 + \beta_1 \bar{X} + \bar{u}) - \bar{X} = \beta_0 + (\beta_1 - 1)\bar{X} + \bar{u}.$$

Taking expectations using the fact that X_is are fixed and $\mathrm{E}(u_i) = 0$, we have:

$$\begin{aligned} E(\tilde{\beta}_0) &= \beta_0 + (\beta_1 - 1)\bar{X} + n^{-1} \sum_{i=1}^{n} \mathrm{E}(u_i) \\ &= \beta_0 + (\beta_1 - 1)\bar{X}. \end{aligned}$$

Therefore, $\tilde{\beta}_0$ is unbiased if either: (i) $\beta_1 = 1$, or (ii) $\bar{X} = 0$.

Question 4

We are interested in investigating the factors governing the precision of regression coefficients. Consider the model

$$\boldsymbol{Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \varepsilon_i}$$

with OLS parameter estimates $\widehat{\beta}_1$, $\widehat{\beta}_2$ and $\widehat{\beta}_3$. Under the Gauss–Markov assumptions, we have

$$\mathbf{Var}\left(\widehat{\beta}_2\right) = \frac{\sigma_\varepsilon^2}{\sum_{i=1}^{n} (X_{2i} - \bar{X}_2)^2} \times \frac{1}{1 - r_{X_2 X_3}^2},$$

where σ_ε^2 is the variance of ε and $r_{X_2 X_3}$ is the sample correlation between X_2 and X_3.

(a) **Provide four factors that help with obtaining more precise parameter estimates for, say, $\widehat{\beta}_2$.**

(4 marks)

(b) Assume that the true value of $\beta_3 = 0$, so that the above model includes an irrelevant variable. Discuss the effect of including this irrelevant variable on the unbiasedness and precision of $\widehat{\beta}_2$.

(4 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Chapter 3.3 (Properties of the multiple linear regression coefficients), Chapter 6.3 (The effect of including a variable that ought not to be included).

Dougherty, C. Subject guide (2016): Chapter 3.

Approaching the question

(a) Candidates should note that the expression of the variance can be rewritten as:

$$\text{Var}(\widehat{\beta}_2) = \frac{\sigma_\varepsilon^2}{n \times MSD(X_2) \times (1 - r_{X_2X_3}^2)}$$

where $MSD(X_2) \equiv n^{-1} \sum (X_{2i} - \bar{X}_2)^2$. The four factors that affect the precision then are: n, $MSD(X_2)$, $r_{X_2X_3}^2$ and σ_ε^2. Therefore, to obtain more precise parameter estimates of β_2 it is desirable to have: (i) small error variance, (ii) large sample size, (iii) large sample variability of the regressor X_2, and (iv) small correlation among the regressors X_2 and X_3.

Some candidates gave a discussion of the Gauss–Markov assumptions which is not the answer.

(b) Concept and consequences of including irrelevant variables are standard bookwork – unbiased but less precise because the irrelevant variable typically is correlated with the included regressor, i.e. $r_{X_2X_3}^2 \neq 0$.

More precisely: Let $\tilde{\beta}_2$ be the OLS estimator in a regression without including X_3 and $\widehat{\beta}_2$ be the OLS estimator in a regression with X_3. If $r_{X_2,X_3} \neq 0$ we have:

$$\text{Var}\left(\widehat{\beta}_2\right) = \frac{\sigma_\varepsilon^2}{\sum_{i=1}^{n} (X_{2i} - \bar{X}_2)^2} \times \frac{1}{1 - r_{X_2X_3}^2} > \frac{\sigma_\varepsilon^2}{\sum_{i=1}^{n} (X_{2i} - \bar{X}_2)^2} \equiv \text{Var}(\tilde{\beta}_2).$$

Question 5

A probit model to explain whether a firm is taken over by another firm during a given year postulates

$$\Pr(takeover = 1 \mid x) = \Phi(\beta_0 + \beta_1 avgprof + \beta_2 mtkval + \beta_3 debtearn + \beta_4 ceoten + \beta_5 ceosal + \beta_6 ceoage)$$

where $\Phi(z)$ is the cumulative standardised normal distribution. *takeover* is a binary response variable, *avgprof* is the firm's average profit margin over several prior years, *mktval* is the market value of the firm, *debtearn* is the debt-to-earnings ratio, and *ceoten*, *ceosal*, and *ceoage* are the tenure, annual salary, and age of the chief executive officer, respectively.

(a) It is argued that using the probit regression model is better than using the linear probability model when explaining the binary variable *takeover*. Discuss the benefits/drawback of using the probit regression model when trying to explain a binary variable.

(5 marks)

(b) **Discuss how you would implement the LR test that variables related to the CEO have no effect on the probability of takeover, other factors being equal. Clearly indicate the null, alternative, test statistic and rejection rule.**

(3 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Chapter 10.1 (The linear probability model), Chapter 10.3 (Probit analysis), and Chapter 10.6 (Introduction to maximum likelihood estimation.

Dougherty, C. Subject guide (2016): Chapter 10.

Approaching the question

(a) The probit model has two main advantages over the linear probability model (LPM): predicted probabilities are restricted to lie in $[0, 1]$ and MLE is (asymptotically) efficient whereas OLS (LPM) will be inefficient given the inherent presence of heteroskedasticity.

The main drawbacks of the probit model relative to the linear probability model is that the coefficients cannot be directly interpreted as the marginal effects of the regressor(s) of interest and it is also computationally more complicated.

(b) Let $\ln L^U$ be the maximised value of the log-likelihood of the unrestricted model and $\ln L^R$ be the maximised value of the log-likelihood of a model that excludes all the CEO-related variables (*ceoten*, *ceosal* and *ceoage*). Perform a likelihood-ratio test.

- $\text{H}_0 : \beta_4 = \beta_5 = \beta_6 = 0$ vs. $\text{H}_1 : \beta_4 \neq 0$ or $\beta_5 \neq 0$ or $\beta_6 \neq 0$.
- Test statistic: $LR = 2 \times (\ln L^U - \ln L^R)$.
- Under H_0, $LR \overset{a}{\sim} \chi^2_3$.
- At the 5% significance level, reject H_0 if $LR > 7.815$.

Section B

Answer three questions from this section.

Question 6

The following question concerns the effects of background characteristics on student's performance in the SAT (Scholastic Assessment Test). The SAT test is used for college admissions in the US.

$$\widehat{sat} = \underset{(6.29)}{1,028.10} + \underset{(3.83)}{19.3}\,hsize + \underset{(.53)}{-2.19}hsize^2 - \underset{(4.29)}{45.09}female - \underset{(12.71)}{169.81}black + \underset{(12.71)}{62.31}female*black$$

$$n = 4,127, \ R^2 = .0858$$

The variable *hsize* is the size of the student's high school graduating class, in hundreds, *female* is a gender dummy variable (1 = female, 0 = male), and *black* is a race dummy variable (1 = black, 0 = otherwise). The standard errors are in parentheses.

(a) **What is the economic rationale for including $hsize^2$ in the above regression? Using this equation, determine for a given gender and race, what the graduating class size would be at which the predicted SAT scores are maximised.**

(5 marks)

(b) **Holding *hsize* fixed, what is the estimated difference in SAT scores between nonblack females and nonblack males? Is this difference statistically significant? Interpret this result.**

(5 marks)

(c) **What is the estimated difference in SAT score between black females and nonblack females? What would you need to do to test whether the difference is statistically significant?**

(5 marks)

(d) **Discuss any problem you may have in estimating the model if all females in your sample are black. What name does this problem have and what can you do to mitigate this problem?**

(5 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Chapter 4.3 (Models with quadratic variables), Chapter 6.5 (Testing a linear restriction), Chapter 5.1–5.3 (Dummy variables).

Dougherty, C. Subject guide (2016): Chapter 5.

Approaching the question

(a) The economic rationale for including a quadratic term in class size is that it allows for diminishing effects of class size on student's performance. The optimal class size is given by:

$$\left.\frac{\partial \widehat{SAT}}{\partial hsize}\right]_{hsize^*} = 0 \quad \Leftrightarrow \quad 19.3 - 2 \times 2.19 \times hsize^* = 0 \quad \Leftrightarrow \quad hsize^* = \frac{19.3}{4.38} \approx 4.4.$$

The optimal graduating class size is about 440 students.

(b) The expected difference in SAT scores between nonblack females and nonblack males is given by:

$$\mathrm{E}(sat \mid hsize, black = 0, female = 1) - \mathrm{E}(sat \mid hsize, black = 0, female = 0) = \beta_3.$$

The estimated difference is -45.03 which means that on average, holding the size of graduating class constant, nonblack females have a SAT score that is 45 points below nonblack males. To test the significance of this difference we perform a hypothesis test:

- $\mathrm{H}_0 : \beta_3 = 0$ vs. $\mathrm{H}_1 : \beta_3 \neq 0$.
- Test statistic: $t = \widehat{\beta}_3 / se(\widehat{\beta}_3) = -45.09/4.29 \approx -10.5$.
- Under H_0 and assuming Gauss–Markov plus normality, $t \sim t_{4121}$.
- At the 5% significance level, we reject H_0 since $|t| > 1.96$.

(c) The expected difference in SAT scores between black females and nonblack females is given by:

$$\mathrm{E}(sat \mid hsize, black = 1, female = 1) - \mathrm{E}(sat \mid hsize, black = 0, female = 1) = \beta_4 + \beta_5.$$

The estimated difference is given by $-169.81 + 62.31 = -107.5$ which means that, holding size of graduating class constant, on average black females have a sat score that is 107.5 points lower than their white counterparts. To perform the test of $\mathrm{H}_0 : \beta_4 + \beta_5 = 0$ vs. $\mathrm{H}_1 : \beta_4 + \beta_5 \neq 0$ one would need either: (i) $\widehat{\mathrm{Cov}(\widehat{\beta}_4, \widehat{\beta}_5)}$ to compute the t statistic, or (ii) the R^2 of the restricted model to compute the F statistic.

(d) If all females are black, then $female_i = female_i \times black_i \; \forall \; i$, so the variables will be perfectly collinear and he would not be able to obtain the OLS estimates. In this case, we could either drop the interaction term from the regression at the expense of not being able to identify any heterogeneity in the effects or obtain more data such that the sample contains some white females.

Question 7

Let us consider the estimation of a hedonic price function for houses. The hedonic price refers to the implicit price of a house given certain attributes (e.g., the number of bedrooms). The data contains the sale price of 546 houses sold in the summer of 1987 in Canada along with their important features. The following characteristics are available: the lot size of the property in square feet (*lotsize*), the numbers of bedrooms (*bedrooms*), the number of full bathrooms (*bathrooms*), and a dummy indicating the presence of airconditioning (*airco*).

Consider the following ordinary least squares results

$$\widehat{\log(price)}_i = \underset{\substack{(.232)\\ [.233]}}{7.094} + \underset{\substack{(.028)\\ [.028]}}{0.400} \log(lotsize)_i + \underset{\substack{(.015)\\ [.017]}}{0.078} bedrooms_i + \underset{\substack{(.023)\\ [.024]}}{0.216} bathrooms_i + \underset{\substack{(.024)\\ [.023]}}{0.212} airco_i \quad n = 546, \ RSS = 32.622 \tag{7.1}$$

The usual standard errors are in parentheses, the heteroskedasticity robust standard errors are in square brackets, and *RSS* measures the residual sum of squares.

(a) Interpret the parameter estimates on log(*lotsize*), *bedrooms*, and *airco*. Briefly discuss the statistical significance of the results.

(5 marks)

(b) Suppose that lot size was measured in square metres rather than square feet. How would this affect the parameter estimates of the slopes and intercept? How would this affect the fitted values? *Note*: the conversion (approximate) $1m^2 = 10ft^2$.

(5 marks)

(c) We are interested in testing the hypothesis $H_0 : \beta_{bedrooms} = \beta_{bathrooms}$ against the alternative $H_1 : \beta_{bedrooms} \neq \beta_{bathrooms}$. Discuss a test for this hypothesis that makes use of the following restricted regression result

$$\widehat{\log(price)}_i = \underset{(.234)}{6.994} + \underset{(.282)}{0.408} \log(lotsize)_i + \underset{(.011)}{0.127} bbrooms_i + \underset{(.024)}{0.215} airco_i \tag{7.2}$$
$$n = 546, \ RSS = 33.758$$

where *bbrooms* = *bedrooms* + *bathrooms*. Clearly indicate the assumptions you are making for this test to be valid.

(5 marks)

(d) You are interested in testing for the presence of heteroskedasticity. Say you are told that the variance is increasing with log(*lotsize*). Discuss how you would test for the presence of heteroskedasticity. What is the name of the test you are proposing?

(5 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Chapter 1.4 (Interpretation of a regression equation – units of measurement), Chapter 4.2 (Logarithmic transformations), Chapter 2.6 (Testing hypotheses relating to the regression coefficients), Chapter 7.1 (Heteroskedasticity and its implications), and Chapter 7.2 (Detection of heteroskedasticity).

Dougherty, C. Subject guide (2016): Chapter 7.

Approaching the question

(a) Clear discussion of interpretation required (units not always clear): On average, holding the remaining variables in the regression constant, (i) a 1% increase in lot size is associated with a 0.4% increase in house price, (ii) each extra bedroom is associated with a 7.8% increase in house price, and (iii) houses with air conditioning are 21.2% more expensive than those without. All estimates are statistically significant at 5% significance levels. Candidates should clearly indicate H_0 and H_1, the test statistic and the rejection rule.

(b) Let $lotsize_i$ be the lot size in square feet and $\widetilde{lotsize}_i$ be the lot size in square metres. We have that $\widetilde{lotsize}_i = (10)^{-1} lotsize_i$ and $\log(\widetilde{lotsize}_i) = \log((10)^{-1}) + \log(lotsize_i)$. Since this is an additive transformation of one of the explanatory variables we have that (i) the regression slopes will not be affected, (ii) the intercept will change to $7.094 - \log((10)^{-1}) \times 0.4$, and (iii) the fitted values will also not be affected. Many candidates made an error here, ignoring the fact that the variable whose measurement was changed entered in log form.

(c) Candidates would have to recognise that (6.2) is a restricted version of (6.1) where $\beta_{bedrooms} = \beta_{bathrooms}$ is imposed. We therefore need to use the F test. Test statistic:

$$F = \frac{RRSS - URSS}{URSS} \times \frac{n-K}{J} = \frac{33.758 - 32.622}{32.622} \times \frac{541}{1} \approx 18.84.$$

Assuming the Gauss–Markov assumptions plus normality of the error term hold: under H_0, $F \sim F_{1,\,541}$. At the 5% signicance level we reject H_0 since $F > 3.86$. Conclusion: The effect of one extra bathroom is different from the effect of one extra bedroom. Some candidates did not recognise this and were proposing a test on the coefficient of *bbrooms* equalling zero which is wrong.

(d) Assuming Gauss–Markov assumptions plus the normality of the error hold, he can use the Goldfeld–Quandt test for heteroskedasticity. For that purpose, we should first order the 546 observations by the magnitude of $\log(lotsize_i)$. Fit one regression for the first n^* observations and another for the last n^* observations (usually n^* equals one-third of the sample). Let RSS_1 and RSS_2 denote the sum of squared residuals in each of these regressions, respectively.

- $\mathrm{H}_0 : \sigma_2^2 = \sigma_1^2$ vs. $\mathrm{H}_1 : \sigma_2^2 > \sigma_1^2$.
- Test statistic: $GQ = RSS_2/RSS_1$.
- Under H_0, $GQ \sim F_{n^*-k,\, n^*-k}$.
- Reject if GQ is greater than the 95th percentile of the F distribution above.

Candidates need to be careful not to state $\mathrm{H}_0 : RSS_2 = RSS_1$ vs. $\mathrm{H}_1 : RSS_2 > RSS_1$. Both RSS_1 and RSS_2 are random variables. Because the sample sizes are identical it is also correct to state $\mathrm{H}_0 : RSS_1$ and RSS_2 are not statistically different. Note that simply by having one sample larger than the other, you could also have a larger residual sum of squares.

Question 8

Let $math10$ denote the percentage of students at a high school receiving a passing score on a standardised math test. We are interested in estimating the effect of per student spending on math performance. A simple model is

$$math10_i = \beta_0 + \beta_1 \log(expend_i) + \beta_2 \log(enroll_i) + \beta_3 poverty_i + u_i \qquad (8.1)$$

where, for each high school i; $poverty_i$ is the percentage of students living in poverty, $expend_i$ is the spending per student and $enroll_i$ the number of registered students. You may assume that this model satisfies all Gauss–Markov assumptions.

You are faced with the fact that data is unavailable on a key variable: *poverty*.

(a) Discuss the properties (unbiasedness and consistency) of the estimators when you drop the variable poverty. Explain your answers.

(5 marks)

You do have information available on a closely related variable: the percentage of students eligible for the federally funded school lunch program, $lnchprg_i$. Let us consider using $lnchprg_i$ as a proxy for $poverty_i$.

(b) Briefly discuss why $lnchprg_i$ is a sensible proxy variable for the unobserved variable $poverty_i$.

(2 marks)

(c) It is unlikely that $lnchprg_i$ is an ideal proxy, in the sense that there is an exact linear relationship between them, instead, we will assume that

$$poverty_i = \alpha_0 + \alpha_1 lnchprg_i + v_i, \quad \alpha_1 \neq 0 \qquad (8.2)$$

Discuss the assumptions you need to make to enable consistent parameter estimators of β_1 and β_2 using your estimable equation

$$math10_i = \gamma_0 + \gamma_1 \log(expend_i) + \gamma_2 \log(enroll_i) + \gamma_3 lnchprg_i + e_i,$$

***Hint*: Consider the relation between the γ and the β parameters and express e_i in terms of u_i and v_i.**

(5 marks)

(d) The OLS results with and without $lnchprg_i$ as an explanatory variable are given by (standard errors in parentheses):

$$\widehat{math10}_i = \underset{(26.72)}{-69.24} + \underset{(3.30)}{11.13} \log expend_i + \underset{(0.615)}{0.022} \log emroll_i$$
$$N = 428, \; R^2 = 0.0297$$

$$\widehat{math10}_i = \underset{(24.99)}{-23.14} + \underset{(3.04)}{7.75} \log expend_i - \underset{(0.58)}{1.26} \log emroll_i - \underset{(0.036)}{0.324} lnchprg_i$$
$$N = 428, \; R^2 = 0.1893$$

i. Interpret the coefficient on $lnchprg$. What does this parameter tell us regarding the parameter of interest β_3?

(4 marks)

ii. Give an intuitive discussion explaining why the effect of expenditures on $math10_i$ is lower in the regression where $lnchprg_i$ is included than where it is excluded.

(4 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Chapter 6.2 (The effect of omitting a variable that ought to be included), Chapter 6.4 (Proxy variables).

Dougherty, C. Subject guide (2016): Chapter 6.

Approaching the question

(a) Consider rewriting (8.1) as:

$$math10_i = \beta_0 + \beta_1 \log(expend_i) + \beta_2 \log(enroll_i) + \varepsilon_i$$

where $\varepsilon_i = \beta_3 poverty_i + u_i$. Assuming $\beta_3 \neq 0$, if poverty is correlated with either (log) expenditures and/or (log) enrollment the model will suffer from endogeneity due to omitted variables and OLS will be biased and inconsistent. That is likely to be the case since schools that have smaller expenditures tend to be located in poorer neighbourhoods and hence to have more students living in poverty conditions.

(b) It is a sensible proxy because it is likely to be correlated with poverty and to capture some of the effect of poverty since usually students eligible for the lunch program tend to be those with low levels of family income.

(c) Plug in (8.2) for poverty in (8.1) to obtain:

$$math10_i = \underbrace{(\beta_0 + \beta_3\alpha_0)}_{\gamma_0} + \underbrace{\beta_1}_{\gamma_1} \log(expend_i) + \underbrace{\beta_2}_{\gamma_2} \log(enroll_i) + \underbrace{\beta_3\alpha_1}_{\gamma_3} lnchprg_i + \underbrace{(\beta_3 v_i + u_i)}_{\varepsilon_i}.$$

Given that the model (8.1) satisfies all the Gauss–Markov assumptions, to obtain consistent estimates of β_1 and β_2 we need that: (i) v_i is uncorrelated with $\log(expend_i)$, $\log(enroll_i)$ and $lnchprg_i$, and (ii) $lnchprg_i$ is uncorrelated with u_i.

(d) i. On average, holding expenditure and enrollment constant, a 1 percentage point increase in the number of students eligible for the lunch program is associated with a 0.324 percentage point fall in the percentage of students receiving a passing score in the standardised math test. Since $\gamma_3 = \alpha_1\beta_3$ and assuming $\alpha_1 > 0$, the direction of the effect (sign) of poverty of $math10$ is the same as the effect of $lnchprg$ on $math10$, in this case, with a negative coefficient on $lnchprg$ we can infer that poverty has a negative effect on $math10$.

ii. Omitting relevant variables will result in the remaining parameters attempting to pick up its effect through the correlation these omitted variables have with the included regressors. We, therefore, expect the effect to be smaller, as part of the effect we attribute to expenditure in the short regression is actually coming from the fact that high schools that have larger expenditures tend to have fewer students eligible for the lunch program and those students tend to perform worse in the standardised math test.

Question 9

Let us consider monthly data on the short-term interest rate (the three month Treasury Bill rate) and on the AAA corporate bond yield in the USA. The data run from January 1950 to December 1999. Let $DUS3MT$ denote the changes in three-month Treasury Bill rate, and $DAAA$ denote the changes in AAA bond rate. We consider the following results (with the standard errors given in parentheses)

$$\widehat{DAAA}_t = \underset{(.007)}{0.006} + \underset{(.015)}{0.275} DUS3MT_t \quad t = 1, \ldots, 600 \qquad (9.1)$$

$$RSS = 17.486; \; DW = 1.447$$

where RSS is the residual sum of squares and DW is the Durbin–Watson test.

A researcher interpreting the residuals suggests that the errors show a positive correlation over time.

(a) **What are the consequences of this correlation for the above regression results?**

(5 marks)

(b) **Use the results above to test for the presence of first-order positive autocorrelation. Clearly specify the null and alternative hypothesis, test statistic, assumptions underlying the test, and the acceptance/rejection rule.**

(5 marks)

(c) **In an attempt to remove the autocorrelation you consider the following specification**

$$\widehat{DAAA}_t = \underset{(.007)}{0.005} + \underset{(.015)}{0.252} DUS3MT_t - \underset{(.018)}{0.080} DUS3MT_{t-1} + \underset{(.040)}{0.290} DAAA_{t-1} \quad (9.2)$$

$$RSS = 16.087; \; DW = 1.897$$

Comment on the following statement 'The Durbin–Watson statistic is closer to 2, indicating that we have succesfully removed the autocorrelation'. If you disagree with this statement, suggest what you would need to do instead.

(5 marks)

(d) **Discuss the Common Factor Test as a model specification suitable for this model. What extra information do you need to conduct this test.**

(5 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Chapter 12.1–12.3 (Definition, consequences and detection of autocorrelation; Fitting a model subject to AR(1) autocorrelation).

Dougherty, C. Subject guide (2016): Chapters 12.

Approaching the question

(a) Standard bookwork discussion of autocorrelation in the setting where there are no lagged dependent variables.

(b) Discussion of Durbin–Watson test expected – bookwork. Candidates should clearly indicate the H_0 and H_1, test statistic and rejection rule. Assumptions underlying the test: tests only AR(1) autocorrelation, no lagged dependent variables (deterministic regressors only) in presence of intercept.

(c) We disagree with the statement because the Durbin–Watson test is not valid in the presence of lagged dependent variables since in this case regressors cannot be strictly exogenous. Instead, we should use the Breusch–Godfrey test which is asymptotically valid with predetermined regressors a weaker requirement than strict exogeneity. Alternatively, the Durbin h test can be proposed.

