Examiners' commentaries 2019

EC2020 Elements of econometrics

Important note

This commentary reflects the examination and assessment arrangements for this course in the academic year 2018–19. The format and structure of the examination may change in future years, and any such changes will be publicised on the virtual learning environment (VLE).

Information about the subject guide and the Essential reading references

Unless otherwise stated, all cross-references will be to the latest version of the subject guide (2016). You should always attempt to use the most recent edition of any Essential reading textbook, even if the commentary and/or online reading list and/or subject guide refer to an earlier edition. If different editions of Essential reading are listed, please check the VLE for reading supplements – if none are available, please use the contents list and index of the new edition to find the relevant section.

General remarks

Learning outcomes

At the end of the course, and having completed the Essential reading and activities, you should be able to:

- describe and apply the classical regression model and its application to cross-section data
- describe and apply the:
  - Gauss–Markov conditions and other assumptions required in the application of the classical regression model
  - reasons for expecting violations of these assumptions in certain circumstances
  - tests for violations
  - potential remedial measures, including, where appropriate, the use of instrumental variables
- recognise and apply the advantages of logit, probit and similar models over regression analysis when fitting binary choice models
- competently use regression, logit and probit analysis to quantify economic relationships using standard regression programmes (Stata and EViews) in simple applications
- describe and explain the principles underlying the use of maximum likelihood estimation
- apply regression analysis to time-series models using stationary time series, with awareness of some of the econometric problems specific to time series applications (for example, autocorrelation) and remedial measures
- recognise the difficulties that arise in the application of regression analysis to nonstationary time series, know how to test for unit roots, and know what is meant by cointegration.

Common mistakes committed by candidates

A large number of candidates are not able to clearly distinguish between sample variance and covariance, and population variance and covariance (this is happening year after year).

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

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

and:

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

(you also may denote $\text{Cov}(X,Y) = \sigma_{XY}$ and $\text{Var}(X) = \sigma_X^2$). They are typically unknown, but fixed, quantities.

The sample covariance and variance are estimators of the population covariance and variance, respectively. They are defined as:

$$\text{Sample Cov}(X,Y) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})$$

and:

$$\text{Sample Var}(X) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$$

(you also may use $\widehat{\sigma}_{XY}$ and $\widehat{\sigma}_X^2$). You can compute them given the data.

With a slight abuse of notation, we often divide by n instead, which is irrelevant if we let n be large. The division by $n-1$ is a finite sample issue only (unbiasedness).

The sample covariance and variance show up in our definition of the OLS estimator of the slope in the simple linear regression model, not the population covariance and variance, as:

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

Treating them as being the same results in incorrect analyses and candidates losing significant marks.

Candidates should realise that $\frac{1}{n}\sum_{i=1}^{n} u_i$ is not the same as $\text{E}(u_i)$. So, while we typically assume $\text{E}(u_i) = 0$, this does not guarantee that $\frac{1}{n}\sum_{i=1}^{n} u_i = 0$. Also, while we may be happy to assume $\text{E}(x_i u_i) = 0$ (uncorrelatedness between the errors and regressors), this does not guarantee that $\frac{1}{n}\sum_{i=1}^{n} x_i u_i = 0$. Note that:

- both $\frac{1}{n}\sum_{i=1}^{n} u_i$ and $\frac{1}{n}\sum_{i=1}^{n} x_i u_i$ are random variables, which take the value 0 with probability 0 (continuous random variables)!
- $\text{E}(u_i) = 0$ and $\text{E}(x_i u_i) = 0$ are fixed, not stochastic!

The differences between sample and population moments need to come across clearly when looking at unbiasedness and making consistency arguments. In both cases, we first simplify our estimator (plug in the true model) to obtain:

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

- For *unbiasedness*, clearly indicate that you want to show that $\mathrm{E}(\widehat{\beta}) = \beta$. Unbiasedness does not follow from $\sum\limits_{i=1}^{n} x_i u_i = 0$, instead it follows from $\mathrm{E}\left(\frac{\sum\limits_{i=1}^{n} x_i u_i}{\sum\limits_{i=1}^{n} x_i^2}\right) = 0$.

  If we treat x_i as fixed, $\mathrm{E}\left(\frac{\sum\limits_{i=1}^{n} x_i u_i}{\sum\limits_{i=1}^{n} x_i^2}\right) \equiv \mathrm{E}\left(\sum\limits_{i=1}^{n} d_i u_i\right) = \sum\limits_{i=1}^{n} d_i\,\mathrm{E}(u_i)$ and then unbiasedness follows as $\mathrm{E}(u_i) = 0$.
- For *consistency*, clearly indicate that you want to show that $\mathrm{plim}(\widehat{\beta}) = \beta$. Using the plim properties, we show:

$$
\begin{aligned}
\mathrm{plim}\ \widehat{\beta} = \beta + \mathrm{plim}\left(\frac{\sum\limits_{i=1}^{n} x_i u_i}{\sum\limits_{i=1}^{n} x_i^2}\right) &= \beta + \frac{\mathrm{plim}\left(\frac{1}{n}\sum\limits_{i=1}^{n} x_i u_i\right)}{\mathrm{plim}\left(\frac{1}{n}\sum\limits_{i=1}^{n} x_i^2\right)} \\
&\equiv \beta + \frac{\mathrm{plim}\,(\text{Sample Cov}\,(x, u))}{\mathrm{plim}\,(\text{Sample Var}\,(x))} \\
&= \beta + \frac{\mathrm{Cov}\,(x, u)}{\mathrm{Var}\,(x)} \quad \text{using the law of large numbers}
\end{aligned}
$$

  where $\mathrm{Cov}(x, u) = 0$ and $\mathrm{Var}(x) > 0$, ensuring we get consistency.
- Remember, the law of large numbers ensures that sample averages converge to their population analogues.

Candidates struggled to give competent answers to the interpretation of empirical results. When interpreting an empirical result you should discuss the significance of the coefficients, magnitude and sign of the coefficients.

Candidates particularly struggled with the difference between percentage points and changes when it comes to the interpretation of coefficients in binary choice models. A coefficient in such a model gives the percentage point difference in the outcome variable when the regressor changes by 1 unit. The % change is a *relative* change which depends on the baseline value of the outcome variable (for example, the realisation when a regressor equals zero in the case of a dummy variable regressor). In the case of dummy variables as regressors, candidates often missed that the coefficients have to be interpreted relative to the base case which is the left out dummy variable (see appropriate questions in the examinations for details).

When conducting hypothesis tests, you should make sure that the Gauss–Markov conditions hold. The Gauss–Markov conditions have to be explicitly specified. Only writing that the Gauss–Markov conditions hold is not sufficient. As good practice, begin your examination by explicitly providing the Gauss–Markov conditions. You can then refer back to them thereafter. Moreover, ensure when conducting hypothesis testing that you clearly indicate the null and alternative hypotheses (in terms of the true parameters, say β_1), the test statistic (in terms of the parameter estimates, here $\widehat{\beta}_1$), its distribution (with degrees of freedom), the rejection rule (one-sided or-two sided) for a given significance level (typically 5%) with suitable critical values, and provide an interpretation of your result.

Exogeneity and endogeneity are often very simply explained as variables 'outside of the model' or 'inside the model', respectively. While this is not wrong, it is not a particularly clear or helpful definition, in particular in general contexts which are not related to simultaneous equations models. Defining these concepts in terms of the relationship between regressor and error term is much clearer and naturally leads to a way to show exogeneity or endogeneity by considering the covariance between regressors and error terms.

Candidates often missed that all exogenous regressors and instruments have to be included in the first stage when performing TSLS estimation.

It was very common to misunderstand the central limit theorem. While it provides the distribution of the *estimator* $\widehat{\beta}$, no matter what the true distribution of the errors is, it does not say that the distribution of the error term is approximately normal in large samples.

When asked about identification of structural form equations, candidates often compared the number of excluded regressors to the number of endogenous variables in general, without pointing out which variables are the excluded regressors and could, therefore, be used as an instrument.

Just as last year, many candidates do not answer all parts of the question. Make sure you read the questions properly and provide all details that are requested. Not answering a question will automatically earn you a zero mark for that question.

Key steps to improvement

Essential reading for this course includes the subject guide and the following:

- Dougherty, C. *Introduction to econometrics.* (Oxford: Oxford University Press, 2016) 5th edition [ISBN 9780199676828]; http://oxfordtextbooks.co.uk/orc/dougherty5e/

Apart from the Essential readings you should do some supplementary reading. One very good book at the same level is:

- Gujarati, D.N. and D.C. Porter *Basic econometrics.* (McGraw–Hill, 2009, International edition) 5th edition [ISBN 9780071276252].

To understand the subject clearly it is important to supplement Dougherty's *Introduction to econometrics* (fifth edition) with the subject guide **EC2020 Elements of econometrics** (2016), especially Chapter 10 which covers maximum likelihood estimation. It is very important to carefully go through the subject guide. The subject guide contains solutions to the questions given in the main textbook and also some additional questions and solutions. Working through these will improve your understanding of the subject.

The chapter in the subject guide on maximum likelihood (Chapter 10) includes some additional theory which has not been covered in the main textbook. It is important to read the additional theory given in the subject guide to have a better understanding of the principles of maximum likelihood and tests based on the likelihood function.

Please check the VLE course page for resources for this subject such as a downloadable copy of the subject guide **EC2020 Elements of econometrics** (2016), PowerPoint slideshows that provide a graphical treatment of the topics covered in the textbook, datasets and statistical tables. Candidates should utilise datasets using standard regression programmes (STATA or EViews). This will help in the understanding of the subject.

Examination revision strategy

Many candidates are disappointed to find that their examination performance is poorer than they expected. This may be due to a number of reasons, but one particular failing is '**question spotting**', that is, confining your examination preparation to a few questions and/or topics which have come up in past papers for the course. This can have serious consequences.

We recognise that candidates might not cover all topics in the syllabus in the same depth, but you need to be aware that examiners are free to set questions on **any aspect** of the syllabus. This means that you need to study enough of the syllabus to enable you to answer the required number of examination questions.

The syllabus can be found in the Course information sheet available on the VLE. You should read the syllabus carefully and ensure that you cover sufficient material in preparation for the examination. Examiners will vary the topics and questions from year to year and may well set questions that have not appeared in past papers. Examination papers may legitimately include questions on any topic in the syllabus. So, although past papers can be helpful during your revision, you cannot assume that topics or specific questions that have come up in past examinations will occur again.

If you rely on a question-spotting strategy, it is likely you will find yourself in difficulties when you sit the examination. We strongly advise you not to adopt this strategy.

Examiners' commentaries 2019

EC2020 Elements of econometrics

Important note

This commentary reflects the examination and assessment arrangements for this course in the academic year 2018–19. The format and structure of the examination may change in future years, and any such changes will be publicised on the virtual learning environment (VLE).

Information about the subject guide and the Essential reading references

Unless otherwise stated, all cross-references will be to the latest version of the subject guide (2016). You should always attempt to use the most recent edition of any Essential reading textbook, even if the commentary and/or online reading list and/or subject guide refer to an earlier edition. If different editions of Essential reading are listed, please check the VLE for reading supplements – if none are available, please use the contents list and index of the new edition to find the relevant section.

Comments on specific questions – Zone A

Candidates should answer **EIGHT** of the following **TEN** questions: **ALL** of the questions in Section A (8 marks each) and **THREE** questions from Section B (20 marks each). **Candidates are strongly advised to divide their time accordingly.**

Section A

Answer all questions from this section.

Question 1

Consider the following regression model:

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 t + \varepsilon_t, \quad t = 1, \ldots, T.$$

Both $\{Y_t\}_{t=1}^{T}$ and $\{X_t\}_{t=1}^{T}$ are trend stationary processes. The errors $\{\varepsilon_t\}_{t=1}^{T}$ are independent random variables with zero mean and constant variance.

(a) **Discuss the concept 'trend stationarity' and contrast it to the concept 'difference stationarity'. In your answer make sure you also explain what stationarity means.**

(4 marks)

(b) **Provide a clear interpretation of the parameter β_1. You are told that you can obtain the estimator for β_1 using 'detrended' variables only. Discuss this statement.**

(4 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Sections 13.1 and 13.2.

Dougherty, C. Subject guide (2016): Chapter 13.

Approaching the question

(a) *Trend stationarity* relates to a process which has a deterministic trend ($y_t = \alpha + \beta t + \varepsilon_t$ where ε_t is weakly dependent). It is a type of non-stationarity which can be removed by detrending. Once we include a trend in the regression, we can use standard statistical inference methods: the errors (and, therefore, Y_t) will be weakly dependent and consistent estimates can be obtained.

