Guidance on answering sample examination paper in subject guide EC2020 Elements of econometrics

It is important to read the following notes before reading the solutions to the questions.

Solutions detail

Some of the solutions below provide more detail than would be expected of a candidate in the examination. The reason for this is that it is important that the solutions should be clear to everybody. In an examination, the guiding principles are (1) to read the question very carefully, (2) to judge from the mark allocation how much time should be given to each part of each question, and (3) to make the answers as clear as possible. A minor, but useful, tip, is to use a lot of paper when answering the examination. Paper is cheap. Leave plenty of space for returning to a question and making corrections, if necessary.

Length of the examination

The sample paper should have stated that the time allowed for the examination is three hours.

Notation

There are many variations in econometric notation and the examination will not necessarily use the same conventions as those adopted by the course text. The most common variations are listed below.

Estimated parameters.

In the course text, if the true relationship between variables Y and X is

$$Y = \beta_1 + \beta_2 X + u$$

The fitted relationship is written

$$\hat{Y} = b_1 + b_2 X$$

where b_1 is the estimate of β_1 and b_2 the estimate of β_2. However, it is very common for the estimated parameter to be written as the true parameter with a caret mark added, with the consequence that the fitted line is written:

$$\hat{Y} = \hat{\beta}_1 + \hat{\beta}_2 X$$

where $\hat{\beta}_1$ is the estimate of β_1 and $\hat{\beta}_2$ the estimate of β_2. Obviously, from a practical viewpoint, it makes no difference. The course text uses the first approach because it makes the equations less cluttered. The EC2020 examinations use the second, which also tends to be adopted by higher-level econometrics texts.

Variance and expectation

The (population) variance of a random variable X can be written as var(X) or V(X). The first convention is used by the course test. The sample paper uses the second convention in Question 1 and the first convention in Question 2. The expectation of a random variable X may be written E(X) or EX. The course text uses the first convention. The sample paper, referring to the disturbance term u, uses the first convention in Questions 1 and 2. In Question 6, it uses both conventions.

The answers provided below use the same conventions as the sample paper. Where the sample paper uses more than one convention, that adopted by the course text is used. It is customary in mathematics and statistics to italicize all variables and parameters. The solutions observe this convention.

Question 1

(a)

$$\begin{aligned}\tilde{\beta} &= \frac{1}{8}\left[Y_6 + Y_5 - Y_2 - Y_1\right] \\ &= \frac{1}{8}\left[(\alpha + \beta X_6 + u_6) + (\alpha + \beta X_5 + u_5) - (\alpha + \beta X_2 + u_2) - (\alpha + \beta X_1 + u1)\right] \\ &= \frac{1}{8}\left[(\alpha + 6\beta + u_6) + (\alpha + 5\beta + u_5) - (\alpha + 2\beta + u_2) - (\alpha + \beta + u1)\right] \\ &= \frac{1}{8}\left[8\beta + u_6 + u_5 - u_2 - u_1\right] \\ &= \beta + \frac{1}{8}\left[u_6 + u_5 - u_2 - u_1\right]\end{aligned}$$

Hence

$$E(\tilde{\beta}) = \beta$$

and

$$\begin{aligned}\operatorname{var}(\tilde{\beta}) &= E\left\{\left[\tilde{\beta} - E(\tilde{\beta})\right]^2\right\} \\ &= E\left\{\left[\frac{1}{8}\left(u_6 + u_5 - u_2 - u_1\right)\right]^2\right\} \\ &= E\left\{\frac{1}{64}\left[u_6^2 + u_5^2 + u_2^2 + u_1^2 + 2u_6u_5 - 2u_6u_2 - 2u_6u_1 - 2u_5u_2 - 2u_5u_1 + 2u_2u_1\right]\right\} \\ &= \frac{1}{64}E\left\{\left[u_6^2 + u_5^2 + u_2^2 + u_1^2\right]\right\} \\ &= \frac{1}{64}\left(\sigma^2 + \sigma^2 + \sigma^2 + \sigma^2\right) = \frac{\sigma^2}{16}\end{aligned}$$

Note that the question states that $E(u_i u_j) = 0$ when $j \neq i$ and that $E(u_i^2) = \sigma^2$ for all *i*.)