(d) Assuming AR(1) autocorrelation and letting $Y_t = DAAA_t$ and $X_t = DUS3MT_t$ we can remove the autocorrelation by rewriting the model as:

$$Y_t = (1-\rho)\beta_1 + \rho Y_{t-1} + \beta_2 X_t - \beta_2 \rho X_{t-1} + \varepsilon_t.$$

This is a restricted version of a more general ADL(1,1) model:

$$Y_t = \lambda_1 + \lambda_2 Y_{t-1} + \lambda_3 X_t + \lambda_4 X_{t-1} + \varepsilon_t$$

under the restriction $\lambda_4 = -\lambda_2\lambda_3$. We can test this restriction using the Common Factor Test. Let RSS_r and RSS_u be the restricted sums of squares of the restricted and unrestricted model, respectively.

- Test statistic: $CF = n \times \log(RSS_r/RSS_u)$.
- Under H_0, $CF \overset{a}{\sim} \chi^2_1$.
- Extra information needed: RSS_r (since RSS_u is given in (9.2)).

Question 10

Consider the model

$$\boldsymbol{y_t = \alpha + \beta x_t + \varepsilon_t, \quad t = 1, \ldots, T} \tag{10.1}$$

where $\boldsymbol{y_t}$ and $\boldsymbol{x_t}$ are both integrated of order one.

(a) **Explain what it means to say that $\boldsymbol{y_t}$ is integrated of order one. Discuss how you would test for this. In your answer make sure that it is clear how to implement your test.**

(6 marks)

(b) **Give an example of an economic variable that is potentially integrated of order one and give an intuitive explanation why you expect this process to be integrated of order 1.**

(2 marks)

(c) It will be important to distinguish whether the above relationship is 'spurious' as opposed to 'cointegrating'.

i. Explain what it means to say that y_t and x_t have a cointegrating relationship and how does that contrast to a spurious relationship.

(4 marks)

ii. Discuss how you can test for evidence of a cointegrating relationship.

(4 marks)

(d) Suppose that

$$\varepsilon_t = \rho\varepsilon_{t-1} + v_t, \quad |\rho| < 1,$$

and v_t is an i.i.d. $(0, \sigma^2)$ innovation which is independent of ε_{t-1}. Show that you can rewrite equation (10.1) in terms of an error correction model:

$$\Delta y_t = \delta_1 \Delta x_t + \delta_2 (y_{t-1} - \alpha - \beta x_{t-1}) + v_t.$$

Clearly indicate the relation between (δ_1, δ_2) and (α, β, ρ). Give an economic intuition behind this result.

(4 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Chapter 13.1 (Stationarity and nonstationarity), Chapter 13.4–13.5 (Tests of nonstationarity), and Chapter 13.6 (Cointegration).

Dougherty, C. Subject guide (2016): Chapter 13.

Approaching the question

(a) Integrated of order one (or simply I(1)) means that the process can be made stationary by differencing once. One could test for this using a Dickey–Fuller or an Augmented Dickey–Fuller test. Standard bookwork. For instance, suppose $y_t = \rho y_{t-1} + \varepsilon_t$. We can conduct the Dickey–Fuller test by running the auxiliary regression of the form:

$$\Delta y_t = \gamma y_{t-1} + \varepsilon_t$$

where $\gamma \equiv \rho - 1$. Provide test statistic and discuss rejection rule, clearly indicating H_0 and H_1.

(b) One example could be CPI due to its clearly trending behaviour over time it can be suggestive of a time-series integrated of order 1.

(c) i. Two variables integrated of order one are said to be cointegrated if there exists a linear combination of them that is integrated of order zero. In those cases, a regression of (10.1) will capture the long-run relationship between y_t and x_t and we should prefer an Error Correction Model to capture both the short-term and the long-term relationships among the two variables. In contrast, if the two series are not cointegrated a regression like (10.1) is likely to indicate a very strong relationship between the two variables even when there is no relationship at all. Such a regression is called a spurious regression.

ii. If y_t and x_t are both found to be integrated of order one, one could test for cointegration as follows:

* Run a regression of y_t on a constant and x_t and collect residuals.
* Perform a DF or ADF test on the residual series.
* If reject the null of non-stationarity conclude the series are cointegrated.

(d) Lag (10.1) by one period and multiply both sides by ρ to obtain:

$$\rho y_{t-1} = \rho\alpha + \rho\beta x_{t-1} + \rho\varepsilon_{t-1}.$$

Subtract the above from (10.1), hence:

$$y_t - \rho y_{t-1} = (1-\rho)\alpha + \beta x_t - \rho\beta x_{t-1} + \underbrace{\varepsilon_t - \rho\varepsilon_{t-1}}_{v_t}.$$

Rearranging:

$$y_t = \alpha + \rho(y_{t-1} - \alpha - \beta x_{t-1}) + \beta x_t + v_t.$$

Subtract $y_{t-1} = \alpha + \beta x_{t-1} + \varepsilon_{t-1}$, hence:

$$\Delta y_t = \rho(y_{t-1} - \alpha - \beta x_{t-1}) + \beta(x_t - x_{t-1}) + v_t - \varepsilon_{t-1}.$$

Finally, using $\varepsilon_{t-1} = y_{t-1} - \alpha - \beta x_{t-1}$ we obtain:

$$\Delta y_t = \beta\Delta x_t + (\rho - 1)(y_{t-1} - \alpha - \beta x_{t-1}) + v_t.$$

Therefore, $\delta_1 = \beta$ and $\delta_2 = \rho - 1$. Economic intuition: this is a process that reverts to the long-run equilibrium after a shock (δ_1 captures the short-run dynamics and δ_2 captures the speed of the reversal to the long-run equilibrium).

Examiners' commentaries 2019

EC2020 Elements of econometrics

Important note

This commentary reflects the examination and assessment arrangements for this course in the academic year 2018–19. The format and structure of the examination may change in future years, and any such changes will be publicised on the virtual learning environment (VLE).

Information about the subject guide and the Essential reading references

Unless otherwise stated, all cross-references will be to the latest version of the subject guide (2016). You should always attempt to use the most recent edition of any Essential reading textbook, even if the commentary and/or online reading list and/or subject guide refer to an earlier edition. If different editions of Essential reading are listed, please check the VLE for reading supplements – if none are available, please use the contents list and index of the new edition to find the relevant section.

General remarks

Learning outcomes

At the end of the course, and having completed the Essential reading and activities, you should be able to:

- describe and apply the classical regression model and its application to cross-section data
- describe and apply the:
  - Gauss–Markov conditions and other assumptions required in the application of the classical regression model
  - reasons for expecting violations of these assumptions in certain circumstances
  - tests for violations
  - potential remedial measures, including, where appropriate, the use of instrumental variables
- recognise and apply the advantages of logit, probit and similar models over regression analysis when fitting binary choice models
- competently use regression, logit and probit analysis to quantify economic relationships using standard regression programmes (Stata and EViews) in simple applications
- describe and explain the principles underlying the use of maximum likelihood estimation
- apply regression analysis to time-series models using stationary time series, with awareness of some of the econometric problems specific to time series applications (for example, autocorrelation) and remedial measures
- recognise the difficulties that arise in the application of regression analysis to nonstationary time series, know how to test for unit roots, and know what is meant by cointegration.

Common mistakes committed by candidates

A large number of candidates are not able to clearly distinguish between sample variance and covariance, and population variance and covariance (this is happening year after year).

The use of $\text{Cov}(X, Y)$ and $\text{Var}(X)$ should be restricted to describing the population covariance and variances, respectively, with definitions:

$$\text{Cov}(X, Y) = \text{E}\left((X - \text{E}(X))(Y - \text{E}(Y))\right) = \text{E}(XY) - \text{E}(X)\,\text{E}(Y)$$

and:

$$\text{Var}(X) = \text{E}((X - \text{E}(X))^2) = \text{E}(X^2) - (\text{E}(X))^2$$

(you also may denote $\text{Cov}(X, Y) = \sigma_{XY}$ and $\text{Var}(X) = \sigma_X^2$). They are typically unknown, but fixed, quantities.

The sample covariance and variance are estimators of the population covariance and variance, respectively. They are defined as:

$$\text{Sample Cov}(X, Y) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})$$

and:

$$\text{Sample Var}(X) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

(you also may use $\widehat{\sigma}_{XY}$ and $\widehat{\sigma}_X^2$). You can compute them given the data.

With a slight abuse of notation, we often divide by n instead, which is irrelevant if we let n be large. The division by $n - 1$ is a finite sample issue only (unbiasedness).

The sample covariance and variance show up in our definition of the OLS estimator of the slope in the simple linear regression model, not the population covariance and variance, as:

$$\widehat{\beta} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \frac{\text{Sample Cov}(X, Y)}{\text{Sample Var}(X)} \neq \frac{\text{Cov}(X, Y)}{\text{Var}(X)}.$$

Treating them as being the same results in incorrect analyses and candidates losing significant marks.

Candidates should realise that $\frac{1}{n}\sum_{i=1}^{n} u_i$ is not the same as $\text{E}(u_i)$. So, while we typically assume $\text{E}(u_i) = 0$, this does not guarantee that $\frac{1}{n}\sum_{i=1}^{n} u_i = 0$. Also, while we may be happy to assume $\text{E}(x_i u_i) = 0$ (uncorrelatedness between the errors and regressors), this does not guarantee that $\frac{1}{n}\sum_{i=1}^{n} x_i u_i = 0$. Note that:

- both $\frac{1}{n}\sum_{i=1}^{n} u_i$ and $\frac{1}{n}\sum_{i=1}^{n} x_i u_i$ are random variables, which take the value 0 with probability 0 (continuous random variables)!
- $\text{E}(u_i) = 0$ and $\text{E}(x_i u_i) = 0$ are fixed, not stochastic!

The differences between sample and population moments need to come across clearly when looking at unbiasedness and making consistency arguments. In both cases, we first simplify our estimator (plug in the true model) to obtain:

$$\widehat{\beta} = \beta + \frac{\sum_{i=1}^{n}(X_i - \bar{X})u_i}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \beta + \frac{\sum_{i=1}^{n} x_i u_i}{\sum_{i=1}^{n} x_i^2} \quad \text{with } x_i = X_i - \bar{X}.$$

- For *unbiasedness*, clearly indicate that you want to show that $\mathrm{E}(\widehat{\beta}) = \beta$. Unbiasedness does not follow from $\sum\limits_{i=1}^{n} x_i u_i = 0$, instead it follows from $\mathrm{E}\left(\frac{\sum_{i=1}^{n} x_i u_i}{\sum_{i=1}^{n} x_i^2}\right) = 0$.

  If we treat x_i as fixed, $\mathrm{E}\left(\frac{\sum_{i=1}^{n} x_i u_i}{\sum_{i=1}^{n} x_i^2}\right) \equiv \mathrm{E}\left(\sum\limits_{i=1}^{n} d_i u_i\right) = \sum\limits_{i=1}^{n} d_i\, \mathrm{E}(u_i)$ and then unbiasedness follows as $\mathrm{E}(u_i) = 0$.
- For *consistency*, clearly indicate that you want to show that $\operatorname{plim}(\widehat{\beta}) = \beta$. Using the plim properties, we show:

$$\begin{aligned}
\operatorname{plim} \widehat{\beta} = \beta + \operatorname{plim}\left(\frac{\sum_{i=1}^{n} x_i u_i}{\sum_{i=1}^{n} x_i^2}\right) &= \beta + \frac{\operatorname{plim}\left(\frac{1}{n}\sum_{i=1}^{n} x_i u_i\right)}{\operatorname{plim}\left(\frac{1}{n}\sum_{i=1}^{n} x_i^2\right)} \\
&\equiv \beta + \frac{\operatorname{plim}\left(\text{Sample Cov}\,(x, u)\right)}{\operatorname{plim}\left(\text{Sample Var}\,(x)\right)} \\
&= \beta + \frac{\operatorname{Cov}(x, u)}{\operatorname{Var}(x)} \quad \text{using the law of large numbers}
\end{aligned}$$

  where $\operatorname{Cov}(x, u) = 0$ and $\operatorname{Var}(x) > 0$, ensuring we get consistency.
- Remember, the law of large numbers ensures that sample averages converge to their population analogues.

Candidates struggled to give competent answers to the interpretation of empirical results. When interpreting an empirical result you should discuss the significance of the coefficients, magnitude and sign of the coefficients.

Candidates particularly struggled with the difference between percentage points and changes when it comes to the interpretation of coefficients in binary choice models. A coefficient in such a model gives the percentage point difference in the outcome variable when the regressor changes by 1 unit. The % change is a *relative* change which depends on the baseline value of the outcome variable (for example, the realisation when a regressor equals zero in the case of a dummy variable regressor). In the case of dummy variables as regressors, candidates often missed that the coefficients have to be interpreted relative to the base case which is the left out dummy variable (see appropriate questions in the examinations for details).

When conducting hypothesis tests, you should make sure that the Gauss–Markov conditions hold. The Gauss–Markov conditions have to be explicitly specified. Only writing that the Gauss–Markov conditions hold is not sufficient. As good practice, begin your examination by explicitly providing the Gauss–Markov conditions. You can then refer back to them thereafter. Moreover, ensure when conducting hypothesis testing that you clearly indicate the null and alternative hypotheses (in terms of the true parameters, say β_1), the test statistic (in terms of the parameter estimates, here $\widehat{\beta}_1$), its distribution (with degrees of freedom), the rejection rule (one-sided or-two sided) for a given significance level (typically 5%) with suitable critical values, and provide an interpretation of your result.

Exogeneity and endogeneity are often very simply explained as variables 'outside of the model' or 'inside the model', respectively. While this is not wrong, it is not a particularly clear or helpful definition, in particular in general contexts which are not related to simultaneous equations models. Defining these concepts in terms of the relationship between regressor and error term is much clearer and naturally leads to a way to show exogeneity or endogeneity by considering the covariance between regressors and error terms.

Candidates often missed that all exogenous regressors and instruments have to be included in the first stage when performing TSLS estimation.

It was very common to misunderstand the central limit theorem. While it provides the distribution of the *estimator* $\widehat{\beta}$, no matter what the true distribution of the errors is, it does not say that the distribution of the error term is approximately normal in large samples.

When asked about identification of structural form equations, candidates often compared the number of excluded regressors to the number of endogenous variables in general, without pointing out which variables are the excluded regressors and could, therefore, be used as an instrument.

Just as last year, many candidates do not answer all parts of the question. Make sure you read the questions properly and provide all details that are requested. Not answering a question will automatically earn you a zero mark for that question.

Key steps to improvement

Essential reading for this course includes the subject guide and the following:

- Dougherty, C. *Introduction to econometrics.* (Oxford: Oxford University Press, 2016) 5th edition [ISBN 9780199676828]; http://oxfordtextbooks.co.uk/orc/dougherty5e/

Apart from the Essential readings you should do some supplementary reading. One very good book at the same level is:

- Gujarati, D.N. and D.C. Porter *Basic econometrics.* (McGraw–Hill, 2009, International edition) 5th edition [ISBN 9780071276252].

To understand the subject clearly it is important to supplement Dougherty's *Introduction to econometrics* (fifth edition) with the subject guide **EC2020 Elements of econometrics** (2016), especially Chapter 10 which covers maximum likelihood estimation. It is very important to carefully go through the subject guide. The subject guide contains solutions to the questions given in the main textbook and also some additional questions and solutions. Working through these will improve your understanding of the subject.

The chapter in the subject guide on maximum likelihood (Chapter 10) includes some additional theory which has not been covered in the main textbook. It is important to read the additional theory given in the subject guide to have a better understanding of the principles of maximum likelihood and tests based on the likelihood function.

Please check the VLE course page for resources for this subject such as a downloadable copy of the subject guide **EC2020 Elements of econometrics** (2016), PowerPoint slideshows that provide a graphical treatment of the topics covered in the textbook, datasets and statistical tables. Candidates should utilise datasets using standard regression programmes (STATA or EViews). This will help in the understanding of the subject.

Examination revision strategy

Many candidates are disappointed to find that their examination performance is poorer than they expected. This may be due to a number of reasons, but one particular failing is '**question spotting**', that is, confining your examination preparation to a few questions and/or topics which have come up in past papers for the course. This can have serious consequences.

We recognise that candidates might not cover all topics in the syllabus in the same depth, but you need to be aware that examiners are free to set questions on **any aspect** of the syllabus. This means that you need to study enough of the syllabus to enable you to answer the required number of examination questions.

The syllabus can be found in the Course information sheet available on the VLE. You should read the syllabus carefully and ensure that you cover sufficient material in preparation for the examination. Examiners will vary the topics and questions from year to year and may well set questions that have not appeared in past papers. Examination papers may legitimately include questions on any topic in the syllabus. So, although past papers can be helpful during your revision, you cannot assume that topics or specific questions that have come up in past examinations will occur again.

If you rely on a question-spotting strategy, it is likely you will find yourself in difficulties when you sit the examination. We strongly advise you not to adopt this strategy.

Examiners' commentaries 2019

EC2020 Elements of econometrics

Important note

This commentary reflects the examination and assessment arrangements for this course in the academic year 2018–19. The format and structure of the examination may change in future years, and any such changes will be publicised on the virtual learning environment (VLE).

Information about the subject guide and the Essential reading references

Unless otherwise stated, all cross-references will be to the latest version of the subject guide (2016). You should always attempt to use the most recent edition of any Essential reading textbook, even if the commentary and/or online reading list and/or subject guide refer to an earlier edition. If different editions of Essential reading are listed, please check the VLE for reading supplements – if none are available, please use the contents list and index of the new edition to find the relevant section.

Comments on specific questions – Zone A

Candidates should answer **EIGHT** of the following **TEN** questions: **ALL** of the questions in Section A (8 marks each) and **THREE** questions from Section B (20 marks each). **Candidates are strongly advised to divide their time accordingly.**

Section A

Answer all questions from this section.

Question 1

Consider the following regression model:

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 t + \varepsilon_t, \quad t = 1, \ldots, T.$$

Both $\{Y_t\}_{t=1}^T$ and $\{X_t\}_{t=1}^T$ are trend stationary processes. The errors $\{\varepsilon_t\}_{t=1}^T$ are independent random variables with zero mean and constant variance.

(a) **Discuss the concept 'trend stationarity' and contrast it to the concept 'difference stationarity'. In your answer make sure you also explain what stationarity means.**

(4 marks)

(b) **Provide a clear interpretation of the parameter β_1. You are told that you can obtain the estimator for β_1 using 'detrended' variables only. Discuss this statement.**

(4 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Sections 13.1 and 13.2.

Dougherty, C. Subject guide (2016): Chapter 13.

Approaching the question

(a) *Trend stationarity* relates to a process which has a deterministic trend ($y_t = \alpha + \beta t + \varepsilon_t$ where ε_t is weakly dependent). It is a type of non-stationarity which can be removed by detrending. Once we include a trend in the regression, we can use standard statistical inference methods: the errors (and, therefore, Y_t) will be weakly dependent and consistent estimates can be obtained.

Difference stationarity relates to a process which can be made stationary by taking first differences: ΔY_t, examples are random walk and random walk with drift. Here Y_t is persistent, strongly dependent, and estimation of the process Y_t on Y_{t-1} would fail to produce consistent parameter estimates.

By stationarity we mean *covariance stationarity*, which requires constant (finite) means and variances, and covariances which only depend on distance in time, i.e. $\mathrm{E}(Y_t, Y_{t-s})$ only depends on s, not on location t.

(b) β_1 is the marginal effect of X_t on Y_t after controlling for a linear time trend, that is holding time constant (*ceteris paribus*).

Frisch–Waugh–Lovell (regression anatomy) argue that including the time trend $\beta_2 t$ explicitly in the regression is the same as working with detrended data where we control for the trend beforehand (implicitly).

Since both Y_t and X_t have a linear time trend, we need to include it in the regression in order to obtain consistent estimates of β_1. Omitting the time trend in this setting would give rise to omitted variable bias (OVB).

Question 2

Consider the following ADL(1, 1) model relating the crime rate in a particular province, *crime$_t$*, to the clear-up rate (percentage of crimes resulting in a conviction):

$$\boldsymbol{crime_t = \alpha + \rho crime_{t-1} + \delta_1 clearup_t + \delta_2 clearup_{t-1} + u_t, \quad \text{with } |\rho| < 1}$$

where u_t is white noise, an i.i.d. innovation that is uncorrelated to anything in the past.

(a) Briefly indicate whether OLS will provide unbiased and consistent parameter estimators.

(2 marks)

(b) Derive the long run relationship between *crime* and *clearup*.

(2 marks)

(c) Rewrite the model above in terms of an error correction model (ECM) and interpret its coefficients.

(4 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Sections 11.3, 11.4 and 11.5.

Dougherty, C. Subject guide (2016): Chapter 11.

Approaching the question

(a) OLS will be consistent because the error is uncorrelated with the regressors. OLS will be biased because of the presence of the lagged endogenous variable. The latter is due to the fact that u_t will be correlated with $crime_{t+s}$ for any $s \geq 0$ (regressor $crime_{t-1}$ in future periods).

(b) Let in equilibrium $X_t = X_{t-1} = \tilde{X}$ for both variables in the model, so that:

$$\widetilde{crime} = \alpha + \rho\widetilde{crime} + \delta_1\widetilde{clearup} + \delta_2\widetilde{clearup}.$$

After rewriting we obtain the long-run relationship:

$$crime_t = \frac{\alpha}{1-\rho} + \frac{\delta_1 + \delta_2}{1-\rho}clearup_t.$$

(c) The error correction model states that the change in *crime* in any period will be governed by the change in *clearup* and an error correction mechanism which reveals the speed with which *crime* changes in response to a disequilibrium. Here, we get:

$$\Delta crime_t = \delta_1 \Delta clearup_t + (\rho - 1)diseq_{t-1} + u_t$$

where:

$$diseq_{t-1} = crime_{t-1} - \frac{\alpha}{1-\rho} - \frac{\delta_1 + \delta_2}{1-\rho}clearup_{t-1}$$

and $(\rho - 1)$ indicates the speed of adjustment.

Question 3

Consider the simple linear regression model:

$$\boldsymbol{Y_i = \beta X_i + u_i, \quad i = 1, \ldots, n.}$$

We assume that the errors $\{u_i\}_{i=1}^n$ are independent normal random variables with zero mean. The regressor $\{X_i\}_{i=1}^n$ is non-stochastic (fixed under repeated sampling). You suspect that the errors exhibit heteroskedasticity.

(a) Explain what we mean by the concept of heteroskedasticity. Enhance your answer with the help of a graphical illustration.

(3 marks)

(b) Suppose you want to test $H_0 : \beta = 1$ against $H_1 : \beta > 1$. Discuss how you would conduct this test based on the OLS estimator, recognising the presence of heteroskedasticity. Please provide a detailed answer.

(5 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Sections 7.1 and 7.2.

Dougherty, C. Subject guide (2016): Chapters 2 and 7.

Approaching the question

(a) Heteroskedasticity means that the error variance depends on the regressors, it violates the assumption that $E(u_i^2) = \sigma^2$.

A graphical illustration should show data points and some depiction of the variance – for example, drawing the distribution for the errors at different levels of X. The graph should be clearly labelled.

(b) We need to propose the t test where the standard error on β needs to be estimated taking the heteroskedasticity into account (the usual t statistic is invalid because of the presence of heteroskedastity):

$$T = \frac{\widehat{\beta} - 1}{\text{s.e.}(\widehat{\beta})} \sim t_{n-1} \text{ under } \mathrm{H}_0.$$

The White robust standard error we use for this is the square root of the following heteroskedasticity-robust variance estimate:

$$\widehat{\mathrm{Var}}(\widehat{\beta}) = \frac{\sum_{i=1}^{n} X_i^2 \widehat{u}_i^2}{\left(\sum_{i=1}^{n} X_i^2\right)^2}.$$

The test then proceeds in the standard way. Using the 5% significance level we reject H_0 if the realisation of our test statistic exceeds the critical value: reject if $t > t_{0.05,\,n-1}$ (with n large, we reject H_0 if $t > 1.645$). Otherwise, fail to reject H_0.

A good number of candidates proposed a Goldfeld–Quandt test of heteroskedasticity instead of a discussion of how to conduct hypothesis testing given heteroskedasticity which is what the question asked. It is important to carefully read what the question asks!

Question 4

Consider the simple linear regression model:

$$\boldsymbol{Y_i = \beta_0 + \beta_1 X_i + u_i, \quad i = 1, \ldots, n.}$$

We assume that the errors $\{u_i\}_{i=1}^n$ are independent random variables with zero mean. The regressor $\{X_i\}_{i=1}^n$ is non-stochastic (fixed under repeated sampling). Under these conditions, the OLS estimator for β_1, $\widehat{\beta}_1$, is unbiased. (You are not asked to derive $\widehat{\beta}_1$).