Difference stationarity relates to a process which can be made stationary by taking first differences: ΔY_t, examples are random walk and random walk with drift. Here Y_t is persistent, strongly dependent, and estimation of the process Y_t on Y_{t-1} would fail to produce consistent parameter estimates.

By stationarity we mean *covariance stationarity*, which requires constant (finite) means and variances, and covariances which only depend on distance in time, i.e. $\mathrm{E}(Y_t, Y_{t-s})$ only depends on s, not on location t.

(b) β_1 is the marginal effect of X_t on Y_t after controlling for a linear time trend, that is holding time constant (*ceteris paribus*).

Frisch–Waugh–Lovell (regression anatomy) argue that including the time trend $\beta_2 t$ explicitly in the regression is the same as working with detrended data where we control for the trend beforehand (implicitly).

Since both Y_t and X_t have a linear time trend, we need to include it in the regression in order to obtain consistent estimates of β_1. Omitting the time trend in this setting would give rise to omitted variable bias (OVB).

Question 2

Consider the following ADL(1, 1) model relating the crime rate in a particular province, $crime_t$, to the clear-up rate (percentage of crimes resulting in a conviction):

$$\boldsymbol{crime_t = \alpha + \rho crime_{t-1} + \delta_1 clearup_t + \delta_2 clearup_{t-1} + u_t, \quad \text{with } |\rho| < 1}$$

where u_t is white noise, an i.i.d. innovation that is uncorrelated to anything in the past.

(a) Briefly indicate whether OLS will provide unbiased and consistent parameter estimators.

(2 marks)

(b) Derive the long run relationship between *crime* and *clearup*.

(2 marks)

(c) Rewrite the model above in terms of an error correction model (ECM) and interpret its coefficients.

(4 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Sections 11.3, 11.4 and 11.5.

Dougherty, C. Subject guide (2016): Chapter 11.

Approaching the question

(a) OLS will be consistent because the error is uncorrelated with the regressors. OLS will be biased because of the presence of the lagged endogenous variable. The latter is due to the fact that u_t will be correlated with $crime_{t+s}$ for any $s \geq 0$ (regressor $crime_{t-1}$ in future periods).

(b) Let in equilibrium $X_t = X_{t-1} = \tilde{X}$ for both variables in the model, so that:

$$\widetilde{crime} = \alpha + \rho\widetilde{crime} + \delta_1 \widetilde{clearup} + \delta_2 \widetilde{clearup}.$$

After rewriting we obtain the long-run relationship:

$$crime_t = \frac{\alpha}{1-\rho} + \frac{\delta_1 + \delta_2}{1-\rho} clearup_t.$$

(c) The error correction model states that the change in *crime* in any period will be governed by the change in *clearup* and an error correction mechanism which reveals the speed with which *crime* changes in response to a disequilibrium. Here, we get:

$$\Delta crime_t = \delta_1 \Delta clearup_t + (\rho - 1) diseq_{t-1} + u_t$$

where:

$$diseq_{t-1} = crime_{t-1} - \frac{\alpha}{1-\rho} - \frac{\delta_1 + \delta_2}{1-\rho} clearup_{t-1}$$

and $(\rho - 1)$ indicates the speed of adjustment.

Question 3

Consider the simple linear regression model:

$$\boldsymbol{Y_i = \beta X_i + u_i, \quad i = 1, \ldots, n.}$$

We assume that the errors $\{u_i\}_{i=1}^n$ are independent normal random variables with zero mean. The regressor $\{X_i\}_{i=1}^n$ is non-stochastic (fixed under repeated sampling). You suspect that the errors exhibit heteroskedasticity.

(a) Explain what we mean by the concept of heteroskedasticity. Enhance your answer with the help of a graphical illustration.

(3 marks)

(b) Suppose you want to test $H_0 : \beta = 1$ against $H_1 : \beta > 1$. Discuss how you would conduct this test based on the OLS estimator, recognising the presence of heteroskedasticity. Please provide a detailed answer.

(5 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Sections 7.1 and 7.2.

Dougherty, C. Subject guide (2016): Chapters 2 and 7.

Approaching the question

(a) Heteroskedasticity means that the error variance depends on the regressors, it violates the assumption that $E(u_i^2) = \sigma^2$.

A graphical illustration should show data points and some depiction of the variance – for example, drawing the distribution for the errors at different levels of X. The graph should be clearly labelled.

(b) We need to propose the t test where the standard error on β needs to be estimated taking the heteroskedasticity into account (the usual t statistic is invalid because of the presence of heteroskedastity):

$$T = \frac{\widehat{\beta} - 1}{\text{s.e.}(\widehat{\beta})} \sim t_{n-1} \text{ under } \text{H}_0.$$

The White robust standard error we use for this is the square root of the following heteroskedasticity-robust variance estimate:

$$\widehat{\text{Var}}(\widehat{\beta}) = \frac{\sum_{i=1}^{n} X_i^2 \widehat{u}_i^2}{\left(\sum_{i=1}^{n} X_i^2\right)^2}.$$

The test then proceeds in the standard way. Using the 5% significance level we reject H_0 if the realisation of our test statistic exceeds the critical value: reject if $t > t_{0.05,\,n-1}$ (with n large, we reject H_0 if $t > 1.645$). Otherwise, fail to reject H_0.

A good number of candidates proposed a Goldfeld–Quandt test of heteroskedasticity instead of a discussion of how to conduct hypothesis testing given heteroskedasticity which is what the question asked. It is important to carefully read what the question asks!

Question 4

Consider the simple linear regression model:

$$\boldsymbol{Y_i = \beta_0 + \beta_1 X_i + u_i, \quad i = 1, \ldots, n.}$$

We assume that the errors $\{u_i\}_{i=1}^n$ are independent random variables with zero mean. The regressor $\{X_i\}_{i=1}^n$ is non-stochastic (fixed under repeated sampling). Under these conditions, the OLS estimator for β_1, $\widehat{\beta}_1$, is unbiased. (You are not asked to derive $\widehat{\beta}_1$).

(a) **Explain the concept of unbiasedness of an estimator.**

(2 marks)

(b) **Let us consider two other estimators for the slope β_1:**

$$\widehat{\beta}_1^{\circ} = \frac{\sum_{i=1}^{n}(Z_i - \bar{Z})Y_i}{\sum_{i=1}^{n}(Z_i - \bar{Z})X_i} \quad \text{and} \quad \widehat{\beta}_1^{*} = \frac{\sum_{i=1}^{n}(Z_i - \bar{Z})Y_i}{\sum_{i=1}^{n}(Z_i - \bar{Z})Z_i}$$

where $Z_i = \sqrt{X_i}$ for all i and $\bar{Z} = \frac{1}{n}\sum_{i=1}^{n} Z_i$. Please indicate whether $\widehat{\beta}_1^{\circ}$ and $\widehat{\beta}_1^{*}$ are unbiased estimators for β_1. Clearly show your derivations.

(4 marks)

(c) **Briefly indicate how you would choose between the three estimators, $\widehat{\beta}_1$, $\widehat{\beta}_1^{\circ}$ and $\widehat{\beta}_1^{*}$.**

(2 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Sections 2.3 and 2.5.

Dougherty, C. Subject guide (2016): Chapter 2.

Approaching the question

(a) Unbiasedness means $\mathrm{E}(\widehat{\beta}_1) = \beta_1$. It means that the expected value of the estimator is the true parameter; that is, we are correct on average in repeated samples. This ensures that we will not make systematic errors when estimating β.

(b) In answering this question, it is easiest to derive results using Z_i rather than working with $Z_i = \sqrt{X_i}$. It is important to note furthermore that $\sum(Z_i - \bar{Z}) = 0$.

When plugging in the true model ($Y_i = \beta_0 + \beta_1 X_i + u_i$), we can then write:

$$\widehat{\beta}_1^{\circ} = \beta_1 + \frac{\sum(Z_i - \bar{Z})u_i}{\sum(Z_i - \bar{Z})X_i}$$

and:

$$\mathrm{E}(\widehat{\beta}_1^{\circ}) = \beta_1 + \frac{\sum(Z_i - \bar{Z})\mathrm{E}(u_i)}{\sum(Z_i - \bar{Z})X_i} = \beta_1$$

(X and Z are assumed to be non-stochastic). Unbiasedness follows as $\mathrm{E}(u_i) = 0$.

For the second estimator, when plugging in the true model, we can write:

$$\widehat{\beta}_1^{*} = \beta_1 + \frac{\sum(Z_i - \bar{Z})X_i}{\sum(Z_i - \bar{Z})Z_i} + \frac{\sum(Z_i - \bar{Z})u_i}{\sum(Z_i - \bar{Z})Z_i}$$

and:

$$\mathrm{E}(\widehat{\beta}_1^{*}) = \beta_1 + \frac{\sum(Z_i - \bar{Z})Z_i^2}{\sum(Z_i - \bar{Z})Z_i} \neq \beta_1$$

using again the fact that Z is non-stochastic and $\mathrm{E}(u_i) = 0$, hence $\widehat{\beta}_1^{*}$ is biased.

Side notes: (i) $\widehat{\beta}_1^{\circ}$ essentially is an IV estimator where Z is used as an instrument for X. Both Z and X are exogenous here. (ii) $\widehat{\beta}_1^{*}$, on the other hand, is the slope parameter of a regression of Z instead of X on Y, and hence should not give us an unbiased estimator of β_1.

Many candidates failed to realise that $\sum(Z_i - \bar{Z}) = 0$ and wrongly concluded that the first estimator $\widehat{\beta}_1^{\circ}$ is biased.

(c) When choosing between $\widehat{\beta}_1$ and $\widehat{\beta}_1^{\circ}$ (both unbiased), efficiency considerations matter. Since the model satisfies the Gauss–Markov assumptions, we know that by the Gauss–Markov theorem, $\widehat{\beta}_1$ is BLUE and should, therefore, be chosen over $\widehat{\beta}_1^{\circ}$. Nevertheless, there may be a trade-off between bias and variance, and the mean squared error of $\widehat{\beta}_1^{*}$ could be smaller than that of $\widehat{\beta}_1$ (which is unbiased).

The bias–variance trade-off was mentioned only by a few candidates.

Question 5

Consider the OLS estimator for β in the linear regression model:

$$Y_i = \beta X_i + \varepsilon_i, \quad i = 1, \ldots, n$$

where $\{(Y_i, X_i)\}_{i=1}^n$ form an i.i.d. sample from a population and the errors are drawn from an <u>unknown</u> distribution with mean zero and variance σ^2. You are told that X_i and ε_i are <u>uncorrelated</u> (not necessarily independent therefore!).

(a) Discuss the importance of convergence in probability.

(4 marks)

(b) Discuss the importance of convergence in distribution.

(4 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Sections R.14 and R.15.

Dougherty, C. Subject guide (2016): Chapter 8.

Approaching the question

(a) Convergence in probability plays an important role when proving *consistency* of estimators. A consistent estimator converges in probability to the true parameter and we either write $\text{plim}(\widehat{\beta}) = \beta$ or $\widehat{\beta} \xrightarrow{p} \beta$.

Proofs of consistency (a large sample property) are particularly important in settings where we cannot show that our estimator is unbiased (a finite sample property). For the OLS estimator:

$$\text{plim}(\widehat{\beta}) = \text{plim}(\beta) + \frac{\text{plim}(\frac{1}{n}\sum X_i\varepsilon_i)}{\text{plim}(\frac{1}{n}\sum X_i^2)} = \beta + \frac{\text{E}(X_i\varepsilon_i)}{\text{E}(X_i^2)} = \beta$$

where we use the properties of the plim operator and the Law of Large Numbers to argue that the sample averages converge in probability to their population equivalents.

(b) Convergence in distribution plays an important role in settings where we do not know the exact sampling distribution of our estimator, such as in the linear regression setting when the distribution of u is unknown. We need a distributional result in order for us to conduct hypothesis testing.

The central limit theorem allows us to obtain an approximate (asymptotic) distribution for $\widehat{\beta}$ even without knowing the true distribution of u. The asymptotic distribution provides a good approximation of the true sampling distribution as long as the sample is sufficiently large. This powerful result enables us to conduct hypothesis testing which is valid assuming our sample is large.

Section B

Answer three questions from this section.

Question 6

Let us consider how workplace smoking bans affect the incidence of smoking. Below, we use data on 10,000 US indoor workers from 1991 to 1993 taken from 'Do Workplace Smoking Bans Reduce Smoking', by Evans et al. (*American Economic Review*, 1999).

Let *smoker* be a dummy variable indicating whether a worker smokes (1 = yes, 0 = no) and *smkban* a dummy variable indicating whether there is a ban on smoking in the workplace (1 = yes, 0 = no).