The variance of $\tilde{\beta}$ is therefore greater than that of $\hat{\beta}$, as it must be, in view of the Gauss–Markov theorem.

(b)

$$Y_t = \alpha + \beta X_t + \beta \sum_{j=1}^{\infty} \lambda^j X_{t-j} + \varepsilon_t$$

Hence

$$\lambda Y_{t-1} = \alpha\lambda + \beta\lambda X_{t-1} + \beta\sum_{j=2}^{\infty}\lambda^j X_{t-j} + \varepsilon_{t-1}$$

$$= \alpha\lambda + \beta\sum_{j=1}^{\infty}\lambda^j X_{t-j} + \varepsilon_{t-1}$$

Subtracting the second equation from the first, we have

$$Y_t - \lambda Y_{t-1} = \alpha(1-\lambda) + \beta X_t + \varepsilon_t - \lambda\varepsilon_{t-1}$$

and so

$$Y_t = \alpha(1-\lambda) + \beta X_t + \lambda Y_{t-1} + \varepsilon_t - \lambda\varepsilon_{t-1}$$

Econometric problems:

If the ε_t are iid, as is customary with this notation, then the model is subject to moving average autocorrelation. The second element, ε_{t-1}, of the compound disturbance term influences Y_{t-1}. Thus there is a violation of the regression model assumption that the disturbance term be distributed independently of the regressors (Assumption C.7 part (1) in the text), and OLS will yield inconsistent estimates. However, if the ε_t are not iid, there is a possibility that the effect of the moving average autocorrelation is so weak that OLS can safely be used after all. See Box 12.2 on page 448 of the text.

(c) Cointegration is a term that is used in the context of two or more nonstationary processes of the same order of integration. When a linear combination of the processes is stationary, the variables are said to be cointegrated. The concept is important because regressions performed on nonstationary processes that are not cointegrated can be expected to yield spurious results.

(d) Let the fitted model be

$$\hat{Y}_t = \hat{\beta} X_t$$

Then e_t, the residual in observation t, is given by

$$e_t = Y_t - \hat{Y}_t = Y_t - \hat{\beta} X_t$$

and RSS, the residual sum of squares, is given by

$$RSS = \sum e_t^2 = \sum \left(Y_t - \hat{\beta} X_t\right)^2$$

Setting the first differential of this with respect to $\hat{\beta}$ equal to zero, we obtain the normal equation

$$\sum -2X_t\left(Y_t - \hat{\beta} X_t\right) = 0$$

This yields

$$\hat{\beta} = \frac{\sum X_t Y_t}{\sum X_t^2}$$

It is a linear function of the data on Y_t since

$$\hat{\beta} = \sum \frac{X_t}{\Delta} Y_t$$

where Δ is the sum the squares of X. Substituting for Y_t from the hypothesized model,

$$\hat{\beta} = \sum \frac{X_t}{\Delta}(\beta X_t + u_t) = \beta \sum \frac{X_t^2}{\Delta} + \sum \frac{X_t}{\Delta} u_t = \frac{\beta}{\Delta} \sum X_t^2 + \sum \frac{X_t}{\Delta} u_t = \beta + \sum \frac{X_t}{\Delta} u_t$$

Hence

$$E(\hat{\beta}) = \beta + \sum \frac{X_t}{\Delta} E(u_t) = \beta$$

and the estimator is unbiased. Note that in the last line we have been told that the X_t are fixed in repeated samples and so we may treat them as nonstochastic.

(e) The Durbin–Watson test is a test for the presence of autocorrelation in the data generation process for the disturbance term. It assumes that, if autocorrelation is present, it takes the form of first-order autoregressive, AR(1), autocorrelation. It further assumes that the regressors are nonstochastic.