(a) Explain the concept of unbiasedness of an estimator.

(2 marks)

(b) Let us consider two other estimators for the slope β_1:

$$\widehat{\beta}_1^{\circ} = \frac{\sum_{i=1}^{n} (Z_i - \bar{Z}) Y_i}{\sum_{i=1}^{n} (Z_i - \bar{Z}) X_i} \quad \text{and} \quad \widehat{\beta}_1^{*} = \frac{\sum_{i=1}^{n} (Z_i - \bar{Z}) Y_i}{\sum_{i=1}^{n} (Z_i - \bar{Z}) Z_i}$$

where $Z_i = \sqrt{X_i}$ for all i and $\bar{Z} = \frac{1}{n}\sum_{i=1}^{n} Z_i$. Please indicate whether $\widehat{\beta}_1^{\circ}$ and $\widehat{\beta}_1^{*}$ are unbiased estimators for β_1. Clearly show your derivations.

(4 marks)

(c) Briefly indicate how you would choose between the three estimators, $\widehat{\beta}_1$, $\widehat{\beta}_1^{\circ}$ and $\widehat{\beta}_1^{*}$.

(2 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Sections 2.3 and 2.5.

Dougherty, C. Subject guide (2016): Chapter 2.

Approaching the question

(a) Unbiasedness means $\mathrm{E}(\widehat{\beta}_1) = \beta_1$. It means that the expected value of the estimator is the true parameter; that is, we are correct on average in repeated samples. This ensures that we will not make systematic errors when estimating β.

(b) In answering this question, it is easiest to derive results using Z_i rather than working with $Z_i = \sqrt{X_i}$. It is important to note furthermore that $\sum(Z_i - \bar{Z}) = 0$.

When plugging in the true model ($Y_i = \beta_0 + \beta_1 X_i + u_i$), we can then write:

$$\widehat{\beta}_1^{\circ} = \beta_1 + \frac{\sum(Z_i - \bar{Z})u_i}{\sum(Z_i - \bar{Z})X_i}$$

and:

$$\mathrm{E}(\widehat{\beta}_1^{\circ}) = \beta_1 + \frac{\sum(Z_i - \bar{Z})\mathrm{E}(u_i)}{\sum(Z_i - \bar{Z})X_i} = \beta_1$$

(X and Z are assumed to be non-stochastic). Unbiasedness follows as $\mathrm{E}(u_i) = 0$.

For the second estimator, when plugging in the true model, we can write:

$$\widehat{\beta}_1^{*} = \beta_1 + \frac{\sum(Z_i - \bar{Z})X_i}{\sum(Z_i - \bar{Z})Z_i} + \frac{\sum(Z_i - \bar{Z})u_i}{\sum(Z_i - \bar{Z})Z_i}$$

and:

$$\mathrm{E}(\widehat{\beta}_1^{*}) = \beta_1 + \frac{\sum(Z_i - \bar{Z})Z_i^2}{\sum(Z_i - \bar{Z})Z_i} \neq \beta_1$$

using again the fact that Z is non-stochastic and $\mathrm{E}(u_i) = 0$, hence $\widehat{\beta}_1^{*}$ is biased.

Side notes: (i) $\widehat{\beta}_1^{\circ}$ essentially is an IV estimator where Z is used as an instrument for X. Both Z and X are exogenous here. (ii) $\widehat{\beta}_1^{*}$, on the other hand, is the slope parameter of a regression of Z instead of X on Y, and hence should not give us an unbiased estimator of β_1.

Many candidates failed to realise that $\sum(Z_i - \bar{Z}) = 0$ and wrongly concluded that the first estimator $\widehat{\beta}_1^{\circ}$ is biased.

(c) When choosing between $\widehat{\beta}_1$ and $\widehat{\beta}_1^{\circ}$ (both unbiased), efficiency considerations matter. Since the model satisfies the Gauss–Markov assumptions, we know that by the Gauss–Markov theorem, $\widehat{\beta}_1$ is BLUE and should, therefore, be chosen over $\widehat{\beta}_1^{\circ}$. Nevertheless, there may be a trade-off between bias and variance, and the mean squared error of $\widehat{\beta}_1^{*}$ could be smaller than that of $\widehat{\beta}_1$ (which is unbiased).

The bias–variance trade-off was mentioned only by a few candidates.

Question 5

Consider the OLS estimator for β in the linear regression model:

$$Y_i = \beta X_i + \varepsilon_i, \quad i = 1, \ldots, n$$

where $\{(Y_i, X_i)\}_{i=1}^n$ form an i.i.d. sample from a population and the errors are drawn from an <u>unknown</u> distribution with mean zero and variance σ^2. You are told that X_i and ε_i are <u>uncorrelated</u> (not necessarily independent therefore!).

(a) Discuss the importance of convergence in probability.

(4 marks)

(b) Discuss the importance of convergence in distribution.

(4 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Sections R.14 and R.15.

Dougherty, C. Subject guide (2016): Chapter 8.

Approaching the question

(a) Convergence in probability plays an important role when proving *consistency* of estimators. A consistent estimator converges in probability to the true parameter and we either write $\text{plim}(\widehat{\beta}) = \beta$ or $\widehat{\beta} \xrightarrow{p} \beta$.

Proofs of consistency (a large sample property) are particularly important in settings where we cannot show that our estimator is unbiased (a finite sample property). For the OLS estimator:

$$\text{plim}(\widehat{\beta}) = \text{plim}(\beta) + \frac{\text{plim}(\frac{1}{n}\sum X_i\varepsilon_i)}{\text{plim}(\frac{1}{n}\sum X_i^2)} = \beta + \frac{\text{E}(X_i\varepsilon_i)}{\text{E}(X_i^2)} = \beta$$

where we use the properties of the plim operator and the Law of Large Numbers to argue that the sample averages converge in probability to their population equivalents.

(b) Convergence in distribution plays an important role in settings where we do not know the exact sampling distribution of our estimator, such as in the linear regression setting when the distribution of u is unknown. We need a distributional result in order for us to conduct hypothesis testing.

The central limit theorem allows us to obtain an approximate (asymptotic) distribution for $\widehat{\beta}$ even without knowing the true distribution of u. The asymptotic distribution provides a good approximation of the true sampling distribution as long as the sample is sufficiently large. This powerful result enables us to conduct hypothesis testing which is valid assuming our sample is large.

Section B

Answer three questions from this section.

Question 6

Let us consider how workplace smoking bans affect the incidence of smoking. Below, we use data on 10,000 US indoor workers from 1991 to 1993 taken from 'Do Workplace Smoking Bans Reduce Smoking', by Evans et al. (*American Economic Review*, 1999).

Let *smoker* be a dummy variable indicating whether a worker smokes (1 = yes, 0 = no) and *smkban* a dummy variable indicating whether there is a ban on smoking in the workplace (1 = yes, 0 = no).

(a) The following OLS regression results were obtained:

$$\widehat{smoker} = \underset{(0.007)}{0.290} - \underset{(0.009)}{0.078}\, smkban \qquad (6.1)$$

$$n = 10000,\ R^2 = 0.0078,\ RSS = 1821.59$$

The standard errors (SEs) are in parentheses. Interpret the parameter estimates of the coefficient on *smkban*. Provide the (approximate) 95% confidence interval for the coefficient on *smkban*. How can we use this confidence interval to test the hypothesis that $\beta_{smkban} = 0$?

(6 marks)

A further specification was considered that included other characteristics of the worker: the age (in years), gender (male/female), ethnicity (black/hispanic/white), and level of education ($E1$ = highschool dropout, $E2$ = highschool graduate, $E3$ = some college, $E4$ = college graduate, $E5$ = Master degree or above). The following OLS regression results were obtained for this multiple linear regression model:

$$\widehat{smoker} = \underset{(0.019)}{0.201} - \underset{(0.009)}{0.045}\, smkban - \underset{(.009)}{0.033} female - \underset{(.0003)}{0.001}\, age - \underset{(.016)}{0.027} black \qquad (6.2)$$

$$-\underset{(.014)}{0.104} hispanic + \underset{(.019)}{0.310} E1 + \underset{(.012)}{0.224} E2 + \underset{(.012)}{0.156} E3 + \underset{(.012)}{0.042} E4$$

$$n = 10000, \; R^2 = 0.0526, \; RSS = 1736.81$$

The SEs are in parentheses.

(b) Compare the coefficient estimates on *smkban* from the simple and multiple regression model in (6.1) and (6.2) and explain why the estimates differ.

(3 marks)

(c) Interpret the estimated parameter on $E2$ (highschool graduate) in (6.2) and indicate how you can obtain its p-value and what information the p-value provides.

(5 marks)

(d) Both the simple and multiple regression model suffer from heteroskedasticity. Explain why. What are the implications of heteroskedasticy for the parameter estimates and the standard errors in (6.1) and (6.2)? What can you do to resolve this problem? Explain your answer.

(6 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Sections 2.6, 3.2, 6.2, 7.3 and 10.1.

Dougherty, C. Subject guide (2016): Chapters 5, 6 and 10.

Approaching the question

(a) This is the linear probability model, where:

$$\mathrm{E}(smoker = 1 \mid X) = P(smoker = 1 \mid X) = \beta_0 + \beta_1 smkban.$$

If the smoking ban is introduced, the probability of a worker smoking decreases by 7.8 percentage points (not 7.8%).

An (approximate) 95% confidence interval is:

$$(\widehat{\beta}_1 - z_{crit,\,0.025} \times \text{s.e.}(\widehat{\beta}_1),\; \widehat{\beta}_1 + z_{crit,\,0.025} \times \text{s.e.}(\widehat{\beta}_1)$$
$$= (-0.078 - 1.96 \times 0.009,\; -0.078 + 1.96 \times 0.009) \approx (-0.096,\; -0.060)$$

where z relates to the standard normal distribution (for large degrees of freedom the t distribution converges to the standard normal).

We can use confidence intervals for testing. For a two-sided $\mathrm{H}_0 : \beta_1 = 0$ test, we reject H_0 at the 5% significance level if zero does not lie in the 95% confidence interval.

(b) The reason why the coefficient estimates on *smkban* differ in (6.1) and (6.2) is due to the *omitted variable bias problem.* When we omit relevant variables (for example, education and gender) which are related to the included variable *smkban*), then the parameter estimates on *smkban* will estimate not only the direct effect that *smkban* has on smoking (the parameter of interest) but the indirect effect of these omitted variables as well. We also call these omitted variables confounders.

The more negative effect found in (a) signals that once the effect of education on smoking (gender, age and ethnicity) is controlled for, the true effect of a smoking ban on the probability of smoking is smaller.

(c) The coefficient on $E2$ means that highschool graduates are 22.4 percentage points more likely to smoke relative to people with a Master's degree or above since $E5$ *is the left out base category.* (Many candidates failed to recognise that the left out category was required to interpret the parameter clearly). The t statistic for this effect is:

$$t = \frac{\widehat{\beta}}{\text{s.e.}(\widehat{\beta})} = \frac{0.224}{0.012} = 18.667$$

which has a p-value smaller than 0.05 (we would reject H_0 at the 5% significance level). The p-value is the lowest level of significance at which we can reject the null hypothesis. Since the sample size is large, we can use the standard normal distribution to obtain the p-value (candidates may provide a graphical discussion instead). For a one-sided test, $p = 1 - \Phi(t)$ and for a two-sided test, $p = 2 \times (1 - \Phi(t))$, assuming $t \geq 0$.

(d) In the linear probability model (LPM) we have to deal with the problem of heteroskedasticity. Conditional on X, $\text{Var}(y \mid X) = p(X)(1 - p(X)) \equiv \text{Var}(u \mid X)$. The implication of heteroskedasticity for the estimated standard errors is that they are wrong (they rely on homoskedasticity) and inference based on them would be invalid. The parameter estimates themselves remain unbiased and consistent. For inference, we should use heteroskedasticity-robust standard errors or apply WLS instead.

Question 7

In question 6 we considered the linear regression model to study how workplace smoking bans affect the incidence of smoking. Here we consider the results from applying a probit regression of *smoker* (1 = yes, 0 = no) on *smkban* (1 = yes, 0 = no) and the other explanatory variables:

```
. probit smoker smkban female age black hispanic E1 E2 E3 E4

Iteration 0:   log likelihood = -5537.1662
Iteration 1:   log likelihood = -5255.1526
Iteration 2:   log likelihood =  -5252.349
Iteration 3:   log likelihood = -5252.3489

Probit regression                               Number of obs     =     10,000
                                                LR chi2(9)        =     569.63
                                                Prob > chi2       =     0.0000
Log likelihood = -5252.3489                     Pseudo R2         =     0.0514
```

smoker	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
smkban	-.1517626	.0289268	-5.25	0.000	-.208458	-.0950671
female	-.1106249	.0287785	-3.84	0.000	-.1670298	-.05422
age	-.0042031	.0011748	-3.58	0.000	-.0065057	-.0019006
black	-.07969	.0525369	-1.52	0.129	-.1826604	.0232804
hispanic	-.3327039	.0476677	-6.98	0.000	-.4261308	-.2392769
E1	1.094231	.0714121	15.32	0.000	.9542663	1.234197
E2	.8518588	.0594747	14.32	0.000	.7352906	.9684271
E3	.6492566	.0606989	10.70	0.000	.530289	.7682241
E4	.2224747	.0649939	3.42	0.001	.0950891	.3498603
_cons	-.9842425	.0756055	-13.02	0.000	-1.132427	-.8360584

(a) **It is argued that using the probit regression model is better than using the linear probability model when explaining the binary variable *smoker*. Discuss**

the benefits/drawbacks of using the Probit model when trying to explain a binary variable.

(5 marks)

(b) **Explain briefly how the Probit estimates are obtained and discuss the properties of the parameter estimates.**

Hint: **You may recall that for the Probit model, we will specify:**

$$\mathbf{Pr}(\boldsymbol{smoker} = 1) = \boldsymbol{\Phi}(\boldsymbol{\beta_0} + \boldsymbol{\beta_1 smkban} + \boldsymbol{\beta_2 female} + \cdots + \boldsymbol{\beta_8 E3} + \boldsymbol{\beta_9 E4})$$

where $\boldsymbol{\Phi}$ is the standard normal CDF (cumulative distribution function).

(5 marks)

(c) **Explain how you can estimate the effect of the smoking ban on the probability of smoking for a 50-year old white, college graduated man. You are not expected to use your calculator, clarity of the computations required is enough.**

(5 marks)

(d) **Discuss how you could test the joint significance of the worker's characteristics (gender, age, ethnicity and level of education) using the likelihood ratio test. Clearly indicate the test statistic, its distribution, the rejection rule and the additional information you would need to implement it.**

(5 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Sections 10.3 and 10.6.

Dougherty, C. Subject guide (2016): Chapter 10.

Approaching the question

(a) The probit model has three main advantages over the linear probability model (LPM): (i) predicted probabilities are restricted to lie in $[0, 1]$, (ii) maximum likelihood estimators are (asymptotically) efficient whereas OLS (LPM) estimators will be inefficient, and (iii) maximum likelihood estimators automatically deal with heteroskedasticity.

The main drawbacks of the probit model relative to the LPM are that (i) the coefficients cannot be directly interpreted as the marginal effects of the regressor(s) of interest, so we need to compute predicted probabilities using the probit specification, and (ii) it is also computationally more complicated.

(b) The parameters are estimated by maximum likelihood estimation, where the log-likelihood function is given by:

$$\begin{aligned}\log L(\beta) &= \sum_{i=1}^{n} \{smoker_i \log(P(smoker_i = 1 \mid X) + (1 - smoker_i) \log(P(smoker_i = 0 \mid X)\} \\ &= \sum_{i=1}^{n} \{smoker_i \log(\Phi(\beta_0 + \beta_1 smkban_i + \beta_2 female_i + \cdots + \beta_9 E4_i) + \\ &\quad (1 - smoker_i) \log(1 - \Phi(\beta_0 + \beta_1 smkban_i + \beta_2 female_i + \cdots + \beta_9 E4_i))\}.\end{aligned}$$

To obtain the parameter estimates the first-order conditions are solved (numerically as no explicit formulae exist).

Under suitable regularity conditions, the estimates are consistent, asymptotically normally, and asymptotically efficient.

(c) The effect is the difference in predicted probabilities between the man with the characteristics given, with *smkban* $= 1$ versus *smkban* $= 0$:

$$\widehat{P}(y_i = 1 \mid X, smkban = 1) - \widehat{P}(y_i = 1 \mid X, smkban = 0)$$

$$= \Phi\left(\sum_k x_{ki}\widehat{\beta}_k + \widehat{\beta}_1\right) - \Phi\left(\sum_k x_{ki}\widehat{\beta}_k\right)$$

$$= \Phi(-0.984 + 0.222 - 0.004 \times 50 - 0.152) - \Phi(-0.984 + 0.222 - 0.004 \times 50).$$

A discussion of marginal effects (ignoring the fact that *smkban* is a dummy variable) is acceptable as well.

(d) We would like to use the LR test, to test:

$$\mathrm{H}_0 : \beta_j = 0 \ \forall \ j \in \{\text{worker characteristics}\}$$

$$\mathrm{H}_1 : \beta_j \neq 0 \text{ for at least one } j \in \{\text{worker characteristics}\}.$$

Estimate the restricted (R) and the unrestricted probit model (U). The restricted model imposes H_0. The unrestricted model is just the originally estimated model above. The LR test makes use of the difference between the two estimated log-likelihood functions (the log ratio of the likelihood functions). The test statistic is:

$$LR = 2(\log L^U - \log L^R) \sim \chi^2_J$$

where J denotes the degrees of freedom which is equal to the number of restrictions; that is, $J = 8$ here. For a given significance level we reject H_0 if its realisation exceeds the critical value given by the χ^2_8 distribution, which equals 15.51 for the 5% significance level.

Question 8

An economist is interested in estimating the production function for widgets which is postulated to follow a Cobb–Douglas specification:

$$\boldsymbol{Y_i = \exp(\beta_0) L_i^{\beta_L} K_i^{\beta_K} \exp(u_i)}$$

where Y_i is a measure of output for firm i; L_i is labour, K_i is capital stock and u_i is an unobserved term that captures technological or managerial efficiency and other external factors (e.g., weather). The parameters to be estimated are $(\beta_0, \beta_L, \beta_K)$. Taking logs:

$$\boldsymbol{\ln Y_i = \beta_0 + \beta_L \ln L_i + \beta_K \ln K_i + u_i.}$$

(a) Provide the interpretation of the parameter β_L. What effect, if any, will changing the units of measurement of labour have on the parameter estimates for $(\beta_0, \beta_L, \beta_K)$? Explain your answer.

(5 marks)

(b) Assume that you have a cross-section of firms and that more productive firms hire less workers (labour). Explain why OLS would not provide consistent estimates for $(\beta_0, \beta_L, \beta_K)$. Would it over- or underestimate β_L on average? Clearly explain your answer.

(5 marks)

Instead of applying OLS, the economist decides to use the average wage paid by firm i, W_i, as an instrument for the (log) quantity of labour employed by that firm, $\ln L_i$.

(c) Describe in detail how you would estimate the parameters of the production function using Two Stage Least Squares (TSLS). What restrictions would be necessary for this researcher to successfully use this instrumental variable in the estimation of the parameters $(\beta_0, \beta_L, \beta_K)$ and what would you need to assume about capital stock?

(7 marks)

(d) If average wages per firm do not vary much by firm (potentially because of unionisation or high mobility of the labour force), how would this affect the properties of the estimation procedure suggested in (c)? Explain your answer.

(3 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Sections 4.2, 6.2 and 9.3.

Dougherty, C. Subject guide (2016): Chapters 4 and 8.

Approaching the question

(a) Since the specification is in logs, β_L is the output *elasticity* with respect to labour. It measures the % change in output when labour is increased by 1%, *ceteris paribus*. Since this is a relative change measure, changing the units of measurement will not change the parameter estimate. The same holds true for β_K, the output elasticity with respect to capital. However, the intercept will be changed by a change in the units of labour. If labour is measured in new units $L_i^* = c \times L_i$, the log specification would give rise to $\ln L_i^* = \ln c \times L_i = \ln c + \ln L_i$ which means the intercept is shifted by $\beta_L \ln c$ units.

(b) The problem with using OLS on the above regression is that there will be correlation between the error u_i and the regressor $\ln L_i$ as more productive firms (higher u_i) are associated with hiring fewer workers (lower L_i). This correlation will make the parameter estimates inconsistent. This problem is a result of the *omission of relevant variables* which creates OVB. The parameter estimates for β_L will be underestimated as it captures the fact that firms with higher technological or managerial efficiency require less labour for the same output.

(c) Two-stage least squares (TSLS) proceeds in two stages. In the first stage, $\ln L_i$ is regressed on all other exogenous regressors and the instrument w. This yields the fitted values:

$$\widehat{\ln L_i} = \widehat{\pi}_0 + \widehat{\pi}_1 w_i + \widehat{\pi}_2 \ln K_i.$$

In the second stage, the original regression is run where the predicted values from the first stage above $\widehat{\ln L_i}$ are used instead of the original regressor $\ln L_i$. We run:

$$\ln Y_i = \beta_0 + \beta_L \widehat{\ln L_i} + \beta_K \ln K_i + u_i.$$

Alternatively, the second stage can be described as performing IV on the original regression where we use $\widehat{\ln L_i}$ as an instrument for $\ln L_i$.

The instrumental variable approach requires the instrument to be relevant and exogenous. Instrument relevance means that the instrument must be correlated with the original regressor, that is $\text{Cov}(\ln L_i, w_i) \neq 0$. Instrument exogeneity requires the instrument to be uncorrelated with the error term; that is, $\text{Cov}(w_i, u_i) = 0$.

For consistent parameter estimates, we require capital stock to also be exogenous, i.e. uncorrelated with the error term.

(d) Candidates should recall the variance of the IV estimator in the simple linear regression setting, revealing the importance that the instrument w is highly correlated with the endogenous regressor $\ln L$ for its precision. If there is not much variation in wages, the instrument is likely not to be very correlated with $\ln L_i$. This is a problem we also call the 'weak instrument problem' as we will have a weak first stage ($\ln L_i$ is not predicted well by the instrument and other exogenous regressors). This problem results in imprecise TSLS estimators.

Question 9

Let us consider the expectations augmented Phillips curve (see also Mankiw, 1994):

$$\mathit{infl}_t - \mathit{infl}_t^e = \beta_1(\mathit{unem}_t - \mu_0) + e_t$$

where μ_0 is the natural rate of unemployment (assumed to be constant over time) and infl_t^e is the expected rate of inflation formed in $t-1$.

This model suggests that there is a trade-off between unanticipated inflation ($\mathit{infl}_t - \mathit{infl}_t^e$) and cyclical unemployment (difference between actual unemployment and the natural rate of unemployment). We assume that e_t (also called supply shock) is an i.i.d. random variable with zero mean.

(a) You are told that expectations are formed as follows:

$$\mathit{infl}_t^e - \mathit{infl}_{t-1}^e = \lambda(\mathit{infl}_{t-1} - \mathit{infl}_{t-1}^e).$$

What name do we give such a process and how should we interpret λ?

(2 marks)

(b) Show that you can rewrite the model as:

$$\Delta \mathit{infl}_t = \gamma_0 + \gamma_1 \mathit{unem}_t + \gamma_2 \mathit{unem}_{t-1} + v_t \qquad (9.1)$$

where $\Delta \mathit{infl}_t = \mathit{infl}_t - \mathit{infl}_{t-1}$. Clearly indicate the relation between $(\gamma_0, \gamma_1, \gamma_2)$ and $(\mu_0, \beta_1, \lambda)$ and show that:

$$v_t = e_t - (1-\lambda)e_{t-1}.$$

***Hint:* If you want you may use the following shorthand notation in your derivations: $y_t = \mathit{infl}_t$, $y_t^e = \mathit{infl}_t^e$ and $x_t = \mathit{unem}_t$.**

(7 marks)

(c) Discuss what assumptions you would like to make about e_t (the supply shock) that will guarantee that the OLS estimator for the parameters in (9.1) is consistent. *Hint:* You may want to give the assumptions you need to make about v_t (composite error term) first.