(a) The following OLS regression results were obtained:

$$\widehat{smoker} = \underset{(0.007)}{0.290} - \underset{(0.009)}{0.078}\, smkban \qquad (6.1)$$

$$n = 10000,\ R^2 = 0.0078,\ RSS = 1821.59$$

The standard errors (SEs) are in parentheses. Interpret the parameter estimates of the coefficient on *smkban*. Provide the (approximate) 95% confidence interval for the coefficient on *smkban*. How can we use this confidence interval to test the hypothesis that $\beta_{smkban} = 0$?

(6 marks)

A further specification was considered that included other characteristics of the worker: the age (in years), gender (male/female), ethnicity (black/hispanic/white), and level of education ($E1$ = highschool dropout, $E2$ = highschool graduate, $E3$ = some college, $E4$ = college graduate, $E5$ = Master degree or above). The following OLS regression results were obtained for this multiple linear regression model:

$$\widehat{smoker} = \underset{(0.019)}{0.201} - \underset{(0.009)}{0.045}\,smkban - \underset{(.009)}{0.033}female - \underset{(.0003)}{0.001}\,age - \underset{(.016)}{0.027}black \qquad (6.2)$$

$$-\underset{(.014)}{0.104}hispanic + \underset{(.019)}{0.310}E1 + \underset{(.012)}{0.224}E2 + \underset{(.012)}{0.156}E3 + \underset{(.012)}{0.042}E4$$

$$n = 10000,\ R^2 = 0.0526,\ RSS = 1736.81$$

The SEs are in parentheses.

(b) Compare the coefficient estimates on *smkban* from the simple and multiple regression model in (6.1) and (6.2) and explain why the estimates differ.

(3 marks)

(c) Interpret the estimated parameter on $E2$ (highschool graduate) in (6.2) and indicate how you can obtain its p-value and what information the p-value provides.

(5 marks)

(d) Both the simple and multiple regression model suffer from heteroskedasticity. Explain why. What are the implications of heteroskedasticy for the parameter estimates and the standard errors in (6.1) and (6.2)? What can you do to resolve this problem? Explain your answer.

(6 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Sections 2.6, 3.2, 6.2, 7.3 and 10.1.

Dougherty, C. Subject guide (2016): Chapters 5, 6 and 10.

Approaching the question

(a) This is the linear probability model, where:

$$\mathrm{E}(smoker = 1 \mid X) = P(smoker = 1 \mid X) = \beta_0 + \beta_1 smkban.$$

If the smoking ban is introduced, the probability of a worker smoking decreases by 7.8 percentage points (not 7.8%).

An (approximate) 95% confidence interval is:

$$(\widehat{\beta}_1 - z_{crit,\,0.025} \times \text{s.e.}(\widehat{\beta}_1),\ \widehat{\beta}_1 + z_{crit,\,0.025} \times \text{s.e.}(\widehat{\beta}_1))$$
$$= (-0.078 - 1.96 \times 0.009,\ -0.078 + 1.96 \times 0.009) \approx (-0.096,\ -0.060)$$

where z relates to the standard normal distribution (for large degrees of freedom the t distribution converges to the standard normal).

We can use confidence intervals for testing. For a two-sided $\text{H}_0 : \beta_1 = 0$ test, we reject H_0 at the 5% significance level if zero does not lie in the 95% confidence interval.

(b) The reason why the coefficient estimates on *smkban* differ in (6.1) and (6.2) is due to the *omitted variable bias problem.* When we omit relevant variables (for example, education and gender) which are related to the included variable *smkban*), then the parameter estimates on *smkban* will estimate not only the direct effect that *smkban* has on smoking (the parameter of interest) but the indirect effect of these omitted variables as well. We also call these omitted variables confounders.

The more negative effect found in (a) signals that once the effect of education on smoking (gender, age and ethnicity) is controlled for, the true effect of a smoking ban on the probability of smoking is smaller.

(c) The coefficient on $E2$ means that highschool graduates are 22.4 percentage points more likely to smoke relative to people with a Master's degree or above since $E5$ *is the left out base category*. (Many candidates failed to recognise that the left out category was required to interpret the parameter clearly). The t statistic for this effect is:

$$t = \frac{\widehat{\beta}}{\text{s.e.}(\widehat{\beta})} = \frac{0.224}{0.012} = 18.667$$

which has a p-value smaller than 0.05 (we would reject H_0 at the 5% significance level). The p-value is the lowest level of significance at which we can reject the null hypothesis. Since the sample size is large, we can use the standard normal distribution to obtain the p-value (candidates may provide a graphical discussion instead). For a one-sided test, $p = 1 - \Phi(t)$ and for a two-sided test, $p = 2 \times (1 - \Phi(t))$, assuming $t \geq 0$.

(d) In the linear probability model (LPM) we have to deal with the problem of heteroskedasticity. Conditional on X, $\text{Var}(y \mid X) = p(X)(1 - p(X)) \equiv \text{Var}(u \mid X)$. The implication of heteroskedasticity for the estimated standard errors is that they are wrong (they rely on homoskedasticity) and inference based on them would be invalid. The parameter estimates themselves remain unbiased and consistent. For inference, we should use heteroskedasticity-robust standard errors or apply WLS instead.

Question 7

In question 6 we considered the linear regression model to study how workplace smoking bans affect the incidence of smoking. Here we consider the results from applying a probit regression of *smoker* (1 = yes, 0 = no) on *smkban* (1 = yes, 0 = no) and the other explanatory variables:

```
. probit smoker smkban female age black hispanic E1 E2 E3 E4

Iteration 0:   log likelihood = -5537.1662
Iteration 1:   log likelihood = -5255.1526
Iteration 2:   log likelihood =  -5252.349
Iteration 3:   log likelihood = -5252.3489

Probit regression                               Number of obs     =     10,000
                                                LR chi2(9)        =     569.63
                                                Prob > chi2       =     0.0000
Log likelihood = -5252.3489                     Pseudo R2         =     0.0514
```

smoker	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
smkban	-.1517626	.0289268	-5.25	0.000	-.208458	-.0950671
female	-.1106249	.0287785	-3.84	0.000	-.1670298	-.05422
age	-.0042031	.0011748	-3.58	0.000	-.0065057	-.0019006
black	-.07969	.0525369	-1.52	0.129	-.1826604	.0232804
hispanic	-.3327039	.0476677	-6.98	0.000	-.4261308	-.2392769
E1	1.094231	.0714121	15.32	0.000	.9542663	1.234197
E2	.8518588	.0594747	14.32	0.000	.7352906	.9684271
E3	.6492566	.0606989	10.70	0.000	.530289	.7682241
E4	.2224747	.0649939	3.42	0.001	.0950891	.3498603
_cons	-.9842425	.0756055	-13.02	0.000	-1.132427	-.8360584

(a) It is argued that using the probit regression model is better than using the linear probability model when explaining the binary variable *smoker*. Discuss

the benefits/drawbacks of using the Probit model when trying to explain a binary variable.

(5 marks)

(b) **Explain briefly how the Probit estimates are obtained and discuss the properties of the parameter estimates.**

Hint: **You may recall that for the Probit model, we will specify:**

$$\mathbf{Pr}(\boldsymbol{smoker} = 1) = \boldsymbol{\Phi}(\boldsymbol{\beta_0} + \boldsymbol{\beta_1 smkban} + \boldsymbol{\beta_2 female} + \cdots + \boldsymbol{\beta_8 E3} + \boldsymbol{\beta_9 E4})$$

where $\boldsymbol{\Phi}$ is the standard normal CDF (cumulative distribution function).

(5 marks)

(c) **Explain how you can estimate the effect of the smoking ban on the probability of smoking for a 50-year old white, college graduated man. You are not expected to use your calculator, clarity of the computations required is enough.**

(5 marks)

(d) **Discuss how you could test the joint significance of the worker's characteristics (gender, age, ethnicity and level of education) using the likelihood ratio test. Clearly indicate the test statistic, its distribution, the rejection rule and the additional information you would need to implement it.**

(5 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Sections 10.3 and 10.6.

Dougherty, C. Subject guide (2016): Chapter 10.

Approaching the question

(a) The probit model has three main advantages over the linear probability model (LPM): (i) predicted probabilities are restricted to lie in $[0, 1]$, (ii) maximum likelihood estimators are (asymptotically) efficient whereas OLS (LPM) estimators will be inefficient, and (iii) maximum likelihood estimators automatically deal with heteroskedasticity.

The main drawbacks of the probit model relative to the LPM are that (i) the coefficients cannot be directly interpreted as the marginal effects of the regressor(s) of interest, so we need to compute predicted probabilities using the probit specification, and (ii) it is also computationally more complicated.

(b) The parameters are estimated by maximum likelihood estimation, where the log-likelihood function is given by:

$$\begin{aligned}\log L(\beta) &= \sum_{i=1}^{n}\{smoker_i \log(P(smoker_i = 1 \,|\, X) + (1 - smoker_i)\log(P(smoker_i = 0 \,|\, X)\} \\ &= \sum_{i=1}^{n}\{smoker_i \log(\Phi(\beta_0 + \beta_1 smkban_i + \beta_2 female_i + \cdots + \beta_9 E4_i) + \\ &\quad (1 - smoker_i)\log(1 - \Phi(\beta_0 + \beta_1 smkban_i + \beta_2 female_i + \cdots + \beta_9 E4_i))\}.\end{aligned}$$

To obtain the parameter estimates the first-order conditions are solved (numerically as no explicit formulae exist).

Under suitable regularity conditions, the estimates are consistent, asymptotically normally, and asymptotically efficient.

(c) The effect is the difference in predicted probabilities between the man with the characteristics given, with *smkban* = 1 versus *smkban* = 0:

$$\widehat{P}(y_i = 1 \mid X, smkban = 1) - \widehat{P}(y_i = 1 \mid X, smkban = 0)$$

$$= \Phi\left(\sum_k x_{ki}\widehat{\beta}_k + \widehat{\beta}_1\right) - \Phi\left(\sum_k x_{ki}\widehat{\beta}_k\right)$$

$$= \Phi(-0.984 + 0.222 - 0.004 \times 50 - 0.152) - \Phi(-0.984 + 0.222 - 0.004 \times 50).$$

A discussion of marginal effects (ignoring the fact that *smkban* is a dummy variable) is acceptable as well.

(d) We would like to use the LR test, to test:

$$\text{H}_0 : \beta_j = 0 \ \forall \ j \in \{\text{worker characteristics}\}$$

$$\text{H}_1 : \beta_j \neq 0 \text{ for at least one } j \in \{\text{worker characteristics}\}.$$

Estimate the restricted (R) and the unrestricted probit model (U). The restricted model imposes H_0. The unrestricted model is just the originally estimated model above. The LR test makes use of the difference between the two estimated log-likelihood functions (the log ratio of the likelihood functions). The test statistic is:

$$LR = 2(\log L^U - \log L^R) \sim \chi^2_J$$

where J denotes the degrees of freedom which is equal to the number of restrictions; that is, $J = 8$ here. For a given significance level we reject H_0 if its realisation exceeds the critical value given by the χ^2_8 distribution, which equals 15.51 for the 5% significance level.

Question 8

An economist is interested in estimating the production function for widgets which is postulated to follow a Cobb–Douglas specification:

$$Y_i = \exp(\beta_0) L_i^{\beta_L} K_i^{\beta_K} \exp(u_i)$$

where Y_i is a measure of output for firm i; L_i is labour, K_i is capital stock and u_i is an unobserved term that captures technological or managerial efficiency and other external factors (e.g., weather). The parameters to be estimated are $(\beta_0, \beta_L, \beta_K)$. Taking logs:

$$\ln Y_i = \beta_0 + \beta_L \ln L_i + \beta_K \ln K_i + u_i.$$

(a) Provide the interpretation of the parameter β_L. What effect, if any, will changing the units of measurement of labour have on the parameter estimates for $(\beta_0, \beta_L, \beta_K)$? Explain your answer.

(5 marks)

(b) Assume that you have a cross-section of firms and that more productive firms hire less workers (labour). Explain why OLS would not provide consistent estimates for $(\beta_0, \beta_L, \beta_K)$. Would it over- or underestimate β_L on average? Clearly explain your answer.

(5 marks)

Instead of applying OLS, the economist decides to use the average wage paid by firm i, W_i, as an instrument for the (log) quantity of labour employed by that firm, $\ln L_i$.

(c) Describe in detail how you would estimate the parameters of the production function using Two Stage Least Squares (TSLS). What restrictions would be necessary for this researcher to successfully use this instrumental variable in the estimation of the parameters $(\beta_0, \beta_L, \beta_K)$ and what would you need to assume about capital stock?

(7 marks)

(d) If average wages per firm do not vary much by firm (potentially because of unionisation or high mobility of the labour force), how would this affect the properties of the estimation procedure suggested in (c)? Explain your answer.