Question 2

(a) The appropriate test is the Goldfeld–Quandt test, conducted in the dimension of X_{t1}. The observations are ordered according to the size of X_{t1}. Let the closest integer to $3T/8$ be denoted n. Create a subsample A from the first n observations, and another subsample B from the last n, and fit OLS regressions to both subsamples. Let RSS_A and RSS_B be the residual sums of squares for the two subsamples. Compute

$$F(n-3, n-3) = \frac{RSS_B / n}{RSS_A / n}$$

Under the null hypothesis of no heteroscedasticity of the type suspected, this will have an F distribution with n–3 and n–3 degrees of freedom. This F statistic should be compared with the critical value of F at the significance level chosen.

(b) The OLS estimators will be inefficient and the standard errors invalid. (However, the OLS estimators remain unbiased.)

(c) One should weight the observations by $\dfrac{1}{\sqrt{X_{t1}}}$. The transformed model is then

$$\frac{Y_t}{\sqrt{X_{1t}}} = \beta_0 \frac{1}{\sqrt{X_{1t}}} + \beta_1 \sqrt{X_{1t}} + \beta_2 \frac{X_{2t}}{\sqrt{X_{1t}}} + \frac{u_t}{\sqrt{X_{1t}}}$$

This is a model with three regressors and no intercept. The variance of the disturbance term in each observation is σ^2, and so it is homoscedastic.

(d) Substituting from the hypothesized model, u_i has normal distribution

$$f(u_i) = \left(2\pi\sigma^2 x_t^2\right)^{-\frac{1}{2}} e^{-\left\{\frac{1}{2\sigma^2}\right\}\left\{\frac{y_t - \alpha x}{x_t}\right\}^2}$$

The likelihood function is the product of these expressions across the sample for $T = 1, ..., T$. Taking logarithms, one obtains the loglikelihood function

$$L = -\frac{T}{2}\log 2\pi - \frac{T}{2}\log \sigma^2 - \frac{1}{2}\log \sum x_t^2 - \left\{\frac{1}{2\sigma^2}\right\}\sum\left(\frac{y_t}{x_t} - \alpha\right)^2$$

Differentiating with respect to α, one first-order condition is

$$-2\sum\left(\frac{y_t}{x_t} - \hat{\alpha}_{\text{MLE}}\right) = 0$$

Hence

$$\hat{\alpha}_{\text{MLE}} = \frac{1}{T}\sum\frac{y_t}{x_t}$$

Differentiating with respect to σ^2, the other first-order condition is

$$-\frac{T}{2\hat{\sigma}_{\text{MLE}}^2} + \left\{\frac{1}{2\hat{\sigma}_{\text{MLE}}^4}\right\}\sum\left(\frac{y_t}{x_t} - \alpha\right)^2 = 0$$

Hence

$$\hat{\sigma}_{\text{MLE}}^2 = \left\{\frac{1}{T}\right\}\sum\left(\frac{y_t}{x_t} - \alpha\right)^2$$

Question 3

(a) A difference-stationary time series process is a nonstationary process that may be rendered stationary by taking its first differences. This results in the random component having the same distribution over time, instead of having an increasing variance. It also eliminates any systematic time trend (for example, the drift in a random walk with drift).

A trend-stationary time series process is one with a deterministic time trend and a stationary random component. The deterministic time trend may be eliminated asymptotically by regressing the series on time and saving the residuals. These are then said to constitute the trend-stationary version of the process. The random component is unaffected and remains stationary.

(b) The exclusion of one or more regressors that ought to be included in the model will in general cause the coefficients of the other variables to be biased and the standard errors and test statistics to be invalid. In the special case where the excluded variables are uncorrelated with the included variables, the coefficients of the latter remain unbiased. If the excluded variables are stationary and not autocorrelated, the standard errors will remain valid, but will not necessarily be normally distributed, and hence tests may remain invalid.

