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 $\mathrm{Cov}(X, Y)$ and $\mathrm{Var}(X)$ should be restricted to describing the population covariance and variances, respectively, with definitions:

$$\mathrm{Cov}(X, Y) = \mathrm{E}\left((X - \mathrm{E}(X))(Y - \mathrm{E}(Y))\right) = \mathrm{E}(XY) - \mathrm{E}(X)\,\mathrm{E}(Y)$$

and:

$$\mathrm{Var}(X) = \mathrm{E}((X - \mathrm{E}(X))^2) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2$$

(you also may denote $\mathrm{Cov}(X, Y) = \sigma_{XY}$ and $\mathrm{Var}(X) = \sigma_X^2$). 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:

$$\widehat{\beta} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \frac{\text{Sample Cov}(X, Y)}{\text{Sample Var}(X)} \neq \frac{\mathrm{Cov}(X, Y)}{\mathrm{Var}(X)}.$$

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 $\mathrm{E}(u_i)$. 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:

$$\widehat{\beta} = \beta + \frac{\sum_{i=1}^{n}(X_i - \bar{X})u_i}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \beta + \frac{\sum_{i=1}^{n} x_i u_i}{\sum_{i=1}^{n} x_i^2} \qquad \text{with } x_i = X_i - \bar{X}.$$

- 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 $\mathrm{E}(X_i) = 1/\lambda$ and $\mathrm{Var}(X_i) = 1/\lambda^2$ 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 $\text{plim}(1/\bar{X}) = 1/(\text{plim}\ \bar{X})$. 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:

$$\left.\frac{\partial \ln L}{\partial \lambda}\right|_{\widehat{\lambda}_{MLE}} = \frac{n}{\widehat{\lambda}_{MLE}} - \sum_{i=1}^{n} X_i = 0 \qquad \text{or} \qquad \widehat{\lambda}_{MLE} = \frac{1}{\bar{X}} \qquad \text{with} \qquad \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

(It should clearly state $\widehat{\lambda}_{MLE} = 1/\bar{X}$, saying $\lambda = 1/\bar{X}$ is incorrect.)

The MLE is *not unbiased*:

$$\mathrm{E}(\widehat{\lambda}_{MLE}) = \mathrm{E}\left(\frac{1}{\bar{X}}\right) \overset{\text{Jensen's Ineq}}{\neq} \frac{1}{\mathrm{E}(\bar{X})} = \frac{1}{\frac{1}{n}\sum \mathrm{E}(X_i)} = \lambda.$$

The MLE is *consistent*:

$$\text{plim}\ \widehat{\lambda}_{MLE} \overset{\text{plim rules}}{=} \frac{1}{\text{plim}(\bar{X})} \overset{\text{LLN}}{=} \frac{1}{\mathrm{E}(X_i)} = \lambda.$$

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 $|\rho| < 1$.
- $DW = \dfrac{\sum_{t=2}^{T} (\widehat{u}_t - \widehat{u}_{t-1})^2}{\sum_{t=2}^{T} \widehat{u}_t{}^2} \simeq 2(1 - \widehat{\rho})$, 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 $\widehat{\rho} = \frac{\sum_{t=2}^{T} \widehat{u}_t \widehat{u}_{t-1}}{\sum_{t=2}^{T} \widehat{u}_t{}^2}$.
- Step 3: Estimate the transformed model:

$$\begin{aligned} & gprice_t - \widehat{\rho} gprice_{t-1} \\ & \quad = \beta_0 (1 - \widehat{\rho}) + \beta_1 (gwage_t - \widehat{\rho} gwage_{t-1}) + \beta_2 (gwage_{t-1} - \widehat{\rho} gwage_{t-2}) + v_t \\ & \qquad \text{for } t = 3, \ldots, T \qquad \text{(lose one observation).} \end{aligned}$$

  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 $\widehat{\beta}/\text{SE}(\widehat{\beta}) = -0.425/0.176 = -2.415$. 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 $\text{H}_0 : \beta_{wave2} = \beta_{wave3} = 0$ against $\text{H}_1 : \beta_{wave2} \neq 0$ 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: $\Lambda(z) = \dfrac{1}{1+\exp(-z)}$.

$\widehat{\Pr}(arr86 = 1 \mid pcnv = 0.25, avgsen = 1, tottime = 1, ptime86 = 0, qemp86 = 2) = \Lambda(z_1)$, 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.