(3 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Sections 4.2, 6.2 and 9.3.

Dougherty, C. Subject guide (2016): Chapters 4 and 8.

Approaching the question

(a) Since the specification is in logs, β_L is the output *elasticity* with respect to labour. It measures the % change in output when labour is increased by 1%, *ceteris paribus*. Since this is a relative change measure, changing the units of measurement will not change the parameter estimate. The same holds true for β_K, the output elasticity with respect to capital. However, the intercept will be changed by a change in the units of labour. If labour is measured in new units $L_i^* = c \times L_i$, the log specification would give rise to $\ln L_i^* = \ln c \times L_i = \ln c + \ln L_i$ which means the intercept is shifted by $\beta_L \ln c$ units.

(b) The problem with using OLS on the above regression is that there will be correlation between the error u_i and the regressor $\ln L_i$ as more productive firms (higher u_i) are associated with hiring fewer workers (lower L_i). This correlation will make the parameter estimates inconsistent. This problem is a result of the *omission of relevant variables* which creates OVB. The parameter estimates for β_L will be underestimated as it captures the fact that firms with higher technological or managerial efficiency require less labour for the same output.

(c) Two-stage least squares (TSLS) proceeds in two stages. In the first stage, $\ln L_i$ is regressed on all other exogenous regressors and the instrument w. This yields the fitted values:

$$\widehat{\ln L_i} = \widehat{\pi}_0 + \widehat{\pi}_1 w_i + \widehat{\pi}_2 \ln K_i.$$

In the second stage, the original regression is run where the predicted values from the first stage above $\widehat{\ln L_i}$ are used instead of the original regressor $\ln L_i$. We run:

$$\ln Y_i = \beta_0 + \beta_L \widehat{\ln L_i} + \beta_K \ln K_i + u_i.$$

Alternatively, the second stage can be described as performing IV on the original regression where we use $\widehat{\ln L_i}$ as an instrument for $\ln L_i$.

The instrumental variable approach requires the instrument to be relevant and exogenous. Instrument relevance means that the instrument must be correlated with the original regressor, that is $\text{Cov}(\ln L_i, w_i) \neq 0$. Instrument exogeneity requires the instrument to be uncorrelated with the error term; that is, $\text{Cov}(w_i, u_i) = 0$.

For consistent parameter estimates, we require capital stock to also be exogenous, i.e. uncorrelated with the error term.

(d) Candidates should recall the variance of the IV estimator in the simple linear regression setting, revealing the importance that the instrument w is highly correlated with the endogenous regressor $\ln L$ for its precision. If there is not much variation in wages, the instrument is likely not to be very correlated with $\ln L_i$. This is a problem we also call the 'weak instrument problem' as we will have a weak first stage ($\ln L_i$ is not predicted well by the instrument and other exogenous regressors). This problem results in imprecise TSLS estimators.

Question 9

Let us consider the expectations augmented Phillips curve (see also Mankiw, 1994):

$$infl_t - infl_t^e = \beta_1(unem_t - \mu_0) + e_t$$

where μ_0 is the natural rate of unemployment (assumed to be constant over time) and $infl_t^e$ is the expected rate of inflation formed in $t-1$.

This model suggests that there is a trade-off between unanticipated inflation ($infl_t - infl_t^e$) and cyclical unemployment (difference between actual unemployment and the natural rate of unemployment). We assume that e_t (also called supply shock) is an i.i.d. random variable with zero mean.

(a) You are told that expectations are formed as follows:

$$infl_t^e - infl_{t-1}^e = \lambda(infl_{t-1} - infl_{t-1}^e).$$

What name do we give such a process and how should we interpret λ?

(2 marks)

(b) Show that you can rewrite the model as:

$$\Delta infl_t = \gamma_0 + \gamma_1 unem_t + \gamma_2 unem_{t-1} + v_t \tag{9.1}$$

where $\Delta infl_t = infl_t - infl_{t-1}$. Clearly indicate the relation between $(\gamma_0, \gamma_1, \gamma_2)$ and $(\mu_0, \beta_1, \lambda)$ and show that:

$$v_t = e_t - (1-\lambda)e_{t-1}.$$

***Hint:* If you want you may use the following shorthand notation in your derivations: $y_t = infl_t$, $y_t^e = infl_t^e$ and $x_t = unem_t$.**

(7 marks)

(c) Discuss what assumptions you would like to make about e_t (the supply shock) that will guarantee that the OLS estimator for the parameters in (9.1) is consistent. *Hint:* You may want to give the assumptions you need to make about v_t (composite error term) first.

(3 marks)

(d) Show how you can obtain a consistent estimator for λ using your consistent estimates for $(\gamma_0, \gamma_1, \gamma_2)$. Provide a proof of its consistency. [*Note:* If you did not manage to get an explicit relation between λ and $(\gamma_0, \gamma_1, \gamma_2)$, consider $\lambda = g(\gamma_0, \gamma_1, \gamma_2)$ where $g(\cdot)$ is some continuous function].

(3 marks)

(e) One of the assumptions provided in (c) is rather unreasonable (*Hint:* future unemployment may be related to current supply shocks). Discuss how you could use IV (TSLS) to obtain a consistent estimator for the parameters in (9.1). Discuss what conditions your instruments need to satisfy and propose a suitable instrument in this setting.

(5 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Sections R.14, 9.3, 11.3, 11.4 and 11.5.

Dougherty, C. Subject guide (2016): Chapters 8 and 11.

Approaching the question

(a) This is an adaptive expectations model. The parameter λ, which should lie between 0 and 1, indicates the speed with which expectations are adjusted in response to deviations between the actual and the expected rate of inflation in the previous period.

(b) To obtain this it is useful to realise that the equation given in part (a) can be rewritten as:

$$y_t^e = (1-\lambda)y_{t-1}^e + \lambda y_{t-1}. \qquad (*)$$

This shows that if we subtract $(1-\lambda)$ times the lagged expectation augmented Phillips curve, we can get rid of the inflation expectation variable. Let us substract from the first equation given in the question, $1-\lambda$ times the equation one period lagged:

$$y_t - y_t^e - (1-\lambda)[y_{t-1} - y_{t-1}^e] = \beta_1(x_t - \mu_0) + e_t - (1-\lambda)[\beta_1(x_{t-1} - \mu_0) + e_{t-1}].$$

Rearranging yields:

$$y_t - (1-\lambda)y_{t-1} - (y_t^e - (1-\lambda)y_{t-1}^e) = \beta_1 x_t - (1-\lambda)\beta_1 x_{t-1} - \lambda\beta_1\mu_0 + e_t - (1-\lambda)e_{t-1}.$$

Using (*) we get:

$$y_t - (1-\lambda)y_{t-1} - \lambda y_{t-1} = \beta_1 x_t - (1-\lambda)\beta_1 x_{t-1} - \lambda\beta_1\mu_0 + v_t$$

which yields:

$$\Delta y_t = -\beta_1\lambda\mu_0 + \beta_1 x_t - (1-\lambda)\beta_1 x_{t-1} + v_t$$

which gives rise to the following relationship between the γs and original parameters:

$$\gamma_0 = -\beta_1\lambda\mu_0, \quad \gamma_1 = \beta_1 \quad \text{and} \quad \gamma_2 = -(1-\lambda)\beta_1.$$

(c) Consistency requires the error term in the new model v_t and both regressors, x_t and x_{t-1} to be uncorrelated. Since e_t has a zero mean this means:

$$\mathrm{E}(x_t(e_t - (1-\lambda)e_{t-1})) = 0 \quad \text{and} \quad \mathrm{E}(x_{t-1}(e_t - (1-\lambda)e_{t-1})) = 0.$$

A sufficient condition for this would be that $\mathrm{E}(e_t x_s) = 0$ for $s \in \{t-1, t, t+1\}$ or uncorrelatedness between x_t and e_s $\forall\ s, t$.

(d) The above relationships can be solved for λ as:

$$\lambda = \frac{\gamma_2}{\gamma_1} + 1.$$

Using the estimators $\widehat{\gamma}_1$ and $\widehat{\gamma}_2$, an estimator $\widehat{\lambda}$ can be constructed accordingly. For consistent estimators $\widehat{\gamma}_1$ and $\widehat{\gamma}_2$, the estimator $\widehat{\lambda}$ is also consistent by the Slutsky theorem:

$$\mathrm{plim}(\widehat{\lambda}) = \mathrm{plim}\left(\frac{\widehat{\gamma}_2}{\widehat{\gamma}_1} + 1\right) = \frac{\mathrm{plim}(\widehat{\gamma}_2)}{\mathrm{plim}(\widehat{\gamma}_1)} + 1 = \frac{\gamma_2}{\gamma_1} + 1 = \lambda.$$

Slutsky's theorem (probability limit rules) guarantees the second equality since these are continuous functions.

(e) If future unemployment is related to current supply shocks, then $\mathrm{Cov}(x_t, e_{t-1}) \neq 0$ and we get inconsistency as that would mean $\mathrm{Cov}(x_t, v_t) \neq 0$.

As long as unemployment is only related to past supply shocks, but not current or future ones x_{t-1} is still uncorrelated with v_t, $\mathrm{Cov}(x_{t-1}, v_t) = 0$.

We should proceed using an *instrumental variable approach* (2SLS), where we need to look for an instrument to deal with the endogeneity of x_t (this could for instance be a demand shock which is unrelated with supply shocks). Call this instrument z_t.

The instrument must satisfy instrument relevance and instrument exogeneity. Instrument relevance means that the instrument must be correlated with the original regressor, that is $\mathrm{Cov}(x_t, z_t) \neq 0$. Instrument exogeneity requires the instrument to be uncorrelated with the error term; that is, $\mathrm{Cov}(z_t, v_t) = 0$. Since v_t is a function of both e_t and e_{t-1}, the latter assumption is satisfied when the instrument is unrelated with any current and past supply shocks.

In this particular case, due to the fact that v_t has an MA(1) error structure, we can also use x_{t-2} as an instrument. We expect there to be correlation between x_t and x_{t-2} and x_{t-2} (the past) is uncorrelated with e_t and e_{t-1} which make up v_t (we cannot use x_{t-1} because that variable is already included in the model).

A TSLS procedure would proceed by running a regression of x_t on the instrument and the other regressors to obtain fitted values $\widehat{x}_t$. In the second stage, the outcome variable Δy_t would be regressed on an intercept, the predicted values $\widehat{x}_t$ and x_{t-1}. Alternatively, we can apply IV in the second stage where we use $\widehat{x}_t$ as our instrument for x_t. This would give rise to consistent TSLS estimators $(\gamma_0^{TSLS}, \gamma_1^{TSLS}, \gamma_2^{TSLS})$.

Question 10

This question is based on 'Capital Accumulation and Growth: A New Look at the Empirical Evidence', by Bond et al. (*Journal of Applied Econometrics*, 2010). In this article, the authors are interested in a regression model for the (logarithm of) output-per-capita y_t in a given country and time period t similar to:

$$y_t = \alpha + \rho y_{t-1} + \beta x_t + \gamma_t + \varepsilon_t, \quad |\rho| < 1 \tag{10.1}$$

where x_t is (the logarithm of) investment-per output. Imagine that investment rates are also affected by current output-per-capita, so that:

$$x_t = \varphi_0 + \varphi_1 y_t + u_t. \tag{10.2}$$

Equations (10.1) and (10.2) then form a simultaneous equation model. Both errors ε_t and u_t have zero mean.

For (a) and (b), we start by assuming that ε_t and u_t are not serially correlated.

(a) Obtain the reduced form equations for y_t and x_t. (Note: the variables t and y_{t-1} are exogenous.) Are equations (10.1) and (10.2) identified? Discuss.

(5 marks)

(b) Assume that $x_t = \varphi_0 + u_t$ (that is $\varphi_1 = 0$). Under what conditions is the OLS estimator for the parameters in the structural equation for output-per-capita (i.e., equation (10.1)) consistent? Discuss.

(5 marks)

(c) Assume that $x_t = \varphi_0 + u_t$ (that is $\varphi_1 = 0$), u_t is not serially correlated, and $\gamma = 0$. How would you test whether ε_t is serially correlated?

(5 marks)

(d) Assume that $x_t = \varphi_0 + u_t$ (that is $\varphi_1 = 0$), $\rho = 0$ and $\gamma = 0$. Using the Dickey–Fuller test, the authors fail to reject the hypothesis of a unit root in y_t and x_t for most of the countries in their sample at usual significance levels (5%). (Note: we are no longer assuming ε_t or u_t to not be serially correlated). Explaining your answers, what can you say about the OLS estimator for α and β applied to equation (10.1):

i. if ε_t is stationary and weakly dependent?

ii. if ε_t is non-stationary and strongly dependent?