If an irrelevant variable is added to the regression model, in general it will not cause the coefficients of the other variables to be biased. However, unless it is uncorrelated with them, it will cause their variances to increase, and this will be reflected in larger standard errors, which remain valid. However, if the irrelevant variable happens to be correlated with the disturbance term, the coefficients will be biased and the standard errors invalid.

(c) Dummy variables are used to allow qualitative characteristics to be included among the regressors in a regression model. If a qualitative variable has *s* possible categories, it is usual to choose one of these categories as the basic category and to define dummy variables for the other *s*–1. Each dummy variable has an unknown, estimable, coefficient in the usual way. If an observation relates to a particular category, the dummy variable for that category is set equal to 1 and all the other dummy variables are set equal to zero. Effectively, for that category, the estimated intercept will be equal to that for the reference category plus the value of its coefficient. For this reason, dummy variables of this type are sometimes described as intercept dummy variables.

Slope dummy variables are defined by interacting an intercept dummy variable with one or more of the continuous variables in the model. This has the effect of allowing the coefficient of the continuous variable to differ according to the category. The estimated coefficient of the continuous variable is its coefficient for the reference category. The coefficient of a slope dummy variable then estimates the extra effect (positive or negative) of the continuous variable in the case of an observation relating to that category.

Question 4

(a) The word “estimate” is ambiguous in this context. It could refer to the coefficient in the probit regression. An explanation of how this is determined is outside the syllabus for the course. Therefore, it is assumed here that it refers to the marginal effect.

In the probit model, the probability of an event occurring is given by a cumulative normal distribution, which will be denoted F(z), where z is the “score”, defined in this model as $z = \beta X$, where X is the explanatory variable.. The marginal effect is the derivative of $F(z)$ with respect to the X. Using the chain rule, this is

$$\frac{\mathrm{d}F(z)}{\mathrm{d}X} = \frac{\mathrm{d}F(z)}{\mathrm{d}z}\frac{\mathrm{d}z}{\mathrm{d}X}$$

The differential of the cumulative normal distribution with respect to z is just the standardized normal distribution. The differential of z with respect to X is β. Hence

$$\frac{\mathrm{d}F(z)}{\mathrm{d}X} = \left\{\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}\right\}\beta$$

This expression requires some assumption about the size of z. The most common procedure is to calculate the value corresponding to the sample mean of X, $\bar{z} = \hat{\beta}\,\bar{X}$ and use this. Hence, in this case, the marginal effect would be computed as

$$\frac{\mathrm{d}F(z)}{\mathrm{d}X} = \left\{ \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\hat{\beta}\bar{X})^2} \right\} \hat{\beta}\,.$$

(b) The t statistics of *mar* for the three models are 2.67, 1.81, and 1.80. That for the OLS regression, although apparently significant at the 1 percent level, is suspect because the disturbance term does not have a normal distribution and is heteroscedastic as well. Those for the logit and probit models, being estimated by maximum likelihood estimation, are valid only asymptotically, but that would not appear to be a problem since the sample size is large. They are significant at the 5 percent level only if one feels justified in performing a one-sided test, ruling out the possibility that *mar* might have a negative coefficient. They are very similar in size, which is common for logit and probit. There is nothing to choose between these models in terms of reliability in this respect.

(c) For the OLS regression, the predicted probability is

$$0.093 + 0.038 * 40 - 0.051 * 16 + 0.024 - 0.068 = 0.75$$

For the probit regression, the score is

$$z = 0.259 + 0.107 * 40 - 0.142 * 16 + 0.063 - 1.593 = 0.737$$

Reading off from the table for the cumulative normal distribution, this implies a probability of 0.77. In this example, the OLS and probit estimates are therefore very similar.

(d) To perform the likelihood ratio test, one computes the statistic given by twice the difference in the loglikelihoods. In this case, this is 189.52. Since there are 5 parameters, this is distributed with a chi-squared distribution with 5 degrees of freedom under the null hypothesis that all the coefficients are zero. The test statistic is greater than 20.52, the critical value at the 0.1 percent significance level, and hence one rejects the null hypothesis.