(3 marks)

(d) Show how you can obtain a consistent estimator for λ using your consistent estimates for $(\gamma_0, \gamma_1, \gamma_2)$. Provide a proof of its consistency. [*Note:* If you did not manage to get an explicit relation between λ and $(\gamma_0, \gamma_1, \gamma_2)$, consider $\lambda = g(\gamma_0, \gamma_1, \gamma_2)$ where $g(\cdot)$ is some continuous function].

(3 marks)

(e) One of the assumptions provided in (c) is rather unreasonable (*Hint:* future unemployment may be related to current supply shocks). Discuss how you could use IV (TSLS) to obtain a consistent estimator for the parameters in (9.1). Discuss what conditions your instruments need to satisfy and propose a suitable instrument in this setting.

(5 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Sections R.14, 9.3, 11.3, 11.4 and 11.5.

Dougherty, C. Subject guide (2016): Chapters 8 and 11.

Approaching the question

(a) This is an adaptive expectations model. The parameter λ, which should lie between 0 and 1, indicates the speed with which expectations are adjusted in response to deviations between the actual and the expected rate of inflation in the previous period.

(b) To obtain this it is useful to realise that the equation given in part (a) can be rewritten as:

$$y_t^e = (1-\lambda)y_{t-1}^e + \lambda y_{t-1}. \tag{*}$$

This shows that if we subtract $(1-\lambda)$ times the lagged expectation augmented Phillips curve, we can get rid of the inflation expectation variable. Let us substract from the first equation given in the question, $1-\lambda$ times the equation one period lagged:

$$y_t - y_t^e - (1-\lambda)[y_{t-1} - y_{t-1}^e] = \beta_1(x_t - \mu_0) + e_t - (1-\lambda)[\beta_1(x_{t-1} - \mu_0) + e_{t-1}].$$

Rearranging yields:

$$y_t - (1-\lambda)y_{t-1} - (y_t^e - (1-\lambda)y_{t-1}^e) = \beta_1 x_t - (1-\lambda)\beta_1 x_{t-1} - \lambda\beta_1\mu_0 + e_t - (1-\lambda)e_{t-1}.$$

Using (*) we get:

$$y_t - (1-\lambda)y_{t-1} - \lambda y_{t-1} = \beta_1 x_t - (1-\lambda)\beta_1 x_{t-1} - \lambda\beta_1\mu_0 + v_t$$

which yields:

$$\Delta y_t = -\beta_1\lambda\mu_0 + \beta_1 x_t - (1-\lambda)\beta_1 x_{t-1} + v_t$$

which gives rise to the following relationship between the γs and original parameters:

$$\gamma_0 = -\beta_1\lambda\mu_0, \quad \gamma_1 = \beta_1 \quad \text{and} \quad \gamma_2 = -(1-\lambda)\beta_1.$$

(c) Consistency requires the error term in the new model v_t and both regressors, x_t and x_{t-1} to be uncorrelated. Since e_t has a zero mean this means:

$$\mathrm{E}(x_t(e_t - (1-\lambda)e_{t-1})) = 0 \quad \text{and} \quad \mathrm{E}(x_{t-1}(e_t - (1-\lambda)e_{t-1})) = 0.$$

A sufficient condition for this would be that $\mathrm{E}(e_t x_s) = 0$ for $s \in \{t-1, t, t+1\}$ or uncorrelatedness between x_t and e_s $\forall\ s, t$.

(d) The above relationships can be solved for λ as:

$$\lambda = \frac{\gamma_2}{\gamma_1} + 1.$$

Using the estimators $\widehat{\gamma}_1$ and $\widehat{\gamma}_2$, an estimator $\widehat{\lambda}$ can be constructed accordingly. For consistent estimators $\widehat{\gamma}_1$ and $\widehat{\gamma}_2$, the estimator $\widehat{\lambda}$ is also consistent by the Slutsky theorem:

$$\mathrm{plim}(\widehat{\lambda}) = \mathrm{plim}\left(\frac{\widehat{\gamma}_2}{\widehat{\gamma}_1} + 1\right) = \frac{\mathrm{plim}(\widehat{\gamma}_2)}{\mathrm{plim}(\widehat{\gamma}_1)} + 1 = \frac{\gamma_2}{\gamma_1} + 1 = \lambda.$$

Slutsky's theorem (probability limit rules) guarantees the second equality since these are continuous functions.

(e) If future unemployment is related to current supply shocks, then $\mathrm{Cov}(x_t, e_{t-1}) \neq 0$ and we get inconsistency as that would mean $\mathrm{Cov}(x_t, v_t) \neq 0$.

As long as unemployment is only related to past supply shocks, but not current or future ones x_{t-1} is still uncorrelated with v_t, $\mathrm{Cov}(x_{t-1}, v_t) = 0$.

We should proceed using an *instrumental variable approach* (2SLS), where we need to look for an instrument to deal with the endogeneity of x_t (this could for instance be a demand shock which is unrelated with supply shocks). Call this instrument z_t.

The instrument must satisfy instrument relevance and instrument exogeneity. Instrument relevance means that the instrument must be correlated with the original regressor, that is $\mathrm{Cov}(x_t, z_t) \neq 0$. Instrument exogeneity requires the instrument to be uncorrelated with the error term; that is, $\mathrm{Cov}(z_t, v_t) = 0$. Since v_t is a function of both e_t and e_{t-1}, the latter assumption is satisfied when the instrument is unrelated with any current and past supply shocks.

In this particular case, due to the fact that v_t has an MA(1) error structure, we can also use x_{t-2} as an instrument. We expect there to be correlation between x_t and x_{t-2} and x_{t-2} (the past) is uncorrelated with e_t and e_{t-1} which make up v_t (we cannot use x_{t-1} because that variable is already included in the model).

A TSLS procedure would proceed by running a regression of x_t on the instrument and the other regressors to obtain fitted values $\widehat{x}_t$. In the second stage, the outcome variable Δy_t would be regressed on an intercept, the predicted values $\widehat{x}_t$ and x_{t-1}. Alternatively, we can apply IV in the second stage where we use $\widehat{x}_t$ as our instrument for x_t. This would give rise to consistent TSLS estimators $(\gamma_0^{TSLS}, \gamma_1^{TSLS}, \gamma_2^{TSLS})$.

Question 10

This question is based on 'Capital Accumulation and Growth: A New Look at the Empirical Evidence', by Bond et al. (*Journal of Applied Econometrics*, 2010). In this article, the authors are interested in a regression model for the (logarithm of) output-per-capita y_t in a given country and time period t similar to:

$$y_t = \alpha + \rho y_{t-1} + \beta x_t + \gamma_t + \varepsilon_t, \quad |\rho| < 1 \tag{10.1}$$

where x_t is (the logarithm of) investment-per output. Imagine that investment rates are also affected by current output-per-capita, so that:

$$x_t = \varphi_0 + \varphi_1 y_t + u_t. \tag{10.2}$$

Equations (10.1) and (10.2) then form a simultaneous equation model. Both errors ε_t and u_t have zero mean.

For (a) and (b), we start by assuming that ε_t and u_t are not serially correlated.

(a) Obtain the reduced form equations for y_t and x_t. (Note: the variables t and y_{t-1} are exogenous.) Are equations (10.1) and (10.2) identified? Discuss.

(5 marks)

(b) Assume that $x_t = \varphi_0 + u_t$ (that is $\varphi_1 = 0$). Under what conditions is the OLS estimator for the parameters in the structural equation for output-per-capita (i.e., equation (10.1)) consistent? Discuss.

(5 marks)

(c) Assume that $x_t = \varphi_0 + u_t$ (that is $\varphi_1 = 0$), u_t is not serially correlated, and $\gamma = 0$. How would you test whether ε_t is serially correlated?

(5 marks)

(d) Assume that $x_t = \varphi_0 + u_t$ (that is $\varphi_1 = 0$), $\rho = 0$ and $\gamma = 0$. Using the Dickey–Fuller test, the authors fail to reject the hypothesis of a unit root in y_t and x_t for most of the countries in their sample at usual significance levels (5%). (Note: we are no longer assuming ε_t or u_t to not be serially correlated). Explaining your answers, what can you say about the OLS estimator for α and β applied to equation (10.1):

i. if ε_t is stationary and weakly dependent?

ii. if ε_t is non-stationary and strongly dependent?

(5 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Sections 9.1, 12.2, 13.2 and 13.6.

Dougherty, C. Subject guide (2016): Chapters 9 and 13.

Approaching the question

(a) Substituting x_t and y_t in appropriately gives the following reduced forms:

$$y_t = \frac{\alpha + \beta\varphi_0}{1-\beta\varphi_1} + \frac{\rho}{1-\beta\varphi_1}y_{t-1} + \frac{\gamma}{1-\beta\varphi_1}t + \frac{\beta u_t + \varepsilon_t}{1-\beta\varphi_1}$$
$$x_t = \frac{\varphi_0 + \varphi_1\alpha}{1-\beta\varphi_1} + \frac{\rho\varphi_1}{1-\beta\varphi_1}y_{t-1} + \frac{\varphi_1\gamma}{1-\beta\varphi_1}t + \frac{\varphi_1\varepsilon_t + u_t}{1-\beta\varphi_1}.$$

Equation (10.2) is identified since the exogenous variables t and y_{t-1} can be used as instruments for the single endogenous variable y_t in this equation. Both exogenous variables are excluded from (10.2). This equation is overidentified. Equation (10.1), though, is not identified since there is an instrument we can use for the single endogenous variable x_t in this equation. There are no exclusions in (10.1), hence this equation is underidentified.

(b) Consistency of the OLS estimator of (10.1) requires that the error ε_t is uncorrelated with the regressors y_{t-1}, x_t, t. We do not need to worry about correlation between ε_t and fixed regressors such at t. If $x_t = \varphi_0 + u_t$, uncorrelatedness between x_t and ε_t requires uncorrelatedness between ε_t and u_t which is given in the question. To ensure that y_{t-1} is uncorrelated with ε_t finally we require the absence of autocorrelation in ε_t which is true by assumption.

(c) We are asked to test for serial correlation in the error ε_t, from the following regression model:

$$y_t = \alpha + \rho y_{t-1} + \beta x_t + \varepsilon_t, \quad |\rho| < 1.$$

Say we postulate an AR(1) process for ε_t we test H_0: no autocorrelation against H_1: AR(1) presence of autocorrelation. We should apply the Breusch–Godfrey test (the Durbin–Watson test cannot be applied because of the presence of the lagged endogenous variable).

Using the OLS residuals, we run the following regression:

$$\widehat{\varepsilon}_t = \gamma_0 + \gamma_1 y_{t-1} + \gamma_2 x_t + \delta\widehat{\varepsilon}_{t-1} + v_t.$$

The test statistic is given by nR^2 of this regression, where n is the sample size. Under the null hypothesis its asymptotic distribution is χ^2_1. We reject H_0 if the realisation exceeds the critical value given by this distribution for a given significance level.

(d) We are asked to consider the following model:

$$y_t = \alpha + \beta x_t + \varepsilon_t$$

in the setting where both y_t and x_t are expected to be integrated of order one.

i. If ε_t is non-stationary and strongly dependent, y_t and x_t are unrelated. Nevertheless, running the regression (10.1) may still give significant parameter estimates since we face the case of spurious regression, wrongly suggesting that there is a relationship between y_t and x_t. This is because the estimates are inconsistent. This is the spurious regression problem.

ii. If ε_t is stationary and weakly dependent, then the two unit root processes y_t and x_t must be cointegrated. This means the parameters α and β can be consistently estimated by OLS.

Examiners' commentaries 2019

EC2020 Elements of econometrics

Important note

This commentary reflects the examination and assessment arrangements for this course in the academic year 2018–19. The format and structure of the examination may change in future years, and any such changes will be publicised on the virtual learning environment (VLE).

Information about the subject guide and the Essential reading references

Unless otherwise stated, all cross-references will be to the latest version of the subject guide (2016). You should always attempt to use the most recent edition of any Essential reading textbook, even if the commentary and/or online reading list and/or subject guide refer to an earlier edition. If different editions of Essential reading are listed, please check the VLE for reading supplements – if none are available, please use the contents list and index of the new edition to find the relevant section.

Comments on specific questions – Zone B

Candidates should answer **EIGHT** of the following **TEN** questions: **ALL** of the questions in Section A (8 marks each) and **THREE** questions from Section B (20 marks each). **Candidates are strongly advised to divide their time accordingly.**

Section A

Answer all questions from this section.

Question 1

Consider the simple linear regression model:

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

We assume that the errors $\{u_i\}_{i=1}^n$ are independent normal random variables with zero mean. The regressor $\{X_i\}_{i=1}^n$ is non-stochastic (fixed under repeated sampling). You suspect that the errors exhibit heteroskedasticity.

(a) **Explain what we mean by the concept of heteroskedasticity. Enhance your answer with the help of a graphical illustration.**

(3 marks)

(b) **Derive the variance of the OLS estimator for $\widehat{\beta}$ in the presence of heteroskedastity and explain how a robust (White) standard error for $\widehat{\beta}$ can be obtained.**

(5 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Sections 7.1 and 7.3.

Dougherty, C. Subject guide (2016): Chapter 7.

Approaching the question

(a) Heteroskedasticity means that the error variance depends on the regressors, it violates the assumption that $\mathrm{E}(u_i^2) = \sigma^2$.

A graphical illustration should show data points and some depiction of the variance – for example, drawing the distribution for the errors at different levels of X. The graph should be clearly labelled.

(b) The variance of the estimator $\widehat{\beta}$ depends on X_i and, given the unbiasedness, is given by:

$$\mathrm{Var}(\widehat{\beta}) = \mathrm{E}\left((\widehat{\beta} - \beta)^2\right) = \mathrm{E}\left(\frac{\left(\sum_i X_i u_i\right)^2}{\left(\sum_i X_i^2\right)^2}\right) = \frac{\sum_i X_i^2 \mathrm{E}(u_i^2)}{\left(\sum_i X_i^2\right)^2} = \frac{\sum_i X_i^2 \sigma_i^2}{\left(\sum_i X_i^2\right)^2}$$

where σ_i^2 is the variance of the error term which depends on i and the second equality follows from fixed X in repeated samples and independence of u_i across observations giving $\mathrm{E}(u_i u_j) = 0 \ \forall\ i \neq j$. Many candidates struggled to derive the variance.

The White robust standard error for $\widehat{\beta}$ is the square root of the estimate of the above variance. The estimate would be constructed using the residuals from an OLS regression on the original model:

$$\widehat{\mathrm{Var}}(\widehat{\beta}) = \frac{\sum_i X_i^2 \widehat{u}_i^2}{\left(\sum_i X_i^2\right)^2}.$$

Question 2

Consider the human capital earnings function given by:

$$\boldsymbol{earnings_i = \beta_0 + \beta_1 educ_i + \beta_2 exper_i + \beta_3 exper_i^2 + u_i, \quad i = 1, \ldots, n}$$

where *earnings* denotes the hourly earnings of an individual and *educ* and *exper* denote the years of schooling and experience, respectively. We assume we have obtained a random sample $\{(earnings_i, educ_i, exper_i)\}_{i=1}^n$ from the population. The errors $\{u_i\}_{i=1}^n$ are i.i.d. normal random variables with zero mean and variance σ^2. We assume independence between the errors and regressors (i.e., we ignore the usual ability bias problem).

(a) Discuss the rationale for including both *exper* and *exper*^{2} in this model. In your answer be explicit about the expected signs for β_2 and β_3.

(2 marks)

(b) Discuss briefly how would you test that *exper* has a significant effect on *earnings*?

(2 marks)

(c) Provide a clear interpretation of the parameter β_1. You are told that you can obtain the estimator for β_1 by running the regression:

$$\boldsymbol{earnings_i^* = \beta_1 educ_i^* + e_i, \quad i = 1, \ldots, n}$$

where $earnings_i^*$ and $educ_i^*$ are obtained from running a regression of *earnings* (and *educ*) on an intercept, *exper* and *exper*^{2}. Explain this statement.

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Chapter 3.2 and 3.5.

Dougherty, C. Subject guide (2016): Chapter 3.

Approaching the question

(a) The rationale is to allow for diminishing returns to experience, β_2 positive and β_3 negative. For other values of the βs a different explanation should be given.

(b) Since both β_2 and β_3 capture experience, such a test needs to be a joint test with the hypotheses:

$$\mathrm{H}_0 : \beta_2 = \beta_3 = 0 \quad \text{vs.} \quad \mathrm{H}_1 : \beta_2 \neq 0 \text{ and/or } \beta_3 \neq 0.$$

The F test can be used for such a joint hypothesis with the critical values on the $F_{2,\,n-4}$ distribution.

Many candidates failed to recognise that this needs to be a joint test since we have two different regressors which capture experience.

(c) The parameter β_1 is the marginal effect of education on earnings; that is, it measures by how many (monetary) units hourly earnings would increase as a response to someone going to school for one more year, *ceteris paribus*.

The parameter estimate in the original model can also be obtained by the given equation where the starred variables denote the residuals from running a regression of the respective variable on the left-out intercept and regressors *exper* and *exper*^{2}. For both *earnings* and *educ*, this would isolate the remaining variation in the variables after controlling for the intercept, *exper* and *exper*^{2}. Hence this procedure in two steps is the same as controlling for the three left-out variables in the original regression equation. The is the Frisch–Waugh–Lovell result (regression anatomy).

Question 3

Consider the OLS estimator for β in the linear regression model:

$$Y_i = \beta X_i + \varepsilon_i, \quad i = 1, \ldots, n$$

where $\{(Y_i, X_i)\}_{i=1}^n$ form an i.i.d. sample from a population and the errors are drawn from an unknown distribution with mean zero and variance σ^2. You are told that X_i and ε_i are uncorrelated (not necessarily independent therefore!).

(a) Discuss the importance of convergence in probability.

(4 marks)

(b) Discuss the importance of convergence in distribution.

(4 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Sections R.14 and R.15.

Dougherty, C. Subject guide (2016): Chapter 8.

Approaching the question

(a) Convergence in probability plays an important role when proving *consistency* of estimators. A consistent estimator converges in probability to the true parameter and we either write $\text{plim}(\widehat{\beta}) = \beta$ or $\widehat{\beta} \xrightarrow{p} \beta$.

Proofs of consistency (a large sample property) are particularly important in settings where we cannot show that our estimator is unbiased (a finite sample property). For the OLS estimator:

$$\text{plim}(\widehat{\beta}) = \text{plim}(\beta) + \frac{\text{plim}(\frac{1}{n}\sum X_i\varepsilon_i)}{\text{plim}(\frac{1}{n}\sum X_i^2)} = \beta + \frac{\text{E}(X_i\varepsilon_i)}{\text{E}(X_i^2)} = \beta$$

where we use the properties of the plim operator and the Law of Large Numbers to argue that the sample averages converge in probability to their population equivalents.

(b) Convergence in distribution plays an important role in settings where we do not know the exact sampling distribution of our estimator, such as in the linear regression setting when the distribution of u is unknown. We need a distributional result in order for us to conduct hypothesis testing.

The central limit theorem allows us to obtain an approximate (asymptotic) distribution for $\widehat{\beta}$ even without knowing the true distribution of u. The asymptotic distribution provides a good approximation of the true sampling distribution as long as the sample is sufficiently large. This powerful result enables us to conduct hypothesis testing which is valid assuming our sample is large.

Question 4

Consider the following regression model:

$$\boldsymbol{Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t, \quad t = 1, \ldots, T.}$$

Both $\{Y_t\}_{t=1}^T$ and $\{X_t\}_{t=1}^T$ are difference stationary processes.

(a) Discuss the concept 'difference stationarity' and contrast it to the concept 'trend stationarity'. In your answer make sure you also explain what stationarity means.

(4 marks)

(b) How can you test whether the above relation is spurious or cointegrating? In your answer you are expected to explain the difference between a spurious and cointegrating relationship.

(4 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Sections 13.1 and 13.4.

Dougherty, C. Subject guide (2016): Chapter 13.

Approaching the question

(a) *Trend stationarity* relates to a process which has a deterministic trend ($y_t = \alpha + \beta t + \varepsilon_t$ where ε_t is weakly dependent). It is a type of non-stationarity which can be removed by detrending. Once we include a trend in the regression, we can use standard statistical inference methods: the errors (and, therefore, Y_t) will be weakly dependent and consistent estimates can be obtained.

Difference stationarity relates to a process which can be made stationary by taking first differences: ΔY_t, examples are random walk and random walk with drift. Here Y_t is persistent, strongly dependent, and estimation of the process Y_t on Y_{t-1} would fail to produce consistent parameter estimates.

By stationarity we mean *covariance stationarity*, which requires constant (finite) means and variances, and covariances which only depend on distance in time, i.e. $\text{E}(Y_t, Y_{t-s})$ only depends on s, not on location t.

(b) If Y_t and X_t are cointegrated, the error term in this model will be covariance stationary and consistent estimates of β_0 and β_1 can be obtained. The existence of a cointegrating relationship indicates the presence of a long-run relationship (existence of an ECM). If the error term has a unit root, however, the problem of a spurious relation will arise (the relationship is meaningless). Even if Y_t and X_t are independent, estimates of β_0 and β_1 may show up significant and suggest there would be a relationship. This is because a unit root error term implies that the parameters cannot consistently be estimated and we cannot use the usual distribution to obtain critical values.

A test which allows us to distinguish between the two cases would be a test whether the error term ε_t has a unit root. We can conduct this by performing a *Dickey–Fuller test* on the differenced residuals $\Delta\widehat{\varepsilon}_t$.

Question 5

Consider the simple linear regression model:

$$\boldsymbol{Y_i = \beta_0 + \beta_1 X_i + u_i, \quad i = 1, \ldots, n.}$$

We assume that the errors $\{u_i\}_{i=1}^n$ are independent random variables with zero mean. The regressor $\{X_i\}_{i=1}^n$ is non-stochastic (fixed under repeated sampling). Under these conditions, the OLS estimator for β_1, $\widehat{\beta}_1$, is unbiased. (You are not asked to derive $\widehat{\beta}_1$).

(a) Explain the concept of unbiasedness of an estimator.

(2 marks)

(b) Let us consider two other estimators for the slope β_1:

$$\widehat{\beta}_1^{\circ} = \frac{\sum_{i=1}^{n}(Z_i - \bar{Z})Y_i}{\sum_{i=1}^{n}(Z_i - \bar{Z})X_i} \quad \text{and} \quad \widehat{\beta}_1^{*} = \frac{\sum_{i=1}^{n}(Z_i - \bar{Z})Y_i}{\sum_{i=1}^{n}(Z_i - \bar{Z})Z_i}$$

where $Z_i = \sqrt{X_i}$ for all i and $\bar{Z} = \frac{1}{n}\sum_{i=1}^{n} Z_i$. Please indicate whether $\widehat{\beta}_1^{\circ}$ and $\widehat{\beta}_1^{*}$ are unbiased estimators for β_1. Clearly show your derivations.

(4 marks)

(c) Briefly indicate how you would choose between the three estimators, $\widehat{\beta}_1$, $\widehat{\beta}_1^{\circ}$ and $\widehat{\beta}_1^{*}$.

(2 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Sections 2.3 and 2.5.

Dougherty, C. Subject guide (2016): Chapter 2.

Approaching the question

(a) Unbiasedness means $\mathrm{E}(\widehat{\beta}_1) = \beta_1$. It means that the expected value of the estimator is the true parameter; that is, we are correct on average in repeated samples. This ensures that we will not make systematic errors when estimating β.

(b) In answering this question, it is easiest to derive results using Z_i rather than working with $Z_i = \sqrt{X_i}$. It is important to note furthermore that $\sum(Z_i - \bar{Z}) = 0$.

When plugging in the true model ($Y_i = \beta_0 + \beta_1 X_i + u_i$), we can then write:

$$\widehat{\beta}_1^\circ = \beta_1 + \frac{\sum(Z_i - \bar{Z})u_i}{\sum(Z_i - \bar{Z})X_i}$$

and:

$$\mathrm{E}(\widehat{\beta}_1^\circ) = \beta_1 + \frac{\sum(Z_i - \bar{Z})\mathrm{E}(u_i)}{\sum(Z_i - \bar{Z})X_i} = \beta_1$$

(X and Z are assumed to be non-stochastic). Unbiasedness follows as $\mathrm{E}(u_i) = 0$.