(5 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Sections 9.1, 12.2, 13.2 and 13.6.

Dougherty, C. Subject guide (2016): Chapters 9 and 13.

Approaching the question

(a) Substituting x_t and y_t in appropriately gives the following reduced forms:

$$y_t = \frac{\alpha + \beta\varphi_0}{1-\beta\varphi_1} + \frac{\rho}{1-\beta\varphi_1}y_{t-1} + \frac{\gamma}{1-\beta\varphi_1}t + \frac{\beta u_t + \varepsilon_t}{1-\beta\varphi_1}$$

$$x_t = \frac{\varphi_0 + \varphi_1\alpha}{1-\beta\varphi_1} + \frac{\rho\varphi_1}{1-\beta\varphi_1}y_{t-1} + \frac{\varphi_1\gamma}{1-\beta\varphi_1}t + \frac{\varphi_1\varepsilon_t + u_t}{1-\beta\varphi_1}.$$

Equation (10.2) is identified since the exogenous variables t and y_{t-1} can be used as instruments for the single endogenous variable y_t in this equation. Both exogenous variables are excluded from (10.2). This equation is overidentified. Equation (10.1), though, is not identified since there is an instrument we can use for the single endogenous variable x_t in this equation. There are no exclusions in (10.1), hence this equation is underidentified.

(b) Consistency of the OLS estimator of (10.1) requires that the error ε_t is uncorrelated with the regressors y_{t-1}, x_t, t. We do not need to worry about correlation between ε_t and fixed regressors such at t. If $x_t = \varphi_0 + u_t$, uncorrelatedness between x_t and ε_t requires uncorrelatedness between ε_t and u_t which is given in the question. To ensure that y_{t-1} is uncorrelated with ε_t finally we require the absence of autocorrelation in ε_t which is true by assumption.

(c) We are asked to test for serial correlation in the error ε_t, from the following regression model:

$$y_t = \alpha + \rho y_{t-1} + \beta x_t + \varepsilon_t, \quad |\rho| < 1.$$

Say we postulate an AR(1) process for ε_t we test H_0: no autocorrelation against H_1: AR(1) presence of autocorrelation. We should apply the Breusch–Godfrey test (the Durbin–Watson test cannot be applied because of the presence of the lagged endogenous variable).

Using the OLS residuals, we run the following regression:

$$\widehat{\varepsilon}_t = \gamma_0 + \gamma_1 y_{t-1} + \gamma_2 x_t + \delta\widehat{\varepsilon}_{t-1} + v_t.$$

The test statistic is given by nR^2 of this regression, where n is the sample size. Under the null hypothesis its asymptotic distribution is χ_1^2. We reject H_0 if the realisation exceeds the critical value given by this distribution for a given significance level.

(d) We are asked to consider the following model:

$$y_t = \alpha + \beta x_t + \varepsilon_t$$

in the setting where both y_t and x_t are expected to be integrated of order one.

i. If ε_t is non-stationary and strongly dependent, y_t and x_t are unrelated. Nevertheless, running the regression (10.1) may still give significant parameter estimates since we face the case of spurious regression, wrongly suggesting that there is a relationship between y_t and x_t. This is because the estimates are inconsistent. This is the spurious regression problem.

ii. If ε_t is stationary and weakly dependent, then the two unit root processes y_t and x_t must be cointegrated. This means the parameters α and β can be consistently estimated by OLS.

Examiners' commentaries 2019

EC2020 Elements of econometrics

Important note

This commentary reflects the examination and assessment arrangements for this course in the academic year 2018–19. The format and structure of the examination may change in future years, and any such changes will be publicised on the virtual learning environment (VLE).

Information about the subject guide and the Essential reading references

Unless otherwise stated, all cross-references will be to the latest version of the subject guide (2016). You should always attempt to use the most recent edition of any Essential reading textbook, even if the commentary and/or online reading list and/or subject guide refer to an earlier edition. If different editions of Essential reading are listed, please check the VLE for reading supplements – if none are available, please use the contents list and index of the new edition to find the relevant section.

Comments on specific questions – Zone B

Candidates should answer **EIGHT** of the following **TEN** questions: **ALL** of the questions in Section A (8 marks each) and **THREE** questions from Section B (20 marks each). **Candidates are strongly advised to divide their time accordingly.**

Section A

Answer all questions from this section.

Question 1

Consider the simple linear regression model:

$$Y_i = \beta X_i + u_i, \quad i = 1, \ldots, n.$$

We assume that the errors $\{u_i\}_{i=1}^n$ are independent normal random variables with zero mean. The regressor $\{X_i\}_{i=1}^n$ is non-stochastic (fixed under repeated sampling). You suspect that the errors exhibit heteroskedasticity.

(a) **Explain what we mean by the concept of heteroskedasticity. Enhance your answer with the help of a graphical illustration.**

(3 marks)

(b) **Derive the variance of the OLS estimator for $\widehat{\beta}$ in the presence of heteroskedastity and explain how a robust (White) standard error for $\widehat{\beta}$ can be obtained.**

(5 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Sections 7.1 and 7.3.

Dougherty, C. Subject guide (2016): Chapter 7.

Approaching the question

(a) Heteroskedasticity means that the error variance depends on the regressors, it violates the assumption that $\mathrm{E}(u_i^2) = \sigma^2$.

A graphical illustration should show data points and some depiction of the variance – for example, drawing the distribution for the errors at different levels of X. The graph should be clearly labelled.

(b) The variance of the estimator $\widehat{\beta}$ depends on X_i and, given the unbiasedness, is given by:

$$\mathrm{Var}(\widehat{\beta}) = \mathrm{E}\left((\widehat{\beta} - \beta)^2\right) = \mathrm{E}\left(\frac{\left(\sum_i X_i u_i\right)^2}{\left(\sum_i X_i^2\right)^2}\right) = \frac{\sum_i X_i^2 \mathrm{E}(u_i^2)}{\left(\sum_i X_i^2\right)^2} = \frac{\sum_i X_i^2 \sigma_i^2}{\left(\sum_i X_i^2\right)^2}$$

where σ_i^2 is the variance of the error term which depends on i and the second equality follows from fixed X in repeated samples and independence of u_i across observations giving $\mathrm{E}(u_i u_j) = 0 \ \forall \ i \neq j$. Many candidates struggled to derive the variance.

The White robust standard error for $\widehat{\beta}$ is the square root of the estimate of the above variance. The estimate would be constructed using the residuals from an OLS regression on the original model:

$$\widehat{\mathrm{Var}}(\widehat{\beta}) = \frac{\sum_i X_i^2 \widehat{u}_i^2}{\left(\sum_i X_i^2\right)^2}.$$

Question 2

Consider the human capital earnings function given by:

$$earnings_i = \beta_0 + \beta_1 educ_i + \beta_2 exper_i + \beta_3 exper_i^2 + u_i, \quad i = 1, \ldots, n$$

where *earnings* denotes the hourly earnings of an individual and *educ* and *exper* denote the years of schooling and experience, respectively. We assume we have obtained a random sample $\{(earnings_i, educ_i, exper_i)\}_{i=1}^n$ from the population. The errors $\{u_i\}_{i=1}^n$ are i.i.d. normal random variables with zero mean and variance σ^2. We assume independence between the errors and regressors (i.e., we ignore the usual ability bias problem).

(a) **Discuss the rationale for including both *exper* and *exper*^{2} in this model. In your answer be explicit about the expected signs for β_2 and β_3.**

(2 marks)

(b) **Discuss briefly how would you test that *exper* has a significant effect on *earnings*?**

(2 marks)

(c) **Provide a clear interpretation of the parameter β_1. You are told that you can obtain the estimator for β_1 by running the regression:**

$$earnings_i^* = \beta_1 educ_i^* + e_i, \quad i = 1, \ldots, n$$

where $earnings_i^*$ and $educ_i^*$ are obtained from running a regression of *earnings* (and *educ*) on an intercept, *exper* and *exper*^{2}. Explain this statement.

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Chapter 3.2 and 3.5.

Dougherty, C. Subject guide (2016): Chapter 3.

Approaching the question

(a) The rationale is to allow for diminishing returns to experience, β_2 positive and β_3 negative. For other values of the βs a different explanation should be given.

(b) Since both β_2 and β_3 capture experience, such a test needs to be a joint test with the hypotheses:

$$\mathrm{H}_0 : \beta_2 = \beta_3 = 0 \quad \text{vs.} \quad \mathrm{H}_1 : \beta_2 \neq 0 \text{ and/or } \beta_3 \neq 0.$$

The F test can be used for such a joint hypothesis with the critical values on the $F_{2,\,n-4}$ distribution.

Many candidates failed to recognise that this needs to be a joint test since we have two different regressors which capture experience.

(c) The parameter β_1 is the marginal effect of education on earnings; that is, it measures by how many (monetary) units hourly earnings would increase as a response to someone going to school for one more year, *ceteris paribus*.

The parameter estimate in the original model can also be obtained by the given equation where the starred variables denote the residuals from running a regression of the respective variable on the left-out intercept and regressors *exper* and *exper*^{2}. For both *earnings* and *educ*, this would isolate the remaining variation in the variables after controlling for the intercept, *exper* and *exper*^{2}. Hence this procedure in two steps is the same as controlling for the three left-out variables in the original regression equation. The is the Frisch–Waugh–Lovell result (regression anatomy).

Question 3

Consider the OLS estimator for β in the linear regression model:

$$Y_i = \beta X_i + \varepsilon_i, \quad i = 1, \ldots, n$$

where $\{(Y_i, X_i)\}_{i=1}^n$ form an i.i.d. sample from a population and the errors are drawn from an unknown distribution with mean zero and variance σ^2. You are told that X_i and ε_i are uncorrelated (not necessarily independent therefore!).

(a) Discuss the importance of convergence in probability.

(4 marks)

(b) Discuss the importance of convergence in distribution.

(4 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Sections R.14 and R.15.

Dougherty, C. Subject guide (2016): Chapter 8.

Approaching the question

(a) Convergence in probability plays an important role when proving *consistency* of estimators. A consistent estimator converges in probability to the true parameter and we either write $\text{plim}(\widehat{\beta}) = \beta$ or $\widehat{\beta} \xrightarrow{p} \beta$.

Proofs of consistency (a large sample property) are particularly important in settings where we cannot show that our estimator is unbiased (a finite sample property). For the OLS estimator:

$$\text{plim}(\widehat{\beta}) = \text{plim}(\beta) + \frac{\text{plim}(\frac{1}{n}\sum X_i\varepsilon_i)}{\text{plim}(\frac{1}{n}\sum X_i^2)} = \beta + \frac{\text{E}(X_i\varepsilon_i)}{\text{E}(X_i^2)} = \beta$$

where we use the properties of the plim operator and the Law of Large Numbers to argue that the sample averages converge in probability to their population equivalents.

(b) Convergence in distribution plays an important role in settings where we do not know the exact sampling distribution of our estimator, such as in the linear regression setting when the distribution of u is unknown. We need a distributional result in order for us to conduct hypothesis testing.

The central limit theorem allows us to obtain an approximate (asymptotic) distribution for $\widehat{\beta}$ even without knowing the true distribution of u. The asymptotic distribution provides a good approximation of the true sampling distribution as long as the sample is sufficiently large. This powerful result enables us to conduct hypothesis testing which is valid assuming our sample is large.

Question 4

Consider the following regression model:

$$\boldsymbol{Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t, \quad t = 1, \ldots, T.}$$

Both $\{Y_t\}_{t=1}^T$ and $\{X_t\}_{t=1}^T$ are difference stationary processes.

(a) Discuss the concept 'difference stationarity' and contrast it to the concept 'trend stationarity'. In your answer make sure you also explain what stationarity means.

(4 marks)

(b) How can you test whether the above relation is spurious or cointegrating? In your answer you are expected to explain the difference between a spurious and cointegrating relationship.

(4 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Sections 13.1 and 13.4.

Dougherty, C. Subject guide (2016): Chapter 13.

Approaching the question

(a) *Trend stationarity* relates to a process which has a deterministic trend ($y_t = \alpha + \beta t + \varepsilon_t$ where ε_t is weakly dependent). It is a type of non-stationarity which can be removed by detrending. Once we include a trend in the regression, we can use standard statistical inference methods: the errors (and, therefore, Y_t) will be weakly dependent and consistent estimates can be obtained.

Difference stationarity relates to a process which can be made stationary by taking first differences: ΔY_t, examples are random walk and random walk with drift. Here Y_t is persistent, strongly dependent, and estimation of the process Y_t on Y_{t-1} would fail to produce consistent parameter estimates.