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:

$$\log W_i = \alpha_1 + \alpha_2 S_i + \gamma Ability_i + \varepsilon_i$$

we obtain:

$$\widehat{\alpha}_2 = \alpha_2 + \frac{\text{Sample Cov}(S_i, Abilty_i)}{\text{Sample Var}(S_i)}\gamma + \frac{\text{Sample Cov}(S_i, \varepsilon_i)}{\text{Sample Var}(S_i)}.$$

To consider the bias, we take expectations. *Assuming that S_i and $Abilty_i$ are non-stochastic:*

$$\text{E}(\widehat{\alpha}_2) = \alpha_2 + \underbrace{\frac{\text{Sample Cov}(S_i, Abilty_i)}{\text{Sample Var}(S_i)}\gamma}_{\text{Omitted Variable Bias}} \neq \alpha_2$$

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*:

$$\begin{aligned}
\text{plim } \widehat{\alpha}_2 &= \alpha_2 + \frac{\text{plim Sample Cov}(S_i, Abilty_i)}{\text{plim Sample Var}(S_i)}\gamma + \frac{\text{plim Sample Cov}(S_i, \varepsilon_i)}{\text{plim Sample Var}(S_i)} \\
&= \alpha_2 + \frac{\text{Cov}(S_i, Abilty_i)}{\text{Var}(S_i)}\gamma + \frac{\text{Cov}(S_i, \varepsilon_i)}{\text{Var}(S_i)} \\
&= \alpha_2 + \frac{\text{Cov}(S_i, Abilty_i)}{\text{Var}(S_i)}\gamma
\end{aligned}$$

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*:

$$\mathrm{E}(\widehat{\lambda}_{MLE}) = \mathrm{E}\left(\frac{1}{\bar{X}}\right) \overset{\text{Jensen's Ineq}}{\neq} \frac{1}{\mathrm{E}(\bar{X})} = \frac{1}{\frac{1}{n}\sum \mathrm{E}(X_i)} = \lambda.$$

The MLE is *consistent*:

$$\operatorname{plim} \widehat{\lambda}_{MLE} \overset{\text{plim rules}}{=} \frac{1}{\operatorname{plim}(\bar{X})} \overset{\text{LLN}}{=} \frac{1}{\mathrm{E}(X_i)} = \lambda.$$

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 $\boldsymbol{\varepsilon_t}$ i.i.d. $\boldsymbol{(0, \sigma^2)}$. Indicate (with explanation) the source(s) of non-stationarity of $\boldsymbol{y_t}$. 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 $\boldsymbol{(\gamma_0, \gamma_1, \rho)}$ and $\boldsymbol{(\beta_0, \beta_1, \beta_2)}$. What name do we give the non-stationary process $\boldsymbol{y_t}$ when $\boldsymbol{\rho = 1}$? 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:

$$\ln \widehat{\text{Price}}_i = \underset{(0.612)}{-8.427} + \underset{(0.091)}{1.334} \ln \text{Area}_i - \underset{(0.128)}{0.165} \text{Aspect Ratio}_i$$

$$N = 430, R^2 = 0.336$$

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

$$\begin{aligned} \widehat{hours} &= \underset{(6.503)}{-17.347} - \underset{(.116)}{2.242} kids + \underset{(.035)}{.938}\, educ + \underset{(.453)}{2.089} age - \underset{(.008)}{.028}\, age^2 - \underset{(.005)}{.075}\, nonmomi \\ n &= 31{,}857 \text{ and } R^2 = .052 \end{aligned} \tag{7.1}$$

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:**

$$\begin{aligned}\widehat{hours} &= \underset{(7.020)}{-16.828} - \underset{(.134)}{2.504}kids + \underset{(.115)}{.916}\,educ + \underset{(.460)}{2.105}age \underset{(.008)}{-.028}\,age^2 - \underset{(.007)}{.076}\,nonmomi \\ n &= 31,857 \text{ and } R^2 = .052 \end{aligned} \qquad (7.2)$$

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 $H_0 : \delta_5 = \delta_6 = 0$) 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: $\Lambda(z) = \dfrac{1}{1+\exp(-z)}$.

$\widehat{\Pr}(arr86 = 1 \,|\, pcnv = 0.25, avgsen = 1, tottime = 1, ptime86 = 0, qemp86 = 2) = \Lambda(z_1)$, 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.

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:**

$$\begin{aligned}\widehat{gfr}_t \;=\;& \underset{(3.27)}{92.52} + \underset{(.119)}{.101}\,pe_t - \underset{(.147)}{.011}\,pe_{t-1} + \underset{(.119)}{.033}\,pe_{t-2} \\ & - \underset{(10.13)}{22.95}\,ww2_t - \underset{(3.76)}{30.83}\,pill_t, \qquad R^2 = .536,\; T = 68.\end{aligned} \tag{9.2}$$

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:

$$u_t = \rho u_{t-1} + e_t \qquad \text{with } |\rho| < 1 \text{ and } e_t \text{ white noise (i.i.d. } (0, \sigma^2)\text{)}.$$

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 $\boldsymbol{y_0 = 0}$. $\mathbf{E}(\boldsymbol{u_t}) = \mathbf{0}$, $\mathbf{E}(\boldsymbol{u_t^2}) = \boldsymbol{\sigma^2}$ and $\mathbf{E}(\boldsymbol{u_t u_s}) = \mathbf{0}$ 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 $\widehat{\gamma}/\mathrm{SE}(\widehat{\gamma}) < \tau$, 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*.