For the second estimator, when plugging in the true model, we can write:

$$\widehat{\beta}_1^* = \beta_1 + \frac{\sum(Z_i - \bar{Z})X_i}{\sum(Z_i - \bar{Z})Z_i} + \frac{\sum(Z_i - \bar{Z})u_i}{\sum(Z_i - \bar{Z})Z_i}$$

and:

$$\mathrm{E}(\widehat{\beta}_1^*) = \beta_1 + \frac{\sum(Z_i - \bar{Z})Z_i^2}{\sum(Z_i - \bar{Z})Z_i} \neq \beta_1$$

using again the fact that Z is non-stochastic and $\mathrm{E}(u_i) = 0$, hence $\widehat{\beta}_1^*$ is biased.

Side notes: (i) $\widehat{\beta}_1^\circ$ essentially is an IV estimator where Z is used as an instrument for X. Both Z and X are exogenous here. (ii) $\widehat{\beta}_1^*$, on the other hand, is the slope parameter of a regression of Z instead of X on Y, and hence should not give us an unbiased estimator of β_1.

Many candidates failed to realise that $\sum(Z_i - \bar{Z}) = 0$ and wrongly concluded that the first estimator $\widehat{\beta}_1^\circ$ is biased.

(c) When choosing between $\widehat{\beta}_1$ and $\widehat{\beta}_1^\circ$ (both unbiased), efficiency considerations matter. Since the model satisfies the Gauss–Markov assumptions, we know that by the Gauss–Markov theorem, $\widehat{\beta}_1$ is BLUE and should, therefore, be chosen over $\widehat{\beta}_1^\circ$. Nevertheless, there may be a trade-off between bias and variance, and the mean squared error of $\widehat{\beta}_1^*$ could be smaller than that of $\widehat{\beta}_1$ (which is unbiased).

The bias–variance trade-off was mentioned only by a few candidates.

Section B

Answer three questions from this section.

Question 6

Let us consider the expectations augmented Phillips curve (see also Mankiw, 1994):

$$\boldsymbol{infl_t - infl_t^e = \beta_1(unem_t - \mu_0) + e_t}$$

where $\boldsymbol{\mu_0}$ is the natural rate of unemployment (assumed to be constant over time) and $\boldsymbol{infl_t^e}$ is the expected rate of inflation formed in $\boldsymbol{t-1}$.

This model suggests that there is a trade-off between unanticipated inflation ($\boldsymbol{infl_t - infl_t^e}$) and cyclical unemployment (difference between actual unemployment and the natural rate of unemployment). We assume that $\boldsymbol{e_t}$ (also called supply shock) is an i.i.d. random variable with zero mean.

(a) You are told that expectations are formed as follows:

$$\boldsymbol{infl_t^e - infl_{t-1}^e = \lambda(infl_{t-1} - infl_{t-1}^e).}$$

What name do we give such a process and how should we interpret $\boldsymbol{\lambda}$?

(2 marks)

(b) Show that you can rewrite the model as:

$$\Delta \mathit{infl}_t = \gamma_0 + \gamma_1 \mathit{unem}_t + \gamma_2 \mathit{unem}_{t-1} + v_t \qquad (6.1)$$

where $\Delta \mathit{infl}_t = \mathit{infl}_t - \mathit{infl}_{t-1}$. Clearly indicate the relation between $(\gamma_0, \gamma_1, \gamma_2)$ and $(\mu_0, \beta_1, \lambda)$ and show that:

$$v_t = e_t - (1 - \lambda)e_{t-1}.$$

***Hint:* If you want you may use the following shorthand notation in your derivations: $y_t = \mathit{infl}_t$, $y_t^e = \mathit{infl}_t^e$ and $x_t = \mathit{unem}_t$.**

(7 marks)

(c) Discuss what assumptions you would like to make about e_t (the supply shock) that will guarantee that the OLS estimator for the parameters in (6.1) is consistent. *Hint:* You may want to give the assumptions you need to make about v_t (composite error term) first.

(3 marks)

(d) Show how you can obtain a consistent estimator for λ using your consistent estimates for $(\gamma_0, \gamma_1, \gamma_2)$. Provide a proof of its consistency. [*Note:* If you did not manage to get an explicit relation between λ and $(\gamma_0, \gamma_1, \gamma_2)$, consider $\lambda = g(\gamma_0, \gamma_1, \gamma_2)$ where $g(\cdot)$ is some continuous function].

(3 marks)

(e) One of the assumptions provided in (c) is rather unreasonable (*Hint:* future unemployment may be related to current supply shocks). Discuss how you could use IV (TSLS) to obtain a consistent estimator for the parameters in (6.1). Discuss what conditions your instruments need to satisfy and propose a suitable instrument.

(5 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Sections R.14, 9.3, 11.3, 11.4 and 11.5.

Dougherty, C. Subject guide (2016): Chapters 8 and 11.

Approaching the question

(a) This is an adaptive expectations model. The parameter λ, which should lie between 0 and 1, indicates the speed with which expectations are adjusted in response to deviations between the actual and the expected rate of inflation in the previous period.

(b) To obtain this it is useful to realise that the equation given in part (a) can be rewritten as:

$$y_t^e = (1 - \lambda)y_{t-1}^e + \lambda y_{t-1}. \qquad (*)$$

This shows that if we subtract $(1 - \lambda)$ times the lagged expectation augmented Phillips curve, we can get rid of the inflation expectation variable. Let us substract from the first equation given in the question, $1 - \lambda$ times the equation one period lagged:

$$y_t - y_t^e - (1 - \lambda)[y_{t-1} - y_{t-1}^e] = \beta_1(x_t - \mu_0) + e_t - (1 - \lambda)[\beta_1(x_{t-1} - \mu_0) + e_{t-1}].$$

Rearranging yields:

$$y_t - (1 - \lambda)y_{t-1} - (y_t^e - (1 - \lambda)y_{t-1}^e) = \beta_1 x_t - (1 - \lambda)\beta_1 x_{t-1} - \lambda\beta_1\mu_0 + e_t - (1 - \lambda)e_{t-1}.$$

Using (*) we get:

$$y_t - (1 - \lambda)y_{t-1} - \lambda y_{t-1} = \beta_1 x_t - (1 - \lambda)\beta_1 x_{t-1} - \lambda\beta_1\mu_0 + v_t$$

which yields:

$$\Delta y_t = -\beta_1 \lambda \mu_0 + \beta_1 x_t - (1-\lambda)\beta_1 x_{t-1} + v_t$$

which gives rise to the following relationship between the γs and original parameters:

$$\gamma_0 = -\beta_1 \lambda \mu_0, \quad \gamma_1 = \beta_1 \quad \text{and} \quad \gamma_2 = -(1-\lambda)\beta_1.$$

(c) Consistency requires the error term in the new model v_t and both regressors, x_t and x_{t-1} to be uncorrelated. Since e_t has a zero mean this means:

$$\text{E}(x_t(e_t - (1-\lambda)e_{t-1})) = 0 \quad \text{and} \quad \text{E}(x_{t-1}(e_t - (1-\lambda)e_{t-1})) = 0.$$

A sufficient condition for this would be that $\text{E}(e_t x_s) = 0$ for $s \in \{t-1, t, t+1\}$ or uncorrelatedness between x_t and e_s $\forall\ s, t$.

(d) The above relationships can be solved for λ as:

$$\lambda = \frac{\gamma_2}{\gamma_1} + 1.$$

Using the estimators $\widehat{\gamma}_1$ and $\widehat{\gamma}_2$, an estimator $\widehat{\lambda}$ can be constructed accordingly. For consistent estimators $\widehat{\gamma}_1$ and $\widehat{\gamma}_2$, the estimator $\widehat{\lambda}$ is also consistent by the Slutsky theorem:

$$\text{plim}(\widehat{\lambda}) = \text{plim}\left(\frac{\widehat{\gamma}_2}{\widehat{\gamma}_1} + 1\right) = \frac{\text{plim}(\widehat{\gamma}_2)}{\text{plim}(\widehat{\gamma}_1)} + 1 = \frac{\gamma_2}{\gamma_1} + 1 = \lambda.$$

Slutsky's theorem (probability limit rules) guarantees the second equality since these are continuous functions.

(e) If future unemployment is related to current supply shocks, then $\text{Cov}(x_t, e_{t-1}) \neq 0$ and we get inconsistency as that would mean $\text{Cov}(x_t, v_t) \neq 0$.

As long as unemployment is only related to past supply shocks, but not current or future ones x_{t-1} is still uncorrelated with v_t, $\text{Cov}(x_{t-1}, v_t) = 0$.

We should proceed using an *instrumental variable approach* (2SLS), where we need to look for an instrument to deal with the endogeneity of x_t (this could for instance be a demand shock which is unrelated with supply shocks). Call this instrument z_t.

The instrument must satisfy instrument relevance and instrument exogeneity. Instrument relevance means that the instrument must be correlated with the original regressor, that is $\text{Cov}(x_t, z_t) \neq 0$. Instrument exogeneity requires the instrument to be uncorrelated with the error term; that is, $\text{Cov}(z_t, v_t) = 0$. Since v_t is a function of both e_t and e_{t-1}, the latter assumption is satisfied when the instrument is unrelated with any current and past supply shocks.

In this particular case, due to the fact that v_t has an MA(1) error structure, we can also use x_{t-2} as an instrument. We expect there to be correlation between x_t and x_{t-2} and x_{t-2} (the past) is uncorrelated with e_t and e_{t-1} which make up v_t (we cannot use x_{t-1} because that variable is already included in the model).

A TSLS procedure would proceed by running a regression of x_t on the instrument and the other regressors to obtain fitted values $\widehat{x}_t$. In the second stage, the outcome variable Δy_t would be regressed on an intercept, the predicted values $\widehat{x}_t$ and x_{t-1}. Alternatively, we can apply IV in the second stage where we use $\widehat{x}_t$ as our instrument for x_t. This would give rise to consistent TSLS estimators $(\gamma_0^{TSLS}, \gamma_1^{TSLS}, \gamma_2^{TSLS})$.

Question 7

This question is based on 'Openness and Inflation: Theory and Evidence', by Romer (*Quarterly Journal of Economics*, 1993). Romer proposed theoretical models of inflation that imply that more 'open' countries should have lower inflation. A simple macroeconomic model is:

$$infl_i = \beta_0 + \beta_1 open_i + \beta_2 \ln(pcinc_i) + u_{1i} \quad (7.1)$$

$$open_i = \alpha_0 + \alpha_1 infl_i + \alpha_2 \ln(pcinc_i) + \alpha_3 \ln(land_i) + u_{2i}, \quad i = 1, \ldots, n \quad (7.2)$$

where *infl* is the average annual inflation rate (since 1973), *open* is the average share of imports in gross domestic (or national) product since 1973, ln(*pcinc*) is the log of per capita income in US dollars, and ln(*land*) is the log of land area in square miles. The variables ln(*pcinc*) and ln(*land*) are treated as exogenous variables; u_{1i} and u_{2i} are serially uncorrelated disturbances with zero mean, variances σ_1^2 and σ_2^2 and covariance σ_{12}.

(a) Explain the concepts of endogenous versus exogenous explanatory variables and show that *open* is an endogenous variable in (7.1).

(5 marks)

(b) Discuss the identification of each structural form equation and explain what it means to say that an equation is exact identified.

(5 marks)

The table below shows the OLS and IV estimation results for equation (7.1). Standard errors are reported in parenthesis.

	OLS	IV
constant	25.109	26.90
	(15.205)	(15.40)
open	−0.215	−0.337
	(0.095)	(0.144)
ln(*pcinc*)	0.018	0.376
	1.975	(2.015)

The IV estimator is obtained by solving the following three conditions (no need to show!):

$$\sum_{i=1}^{n} \left(infl_i - \widehat{\beta}_0^{IV} - \widehat{\beta}_1^{IV} open_i - \widehat{\beta}_1^{IV} \ln(pcinc_i) \right) = 0$$

$$\sum_{i=1}^{n} \left(infl_i - \widehat{\beta}_0^{IV} - \widehat{\beta}_1^{IV} open_i - \widehat{\beta}_1^{IV} \ln(pcinc_i) \right) \ln(pcinc_i) = 0$$

$$\sum_{i=1}^{n} \left(infl_i - \widehat{\beta}_0^{IV} - \widehat{\beta}_1^{IV} open_i - \widehat{\beta}_1^{IV} \ln(pcinc_i) \right) \ln(land_i) = 0.$$

(c) Describe in detail how you would estimate the parameters of (7.1) using Two Stage Least Squares (TSLS). How would these estimates compare with the reported IV estimates?

(5 marks)

(d) Conduct a test to see whether you can find evidence that more 'open' countries have lower inflation. Clearly indicate the null and alternative hypothesis, test statistic and its (asymptotic) distribution under the null. Briefly indicate how you could test whether the OLS and IV results are significantly different (so that it matters whether you use the OLS or IV parameter estimates).

(5 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Sections 2.6, 9.1 and 9.3.

Dougherty, C. Subject guide (2016): Chapters 8 and 9.

Approaching the question

(a) Exogenous variables are uncorrelated with the error term ('outside of the model') while endogenous variables are correlated with the error term ('inside the model'). This has implications for the estimates of the coefficients of these variables. While we can obtain consistent estimates under weak exogeneity and unbiased estimates under strong exogeneity, endogeneity implies that the estimates are generally inconsistent and biased. Since we face a simultaneity problem between $infl_i$ and $open_i$ in this problem, both will be endogenous. To show that $open_i$ is an endogenous variable in (7.1), i.e. that $\text{Cov}(open_i, u_{1i}) \neq 0$, candidates should first derive the reduced form for $open_i$, which is:

$$\begin{aligned} open_i &= \alpha_0 + \alpha_1(\beta_0 + \beta_1 open_i + \beta_2 \ln(pcinc_i) + u_{1i}) + \alpha_2 \ln(pcinc_i) \\ &\quad + \alpha_3 \ln(land_i) + u_{2i} \\ (1 - \alpha_1\beta_1) open_i &= (\alpha_0 + \alpha_1\beta_0) + (\alpha_1\beta_2 + \alpha_2) \ln(pcinc_i) + \alpha_3 \ln(land_i) + \alpha_1 u_{1i} + u_{2i} \\ open_i &= \frac{\alpha_0 + \alpha_1\beta_0}{1 - \alpha_1\beta_1} + \frac{\alpha_1\beta_2 + \alpha_2}{1 - \alpha_1\beta_1} \ln(pcinc_i) + \frac{\alpha_3}{1 - \alpha_1\beta_1} \ln(land_i) + \frac{\alpha_1 u_{1i} + u_{2i}}{1 - \alpha_1\beta_1}. \end{aligned}$$

Using the exogeneity of $\ln(pcinc_i)$ and $\ln(land_i)$ we can then show:

$$\text{Cov}(open_i, u_{1i}) = \text{Cov}\left(\frac{\alpha_1 u_{1i} + u_{2i}}{1 - \alpha_1\beta_1}, u_{1i}\right) = \frac{\alpha_1\sigma_1^2 + \sigma_{12}}{1 - \alpha_1\beta_1} \neq 0 \quad \text{generally.}$$

(b) Identification of equation (7.1) requires that we have at least one variable which we can use for the single endogenous variable in this equation $open_i$ $(G - 1)$. Here we have exactly one such variable $\ln(land_i)$ (R) which gives us exact identification $(G - 1 = R)$. The reduced form reveals the relevance of this instrument, the validity of the instrument is by assumption, and finally the equation is excluded from (7.1) itself. Hence (7.1) is exactly identified. In equation (7.2) there is also a single $(G - 1)$ endogenous variable, $infl_i$, for which at least one instrument is required. Unfortunately, there is no exogenous variable excluded from (7.2) which we can use. Hence (7.2) is underidentified $(R < G - 1)$.

(c) Using $land_i$ as an instrument for $open_i$, we would proceed in two stages. In the first stage, we regress $open_i$ on all remaining exogenous regressors in the original specification and the instrument to obtain fitted values of $open_i$:

$$\widehat{open}_i = \widehat{\pi}_0 + \widehat{\pi}_1 \ln(pcinc_i) + \widehat{\pi}_2 \ln(land_i).$$

For the second stage, we regress the outcome variable $infl_i$ on the original exogenous regressors and the fitted values of $open_i$:

$$infl_i = \beta_0 + \beta_1 \widehat{open}_i + \beta_2 \ln(pcinc_i) + u_{1i}.$$

Alternatively, the second stage applies IV to the original equation using the fitted values of $open_i$ as an instrument for $open_i$.

Under exact identification, the two-stage least squares estimator coincides with the IV estimator which uses the moment conditions, $\widehat{\beta}_1^{TSLS} = \widehat{\beta}_1^{IV}$.

(d) This is a one-sided test on whether the IV estimator of β_1 is significantly smaller than zero or not, that is:

$$\text{H}_0 : \beta_1 = 0 \quad \text{vs.} \quad \text{H}_1 : \beta_1 < 0.$$

The test uses the test statistic:

$$T = \frac{\widehat{\beta}_1^{IV}}{\text{s.e.}(\widehat{\beta}_1^{IV})} \xrightarrow{d} N(0, 1) \text{ under H}_0.$$

We would evaluate the test statistic using the estimates above to obtain:

$$t = \frac{-0.337}{0.144} = -2.34$$

and we should reject H_0 if it takes values lower than the critical value of the standard normal distribution, -1.645 at the 5% significance level (one-sided). We reject H_0 if $t < z$ and fail to reject otherwise. Here, we would reject H_0. This test only works (approximately) for large samples since we are using the asymptotic distribution of the test statistic (IV estimators only have nice large sample properties).

We should conduct a Hausman test to detect whether the OLS and IV estimates differ, $H_0 : \text{Cov}(open_i, u_{1i}) = 0$ against $H_1 : \text{Cov}(open_i, u_{1i}) \neq 0$. Under the null hypothesis both IV and OLS are consistent, whereas under the alternative only IV is consistent. The test is:

$$\left(\widehat{\beta}^{OLS} - \widehat{\beta}^{IV}\right)' V\left(\widehat{\beta}^{OLS} - \widehat{\beta}^{IV}\right)^{-1}\left(\widehat{\beta}^{OLS} - \widehat{\beta}^{IV}\right) \xrightarrow{d} \chi^2_3 \text{ under } H_0.$$

For a given significance level we reject H_0 if the test statistic takes values which are too large relative to the critical values given by the χ^2_3 distribution.

Question 8

Let us consider how workplace smoking bans affect the incidence of smoking. Below, we use data on 10,000 US indoor workers from 1991 to 1993 taken from 'Do Workplace Smoking Bans Reduce Smoking', by Evans et al. (*American Economic Review*, 1999).

Let *smoker* be a dummy variable indicating whether a worker smokes (1 = yes, 0 = no) and *smkban* a dummy variable indicating whether there is a ban on smoking in the workplace (1 = yes, 0 = no).

(a) The following OLS regression results were obtained:

$$\widehat{smoker} = \underset{(0.007)}{0.290} - \underset{(0.009)}{0.078}\, smkban \tag{8.1}$$

$$n = 10000,\ R^2 = 0.0078,\ RSS = 1821.59$$

The standard errors (SEs) are in parentheses. Interpret the parameter estimates of the coefficient on *smkban*. Provide the (approximate) 95% confidence interval for the coefficient on *smkban*. How can we use this confidence interval to test the hypothesis that $\beta_{smkban} = 0$?

(6 marks)

A further specification was considered that included other characteristics of the worker: the age (in years), gender (male/female), ethnicity (black/hispanic/white), and level of education ($E1$ = highschool dropout, $E2$ = highschool graduate, $E3$ = some college, $E4$ = college graduate, $E5$ = Master degree or above). The following OLS regression results were obtained for this multiple linear regression model:

$$\widehat{smoker} = \underset{(0.019)}{0.201} - \underset{(0.009)}{0.045}\, smkban - \underset{(.009)}{0.033} female - \underset{(.0003)}{0.001}\, age - \underset{(.016)}{0.027} black \tag{8.2}$$

$$-\underset{(.014)}{0.104} hispanic + \underset{(.019)}{0.310} E1 + \underset{(.012)}{0.224} E2 + \underset{(.012)}{0.156} E3 + \underset{(.012)}{0.042} E4$$

$$n = 10000,\ R^2 = 0.0526,\ RSS = 1736.81$$

The SEs are in parentheses.

(b) Compare the coefficient estimates on *smkban* from the simple and multiple regression model in (8.1) and (8.2) and explain why the estimates differ.

(3 marks)

(c) **Interpret the estimated parameter on $E2$ (highschool graduate) in (8.2) and indicate how you can obtain its p-value and what information the p-value provides.**

(5 marks)

(d) **Both the simple and multiple regression model suffer from heteroskedasticity. Explain why. What are the implications of heteroskedasticy for the parameter estimates and the standard errors in (8.1) and (8.2)? What can you do to resolve this problem? Explain your answer.**

(6 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Sections 2.6, 3.2, 6.2, 7.3 and 10.1.

Dougherty, C. Subject guide (2016): Chapters 5, 6 and 10.

Approaching the question

(a) This is the linear probability model, where:

$$\mathrm{E}(smoker = 1 \,|\, X) = P(smoker = 1 \,|\, X) = \beta_0 + \beta_1 smkban.$$

If the smoking ban is introduced, the probability of a worker smoking decreases by 7.8 percentage points (not 7.8%).

An (approximate) 95% confidence interval is:

$$\begin{aligned}&(\widehat{\beta}_1 - z_{crit,\,0.025} \times \text{s.e.}(\widehat{\beta}_1),\ \widehat{\beta}_1 + z_{crit,\,0.025} \times \text{s.e.}(\widehat{\beta}_1)) \\ &= (-0.078 - 1.96 \times 0.009,\ -0.078 + 1.96 \times 0.009) \approx (-0.096,\ -0.060)\end{aligned}$$

where z relates to the standard normal distribution (for large degrees of freedom the t distribution converges to the standard normal).

We can use confidence intervals for testing. For a two-sided $\mathrm{H}_0 : \beta_1 = 0$ test, we reject H_0 at the 5% significance level if zero does not lie in the 95% confidence interval.

(b) The reason why the coefficient estimates on *smkban* differ in (8.1) and (8.2) is due to the *omitted variable bias problem.* When we omit relevant variables (for example, education and gender) which are related to the included variable *smkban*), then the parameter estimates on *smkban* will estimate not only the direct effect that *smkban* has on smoking (the parameter of interest) but the indirect effect of these omitted variables as well. We also call these omitted variables confounders.

The more negative effect found in (a) signals that once the effect of education on smoking (gender, age and ethnicity) is controlled for, the true effect of a smoking ban on the probability of smoking is smaller.

(c) The coefficient on $E2$ means that highschool graduates are 22.4 percentage points more likely to smoke relative to people with a Master's degree or above since $E5$ *is the left out base category.* (Many candidates failed to recognise that the left out category was required to interpret the parameter clearly). The t statistic for this effect is:

$$t = \frac{\widehat{\beta}}{\text{s.e.}(\widehat{\beta})} = \frac{0.224}{0.012} = 18.667$$

which has a p-value smaller than 0.05 (we would reject H_0 at the 5% significance level). The p-value is the lowest level of significance at which we can reject the null hypothesis. Since the sample size is large, we can use the standard normal distribution to obtain the p-value (candidates may provide a graphical discussion instead). For a one-sided test, $p = 1 - \Phi(t)$ and for a two-sided test, $p = 2 \times (1 - \Phi(t))$, assuming $t \geq 0$.

(d) In the linear probability model (LPM) we have to deal with the problem of heteroskedasticity. Conditional on X, $\text{Var}(y \mid X) = p(X)(1 - p(X)) \equiv \text{Var}(u \mid X)$. The implication of heteroskedasticity for the estimated standard errors is that they are wrong (they rely on homoskedasticity) and inference based on them would be invalid. The parameter estimates themselves remain unbiased and consistent. For inference, we should use heteroskedasticity-robust standard errors or apply WLS instead.