By stationarity we mean *covariance stationarity*, which requires constant (finite) means and variances, and covariances which only depend on distance in time, i.e. $\text{E}(Y_t, Y_{t-s})$ only depends on s, not on location t.

(b) If Y_t and X_t are cointegrated, the error term in this model will be covariance stationary and consistent estimates of β_0 and β_1 can be obtained. The existence of a cointegrating relationship indicates the presence of a long-run relationship (existence of an ECM). If the error term has a unit root, however, the problem of a spurious relation will arise (the relationship is meaningless). Even if Y_t and X_t are independent, estimates of β_0 and β_1 may show up significant and suggest there would be a relationship. This is because a unit root error term implies that the parameters cannot consistently be estimated and we cannot use the usual distribution to obtain critical values.

A test which allows us to distinguish between the two cases would be a test whether the error term ε_t has a unit root. We can conduct this by performing a *Dickey–Fuller test* on the differenced residuals $\Delta\widehat{\varepsilon}_t$.

Question 5

Consider the simple linear regression model:

$$\boldsymbol{Y_i = \beta_0 + \beta_1 X_i + u_i, \quad i = 1, \ldots, n.}$$

We assume that the errors $\{u_i\}_{i=1}^n$ are independent random variables with zero mean. The regressor $\{X_i\}_{i=1}^n$ is non-stochastic (fixed under repeated sampling). Under these conditions, the OLS estimator for β_1, $\widehat{\beta}_1$, is unbiased. (You are not asked to derive $\widehat{\beta}_1$).

(a) Explain the concept of unbiasedness of an estimator.

(2 marks)

(b) Let us consider two other estimators for the slope β_1:

$$\widehat{\beta}_1^{\circ} = \frac{\sum_{i=1}^n (Z_i - \bar{Z})Y_i}{\sum_{i=1}^n (Z_i - \bar{Z})X_i} \quad \text{and} \quad \widehat{\beta}_1^{*} = \frac{\sum_{i=1}^n (Z_i - \bar{Z})Y_i}{\sum_{i=1}^n (Z_i - \bar{Z})Z_i}$$

where $Z_i = \sqrt{X_i}$ for all i and $\bar{Z} = \frac{1}{n}\sum_{i=1}^n Z_i$. Please indicate whether $\widehat{\beta}_1^{\circ}$ and $\widehat{\beta}_1^{*}$ are unbiased estimators for β_1. Clearly show your derivations.

(4 marks)

(c) Briefly indicate how you would choose between the three estimators, $\widehat{\beta}_1$, $\widehat{\beta}_1^{\circ}$ and $\widehat{\beta}_1^{*}$.

(2 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Sections 2.3 and 2.5.

Dougherty, C. Subject guide (2016): Chapter 2.

Approaching the question

(a) Unbiasedness means $\mathrm{E}(\widehat{\beta}_1) = \beta_1$. It means that the expected value of the estimator is the true parameter; that is, we are correct on average in repeated samples. This ensures that we will not make systematic errors when estimating β.

(b) In answering this question, it is easiest to derive results using Z_i rather than working with $Z_i = \sqrt{X_i}$. It is important to note furthermore that $\sum(Z_i - \bar{Z}) = 0$.

When plugging in the true model ($Y_i = \beta_0 + \beta_1 X_i + u_i$), we can then write:

$$\widehat{\beta}_1^\circ = \beta_1 + \frac{\sum(Z_i - \bar{Z})u_i}{\sum(Z_i - \bar{Z})X_i}$$

and:

$$\mathrm{E}(\widehat{\beta}_1^\circ) = \beta_1 + \frac{\sum(Z_i - \bar{Z})\mathrm{E}(u_i)}{\sum(Z_i - \bar{Z})X_i} = \beta_1$$

(X and Z are assumed to be non-stochastic). Unbiasedness follows as $\mathrm{E}(u_i) = 0$.

For the second estimator, when plugging in the true model, we can write:

$$\widehat{\beta}_1^* = \beta_1 + \frac{\sum(Z_i - \bar{Z})X_i}{\sum(Z_i - \bar{Z})Z_i} + \frac{\sum(Z_i - \bar{Z})u_i}{\sum(Z_i - \bar{Z})Z_i}$$

and:

$$\mathrm{E}(\widehat{\beta}_1^*) = \beta_1 + \frac{\sum(Z_i - \bar{Z})Z_i^2}{\sum(Z_i - \bar{Z})Z_i} \neq \beta_1$$

using again the fact that Z is non-stochastic and $\mathrm{E}(u_i) = 0$, hence $\widehat{\beta}_1^*$ is biased.

Side notes: (i) $\widehat{\beta}_1^\circ$ essentially is an IV estimator where Z is used as an instrument for X. Both Z and X are exogenous here. (ii) $\widehat{\beta}_1^*$, on the other hand, is the slope parameter of a regression of Z instead of X on Y, and hence should not give us an unbiased estimator of β_1.

Many candidates failed to realise that $\sum(Z_i - \bar{Z}) = 0$ and wrongly concluded that the first estimator $\widehat{\beta}_1^\circ$ is biased.

(c) When choosing between $\widehat{\beta}_1$ and $\widehat{\beta}_1^\circ$ (both unbiased), efficiency considerations matter. Since the model satisfies the Gauss–Markov assumptions, we know that by the Gauss–Markov theorem, $\widehat{\beta}_1$ is BLUE and should, therefore, be chosen over $\widehat{\beta}_1^\circ$. Nevertheless, there may be a trade-off between bias and variance, and the mean squared error of $\widehat{\beta}_1^*$ could be smaller than that of $\widehat{\beta}_1$ (which is unbiased).

The bias–variance trade-off was mentioned only by a few candidates.

Section B

Answer three questions from this section.

Question 6

Let us consider the expectations augmented Phillips curve (see also Mankiw, 1994):

$$\boldsymbol{infl_t - infl_t^e = \beta_1(unem_t - \mu_0) + e_t}$$

where μ_0 is the natural rate of unemployment (assumed to be constant over time) and $infl_t^e$ is the expected rate of inflation formed in $t-1$.

This model suggests that there is a trade-off between unanticipated inflation ($infl_t - infl_t^e$) and cyclical unemployment (difference between actual unemployment and the natural rate of unemployment). We assume that e_t (also called supply shock) is an i.i.d. random variable with zero mean.

(a) You are told that expectations are formed as follows:

$$\boldsymbol{infl_t^e - infl_{t-1}^e = \lambda(infl_{t-1} - infl_{t-1}^e).}$$

What name do we give such a process and how should we interpret λ?

(2 marks)

(b) Show that you can rewrite the model as:

$$\boldsymbol{\Delta infl_t = \gamma_0 + \gamma_1 unem_t + \gamma_2 unem_{t-1} + v_t} \qquad (6.1)$$

where $\Delta infl_t = infl_t - infl_{t-1}$. Clearly indicate the relation between $(\gamma_0, \gamma_1, \gamma_2)$ and $(\mu_0, \beta_1, \lambda)$ and show that:

$$\boldsymbol{v_t = e_t - (1-\lambda)e_{t-1}.}$$

***Hint:* If you want you may use the following shorthand notation in your derivations: $y_t = infl_t$, $y_t^e = infl_t^e$ and $x_t = unem_t$.**

(7 marks)

(c) Discuss what assumptions you would like to make about e_t (the supply shock) that will guarantee that the OLS estimator for the parameters in (6.1) is consistent. *Hint:* You may want to give the assumptions you need to make about v_t (composite error term) first.

(3 marks)

(d) Show how you can obtain a consistent estimator for λ using your consistent estimates for $(\gamma_0, \gamma_1, \gamma_2)$. Provide a proof of its consistency. [*Note:* If you did not manage to get an explicit relation between λ and $(\gamma_0, \gamma_1, \gamma_2)$, consider $\lambda = g(\gamma_0, \gamma_1, \gamma_2)$ where $g(\cdot)$ is some continuous function].

(3 marks)

(e) One of the assumptions provided in (c) is rather unreasonable (*Hint:* future unemployment may be related to current supply shocks). Discuss how you could use IV (TSLS) to obtain a consistent estimator for the parameters in (6.1). Discuss what conditions your instruments need to satisfy and propose a suitable instrument.

(5 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Sections R.14, 9.3, 11.3, 11.4 and 11.5.

Dougherty, C. Subject guide (2016): Chapters 8 and 11.

Approaching the question

(a) This is an adaptive expectations model. The parameter λ, which should lie between 0 and 1, indicates the speed with which expectations are adjusted in response to deviations between the actual and the expected rate of inflation in the previous period.

(b) To obtain this it is useful to realise that the equation given in part (a) can be rewritten as:

$$y_t^e = (1-\lambda)y_{t-1}^e + \lambda y_{t-1}. \qquad (*)$$

This shows that if we subtract $(1-\lambda)$ times the lagged expectation augmented Phillips curve, we can get rid of the inflation expectation variable. Let us substract from the first equation given in the question, $1-\lambda$ times the equation one period lagged:

$$y_t - y_t^e - (1-\lambda)[y_{t-1} - y_{t-1}^e] = \beta_1(x_t - \mu_0) + e_t - (1-\lambda)[\beta_1(x_{t-1} - \mu_0) + e_{t-1}].$$

Rearranging yields:

$$y_t - (1-\lambda)y_{t-1} - (y_t^e - (1-\lambda)y_{t-1}^e) = \beta_1 x_t - (1-\lambda)\beta_1 x_{t-1} - \lambda\beta_1\mu_0 + e_t - (1-\lambda)e_{t-1}.$$

Using (*) we get:

$$y_t - (1-\lambda)y_{t-1} - \lambda y_{t-1} = \beta_1 x_t - (1-\lambda)\beta_1 x_{t-1} - \lambda\beta_1\mu_0 + v_t$$

which yields:

$$\Delta y_t = -\beta_1\lambda\mu_0 + \beta_1 x_t - (1-\lambda)\beta_1 x_{t-1} + v_t$$

which gives rise to the following relationship between the γs and original parameters:

$$\gamma_0 = -\beta_1\lambda\mu_0, \quad \gamma_1 = \beta_1 \quad \text{and} \quad \gamma_2 = -(1-\lambda)\beta_1.$$

(c) Consistency requires the error term in the new model v_t and both regressors, x_t and x_{t-1} to be uncorrelated. Since e_t has a zero mean this means:

$$\mathrm{E}(x_t(e_t - (1-\lambda)e_{t-1})) = 0 \quad \text{and} \quad \mathrm{E}(x_{t-1}(e_t - (1-\lambda)e_{t-1})) = 0.$$

A sufficient condition for this would be that $\mathrm{E}(e_t x_s) = 0$ for $s \in \{t-1, t, t+1\}$ or uncorrelatedness between x_t and e_s $\forall\, s, t$.

(d) The above relationships can be solved for λ as:

$$\lambda = \frac{\gamma_2}{\gamma_1} + 1.$$

Using the estimators $\widehat{\gamma}_1$ and $\widehat{\gamma}_2$, an estimator $\widehat{\lambda}$ can be constructed accordingly. For consistent estimators $\widehat{\gamma}_1$ and $\widehat{\gamma}_2$, the estimator $\widehat{\lambda}$ is also consistent by the Slutsky theorem:

$$\mathrm{plim}(\widehat{\lambda}) = \mathrm{plim}\left(\frac{\widehat{\gamma}_2}{\widehat{\gamma}_1} + 1\right) = \frac{\mathrm{plim}(\widehat{\gamma}_2)}{\mathrm{plim}(\widehat{\gamma}_1)} + 1 = \frac{\gamma_2}{\gamma_1} + 1 = \lambda.$$

Slutsky's theorem (probability limit rules) guarantees the second equality since these are continuous functions.

(e) If future unemployment is related to current supply shocks, then $\mathrm{Cov}(x_t, e_{t-1}) \neq 0$ and we get inconsistency as that would mean $\mathrm{Cov}(x_t, v_t) \neq 0$.

As long as unemployment is only related to past supply shocks, but not current or future ones x_{t-1} is still uncorrelated with v_t, $\mathrm{Cov}(x_{t-1}, v_t) = 0$.

We should proceed using an *instrumental variable approach* (2SLS), where we need to look for an instrument to deal with the endogeneity of x_t (this could for instance be a demand shock which is unrelated with supply shocks). Call this instrument z_t.

The instrument must satisfy instrument relevance and instrument exogeneity. Instrument relevance means that the instrument must be correlated with the original regressor, that is $\mathrm{Cov}(x_t, z_t) \neq 0$. Instrument exogeneity requires the instrument to be uncorrelated with the error term; that is, $\mathrm{Cov}(z_t, v_t) = 0$. Since v_t is a function of both e_t and e_{t-1}, the latter assumption is satisfied when the instrument is unrelated with any current and past supply shocks.