Question 9

In question 8 we considered the linear regression model to study how workplace smoking bans affect the incidence of smoking. Here we consider the results from applying a probit regression of *smoker* (1 = yes, 0 = no) on *smkban* (1 = yes, 0 = no) and the other explanatory variables:

```
. probit smoker smkban female age black hispanic E1 E2 E3 E4

Iteration 0:   log likelihood = -5537.1662
Iteration 1:   log likelihood = -5255.1526
Iteration 2:   log likelihood =  -5252.349
Iteration 3:   log likelihood = -5252.3489

Probit regression                               Number of obs     =     10,000
                                                LR chi2(9)        =     569.63
                                                Prob > chi2       =     0.0000
Log likelihood = -5252.3489                     Pseudo R2         =     0.0514
```

smoker	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
smkban	-.1517626	.0289268	-5.25	0.000	-.208458	-.0950671
female	-.1106249	.0287785	-3.84	0.000	-.1670298	-.05422
age	-.0042031	.0011748	-3.58	0.000	-.0065057	-.0019006
black	-.07969	.0525369	-1.52	0.129	-.1826604	.0232804
hispanic	-.3327039	.0476677	-6.98	0.000	-.4261308	-.2392769
E1	1.094231	.0714121	15.32	0.000	.9542663	1.234197
E2	.8518588	.0594747	14.32	0.000	.7352906	.9684271
E3	.6492566	.0606989	10.70	0.000	.530289	.7682241
E4	.2224747	.0649939	3.42	0.001	.0950891	.3498603
_cons	-.9842425	.0756055	-13.02	0.000	-1.132427	-.8360584

(a) It is argued that using the probit regression model is better than using the linear probability model when explaining the binary variable *smoker*. Discuss the benefits/drawbacks of using the Probit model when trying to explain a binary variable.

(5 marks)

(b) Explain briefly how the Probit estimates are obtained and discuss the properties of the parameter estimates.

***Hint:* You may recall that for the Probit model, we will specify:**

$$\Pr(smoker = 1) = \Phi(\beta_0 + \beta_1 smkban + \beta_2 female + \cdots + \beta_8 E3 + \beta_9 E4)$$

where Φ is the standard normal CDF (cumulative distribution function).

(5 marks)

(c) Explain how you can estimate the effect of the smoking ban on the probability of smoking for a 50-year old white, college graduated man. You are not expected to use your calculator, clarity of the computations required is enough.

(5 marks)

(d) **Discuss how you could test the joint significance of the worker's characteristics (gender, age, ethnicity and level of education) using the likelihood ratio test. Clearly indicate the test statistic, its distribution, the rejection rule and the additional information you would need to implement it.**

(5 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Sections 10.3 and 10.6.

Dougherty, C. Subject guide (2016): Chapter 10.

Approaching the question

(a) The probit model has three main advantages over the linear probability model (LPM): (i) predicted probabilities are restricted to lie in $[0, 1]$, (ii) maximum likelihood estimators are (asymptotically) efficient whereas OLS (LPM) estimators will be inefficient, and (iii) maximum likelihood estimators automatically deal with heteroskedasticity.

The main drawbacks of the probit model relative to the LPM are that (i) the coefficients cannot be directly interpreted as the marginal effects of the regressor(s) of interest, so we need to compute predicted probabilities using the probit specification, and (ii) it is also computationally more complicated.

(b) The parameters are estimated by maximum likelihood estimation, where the log-likelihood function is given by:

$$\begin{aligned}\log L(\beta) &= \sum_{i=1}^{n}\{smoker_i \log(P(smoker_i = 1 \,|\, X) + (1 - smoker_i)\log(P(smoker_i = 0 \,|\, X)\} \\ &= \sum_{i=1}^{n}\{smoker_i \log(\Phi(\beta_0 + \beta_1 smkban_i + \beta_2 female_i + \cdots + \beta_9 E4_i) + \\ &\quad (1 - smoker_i)\log(1 - \Phi(\beta_0 + \beta_1 smkban_i + \beta_2 female_i + \cdots + \beta_9 E4_i))\}.\end{aligned}$$

To obtain the parameter estimates the first-order conditions are solved (numerically as no explicit formulae exist).

Under suitable regularity conditions, the estimates are consistent, asymptotically normally, and asymptotically efficient.

(c) The effect is the difference in predicted probabilities between the man with the characteristics given, with $smkban = 1$ versus $smkban = 0$:

$$\begin{aligned}&\widehat{P}(y_i = 1 \,|\, X, smkban = 1) - \widehat{P}(y_i = 1 \,|\, X, smkban = 0) \\ &= \Phi\left(\sum_k x_{ki}\widehat{\beta}_k + \widehat{\beta}_1\right) - \Phi\left(\sum_k x_{ki}\widehat{\beta}_k\right) \\ &= \Phi(-0.984 + 0.222 - 0.004 \times 50 - 0.152) - \Phi(-0.984 + 0.222 - 0.004 \times 50).\end{aligned}$$

A discussion of marginal effects (ignoring the fact that *smkban* is a dummy variable) is acceptable as well.

(d) We would like to use the LR test, to test:

$$\mathrm{H}_0 : \beta_j = 0 \ \forall\ j \in \{\text{worker characteristics}\}$$

$$\mathrm{H}_1 : \beta_j \neq 0 \text{ for at least one } j \in \{\text{worker characteristics}\}.$$

Estimate the restricted (R) and the unrestricted probit model (U). The restricted model imposes H_0. The unrestricted model is just the originally estimated model above. The LR test makes use of the difference between the two estimated log-likelihood functions (the log ratio of the likelihood functions). The test statistic is:

$$LR = 2(\log L^U - \log L^R) \sim \chi^2_J$$

where J denotes the degrees of freedom which is equal to the number of restrictions; that is, $J = 8$ here. For a given significance level we reject H_0 if its realisation exceeds the critical value given by the χ^2_8 distribution, which equals 15.51 for the 5% significance level.

Question 10

This question is based on 'An Empirical Comparison of Alternative Models of the Short-Term Interest Rate', by Chan et al. (*The Journal of Finance*, 1992). In this article, the authors are interested in a regression model for the short-term interest rate r_t given by:

$$r_t - r_{t-1} = \beta_0 + \beta_1 r_{t-1} + \varepsilon_t. \qquad (10.1)$$

Assume that the errors ε_t are not serially correlated.

(a) Provide sufficient conditions (on β_1) for r_t to be stationary and weakly dependent. Will the OLS estimator for a linear regression of r_t on r_{t-1} in this case be unbiased and consistent? Explain your answer.

(5 marks)

(b) A few well-known models in finance hypothesise that $\beta_1 = 0$. Explain how you would test this hypothesis. What name do we give the process r_t when $\beta_1 = 0$?

(5 marks)

Suppose you are interested in the following relation between long-term interest rates R_t and short-term interest rates r_t:

$$R_t - r_t = \alpha_0 + \alpha_1 r_t + u_t \qquad (10.2)$$

where u_t is an error term that has zero mean.

(c) Let us assume $r_t = \beta_0 + \varepsilon_t$ (that is $\beta_1 = -1$). How would you test whether u_t is serially correlated? Explain your answer.

(5 marks)

(d) Let us assume $\beta_1 = 0$. Explaining your answers, what can you say about the OLS estimator for α_0 and α_1 applied to equation (10.2):

i. if u_t is stationary and weakly dependent?

ii. if u_t is non-stationary and strongly dependent?

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Sections 12.2, 13.1, 13.2, 13.4 and 13.6.

Dougherty, C. Subject guide (2016): Chapter 13.

Approaching the question

(a) The model can be rewritten in the classic AR(1) structure:

$$r_t = \beta_0 + (1 + \beta_1)r_{t-1} + \varepsilon_t.$$

Stationarity in an AR(1) process requires the AR coefficient to be smaller than 1 implying $\beta_1 < 0$. We have weak dependence here. Under this assumption and given the absence of serial correlation in ε_t, there is no correlation between the error and the regressor r_{t-1}, hence we can obtain consistent parameter estimates for β_1. Nevertheless, we will not get unbiasedness as ε_t will be correlated with future r_{t+j} $\forall\ j \geq 0$ due to the presence of lagged endogenous variables.

(b) Under $\beta_1 = 0$, the AR coefficient is 1 which indicates that r_t is a random walk with drift, and r_t is difference stationary, that is $\Delta r_t := r_t - r_{t-1}$ is stationary. Here we get strong dependence. To test whether we have a unit root, we can carry out the Dickey–Fuller test. Under $\mathrm{H}_0 : \beta_1 = 0$, r_t has a unit root, whereas under $\mathrm{H}_1 : \beta_1 < 0$, r_t is stationary (and weakly dependent). Using the Dickey–Fuller statistical tables, for a given significance level, we reject H_0 if the test statistic $\widehat{\beta}_1/\text{s.e.}(\widehat{\beta}_1)$ exceeds the appropriate critical value.

(c) If $r_t = \beta_0 + \varepsilon_t$, it is just a random error plus a constant. The proposed equation (10.2) can be estimated using OLS and yields consistent parameter estimates assuming u_t is uncorrelated with r_t.

Say we postulate an AR(1) processs for ε_t. We then test H_0: no autocorrelation against H_1: AR(1) presence of autocorrelation. We should apply the Breusch–Godfrey test (the Durbin–Watson test cannot be applied because of the presence of stochastic regressors).

Using the OLS residuals, we run the following regression:

$$\widehat{u}_t = \gamma_0 + \gamma_1 r_t + \delta \widehat{u}_{t-1} + v_t.$$

The test statistic is given by nR^2 of this regression, where n is the sample size. Under the null hypothesis its asymptotic distribution is χ^2_1. We reject H_0 if the realisation exceeds the critical value given by this distribution for a given significance level.

(d) Under $\beta_1 = 0$, r_t is integrated of order 1 and by implication so will be R_t.

i. If u_t is stationary and weakly dependent, then the two unit root processes R_t and r_t must be cointegrated. This means the parameters α and β can be consistently estimated by OLS.

ii. If u_t is non-stationary and strongly dependent, R_t and r_t are unrelated. Nevertheless, running the regression (10.1) may still give significant parameter estimates since we face the case of spurious regressions, wrongly suggesting that there is a relationship between R_t and r_t. This is because the estimates are inconsistent. This is the spurious regression problem.

Examiners' commentaries 2020

EC2020 Elements of econometrics

Important note

This commentary reflects the examination and assessment arrangements for this course in the academic year 2019–20. The format and structure of the examination may change in future years, and any such changes will be publicised on the virtual learning environment (VLE).

Information about the subject guide and the Essential reading references

Unless otherwise stated, all cross-references will be to the latest version of the subject guide (2016). You should always attempt to use the most recent edition of any Essential reading textbook, even if the commentary and/or online reading list and/or subject guide refer to an earlier edition. If different editions of Essential reading are listed, please check the VLE for reading supplements – if none are available, please use the contents list and index of the new edition to find the relevant section.

General remarks

Learning outcomes

At the end of the course, and having completed the Essential reading and activities, you should be able to:

- describe and apply the classical regression model and its application to cross-section data
- describe and apply the:
  - Gauss–Markov conditions and other assumptions required in the application of the classical regression model
  - reasons for expecting violations of these assumptions in certain circumstances
  - tests for violations
  - potential remedial measures, including, where appropriate, the use of instrumental variables
- recognise and apply the advantages of logit, probit and similar models over regression analysis when fitting binary choice models
- competently use regression, logit and probit analysis to quantify economic relationships using standard regression programmes (Stata and EViews) in simple applications
- describe and explain the principles underlying the use of maximum likelihood estimation
- apply regression analysis to time-series models using stationary time series, with awareness of some of the econometric problems specific to time series applications (for example, autocorrelation) and remedial measures
- recognise the difficulties that arise in the application of regression analysis to nonstationary time series, know how to test for unit roots, and know what is meant by cointegration.

Common mistakes committed by candidates

Due to the impact of the Covid-19 pandemic, there was a shift from in-person examinations to open-book examinations completed within a 24-hour window. Despite the challenges, many candidates performed extremely well. We discuss below common mistakes made by candidates, some of them are repeated year after year.

Candidates should realise that $\frac{1}{n}\sum_{i=1}^{n} u_i$ is not the same as $\mathrm{E}(u_i)$. So, while we typically assume $\mathrm{E}(u_i) = 0$, this does not guarantee that $\frac{1}{n}\sum_{i=1}^{n} u_i = 0$. Also, while we may be happy to assume $\mathrm{E}(x_i u_i) = 0$ (uncorrelatedness between the errors and regressors), this does not guarantee that $\frac{1}{n}\sum_{i=1}^{n} x_i u_i = 0$. Note that:

- both $\frac{1}{n}\sum_{i=1}^{n} u_i$ and $\frac{1}{n}\sum_{i=1}^{n} x_i u_i$ are random variables, which take the value 0 with probability 0 (continuous random variables)!
- $\mathrm{E}(u_i) = 0$ and $\mathrm{E}(x_i u_i) = 0$ are fixed, not stochastic!

The differences between sample and population moments need to come across clearly when looking at unbiasedness and making consistency arguments. In both cases, we first simplify our estimator (plug in the true model) to obtain:

$$\widehat{\beta} = \beta + \frac{\sum_{i=1}^{n}(X_i - \bar{X})u_i}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \beta + \frac{\sum_{i=1}^{n} x_i u_i}{\sum_{i=1}^{n} x_i^2} \quad \text{with } x_i = X_i - \bar{X}.$$

- For *unbiasedness*, clearly indicate that you want to show that $\mathrm{E}(\widehat{\beta}) = \beta$. Unbiasedness does not follow from $\sum_{i=1}^{n} x_i u_i = 0$, instead it follows from $\mathrm{E}\left(\frac{\sum_{i=1}^{n} x_i u_i}{\sum_{i=1}^{n} x_i^2}\right) = 0$.

  If we treat x_i as fixed:

  $$\mathrm{E}\left(\frac{\sum_{i=1}^{n} x_i u_i}{\sum_{i=1}^{n} x_i^2}\right) \equiv \mathrm{E}\left(\sum_{i=1}^{n} d_i u_i\right) = \sum_{i=1}^{n} d_i \mathrm{E}(u_i)$$

  and then unbiasedness follows as $\mathrm{E}(u_i) = 0$.

- For *consistency*, clearly indicate that you want to show that $\operatorname{plim}(\widehat{\beta}) = \beta$. Using the plim properties, we show:

  $$\begin{aligned}\operatorname{plim}\widehat{\beta} = \beta + \operatorname{plim}\left(\frac{\sum_{i=1}^{n} x_i u_i}{\sum_{i=1}^{n} x_i^2}\right) &= \beta + \frac{\operatorname{plim}\left(\frac{1}{n}\sum_{i=1}^{n} x_i u_i\right)}{\operatorname{plim}\left(\frac{1}{n}\sum_{i=1}^{n} x_i^2\right)} \\ &\equiv \beta + \frac{\operatorname{plim}(\text{Sample Cov}(x, u))}{\operatorname{plim}(\text{Sample Var}(x))} \\ &= \beta + \frac{\mathrm{Cov}(x, u)}{\mathrm{Var}(x)} \quad \text{using the law of large numbers}\end{aligned}$$

  where $\mathrm{Cov}(x, u) = 0$ and $\mathrm{Var}(x) > 0$, ensuring we get consistency.

- Remember, the law of large numbers ensures that sample averages converge to their population analogues.

A large number of candidates are not able to clearly distinguish between sample variance and covariance, and population variance and covariance.

The use of $\text{Cov}(X, Y)$ and $\text{Var}(X)$ should be restricted to describing the population covariance and variances, respectively, with definitions:

$$\text{Cov}(X, Y) = \text{E}((X - \text{E}(X))(Y - \text{E}(Y))) = \text{E}(XY) - \text{E}(X)\,\text{E}(Y)$$

and:

$$\text{Var}(X) = \text{E}((X - \text{E}(X))^2) = \text{E}(X^2) - (\text{E}(X))^2$$

(you also may denote $\text{Cov}(X, Y) = \sigma_{XY}$ and $\text{Var}(X) = \sigma_X^2$). They are typically unknown, but fixed, quantities.

The sample covariance and variance are estimators of the population covariance and variance, respectively. They are defined as:

$$\text{Sample Cov}(X, Y) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})$$

and:

$$\text{Sample Var}(X) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

(you also may use $\widehat{\sigma}_{XY}$ and $\widehat{\sigma}_X^2$). You can compute them given the data.

With a slight abuse of notation, we often divide by n instead, which is irrelevant if we let n be large. The division by $n - 1$ is a finite sample issue only (unbiasedness).

The sample covariance and variance show up in our definition of the OLS estimator of the slope in the simple linear regression model, not the population covariance and variance, as:

$$\widehat{\beta} = \frac{\sum\limits_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum\limits_{i=1}^{n}(X_i - \bar{X})^2} = \frac{\text{Sample Cov}(X, Y)}{\text{Sample Var}(X)} \neq \frac{\text{Cov}(X, Y)}{\text{Var}(X)}.$$

Treating them as being the same results in incorrect analyses and candidates lose significant marks. Similarly, a handful of candidates lost marks in Question 1 (b) for using the population variances instead of sample variances in the IV estimate formula:

$$\widehat{\beta}_{IV} = \frac{\sum\limits_{i=1}^{n}(Z_i - \bar{Z})(Y_i - \bar{Y})}{\sum\limits_{i=1}^{n}(Z_i - \bar{Z})(X_i - \bar{X})} = \frac{\text{Sample Cov}(Z_i, Y_i)}{\text{Sample Cov}(Z_i, X_i)} \neq \frac{\text{Cov}(Z_i, Y_i)}{\text{Cov}(Z_i, X_i)}.$$

Candidates often made mistakes in the last steps to correctly derive the variance of the estimator β in Question 1 (a). In particular, candidates often mistook $\sum\limits_{i=1}^{n}\sigma^2$ as simply σ^2; whereas we must have:

$$\text{Var}(\widehat{\beta}) = \frac{\sum\limits_{i=1}^{n}\sigma^2}{\left(\sum\limits_{i=1}^{n}X_i^2\right)^2} = \frac{\sigma^2 + \sigma^2 + \cdots + \sigma^2}{\left(\sum\limits_{i=1}^{n}X_i^2\right)^2} = \frac{n\sigma^2}{\left(\sum\limits_{i=1}^{n}X_i^2\right)^2}.$$

A number of candidates seemed to be confused between the conditions for a valid instrument (exclusion restriction and relevance) and factors related to obtaining a more precise IV estimate (a higher sample size, a higher sample variability of X_i, a lower variance of the error term, and a stronger correlation between X_i and Z_i). Candidates describing properties required of instrumental variables, only received 1 mark for discussing the stronger correlation point (analogous to the relevance of an instrument).

When discussing the importance of distinguishing trend stationary and difference stationary processes, two points are expected: discussing the implication for the long-run behaviour of the processes, and discussing statistical implications (the consistency of the estimators). Most candidates discussed the former very well, only a few touched on the latter.

Many candidates simply described multicollinearity as independent variables being correlated. Although we accept answers discussing high correlation among the independent variables as an alternative to a close linear relation among the regressors, simply taking correlation among regressors as the definition of multicollinearity is incorrect. In attempting True/False questions with a specific instruction to explain your answers, several candidates failed to demonstrate adequate understanding of the course material and simply reproduced the questions at hand.

Candidates struggled to give competent answers to the interpretation of empirical results. When interpreting an empirical result you should discuss the significance of the coefficients, magnitude and sign of the coefficients. Relatedly, *ceteris paribus* (or holding other factors constant) is needed to ensure that the estimate can be interpreted as the marginal change in the dependent variable given a change in the independent variable of interest. Many candidates lost marks from these mistakes.

Candidates particularly struggled with the difference between percentage points and % changes when it comes to the interpretation of coefficients in binary choice models. A coefficient in such a model gives the percentage point difference in the outcome variable when the regressor changes by 1 unit. The % change is a *relative* change that depends on the baseline value of the outcome variable (for example, the realisation when a regressor equals zero in the case of a dummy variable regressor). For example, if we have initially 50% of a sample buying ecological apples, a 10% increase will mean now $(1 + 0.1) \times 50 = 55\%$ of the sample buys ecological apples. A 10 percentage point increase instead will result in $50 + 10 = 60\%$ of the sample choosing to buy ecological apples.

It was very common to misunderstand the Central Limit Theorem. While it provides the distribution of the *estimator* $\widehat{\beta}$, no matter what the true distribution of the errors is, it does *not* say that the distribution of the error term is approximately normal in large samples.

When discussing the marginal effect after a Probit estimation, candidates often did not specify the term z in the pdf $\phi(z)$.

A number of candidates described the Dickey–Fuller test when an Augmented Dickey–Fuller (ADF) test was asked. The difference between the two tests is in the test equation. In the augmented specification, we add sufficient lagged differences of the residuals to ensure that there is no autocorrelation left in the test equation regression.

Exogeneity and endogeneity are often very simply explained as variables 'outside of the model' or 'inside the model', respectively. While this approach is not wrong; it is not a particularly clear or helpful definition in general contexts that are not related to simultaneous equations models. Defining these concepts in terms of the relationship between regressor and error term is much clearer and naturally leads to a way to show exogeneity or endogeneity by considering the covariance between regressors and error terms.

Candidates often missed that all exogenous regressors and instruments have to be included in the first stage when performing a TSLS estimation. Some candidates used the fitted values of the endogenous variable obtained from the first stage as the dependent variable in the second stage. This mistake is a serious misunderstanding. Other candidates regressed the reduced forms of the demand and supply equations in each stage as a TSLS estimation. This method is not helpful as we are interested in the relationship between the quantity and the price (between the two endogenous variables), not between each endogenous variable and the exogenous variables.

When asked about identification of structural form equations, candidates often compared the number of excluded regressors to the number of endogenous variables in general, without pointing out which variables are the excluded regressors and could therefore be used as an instrument. Candidates were also often unclear of the definitions of over-identification, under-identification and exact-identification. For example, some candidates discussed the concepts by comparing the number of variables and the number of numerical values of the parameters that can be obtained from the structural equations. This point is technically correct in the sense that we do not have enough data on the instruments to identify the parameters in the under-identification case. Candidates, however, often did not make that discussion clear.

When deriving the reduced forms, candidates should check conditions for their existence, in particular, whether the denominator of the expression is different from zero. Only a few candidates made this check.

Gauss–Markov conditions were used in many places in the examination, for example, when discussing the BLUE property of an estimator or when conducting hypothesis tests. The Gauss–Markov conditions have to be explicitly specified. Only writing that the Gauss–Markov conditions hold is not sufficient. A good practice that many candidates have adopted is: begin the script by explicitly providing the Gauss–Markov conditions and refer back to them thereafter. Many candidates did not mention either the Gauss–Markov conditions or the normality assumption when conducting t tests.

When conducting hypothesis testing, ensure that you clearly indicate the null and alternative hypotheses (in terms of the true parameters, say β_1), the test statistic (in terms of the parameter estimates, here $\widehat{\beta}_1$), its distribution (with degrees of freedom), the rejection rule (one-sided or two-sided) for a given significance level (typically 5%) with suitable critical values, and provide an interpretation of your result. Candidates often lost marks for missing parts of these steps. Choosing the critical value appropriate to the sample size is also important. Several candidates used 1.96 (a critical value for a very large number of degrees of freedom) in a small sample (30) instead of the appropriate one (2.042). This mistake also cost marks.

When using confidence intervals in a hypothesis test, candidates should decide if the hypothesised value for the population parameter of interest β lies within the confidence interval. Many candidates used the default value of zero when discussing their confidence interval regardless of the population parameter in question.

Another common mistake was to conduct a two-sided test when the statement was worded with a clear direction of the relationship. In particular, in Question 10 (d), the chef's claim had such a hint of the direction ('less elastic') so that a one-sided test was appropriate. Many candidates conducted a two-sided test and lost marks.

Just as last year, many candidates did not answer all parts of a question. Make sure you read the questions properly and provide all details that are requested. Not answering a question will automatically earn you zero marks for that question.

Key steps to improvement

Essential reading for this course includes the subject guide and the following:

- Dougherty, C. *Introduction to econometrics.* (Oxford: Oxford University Press, 2016) 5th edition [ISBN 9780199676828]; http://oxfordtextbooks.co.uk/orc/dougherty5e/

Apart from the Essential readings you should do some supplementary reading. One very good book at the same level is:

- Gujarati, D.N. and D.C. Porter *Basic econometrics.* (McGraw–Hill, 2009, International edition) 5th edition [ISBN 9780071276252].

To understand the subject clearly it is important to supplement Dougherty's *Introduction to econometrics* (fifth edition) with the subject guide **EC2020 Elements of econometrics** (2016), especially Chapter 10 which covers maximum likelihood estimation. It is very important to carefully go through the subject guide. The subject guide contains solutions to the questions given in the main textbook and also some additional questions and solutions. Working through these will improve your understanding of the subject.

The chapter in the subject guide on maximum likelihood (Chapter 10) includes some additional theory which has not been covered in the main textbook. It is important to read the additional theory given in the subject guide to have a better understanding of the principles of maximum likelihood and tests based on the likelihood function.

Please check the VLE course page for resources for this subject such as a downloadable copy of the subject guide **EC2020 Elements of econometrics** (2016), PowerPoint slideshows that provide a graphical treatment of the topics covered in the textbook, datasets and statistical tables. Candidates should utilise datasets using standard regression programmes (STATA or EViews). This will help in the understanding of the subject.

Examination revision strategy

Many candidates are disappointed to find that their examination performance is poorer than they expected. This may be due to a number of reasons, but one particular failing is '**question spotting**', that is, confining your examination preparation to a few questions and/or topics which have come up in past papers for the course. This can have serious consequences.

We recognise that candidates might not cover all topics in the syllabus in the same depth, but you need to be aware that examiners are free to set questions on **any aspect** of the syllabus. This means that you need to study enough of the syllabus to enable you to answer the required number of examination questions.

The syllabus can be found in the Course information sheet available on the VLE. You should read the syllabus carefully and ensure that you cover sufficient material in preparation for the examination. Examiners will vary the topics and questions from year to year and may well set questions that have not appeared in past papers. Examination papers may legitimately include questions on any topic in the syllabus. So, although past papers can be helpful during your revision, you cannot assume that topics or specific questions that have come up in past examinations will occur again.

If you rely on a question-spotting strategy, it is likely you will find yourself in difficulties when you sit the examination. We strongly advise you not to adopt this strategy.

Examiners' commentaries 2020

EC2020 Elements of econometrics

Important note

This commentary reflects the examination and assessment arrangements for this course in the academic year 2019–20. The format and structure of the examination may change in future years, and any such changes will be publicised on the virtual learning environment (VLE).

Information about the subject guide and the Essential reading references

Unless otherwise stated, all cross-references will be to the latest version of the subject guide (2016). You should always attempt to use the most recent edition of any Essential reading textbook, even if the commentary and/or online reading list and/or subject guide refer to an earlier edition. If different editions of Essential reading are listed, please check the VLE for reading supplements – if none are available, please use the contents list and index of the new edition to find the relevant section.

Comments on specific questions

Candidates should answer **EIGHT** of the following **TEN** questions: all **FIVE** questions in Section A (8 marks each) and **THREE** questions from Section B (20 marks each). Candidates are strongly advised to divide their time accordingly. **If more than EIGHT questions are answered, only the first EIGHT questions attempted will be counted.**

Section A

Answer all questions from this section.

Question 1

Consider the following regression model:

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

The error term has a zero mean, variance equal to σ^2/X_i^2, and $\mathrm{E}(u_i u_j) = 0$ for $i \neq j$. You are given a sample of observations $\{(Y_i, X_i)\}_{i=1}^n$. You may treat X_i as being non-stochastic. Clearly annotating your answers:

(a) **The OLS estimator of β can be shown to be unbiased in the presence of heteroskedasticity (you are not asked to show this). Derive the variance of the OLS estimator of β. You are expected to clearly define the OLS estimator of β.**

(5 marks)

(b) **Discuss how you can obtain the Best Linear Unbiased Estimator (BLUE) of β given the heteroskedasticity.**

(3 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Sections 2.3, 2.4 and 7.3.

Dougherty, C. Subject guide (2016): Chapters 2 and 7, and Section 15.15.

Approaching the question

(a) The OLS estimator of β is:

$$\widehat{\beta} = \frac{\sum_{i=1}^{n} X_i Y_i}{\sum_{i=1}^{n} X_i^2} = \beta + \sum_{i=1}^{n} d_i u_i$$

where $d_i = X_i / \sum_{i=1}^{n} X_i^2$ are constants for $i = 1, \ldots, n$. We have:

$$\begin{aligned}
\text{Var}(\widehat{\beta}) &= \text{Var}\left(\beta + \sum_{i=1}^{n} d_i u_i\right) \\
&= \text{Var}\left(\sum_{i=1}^{n} d_i u_i\right) && \text{(additive constant does not carry any variability)} \\
&= \sum_{i=1}^{n} d_i^2 \text{Var}(u_i) + \sum_{i \neq j}^{n} d_i d_j \text{Cov}(u_i, u_j) && \text{(because } d_i \text{ is non-stochastic)} \\
&= \sum_{i=1}^{n} d_i^2 \left(\frac{\sigma^2}{X_i^2}\right) + 0 && \text{(uses properties of } u_i \text{ (heteroskedastic, no autocorrelation))} \\
&= \sum_{i=1}^{n} \left(\frac{X_i}{\sum_{i=1}^{n} X_i^2}\right)^2 \left(\frac{\sigma^2}{X_i^2}\right) \\
&= \frac{\sum_{i=1}^{n} \sigma^2}{\left(\sum_{i=1}^{n} X_i^2\right)^2} \\
&= \frac{n\sigma^2}{\left(\sum_{i=1}^{n} X_i^2\right)^2}.
\end{aligned}$$

(b) Because the variances depend on X_i, we have heteroskedasticity in the regression model. An OLS on the above model will not be the BLUE of β. To get the BLUE estimator of β, we need to transform the model so that it satisfies the Gauss–Markov conditions (which have to be specified somewhere in the script). Specifically, we should consider:

$$Y_i X_i = \beta X_i^2 + u_i X_i \quad \text{for } i = 1, \ldots, n.$$

Because $\text{E}(u_i X_i) = 0$ and $\text{Var}(u_i X_i) = \sigma^2$, there is no autocorrelation in this transformed model. An OLS on this model now gives us the BLUE. The estimator of β is:

$$\widehat{\beta} = \frac{\sum_{i=1}^{n} X_i^3 Y_i}{\sum_{i=1}^{n} X_i^4}.$$

Comments:

- In part (a), several candidates suggested the use of GLS, which was wrong and given zero marks. Many candidates failed to correctly derive the final answer.
- In part (b), many candidates correctly identified the need to use a weighted least squares estimator (WLS) and were well-rewarded. Some suggested the use of GLS to correct for heteroskedasticity and obtain GLS. Both estimators are equivalent in the presence of heteroskedasticity.

Question 2

Consider the simple linear regression model:

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

in the presence of correlation between the errors and regressors. The regressors exhibit variability in the sample, i.e. $\sum_{i=1}^{n}(X_i - \bar{X})^2 \neq 0$. 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):

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

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 $\widehat{\beta}_{IV}$ (you are not asked to derive it).

(1 mark)

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

(4 marks)

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

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Sections 2.5, 8.2, and 8.5.

Dougherty, C. Subject guide (2016): Chapter 8 and Section 15.16.

Approaching the question

(a) There are two ways to write the IV estimate formula:

$$\widehat{\beta}_{IV} = \frac{\sum_{i=1}^{n}(Z_i - \bar{Z})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(Z_i - \bar{Z})(X_i - \bar{X})} = \frac{\text{Sample Cov}(Z_i, Y_i)}{\text{Sample Cov}(Z_i, X_i)}.$$

Note that sample variances have to be used instead of their population counterparts.

(b) We can rewrite the definition of the variance as follows:

$$\text{Var}(\widehat{\beta}_{IV}) = \frac{\sigma^2}{n \times \text{Sample Var}(X_i)} \times \frac{1}{r_{XZ}^2}$$

which shows that we can get more precise estimates by:

- increasing the sample size
- increasing the sample variability of X_i
- reducing the variance of the error term σ^2
- increasing the correlation between X_i and Z_i.

The importance of the precision is that it is useful when conducting hypothesis testing. It makes it easier to detect when the null hypothesis is false. Precision is good when the dispersion of our parameter estimates over different samples is quite small, giving more credence to a particular point estimate.

(c) If X_i is not correlated with u_i, the instrument that has the highest correlation with X_i is X_i itself (perfect correlation) – hence the highest precision as $r_{XZ}^2 = 1$ (and r_{XZ}^2 cannot be larger than 1). If we use X as an instrument for X, we get the OLS estimator of β. OLS is the BLUE estimator in this setting.

Comments:

- In part (b), candidates who discussed the conditions for a valid instrument often received one mark for highlighting the need for a strong correlation between X_i and Z_i. Many candidates did not discuss the importance of the IV estimate precision asked by the question. Some used the IV formula to discuss the conditions and were well-rewarded.
- In part (c), some candidates correctly realised that since there is no correlation between the error term and the regressor, there is no endogeneity present in the model and we could use the OLS as the BLUE. To get the full three marks using this approach, candidates should at least discuss the Gauss–Markov conditions or relate to how the OLS estimate is more efficient in this case.

Question 3

Consider a linear regression model:

$$y_t = \alpha + \beta x_t + \varepsilon_t, \quad t = 1, \ldots, T$$

where the zero mean error ε_t exhibits autocorrelation of an unknown form. We assume our processes are covariance stationary and exhibit weak dependence. The regressor and error may be assumed to be independent.

(a) Explain what it means to say that $\{\varepsilon_t\}_{t=1}^T$ is covariance stationary. Provide an intuitive discussion of the requirements and indicate why these requirements are desirable.

(3 marks)

(b) Recognising that the errors exhibit autocorrelation, discuss how we can conduct statistical inference on β using the OLS estimator. Specifically, discuss how you can test the hypothesis $\text{H}_0 : \beta = 0.7$ against the alternative $\text{H}_1 : \beta < 0.7$ using the OLS estimator.

(5 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Sections 7.1, 7.2, 12.1 and 12.3.

Dougherty, C. Subject guide (2016): Chapters 7 and 12.

Approaching the question

(a) Candidates were expected to discuss the three following conditions, without a need to specifically mention the finite requirement.

- The expected value of ε_t does not change over time (and is finite), i.e. $\mathrm{E}(\varepsilon_t) = \mu$ for all t.
- The variance of ε_t exists and does not change over time (and is finite), i.e. $\mathrm{Var}(\varepsilon_t) = \sigma^2$ for all t.
- The covariance of ε_t and ε_{t+h} only depends on the distance of time h, i.e. $\mathrm{Cov}(\varepsilon_t, \varepsilon_{t+h}) = \mathrm{Cov}(\varepsilon_s, \varepsilon_{s+h})$ for all s and t.

For the intuition, candidates were expected to explain that non-stationarity is a problem that may need to be dealt with (due to consistency/spurious regression). One of the following discussions works well.

- These assumptions replace the assumption typically made in a cross-sectional setting in which we assume the errors to be i.i.d. (independent and identically distributed). These assumptions are more reasonable in time series, where generally there is dependence.
- The fact that the mean and the variance do not change over time ensures that the process (or relationship) is stable over time. The second assumption provides a homoskedastic setting, which ensures the standard errors are valid for testing.

(b) Candidates are expected to discuss the use of Newey–West standard errors (or heteroskedasticity and autocorrelation consistent (HAC) standard errors) to obtain the valid standard error for statistical tests. A complete testing procedure with usual components is also expected. One way to approach the question is the following.

In the presence of autocorrelation, we cannot rely on the usual standard errors of our OLS estimators. Instead, we need to use the Newey–West (HAC) robust standard errors for hypothesis testing. Using the robust standard errors, we obtain our t test statistic:

$$\frac{\widehat{\beta} - 0.7}{\text{Robust SE}(\widehat{\beta})}.$$

Using a 5% significance level we will reject H_0 if:

$$\frac{\widehat{\beta} - 0.7}{\text{Robust SE}(\widehat{\beta})} < -1.645.$$

Candidates should also note that the test is an asymptotic t test (or z test) with a test distribution of $N(0, 1)$ (a t_{n-2} distribution is also accepted).

Comments:

- In part (a), candidates often missed one of the requirements (particularly the one regarding the covariance).
- In part (b), many candidates lost marks for an incorrect test statistic or not providing the test statistic distribution. Several suggested to use GLS instead, which was not correct as the question specifically asked for a hypothesis test using the OLS estimator.

Question 4

Consider the following time series model for $\{y_t\}_{t=1}^{T}$:

$$\boldsymbol{y_t = \alpha + \beta t + u_t, \quad t = 1, \ldots, T}$$

$$\textbf{with} \quad \boldsymbol{u_t = \rho u_{t-1} + \varepsilon_t} \quad \textbf{and} \quad \boldsymbol{|\rho| \leq 1}$$

where ε_t is an i.i.d. $(0, \sigma_\varepsilon^2)$ error that is uncorrelated with anything in the past (white noise).

(a) Show that y_t is trend stationary when $|\rho| < 1$.

(3 marks)

(b) Show that y_t is difference stationary when $\rho = 1$.

(3 marks)

(c) Discuss the importance of distinguishing between trend stationary and difference stationary processes.

(2 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Sections 13.1 and 13.2.

Dougherty, C. Subject guide (2016): Chapter 13.

Approaching the question

(a) Candidates were expected to discuss the following points.

y_t is trend stationary if by detrending we can make y_t stationary. By detrending y_t, we have:

$$y_t - \beta t = \alpha + u_t.$$

This would be the case provided u_t is stationary. If $|\rho| < 1$, we know that AR(1) is a stationary process, hence y_t is trend stationary under this condition.

(b) Candidates were expected to discuss the following points.

y_t is difference stationary if by differencing we can make y_t stationary. Taking the difference of $y_t - y_{t-1}$, we have:

$$\begin{array}{ll} y_t & = \alpha + \beta t + u_t \\ y_{t-1} & = \alpha + \beta(t-1) + u_{t-1} \\ \hline y_t - y_{t-1} & = \beta + u_t - u_{t-1} \\ \Delta y_t & = \beta + \varepsilon_t \end{array}$$

where the latter line recognises that if $\rho = 1$ then $u_t - u_{t-1} = \varepsilon_t$. Since ε_t is i.i.d., we know ε_t is stationary and so $\beta + \varepsilon_t$ also is, hence y_t is difference stationary.

(c) Candidates are expected to discuss the two following points:

- Implication for the long-run behaviour of a process: Time series that are trend stationary always revert to the trend in the long run (the effects of shocks are eventually eliminated). That is, they exhibit weak dependence. Time series which are difference stationary never recover from shocks to the system (the effects of shocks are permanent). That is, they exhibit strong dependence or persistence of the effect.
- There are also statistical implications regarding consistency of the estimators if the assumptions are not satisfied.

Comments:

- Candidates should explicitly discuss the definitions of trend stationary and difference stationary somewhere in the answers.
- In part (b), many candidates lost marks for stating explicitly that the reason Δy_t is stationary follows from the fact that ε_t is i.i.d. (independent and identically distributed).
- In part (c), many candidates focused on the long-run behaviour but did not discuss the statistical implications when the requirements were not met.

Question 5

(a) **Define the term 'multicollinearity' and provide a real life example where this problem is likely to occur.**

(2 marks)

(b) **Examine whether the following statements are true or false. Give an explanation.**

i. **In multiple regression, multicollinearity implies that the least squares estimators of the coefficients are biased and standard errors invalid.**

(3 marks)

ii. **If the coefficient estimates in an equation have high standard errors, this is evidence of high multicollinearity.**

(3 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Section 3.4.

Dougherty, C. Subject guide (2016): Chapter 3.

Approaching the question

(a) Perfect multicollinearity means that some of the regressors can be written as a linear combination of other regressors. For example, $X_i = \sum_{j \neq i}^{n} \lambda_j X_j$, where the X_js are regressors.

Multicollinearity (often referred to as near multicollinearity) indicates that there is a close linear relation between the regressors. Answers stating that regressors exhibit high correlations were also accepted.

A real-life example of multicollinearity is in a setting where we try to explain the determinants of educational attainment and we use various measures of cognitive skills, which are highly correlated.

(b) Candidates were expected to give clear explanations to their answers.

i. Presence of near multicollinearity is not a violation of the Gauss–Markov assumptions. In the presence of multicollinearity, the OLS estimators remain unbiased and their standard errors remain valid as long as all the Gauss–Markov assumptions are satisfied. The statement is false.

ii. It is true that the presence of near multicollinearity gives rise to high standard errors, but having high standard errors could also be due to the fact that we have a very small sample or high variance of the error term. The statement is false.

Comments:

- In part (a), candidates often lost marks for describing multicollinearity as the issue of regressors being simply *correlated*. Some provided a real-life example of the dummy variable trap, which is a special case of perfect multicollinearity. Many candidates did not provide specific examples with a clear regression. They simply discussed two highly correlated (potential) regressors without pointing to the regression at hand.
- In part (b), many candidates lost marks for not explaining their answers but rather repeating the statements.

Section B

Answer three questions from this section.

Question 6

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 [.068]	1.088 (.206)	−0.244 (.474)
regprc	0.735 [.132]	2.029 (.378)	2.030 (.381)
ecoprc	−0.845 [.106]	−2.344 (.318)	−2.267 (.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.

(5 marks)

(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.

(5 marks)

(c) The Probit model B postulates that:

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

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

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

(5 marks)

(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.**

(2 marks)

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.**

(3 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Sections 10.1, 10.3 and 10.6.

Dougherty, C. Subject guide (2016): Chapter 10.

Approaching the question

(a) Candidates are expected to correctly interpret the coefficient of ecologically produced apple price before conducting a two-sided test for its statistical significance. An excellent answer would have the following points.

The coefficient is -0.845, hence when the price of ecologically produced apples increases one dollar, the probability that households buy ecologically produced apples decreases by 84.5% (or a drop in probability of 0.845), *ceteris paribus*.

Conduct a t test with the t statistic of:

$$\frac{\widehat{\beta}_{ecoprc}}{\mathrm{SE}(\widehat{\beta}_{ecoprc})}.$$

We can see that the null hypothesis is rejected, suggesting that the estimate is significant. The null hypothesis, the test statistic, the test statistic distribution, the critical value, and the rejection rule were expected.

The key reason to use robust standard errors is to deal with the heteroskedasticity in the error term inherent in the linear probability model (LPM).

(b) Candidates should explain the difference in terms of linearity between OLS and LPM estimators somewhere in the answer before discussing the advantages and drawbacks of the Probit estimation.

- The LPM is a special case of a binary regression model. When the dependent variable is binary $(0, 1)$, we have $\mathrm{E}(y \,|\, x) = \Pr(y = 1 \,|\, x)$. If we assume the probability is linear in the parameters, the model can be estimated using least squares.
- The Probit model has various advantages over the LPM: Predicted probabilities are restricted to lie in $[0, 1]$, maximum likelihood estimation (MLE) is (asymptotically) efficient whereas least squares estimators such as OLS (LPM) will be inefficient. Second, MLE automatically deals with heteroskedasticity inherent in the LPM, the marginal effect obtained from Probit also is allowed to depend on individual characteristics.
- The main drawback(s) of the Probit model relative to the linear probability model is that the coefficients cannot be directly interpreted as the marginal effects of the regressor(s) of interest. Instead, we need to compute predicted probabilities using the Probit specification, and it is also computationally more complicated.

(c) Candidates are expected to only test the joint significance of β_3, β_4, β_5 and β_6. This joint significance test should be clearly notated. An excellent answer should cover the following points:

We are asked to test:

$$\mathrm{H}_0 : \beta_3 = \beta_4 = \beta_5 = \beta_6 = 0 \quad \text{vs.} \quad \mathrm{H}_1 : \exists\ i \in [3, 4, 5, 6], \beta_i \neq 0.$$

We need to use the likelihood ratio (LR) test for hypothesis testing. This requires us to estimate the restricted Probit (where only *regprc* and *ecoprc* are used, i.e. Probit A) and unrestricted Probit (where all variables are included, i.e. Probit B). The test statistic is given by:

$$LR = 2(\log L^U - \log L^R) \overset{a}{\sim} \chi^2_4 \quad \text{under } \mathrm{H}_0.$$

The χ^2 distribution has four degrees of freedom because we are testing four restrictions. We should reject at the 5% significance level if $LR > \chi^2_{0.05,\,4} = 9.488$. We note that our test statistic value is:

$$2 \times (-399.04 - (-407.60)) = 17.12$$

which exceeds the critical value. We reject the joint significance of non-price variables. (The test uses the property that our maximum likelihood estimators are asymptotically normally distributed.)

(d) Candidates are expected to note down the marginal effect formula for *ecoprc* and also explain the formula somewhere in the answer.

i. To get the marginal effect of *ecoprc* we need:

$$\phi(z)\widehat{\beta}_{ecoprc} \quad \text{with } z = \widehat{\beta}_0 + \widehat{\beta}_1 \mathit{regprc} + \widehat{\beta}_2 \mathit{ecoprc}$$

and $\phi(\cdot)$ is the pdf of $N(0, 1)$. Specifically:

$$\phi(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right).$$

ii. In this setting, we have:

$$z = 1.088 + 2.029 \times 0.884 - 2.344 \times 1.082.$$

Now, we have two ways to obtain the marginal effect. First, we can use the fact that the marginal effect of a one-unit change is given by:

$$\phi(1.088 + 2.029 \times 0.884 - 2.344 \times 1.082) \times -2.344.$$

To get the effect of a \$0.10 reduction, we will then need to multiply the above by -0.10.

Second, candidates could opt for computing:

$$\Phi(1.088 + 2.029 \times 0.884 - 2.344 \times (1.082 - 0.10)) - \Phi(1.088 + 2.029 \times 0.884 - 2.344 \times 1.082).$$

Comments:

- Most candidates lost marks in part (a) for either forgetting the *ceteris paribus* condition or missing several parts required for hypothesis testing.
- Several candidates lost marks in part (c) for choosing the wrong degrees of freedom or misspecifying the null hypothesis.
- Some candidates found part (d) particularly challenging. Note that the question did not require the use of a calculator but some very good candidates went beyond and computed the exact marginal effect.

Question 7

Let us consider the following ADL(1, 1) model:

$$\boldsymbol{Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 X_t + \beta_3 X_{t-1} + u_t, \quad |\beta_1| < 1}$$

where X_t and Y_t are $I(0)$ variables and u_t is an i.i.d. $(0, \sigma_u^2)$ error term that is uncorrelated with $Y_{t-1}, Y_{t-2}, Y_{t-3}, \ldots$. You may assume that u_t is independent of $X_{t'}$ for all t, t'.

(a) Provide the short-run and long-run effect of X on Y. Explain the difference between these effects.

(3 marks)

(b) Discuss what properties your OLS estimators for the ADL(1, 1) parameters will have in the presence of the lagged dependent variable. In particular, (i) are the estimators unbiased and consistent, and (ii) should we use robust standard errors? Provide supportive arguments for your answers.

(7 marks)

(c) Show that when you omit the relevant variable Y_{t-1} in the above model, you will get evidence of autocorrelation. Explain the result.

***Hint:* You are expected to reformulate your model as:**

$$\boldsymbol{Y_t = \beta_0 + \beta_2 X_t + \beta_3 X_{t-1} + v_t}$$

where $v_t = \beta_1 Y_{t-1} + u_t$, and show that, for example, $\text{Cov}(v_{t+1}, v_t) \neq 0$.

(5 marks)

(d) Discuss how you would proceed to test for the presence of autocorrelation in the model in (b) using the Breusch–Godfrey test. You may assume that under the alternative, u_t is (or, can be suitably well approximated by) a stationary AR(1).

(5 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Sections 11.3, 11.4, 11.5 and 12.3.

Dougherty, C. Subject guide (2016): Chapters 11 and 12.

Approaching the question

(a) Candidates were expected to clearly indicate the estimates for the short-run and long-run effects of X on Y before discussing their differences as follows.