In this particular case, due to the fact that v_t has an MA(1) error structure, we can also use x_{t-2} as an instrument. We expect there to be correlation between x_t and x_{t-2} and x_{t-2} (the past) is uncorrelated with e_t and e_{t-1} which make up v_t (we cannot use x_{t-1} because that variable is already included in the model).

A TSLS procedure would proceed by running a regression of x_t on the instrument and the other regressors to obtain fitted values $\widehat{x}_t$. In the second stage, the outcome variable Δy_t would be regressed on an intercept, the predicted values $\widehat{x}_t$ and x_{t-1}. Alternatively, we can apply IV in the second stage where we use $\widehat{x}_t$ as our instrument for x_t. This would give rise to consistent TSLS estimators $(\gamma_0^{TSLS}, \gamma_1^{TSLS}, \gamma_2^{TSLS})$.

Question 7

This question is based on 'Openness and Inflation: Theory and Evidence', by Romer (*Quarterly Journal of Economics*, 1993). Romer proposed theoretical models of inflation that imply that more 'open' countries should have lower inflation. A simple macroeconomic model is:

$$infl_i = \beta_0 + \beta_1 open_i + \beta_2 \ln(pcinc_i) + u_{1i} \tag{7.1}$$

$$open_i = \alpha_0 + \alpha_1 infl_i + \alpha_2 \ln(pcinc_i) + \alpha_3 \ln(land_i) + u_{2i}, \quad i = 1, \ldots, n \tag{7.2}$$

where *infl* is the average annual inflation rate (since 1973), *open* is the average share of imports in gross domestic (or national) product since 1973, ln(*pcinc*) is the log of per capita income in US dollars, and ln(*land*) is the log of land area in square miles. The variables ln(*pcinc*) and ln(*land*) are treated as exogenous variables; u_{1i} and u_{2i} are serially uncorrelated disturbances with zero mean, variances σ_1^2 and σ_2^2 and covariance σ_{12}.

(a) Explain the concepts of endogenous versus exogenous explanatory variables and show that *open* is an endogenous variable in (7.1).

(5 marks)

(b) Discuss the identification of each structural form equation and explain what it means to say that an equation is exact identified.

(5 marks)

The table below shows the OLS and IV estimation results for equation (7.1). Standard errors are reported in parenthesis.

	OLS	IV
constant	25.109	26.90
	(15.205)	(15.40)
open	−0.215	−0.337
	(0.095)	(0.144)
ln(*pcinc*)	0.018	0.376
	1.975	(2.015)

The IV estimator is obtained by solving the following three conditions (no need to show!):

$$\sum_{i=1}^{n}\left(infl_i - \widehat{\beta}_0^{IV} - \widehat{\beta}_1^{IV} open_i - \widehat{\beta}_1^{IV}\ln(pcinc_i)\right) = 0$$

$$\sum_{i=1}^{n}\left(infl_i - \widehat{\beta}_0^{IV} - \widehat{\beta}_1^{IV} open_i - \widehat{\beta}_1^{IV}\ln(pcinc_i)\right)\ln(pcinc_i) = 0$$

$$\sum_{i=1}^{n}\left(infl_i - \widehat{\beta}_0^{IV} - \widehat{\beta}_1^{IV} open_i - \widehat{\beta}_1^{IV}\ln(pcinc_i)\right)\ln(land_i) = 0.$$

(c) Describe in detail how you would estimate the parameters of (7.1) using Two Stage Least Squares (TSLS). How would these estimates compare with the reported IV estimates?

(5 marks)

(d) Conduct a test to see whether you can find evidence that more 'open' countries have lower inflation. Clearly indicate the null and alternative hypothesis, test statistic and its (asymptotic) distribution under the null. Briefly indicate how you could test whether the OLS and IV results are significantly different (so that it matters whether you use the OLS or IV parameter estimates).

(5 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Sections 2.6, 9.1 and 9.3.

Dougherty, C. Subject guide (2016): Chapters 8 and 9.

Approaching the question

(a) Exogenous variables are uncorrelated with the error term ('outside of the model') while endogenous variables are correlated with the error term ('inside the model'). This has implications for the estimates of the coefficients of these variables. While we can obtain consistent estimates under weak exogeneity and unbiased estimates under strong exogeneity, endogeneity implies that the estimates are generally inconsistent and biased. Since we face a simultaneity problem between $infl_i$ and $open_i$ in this problem, both will be endogenous. To show that $open_i$ is an endogenous variable in (7.1), i.e. that $\text{Cov}(open_i, u_{1i}) \neq 0$, candidates should first derive the reduced form for $open_i$, which is:

$$\begin{aligned} open_i &= \alpha_0 + \alpha_1(\beta_0 + \beta_1 open_i + \beta_2 \ln(pcinc_i) + u_{1i}) + \alpha_2 \ln(pcinc_i) \\ &\quad + \alpha_3 \ln(land_i) + u_{2i} \\ (1 - \alpha_1\beta_1) open_i &= (\alpha_0 + \alpha_1\beta_0) + (\alpha_1\beta_2 + \alpha_2)\ln(pcinc_i) + \alpha_3 \ln(land_i) + \alpha_1 u_{1i} + u_{2i} \\ open_i &= \frac{\alpha_0 + \alpha_1\beta_0}{1 - \alpha_1\beta_1} + \frac{\alpha_1\beta_2 + \alpha_2}{1 - \alpha_1\beta_1}\ln(pcinc_i) + \frac{\alpha_3}{1 - \alpha_1\beta_1}\ln(land_i) + \frac{\alpha_1 u_{1i} + u_{2i}}{1 - \alpha_1\beta_1}. \end{aligned}$$

Using the exogeneity of $\ln(pcinc_i)$ and $\ln(land_i)$ we can then show:

$$\text{Cov}(open_i, u_{1i}) = \text{Cov}\left(\frac{\alpha_1 u_{1i} + u_{2i}}{1 - \alpha_1\beta_1}, u_{1i}\right) = \frac{\alpha_1\sigma_1^2 + \sigma_{12}}{1 - \alpha_1\beta_1} \neq 0 \quad \text{generally.}$$

(b) Identification of equation (7.1) requires that we have at least one variable which we can use for the single endogenous variable in this equation $open_i$ $(G - 1)$. Here we have exactly one such variable $\ln(land_i)$ (R) which gives us exact identification $(G - 1 = R)$. The reduced form reveals the relevance of this instrument, the validity of the instrument is by assumption, and finally the equation is excluded from (7.1) itself. Hence (7.1) is exactly identified. In equation (7.2) there is also a single $(G - 1)$ endogenous variable, $infl_i$, for which at least one instrument is required. Unfortunately, there is no exogenous variable excluded from (7.2) which we can use. Hence (7.2) is underidentified $(R < G - 1)$.

(c) Using $land_i$ as an instrument for $open_i$, we would proceed in two stages. In the first stage, we regress $open_i$ on all remaining exogenous regressors in the original specification and the instrument to obtain fitted values of $open_i$:

$$\widehat{open}_i = \widehat{\pi}_0 + \widehat{\pi}_1 \ln(pcinc_i) + \widehat{\pi}_2 \ln(land_i).$$

For the second stage, we regress the outcome variable $infl_i$ on the original exogenous regressors and the fitted values of $open_i$:

$$infl_i = \beta_0 + \beta_1 \widehat{open}_i + \beta_2 \ln(pcinc_i) + u_{1i}.$$

Alternatively, the second stage applies IV to the original equation using the fitted values of $open_i$ as an instrument for $open_i$.

Under exact identification, the two-stage least squares estimator coincides with the IV estimator which uses the moment conditions, $\widehat{\beta}_1^{TSLS} = \widehat{\beta}_1^{IV}$.

(d) This is a one-sided test on whether the IV estimator of β_1 is significantly smaller than zero or not, that is:

$$\text{H}_0 : \beta_1 = 0 \quad \text{vs.} \quad \text{H}_1 : \beta_1 < 0.$$

The test uses the test statistic:

$$T = \frac{\widehat{\beta}_1^{IV}}{\text{s.e.}(\widehat{\beta}_1^{IV})} \xrightarrow{d} N(0,\, 1) \text{ under H}_0.$$

We would evaluate the test statistic using the estimates above to obtain:

$$t = \frac{-0.337}{0.144} = -2.34$$

and we should reject H_0 if it takes values lower than the critical value of the standard normal distribution, -1.645 at the 5% significance level (one-sided). We reject H_0 if $t < z$ and fail to reject otherwise. Here, we would reject H_0. This test only works (approximately) for large samples since we are using the asymptotic distribution of the test statistic (IV estimators only have nice large sample properties).

We should conduct a Hausman test to detect whether the OLS and IV estimates differ, $H_0 : \text{Cov}(open_i, u_{1i}) = 0$ against $H_1 : \text{Cov}(open_i, u_{1i}) \neq 0$. Under the null hypothesis both IV and OLS are consistent, whereas under the alternative only IV is consistent. The test is:

$$\left(\widehat{\beta}^{OLS} - \widehat{\beta}^{IV}\right)' V \left(\widehat{\beta}^{OLS} - \widehat{\beta}^{IV}\right)^{-1} \left(\widehat{\beta}^{OLS} - \widehat{\beta}^{IV}\right) \xrightarrow{d} \chi^2_3 \text{ under } H_0.$$

For a given significance level we reject H_0 if the test statistic takes values which are too large relative to the critical values given by the χ^2_3 distribution.

Question 8

Let us consider how workplace smoking bans affect the incidence of smoking. Below, we use data on 10,000 US indoor workers from 1991 to 1993 taken from 'Do Workplace Smoking Bans Reduce Smoking', by Evans et al. (*American Economic Review*, 1999).

Let *smoker* be a dummy variable indicating whether a worker smokes (1 = yes, 0 = no) and *smkban* a dummy variable indicating whether there is a ban on smoking in the workplace (1 = yes, 0 = no).

(a) The following OLS regression results were obtained:

$$\widehat{smoker} = \underset{(0.007)}{0.290} - \underset{(0.009)}{0.078}\, smkban \qquad (8.1)$$

$$n = 10000,\ R^2 = 0.0078,\ RSS = 1821.59$$

The standard errors (SEs) are in parentheses. Interpret the parameter estimates of the coefficient on *smkban*. Provide the (approximate) 95% confidence interval for the coefficient on *smkban*. How can we use this confidence interval to test the hypothesis that $\beta_{smkban} = 0$?

(6 marks)

A further specification was considered that included other characteristics of the worker: the age (in years), gender (male/female), ethnicity (black/hispanic/white), and level of education ($E1$ = highschool dropout, $E2$ = highschool graduate, $E3$ = some college, $E4$ = college graduate, $E5$ = Master degree or above). The following OLS regression results were obtained for this multiple linear regression model:

$$\widehat{smoker} = \underset{(0.019)}{0.201} - \underset{(0.009)}{0.045}\, smkban - \underset{(.009)}{0.033} female - \underset{(.0003)}{0.001}\, age - \underset{(.016)}{0.027} black \qquad (8.2)$$

$$- \underset{(.014)}{0.104} hispanic + \underset{(.019)}{0.310} E1 + \underset{(.012)}{0.224} E2 + \underset{(.012)}{0.156} E3 + \underset{(.012)}{0.042} E4$$

$$n = 10000,\ R^2 = 0.0526,\ RSS = 1736.81$$

The SEs are in parentheses.

(b) Compare the coefficient estimates on *smkban* from the simple and multiple regression model in (8.1) and (8.2) and explain why the estimates differ.

(3 marks)

(c) Interpret the estimated parameter on $E2$ (highschool graduate) in (8.2) and indicate how you can obtain its p-value and what information the p-value provides.

(5 marks)

(d) Both the simple and multiple regression model suffer from heteroskedasticity. Explain why. What are the implications of heteroskedasticy for the parameter estimates and the standard errors in (8.1) and (8.2)? What can you do to resolve this problem? Explain your answer.

(6 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Sections 2.6, 3.2, 6.2, 7.3 and 10.1.

Dougherty, C. Subject guide (2016): Chapters 5, 6 and 10.

Approaching the question

(a) This is the linear probability model, where:

$$\mathrm{E}(smoker = 1 \,|\, X) = P(smoker = 1 \,|\, X) = \beta_0 + \beta_1 smkban.$$

If the smoking ban is introduced, the probability of a worker smoking decreases by 7.8 percentage points (not 7.8%).