The short-run effect is β_2. The long-run effect is:

$$\frac{\beta_2 + \beta_3}{1 - \beta_1}.$$

We obtain this by solving the model at equilibrium:

$$Y^e = \beta_0 + \beta_1 Y^e + \beta_2 X^e + \beta_3 X^e \quad \Rightarrow \quad Y^e = \frac{\beta_0}{1 - \beta_1} + \underbrace{\frac{\beta_2 + \beta_3}{1 - \beta_1}}_{\text{long-run effect}} X^e.$$

The short-run effect gives the immediate effect that a change (potentially just temporary) in X has on Y (temporary change); whereas the long-run effect indicates what in equilibrium the effect would be of a permanent change in X.

(b) Candidates were expected to clearly indicate the biasedness of the estimator due to the presence of Y_{t-1} and further discuss the assumptions for consistency.

Because of the presence of Y_{t-1} in the above model, the OLS estimator is biased. The assumptions for consistency are satisfied because for this we only need to ensure that u_t is uncorrelated with the regressors Y_{t-1}, X_t and X_{t-1}. Because the X variables are independent of the errors by assumption, they are uncorrelated. As long as there is no autocorrelation, we will not have any correlation between Y_{t-1} and u_t. The requirement for unbiasedness is stronger than before. We do not need to use robust standard errors because there is no heteroskedasticity or autocorrelation.

(c) Candidates were expected to show that $\text{Cov}(v_{t+1}, v_t) \neq 0$. Although some candidates chose to discuss the intuitions underlying these derivations, it is clearer to combine equations and the intuitions as follows.

Using our definition of v_t and v_{t+1}, we note:

$$\begin{aligned}\text{Cov}(v_{t+1}, v_t) &= \text{Cov}(\beta_1 Y_t + u_{t+1}, \beta_1 Y_{t-1} + u_t)\\ &= \beta_1^2 \text{Cov}(Y_t, Y_{t-1}) + \beta_1 \text{Cov}(Y_t, u_t) + \beta_1 \text{Cov}(u_{t+1}, Y_{t-1}) + \text{Cov}(u_{t+1}, u_t).\end{aligned}$$

Because u_t is white noise (i.i.d.), the errors are unrelated to anything in the past. As a result, $\text{Cov}(u_{t+1}, u_t) = 0$ and $\text{Cov}(u_{t+1}, Y_{t-1}) = 0$.

To show that $\text{Cov}(v_{t+1}, v_t) \neq 0$, it is now sufficient to point out that $\text{Cov}(Y_t, u_t) \neq 0$. This follows quickly when noting that as our model is given by:

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 X_t + \beta_3 X_{t-1} + u_t, \quad |\beta_1| < 1$$

this expression is a direct relationship between Y_t and u_t. We immediately have that $\text{Cov}(Y_t, u_t) \neq 0$ because $\text{Cov}(Y_t, Y_{t-1}) \neq 0$ as $\beta_1 \neq 0$.

(d) Candidates were expected to perform the following steps to get full marks, in particular, the test equation. A standard testing procedure was expected as follows.

We should test $\text{H}_0 : \rho = 0$ against $\text{H}_1 : \rho \neq 0$ with $u_t = \rho u_{t-1} + e_t$, where e_t is white noise. Run OLS on the original equation to obtain the residuals $\widehat{u}_t$:

$$\widehat{u}_t = Y_t - \widehat{\beta}_0 - \widehat{\beta}_1 Y_{t-1} - \widehat{\beta}_2 X_t - \widehat{\beta}_3 X_{t-1}.$$

By the assumption that the alternative is a stationary AR(1) process, the test equation is:

$$\widehat{u}_t = \alpha_0 + \alpha_1 Y_{t-1} + \alpha_2 X_t + \alpha_3 X_{t-1} + \rho \widehat{u}_{t-1} + e_t.$$

Using the sample size of the latter regression, n, and its R^2 (goodness-of-fit measure), we obtain our test statistic $nR^2 \overset{a}{\sim} \chi_1^2$. We should reject the null hypothesis if nR^2 exceeds the critical value of the χ_1^2 distribution.

Comments:

- In part (a), several candidates failed to recognise β_2 as the short-run effect; others did not clearly discuss the fact that the long-run effect pertains to the permanent change in X.
- In part (b), only a few candidates discussed the fact that the estimator remains consistent because u_t is uncorrelated with the regressors.
- In part (c), only a few candidates did well in this subquestion. Several did not follow the hint and relied on an incorrect covariance: $\text{Cov}(Y_{t-1}, u_t)$ or $\text{Cov}(Y_t, u_t)$.
- In part (d), like in other questions on hypothesis testing, candidates lost marks for missing one or more parts of the testing process. Several candidates did not include $\rho\widehat{u}_{t-1}$ in the testing equation or failed to indicate the correct degrees of freedom.

Question 8

Let us consider the following ADL(1, 1) model:

$$\boldsymbol{Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 X_t + \beta_3 X_{t-1} + u_t}$$

where X_t and Y_t are $I(1)$ variables and $|\beta_1| < 1$. Let us assume that the error term has zero mean and that the error is uncorrelated with Y_{t-1}, X_t and X_{t-1}.

(a) It is important to distinguish whether the above model is spurious or co-integrating. Explain these concepts clearly and highlight their differences.

(5 marks)

(b) Discuss how you can test whether the above specification is spurious (null hypothesis) or co-integrating (alternative) using a standard Augmented Dickey–Fuller test. Clearly indicate the test equation, test statistic and the rejection rule. The critical value provided by Dickey and Fuller for this test at the 5% level of significance is equal to −2.86.

(5 marks)

(c) You are told that there exists an Error Correction Model (ECM) that describes the short-run and long-run dynamics between Y_t and X_t. What does this tell you as it relates to your answers in (a) and (b)? Provide the Error Correction Model and interpret the various components of the ECM.

(5 marks)

(d) Discuss how the ECM can be estimated using Ordinary Least Squares.

(5 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Sections 13.1, 13.2, 13.4, and 13.6.

Dougherty, C. Subject guide (2016): Chapter 13.

Approaching the question

(a) Candidates were expected to discuss the following points.

Spurious regression refers to the scenario where *two unrelated series* give the impression that they are related, purely because of the fact that both are trending. A classic example is where Y_t and X_t are both unrelated $I(1)$ series (random walk or random walk with drift). Running a regression of Y_t on X_t (a spurious relationship) has a large R^2 and a seemingly significant slope parameter. The problem is that here the error term is $I(1)$ as well and statistical inference is invalid. The regression is meaningless!

Another example of spurious regression is where two completely unrelated trend stationary series may give the impression that they are related purely because of the omission of the 'time' variable (OVB). Once we account for the deterministic trend, the regression will be able to detect that there is no relationship between X and Y. By omitting the time trend in the regression, the parameter on X will capture their joint trending behaviour.

If, on the other hand, we have co-integration then we do have a long-run relationship between our $I(1)$ variables X_t and Y_t. We have a linear combination of $I(1)$ variables that becomes stationary $I(0)$. For example, if X_t and Y_t are $I(1)$ processes, but $Y_t - \beta X_t$ is stationary, we can say X and Y are co-integrated, meaning there is a long-run relationship between these two. In that case, there will be an ECM representation – see part (c).

If we have a spurious relationship, the OLS parameter estimates will not be consistent, the R^2 will be high and the DW test statistic small (as evidence of strong dependence in the errors). For a co-integrating relationship, the OLS parameter estimates will be consistent.

(b) Candidates were expected to perform the augmented version of the Dickey–Fuller test (the ADF test). To test for co-integration, candidates were expected to evaluate whether the disturbance term is a stationary process (co-integrating) or contains unit roots (spurious). With this intuition in mind, candidates should be able to perform an ADF test on any type of model equations. The following steps were expected.

We run the above equation to obtain the residuals:

$$\widehat{u}_t = Y_t - \widehat{\beta}_0 - \widehat{\beta}_1 Y_{t-1} - \widehat{\beta}_2 X_t - \widehat{\beta}_3 X_{t-1}.$$

The ADF specification we can use to test for a spurious relationship is given by:

$$\Delta\widehat{u}_t = \rho\widehat{u}_{t-1} + \delta\Delta\widehat{u}_{t-1} + e_t$$

where by having added sufficient lagged differences we ensure that e_t does not exhibit autocorrelation.

We now test $\mathrm{H}_0 : \rho = 0$ (spurious, nonstationary) against $\mathrm{H}_1 : \rho < 0$ (co-integrating, stationary). The test statistic is:

$$\frac{\widehat{\rho}}{\mathrm{se}(\widehat{\rho})}$$

and we reject at the 5% level of significance if this value is smaller than -2.86.

(c) Candidates were expected to discuss the implication of the existence of an ECM in the model before discussing the error correction model.

If there is an ECM model for X and Y, then X and Y must be co-integrated and the test conducted in (b) should have rejected the null hypothesis. The error correction model can be written as:

$$\Delta Y_t = (\beta_1 - 1)\left(Y_{t-1} - \frac{\beta_0}{1-\beta_1} - \frac{\beta_2+\beta_3}{1-\beta_1}X_{t-1}\right) + \beta_2\Delta X_t + u_t$$

where:

- the error correction term is:

$$Y_{t-1} - \frac{\beta_0}{1-\beta_1} - \frac{\beta_2+\beta_3}{1-\beta_1}X_{t-1}$$

  measuring the disequilibrium at time $t-1$
- $(\beta_1 - 1)$ indicates how fast the two processes are coming back to the steady state
- β_2 is the direct effect of a change in X on Y.

(d) Candidates were expected to discuss the following Engel–Granger two-step estimation.

Firstly, we estimate the co-integrating relationship between X and Y by OLS to obtain residuals:

$$\widehat{diseq}_{t-1} = Y_{t-1} - \widehat{\gamma}_1 - \widehat{\gamma}X_{t-1}.$$

Secondly, using $\widehat{diseq}_{t-1}$ we run OLS on the error correction equation:

$$\Delta Y_t = (\beta_1 - 1)\widehat{diseq}_{t-1} + \beta_2\Delta X_t + u_t.$$

Comments:

- In part (a), only a few candidates clearly distinguished the two concepts and their implications for an OLS estimator (particularly the consistency).
- In part (b), several candidates used the Dickey–Fuller version of the test (without the lagged differences of the residuals) instead of the augmented version. Some candidates used Y_t instead of the residuals $\widehat{u}_t$.
- Candidates did very well in parts (c) and (d). A few made mistakes when rewriting the ECM model and failed to interpret the error correction term.

Question 9

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.**

(5 marks)

(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).**

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

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.

(5 marks)

(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.**

(5 marks)

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

(5 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Sections 2.6, 9.1, and 9.3.

Dougherty, C. Subject guide (2016): Chapters 8 and 9, and Section 15.15.

Approaching the question

(a) To derive the reduced form for P_t^w, candidates were expected to explicitly combine the demand curve and supply curve and set the quantity of demand equal to the quantity of supply as follows:

$$\alpha_0 + \alpha_1 P_t^w + \alpha_2 P_t^b + \alpha_3 Y_t + \alpha_4 A_t + u_t = \beta_0 + \beta_1 P_t^w + \beta_3 S_t + v_t.$$

We then rewrite to get an expression of P_t^w in terms of the exogenous regressors and errors only:

$$\begin{aligned}P_t^w &= \frac{\beta_0 + \beta_3 S_t + v_t - \alpha_0 - \alpha_2 P_t^b - \alpha_3 Y_t - \alpha_4 A_t - u_t}{\alpha_1 - \beta_1} \\ &= \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1} + \frac{\beta_3}{\alpha_1 - \beta_1} S_t - \frac{\alpha_2}{\alpha_1 - \beta_1} P_t^b - \frac{\alpha_3}{\alpha_1 - \beta_1} Y_t - \frac{\alpha_4}{\alpha_1 - \beta_1} A_t + \frac{v_t - u_t}{\alpha_1 - \beta_1}.\end{aligned}$$

Candidates were expected to discuss the fact that $\alpha_1 \neq \beta_1$, which allows us to derive the last step (demand/supply slope different). Therefore, the reduced form equation equals:

$$P_t^w = \pi_0 + \pi_1 S_t + \pi_2 P_t^b + \pi_3 Y_t + \pi_4 A_t + V_t$$

where π_j are our reduced-form parameters.

(b) Candidates were expected to discuss the endogeneity present in the structural equations and relate the problems with the estimation model. Points for discussion are the following.

The key reason that the OLS gives rise to these counterintuitive results is the endogeneity of P_t^w, which renders our OLS inconsistent and biased. Q_t and P_t^w are jointly determined in the system of simultaneous equations, and this joint determination results in the endogeneity problem as $\text{Cov}(P_t^w, u_t) \neq 0$.

Our parameter estimates above therefore suffer from simultaneity bias which may indeed even lead to parameter estimates of the wrong sign. Using the result from (a), we have:

$$\text{Cov}(P_t^w, u_t) = \text{Cov}(\pi_0 + \pi_1 S_t + \pi_2 P_t^b + \pi_3 Y_t + \pi_4 A_t + V_t, u_t)$$

where:

$$V_t = \frac{v_t - u_t}{\alpha_1 - \beta_1}.$$

As we are told that S_t, P_t^b, Y_t, and A_t are exogenous, we get:

$$\text{Cov}(P_t^w, u_t) = \text{Cov}\left(\frac{v_t - u_t}{\alpha_1 - \beta_1}, u_t\right) = \frac{1}{\alpha_1 - \beta_1}(\sigma_{vu} - \sigma_u^2) \neq 0.$$

This result demonstrates the correlation between the regressor and the error term in the structural equations.

(c) Candidates were expected to give clear definitions of over-identification and discuss how it differs from under-identification and exact-identification. Excellent answers should specifically recognise that two parameters in the demand equation could be made zero for the supply equation to become exactly identified.

The supply equation indeed is overidentified because we have three instrumental variables P_t^b, Y_t and A_t ($k = 3$) and only one endogenous variable P_t^w ($G - 1 = 1$). The scenario that the number of instrumental variables is larger than the number of endogenous variables is regarded as overidentification.

If in the demand equation we have, say, $\alpha_3 = \alpha_4 = 0$, the supply equation becomes exactly identified, because then we would only have one instrument P_t^b ($k = 1$) available for the endogenous variable P_t^w ($G - 1 = 1$). Other assumptions may be offered as well.

Set the number of instrumental variables as K and the number of endogenous variables as E. We can distinguish the three cases by the following criteria:

(1) $K > E$ means overidentification.

(2) $K = E$ means exact identification.

(3) $K < E$ means underidentification.

(d) Candidates were expected to perform a Two Stage Least Squares estimation using all of the available instruments as follows.

First, we estimate the reduced form of P_t^w and get the fitted value $\widehat{P}_t^w$:

$$\widehat{P}_t^w = \widehat{\pi}_0 + \widehat{\pi}_1 S_t + \widehat{\pi}_2 P_t^b + \widehat{\pi}_3 Y_t + \widehat{\pi}_4 A_t.$$

Second, we plug $\widehat{P}_t^w$ instead of P_t^w into the supply equation and run OLS to estimate the β parameters, i.e. we have:

$$Q_t = \beta_0 + \beta_1 \widehat{P}_t^w + \beta_2 S_t + \varepsilon_t.$$

Alternatively, the second step would involve performing an IV estimation on the supply equation where we use $\widehat{P}_t^w$ as instrument for P_t^w. Candidates could use the IV estimate formula at this step. Notice that since the second stage equation also includes S_t, candidates should not use the simple linear regression formula, which is correct only in the bivariate case.

Comments:

- In part (a), only a few candidates received full marks for this question. Most candidates failed to explicitly state that $\alpha_1 \neq \beta_1$ so that the reduced form equation exists. Others failed to clearly notate their derivations, in particular explicitly stating the first step of equating the demand and supply equations.
- In part (b), some candidates did not explicitly mention the bias caused by the endogeneity (note a very small sample in this question) and discuss the consistency of the estimator instead. This discussion still received marks. Most candidates failed to use the result in (a) to support their discussion of the correlation between the error term and the regressor.
- In part (c), some candidates discussed the assumptions for a valid instrument but failed to explicitly explain when the supply equation can be exactly identified. Some candidates mistook the number of exogenous variables and endogenous variables available in the model even though they correctly identified that the model was over-identified.
- In part (d), most candidates did very well on this question. There were three common mistakes. First, some candidates omitted some variables from the first-stage regression. Second, a handful of candidates estimated the demand and supply equations instead. Third, instead of running the first stage as described above, some candidates constructed a new instrument (Z_t) as the linear combination of the instruments ($Z_t = \alpha S_t + \beta P_t^b + \gamma Y_t$) to maximise the correlation between this new instrument and the endogenous variable, i.e. max corr(Z_t, P_t^w). The candidates then used this instrument Z_i in a simple linear regression formula with Q_t as the second stage regression. While the idea underlying this method provides a consistent estimator, the candidates failed to discuss how such maximisation could be done. Furthermore, this is not the TSLS procedure.

Question 10

To investigate the relationship between the price of wine and consumption of wine, an economist runs the following regression on a sample of 32 individuals for one week in 2013:

$$\widehat{\log(wine)} = \underset{(0.8911)}{4.2514} - \underset{(0.0031)}{0.8328} \log(price)$$

$$n = 32, \; R^2 = 0.89.$$

***wine* denotes the amount of wine consumed per week in millilitres (a medium glass contains 175ml) and *price* denotes the average price of a selection of wines during the week in GBP (£). The numbers in parentheses are the standard errors.**

(a) **Discuss what would happen to the parameter estimate of the slope coefficient if we had measured the amount of wine consumed per week in number of medium glasses instead of millilitres. Explain your answer.**

(3 marks)

(b) **You are asked to test the hypothesis that the demand for wine has an elasticity equal to −1 against a two-sided alternative, using a 5% level of significance. Clearly stating any assumptions you may need, carry out this test.**

(5 marks)

(c) **Construct a 95% confidence interval for the price elasticity of demand and discuss how this interval can be used to carry out the test in (b).**

(5 marks)

(d) **A famous TV chef suggests in a talk show that the demand for wine is less elastic (i.e. less negative) for people who have eaten at a restaurant during the week, arguing that eating in a restaurant encourages people to drink wine regardless of the price. To test this theory, the economist defines a dummy variable D_i that takes the value 1 if individual i ate at a restaurant during the week, and 0 otherwise. She obtains the following regression result:**

$$\log(\widehat{wine}) = \underset{(0.8911)}{4.2133} - \underset{(0.0028)}{0.8218}\log(price) + \underset{(0.0011)}{0.0889}D \times \log(price)$$

$$n = 32, \; R^2 = 0.92.$$

i. **How does this regression help in assessing the TV chef's claim?**

(3 marks)

ii. **Conduct a test that may offer support for the TV chef's claim. Clearly specify the null and the alternative hypothesis. Explain clearly why your test is effective in answering the question of interest.**

(4 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Sections R12, R13, 1.4, 2.6, 4.2 and 5.2.

Dougherty, C. Subject guide (2016): Chapters 2, 4 and 5.

Approaching the question

(a) Candidates were expected to realise that the slope coefficient can be interpreted as the price elasticity of wine consumption and hence changes in measurement will not affect its interpretation. An excellent discussion could follow the points below.

If we measure the amount of wine consumed per week in medium glasses instead of millilitres, nothing will happen to the parameter estimate of the slope coefficient since this coefficient refers to the elasticity of wine consumption with respect to price. A 1.0% increase in price will result in a 0.8328% decline in amount of wine consumed. This elasticity does not vary with the way we measure the wine consumption (or prices for that matter).

(b) Candidates were expected to perform a two-sided test with all necessary components of a testing procedure. An excellent answer should cover the following points.

Suppose the true model is:

$$\log(wine) = \alpha + \beta \log(price) + u.$$

We perform a two-sided test as follows. We formulate the null hypothesis as $\mathrm{H}_0 : \beta = -1$ against $\mathrm{H}_1 : \beta \neq -1$. Under the Gauss–Markov assumptions, whose details must be provided somewhere in the script, and under normality of the errors assumption, we can use the t test. The test statistic is:

$$\frac{\widehat{\beta}+1}{\mathrm{SE}(\widehat{\beta})} \sim t_{30}$$

under H_0. Decision rule: we should reject at the 5% level of significance if:

$$\left|\frac{\widehat{\beta}+1}{\mathrm{SE}(\widehat{\beta})}\right| > 2.042.$$

Because our t statistic equals $(-0.8328+1)/0.0031 = 53.9355$, which is far larger than any reasonable critical value, we reject the null hypothesis, in favour of the alternative hypothesis.

(c) Candidates were expected to compute the confidence interval with the appropriate critical value, then decide whether we could reject the null hypothesis specified in part (b). An excellent answer should cover the following points.

Using the critical value of 2.042 for $n = 32$, the 95% confidence interval is given by:

$$\left[\widehat{\beta} - 2.042 \times \mathrm{SE}(\widehat{\beta}), \widehat{\beta} + 2.042 \times \mathrm{SE}(\widehat{\beta})\right].$$

Plugging in the numbers, we have:

- the lower bound: $-0.8328 - 2.042 \times 0.0031 = -0.839$
- the upper bound: $-0.8328 + 2.042 \times 0.0031 = -0.826$.

Therefore, the 95% confidence interval equals $[-0.839, -0.826]$. Note that we are using the critical value of 2.042 for the sample size of 30 instead of 1.96, which is appropriate for a large (infinite) sample size.

As -1 does not lie in this interval, we should reject H_0 at the 5% level of significance.

(d) Candidates were expected to discuss the interaction term in part i. and perform a one-sided t test in part ii.

i. Rewrite the regression equation in the following way:

$$\log(wine_i) = \alpha + \beta_1 \log(price_i) + \beta_2 D_i \log(price_i) + \varepsilon_i.$$

β_2 allows us to detect whether there are differences in the elasticity for individuals who went to the restaurant and those who did not. Candidates could provide the following expressions to support the discussion:

$$\mathrm{E}(\log(wine) \mid D = 1, \log(price)) = \alpha + (\beta_1 + \beta_2) \log(price) + \varepsilon$$

$$\mathrm{E}(\log(wine) \mid D = 0, \log(price)) = \alpha + \beta_1 \log(price) + \varepsilon.$$

ii. Notice that the TV chef clearly indicated the direction of the elasticity in their claim (less negative). Hence candidates were expected to perform a one-sided test to evaluate the claim of the TV chef instead of a two-sided test. Candidates were also expected to describe all components of the testing procedure.

We should test the null hypotheses: $\mathrm{H}_0 : \beta_2 = 0$, against the one-sided alternative $\mathrm{H}_1 : \beta_2 > 0$. Candidates lost marks for an alternative hypothesis for a two-sided test.

Under the Gauss–Markov assumptions and the normality assumption, we can use the t test. The t statistic and the test statistic distribution are:

$$\frac{\widehat{\beta}_2}{\mathrm{SE}(\widehat{\beta}_2)} \sim t_{n-3}.$$

We should reject the null hypothesis at the 5% level of significance if:

$$\frac{\widehat{\beta}_2}{\mathrm{SE}(\widehat{\beta}_2)} > 1.699.$$

Interpretation: Because:

$$\frac{\widehat{\beta}_2}{\text{SE}(\widehat{\beta}_2)} = \frac{0.0889}{0.0011} = 80.818 > 1.699$$

we reject the null hypothesis, in favour of the alternative hypothesis. We found evidence that the TV chef's claim is true in our sample.

Comments:

- In part (a), candidates could provide either a discussion on the elasticity interpretation of the estimate or a formal derivation of the new regression equation. If the latter method is chosen, careful steps should be taken to derive the new intercept correctly, which should go up as a result of the transformation. Some candidates derived the new regression equation but mistakenly interpreted the new intercept and had it gone down.
- In part (b), many candidates lost marks for not explicitly stating the Gauss–Markov assumptions (or forgetting the normality assumption).
- In part (c), there were two common mistakes. First, some candidates used the critical value 1.96 instead of 2.042 for a small sample of $n = 32$. Second, many candidates did not discuss the use of the confidence interval in deciding to reject/not reject the null hypothesis. Some used the default value of zero instead of the estimate of interest at hand (-1).
- In part (d) i., most candidates correctly commented that the interaction term helps indicate the differences between those who went to the restaurant and those who did not. In part ii. candidates struggled. Many used a two-sided t test while a one-sided test was expected; others omitted some components of the testing procedure. Some candidates opted for an F test, which tested whether the unrestricted model with the interaction $D_i \log(price_i)$ was equivalent to the restricted model without the interaction. This method does not directly answer the question at hand.