An (approximate) 95% confidence interval is:

$$\begin{aligned}&(\widehat{\beta}_1 - z_{crit,\,0.025} \times \text{s.e.}(\widehat{\beta}_1),\ \widehat{\beta}_1 + z_{crit,\,0.025} \times \text{s.e.}(\widehat{\beta}_1))\\ &= (-0.078 - 1.96 \times 0.009,\ -0.078 + 1.96 \times 0.009) \approx (-0.096,\ -0.060)\end{aligned}$$

where z relates to the standard normal distribution (for large degrees of freedom the t distribution converges to the standard normal).

We can use confidence intervals for testing. For a two-sided $\mathrm{H}_0 : \beta_1 = 0$ test, we reject H_0 at the 5% significance level if zero does not lie in the 95% confidence interval.

(b) The reason why the coefficient estimates on *smkban* differ in (8.1) and (8.2) is due to the *omitted variable bias problem.* When we omit relevant variables (for example, education and gender) which are related to the included variable *smkban*), then the parameter estimates on *smkban* will estimate not only the direct effect that *smkban* has on smoking (the parameter of interest) but the indirect effect of these omitted variables as well. We also call these omitted variables confounders.

The more negative effect found in (a) signals that once the effect of education on smoking (gender, age and ethnicity) is controlled for, the true effect of a smoking ban on the probability of smoking is smaller.

(c) The coefficient on $E2$ means that highschool graduates are 22.4 percentage points more likely to smoke relative to people with a Master's degree or above since $E5$ *is the left out base category.* (Many candidates failed to recognise that the left out category was required to interpret the parameter clearly). The t statistic for this effect is:

$$t = \frac{\widehat{\beta}}{\text{s.e.}(\widehat{\beta})} = \frac{0.224}{0.012} = 18.667$$

which has a p-value smaller than 0.05 (we would reject H_0 at the 5% significance level). The p-value is the lowest level of significance at which we can reject the null hypothesis. Since the sample size is large, we can use the standard normal distribution to obtain the p-value (candidates may provide a graphical discussion instead). For a one-sided test, $p = 1 - \Phi(t)$ and for a two-sided test, $p = 2 \times (1 - \Phi(t))$, assuming $t \geq 0$.

(d) In the linear probability model (LPM) we have to deal with the problem of heteroskedasticity. Conditional on X, $\text{Var}(y \mid X) = p(X)(1 - p(X)) \equiv \text{Var}(u \mid X)$. The implication of heteroskedasticity for the estimated standard errors is that they are wrong (they rely on homoskedasticity) and inference based on them would be invalid. The parameter estimates themselves remain unbiased and consistent. For inference, we should use heteroskedasticity-robust standard errors or apply WLS instead.

Question 9

In question 8 we considered the linear regression model to study how workplace smoking bans affect the incidence of smoking. Here we consider the results from applying a probit regression of *smoker* (1 = yes, 0 = no) on *smkban* (1 = yes, 0 = no) and the other explanatory variables:

```
. probit smoker smkban female age black hispanic E1 E2 E3 E4

Iteration 0:   log likelihood = -5537.1662
Iteration 1:   log likelihood = -5255.1526
Iteration 2:   log likelihood =  -5252.349
Iteration 3:   log likelihood = -5252.3489

Probit regression                               Number of obs     =     10,000
                                                LR chi2(9)        =     569.63
                                                Prob > chi2       =     0.0000
Log likelihood = -5252.3489                     Pseudo R2         =     0.0514
```

smoker	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
smkban	-.1517626	.0289268	-5.25	0.000	-.208458	-.0950671
female	-.1106249	.0287785	-3.84	0.000	-.1670298	-.05422
age	-.0042031	.0011748	-3.58	0.000	-.0065057	-.0019006
black	-.07969	.0525369	-1.52	0.129	-.1826604	.0232804
hispanic	-.3327039	.0476677	-6.98	0.000	-.4261308	-.2392769
E1	1.094231	.0714121	15.32	0.000	.9542663	1.234197
E2	.8518588	.0594747	14.32	0.000	.7352906	.9684271
E3	.6492566	.0606989	10.70	0.000	.530289	.7682241
E4	.2224747	.0649939	3.42	0.001	.0950891	.3498603
_cons	-.9842425	.0756055	-13.02	0.000	-1.132427	-.8360584

(a) **It is argued that using the probit regression model is better than using the linear probability model when explaining the binary variable *smoker*. Discuss the benefits/drawbacks of using the Probit model when trying to explain a binary variable.**

(5 marks)

(b) **Explain briefly how the Probit estimates are obtained and discuss the properties of the parameter estimates.**

***Hint:* You may recall that for the Probit model, we will specify:**

$$\Pr(smoker = 1) = \Phi(\beta_0 + \beta_1 smkban + \beta_2 female + \cdots + \beta_8 E3 + \beta_9 E4)$$

where Φ is the standard normal CDF (cumulative distribution function).

(5 marks)

(c) **Explain how you can estimate the effect of the smoking ban on the probability of smoking for a 50-year old white, college graduated man. You are not expected to use your calculator, clarity of the computations required is enough.**

(5 marks)

(d) Discuss how you could test the joint significance of the worker's characteristics (gender, age, ethnicity and level of education) using the likelihood ratio test. Clearly indicate the test statistic, its distribution, the rejection rule and the additional information you would need to implement it.

(5 marks)

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Sections 10.3 and 10.6.

Dougherty, C. Subject guide (2016): Chapter 10.

Approaching the question

(a) The probit model has three main advantages over the linear probability model (LPM): (i) predicted probabilities are restricted to lie in $[0, 1]$, (ii) maximum likelihood estimators are (asymptotically) efficient whereas OLS (LPM) estimators will be inefficient, and (iii) maximum likelihood estimators automatically deal with heteroskedasticity.

The main drawbacks of the probit model relative to the LPM are that (i) the coefficients cannot be directly interpreted as the marginal effects of the regressor(s) of interest, so we need to compute predicted probabilities using the probit specification, and (ii) it is also computationally more complicated.

(b) The parameters are estimated by maximum likelihood estimation, where the log-likelihood function is given by:

$$\begin{aligned}\log L(\beta) &= \sum_{i=1}^{n}\{smoker_i \log(P(smoker_i = 1 \mid X) + (1 - smoker_i)\log(P(smoker_i = 0 \mid X)\} \\ &= \sum_{i=1}^{n}\{smoker_i \log(\Phi(\beta_0 + \beta_1 smkban_i + \beta_2 female_i + \cdots + \beta_9 E4_i) + \\ &\quad (1 - smoker_i)\log(1 - \Phi(\beta_0 + \beta_1 smkban_i + \beta_2 female_i + \cdots + \beta_9 E4_i))\}.\end{aligned}$$

To obtain the parameter estimates the first-order conditions are solved (numerically as no explicit formulae exist).

Under suitable regularity conditions, the estimates are consistent, asymptotically normally, and asymptotically efficient.

(c) The effect is the difference in predicted probabilities between the man with the characteristics given, with $smkban = 1$ versus $smkban = 0$:

$$\begin{aligned}&\widehat{P}(y_i = 1 \mid X, smkban = 1) - \widehat{P}(y_i = 1 \mid X, smkban = 0) \\ &= \Phi\left(\sum_k x_{ki}\widehat{\beta}_k + \widehat{\beta}_1\right) - \Phi\left(\sum_k x_{ki}\widehat{\beta}_k\right) \\ &= \Phi(-0.984 + 0.222 - 0.004 \times 50 - 0.152) - \Phi(-0.984 + 0.222 - 0.004 \times 50).\end{aligned}$$

A discussion of marginal effects (ignoring the fact that *smkban* is a dummy variable) is acceptable as well.

(d) We would like to use the LR test, to test:

$$\mathrm{H}_0 : \beta_j = 0 \ \forall\ j \in \{\text{worker characteristics}\}$$

$$\mathrm{H}_1 : \beta_j \neq 0 \text{ for at least one } j \in \{\text{worker characteristics}\}.$$

Estimate the restricted (R) and the unrestricted probit model (U). The restricted model imposes H_0. The unrestricted model is just the originally estimated model above. The LR test makes use of the difference between the two estimated log-likelihood functions (the log ratio of the likelihood functions). The test statistic is:

$$LR = 2(\log L^U - \log L^R) \sim \chi^2_J$$

where J denotes the degrees of freedom which is equal to the number of restrictions; that is, $J = 8$ here. For a given significance level we reject H_0 if its realisation exceeds the critical value given by the χ^2_8 distribution, which equals 15.51 for the 5% significance level.

Question 10

This question is based on 'An Empirical Comparison of Alternative Models of the Short-Term Interest Rate', by Chan et al. (*The Journal of Finance*, 1992). In this article, the authors are interested in a regression model for the short-term interest rate r_t given by:

$$r_t - r_{t-1} = \beta_0 + \beta_1 r_{t-1} + \varepsilon_t. \tag{10.1}$$

Assume that the errors ε_t are not serially correlated.

(a) Provide sufficient conditions (on β_1) for r_t to be stationary and weakly dependent. Will the OLS estimator for a linear regression of r_t on r_{t-1} in this case be unbiased and consistent? Explain your answer.

(5 marks)

(b) A few well-known models in finance hypothesise that $\beta_1 = 0$. Explain how you would test this hypothesis. What name do we give the process r_t when $\beta_1 = 0$?

(5 marks)

Suppose you are interested in the following relation between long-term interest rates R_t and short-term interest rates r_t:

$$R_t - r_t = \alpha_0 + \alpha_1 r_t + u_t \tag{10.2}$$

where u_t is an error term that has zero mean.

(c) Let us assume $r_t = \beta_0 + \varepsilon_t$ (that is $\beta_1 = -1$). How would you test whether u_t is serially correlated? Explain your answer.

(5 marks)

(d) Let us assume $\beta_1 = 0$. Explaining your answers, what can you say about the OLS estimator for α_0 and α_1 applied to equation (10.2):

i. if u_t is stationary and weakly dependent?

ii. if u_t is non-stationary and strongly dependent?

Reading for this question

Dougherty, C. *Introduction to econometrics* (fifth edition): Sections 12.2, 13.1, 13.2, 13.4 and 13.6.

Dougherty, C. Subject guide (2016): Chapter 13.

Approaching the question

(a) The model can be rewritten in the classic AR(1) structure:

$$r_t = \beta_0 + (1 + \beta_1) r_{t-1} + \varepsilon_t.$$

Stationarity in an AR(1) process requires the AR coefficient to be smaller than 1 implying $\beta_1 < 0$. We have weak dependence here. Under this assumption and given the absence of serial correlation in ε_t, there is no correlation between the error and the regressor r_{t-1}, hence we can obtain consistent parameter estimates for β_1. Nevertheless, we will not get unbiasedness as ε_t will be correlated with future r_{t+j} $\forall\ j \geq 0$ due to the presence of lagged endogenous variables.

(b) Under $\beta_1 = 0$, the AR coefficient is 1 which indicates that r_t is a random walk with drift, and r_t is difference stationary, that is $\Delta r_t := r_t - r_{t-1}$ is stationary. Here we get strong dependence. To test whether we have a unit root, we can carry out the Dickey–Fuller test. Under $\text{H}_0 : \beta_1 = 0$, r_t has a unit root, whereas under $\text{H}_1 : \beta_1 < 0$, r_t is stationary (and weakly dependent). Using the Dickey–Fuller statistical tables, for a given significance level, we reject H_0 if the test statistic $\widehat{\beta}_1/\text{s.e.}(\widehat{\beta}_1)$ exceeds the appropriate critical value.

(c) If $r_t = \beta_0 + \varepsilon_t$, it is just a random error plus a constant. The proposed equation (10.2) can be estimated using OLS and yields consistent parameter estimates assuming u_t is uncorrelated with r_t.

Say we postulate an AR(1) processs for ε_t. We then test H_0: no autocorrelation against H_1: AR(1) presence of autocorrelation. We should apply the Breusch–Godfrey test (the Durbin–Watson test cannot be applied because of the presence of stochastic regressors).

Using the OLS residuals, we run the following regression:

$$\widehat{u}_t = \gamma_0 + \gamma_1 r_t + \delta \widehat{u}_{t-1} + v_t.$$

The test statistic is given by nR^2 of this regression, where n is the sample size. Under the null hypothesis its asymptotic distribution is χ^2_1. We reject H_0 if the realisation exceeds the critical value given by this distribution for a given significance level.

(d) Under $\beta_1 = 0$, r_t is integrated of order 1 and by implication so will be R_t.

i. If u_t is stationary and weakly dependent, then the two unit root processes R_t and r_t must be cointegrated. This means the parameters α and β can be consistently estimated by OLS.

ii. If u_t is non-stationary and strongly dependent, R_t and r_t are unrelated. Nevertheless, running the regression (10.1) may still give significant parameter estimates since we face the case of spurious regressions, wrongly suggesting that there is a relationship between R_t and r_t. This is because the estimates are inconsistent. This is the spurious regression problem.