Question 5

(a) A disturbance term in a regression model using time series data is said to be autocorrelated if any of its values in the process is not distributed independently of every other. It commonly arises as a consequence of temporal persistence of the effects of the influences of excluded minor factors that are responsible for the presence of the disturbance term.

(b) (i) The t statistic for the null hypothesis is

$$t = \frac{1.021 - 1}{0.118} = 0.18$$

Hence, subject to the assumptions listed in (iii), it is not significantly different from zero at any significance level.

(ii) There are 33 degrees of freedom and hence the critical value of t at the 5 percent significance level is 2.03, so the 95 confidence interval is given by

$$-2.40 - 2.03 * 0.132 \le \beta_{pc} \le -2.40 + 2.03 * 0.132$$

which is

$$-5.08 \le \beta_{pc} \le 0.28$$

(iii) The regression model should not be subject to autocorrelation. The time series should not be nonstationary processes. (Plus the usual assumptions of correct specification, independence of the disturbance term and the regressors, homoscedastic disturbance term, and normal distribution for the disturbance term; these assumptions would appear not to be required as part of the answer.)

(iv) Performing the Durbin–Watson test for AR(1) autocorrelation, with 37 observations and three explanatory variables, the critical value of d_L is 1.11 at the 1 percent significance level. Hence one would reject the null hypothesis of no autocorrelation at that level, invalidating the answers to (i) and (ii).

(v) None, since the regression model is subject to autocorrelation and there has been no attempt to test for cointegration. Very probably, the time series are nonstationary.

Question 6

(a), (b) Substituting from the second equation into the first, we have

$$y_t = \beta x_t^* + u_t - \beta v_t = \beta x_t^* + z_t$$

where $z_t = u_t - \beta v_t$. Using OLS,

$$\hat{\beta} = \frac{\sum x_t^* y_t}{\sum x_t^{*2}} = \beta + \frac{\sum x_t^* z_t}{\sum x_t^{*2}} = \beta + \frac{\sum (x_t + v_t)(u_t - \beta v_t)}{\sum (x_t + v_t)^2}$$

$$= \beta + \frac{\frac{1}{T}\sum (x_t + v_t)(u_t - \beta v_t)}{\frac{1}{T}\sum (x_t + v_t)^2}$$

Now, given the regression model assumptions,

$$\operatorname{plim} \frac{1}{T}\sum (x_t + v_t)(u_t - \beta v_t) = -\beta\sigma_v^2$$

$$\operatorname{plim} \frac{1}{T}\sum (x_t + v_t)^2 = \sigma_x^2 + \sigma_v^2$$

(Note that the question states that x and v have zero population means.) Hence

$$\operatorname{plim} \hat{\beta} = \beta - \beta \frac{\sigma_v^2}{\sigma_x^2 + \sigma_v^2}$$

and the OLS estimator is inconsistent. The sign of $\operatorname{plim}\left(\hat{\beta} - \beta\right)$ is the opposite of the sign of β.

(c) In this case, the model may be written

$$y_t^* = \beta x_t + u_t + w_t = \beta x_t + z_t$$

where $z_t = u_t + w_t$. Using OLS,

$$\hat{\beta} = \frac{\sum x_t y_t^*}{\sum x_t^2} = \beta + \frac{\sum x_t z_t}{\sum x_t^2} = \beta + \frac{\sum x_t (u_t + w_t)}{\sum x_t^2}$$

$$= \beta + \frac{\frac{1}{T}\sum x_t (u_t + w_t)}{\frac{1}{T}\sum x_t^2}$$

Now, given the regression model assumptions,

$$\operatorname{plim} \frac{1}{T}\sum x_t (u_t + w_t) = 0$$

$$\operatorname{plim} \frac{1}{T}\sum x_t^2 = \sigma_x^2$$

Hence

$$\operatorname{plim} \hat{\beta} = \beta$$

and the OLS estimator is consistent.

(d) The measurement error in *x* will give rise to inconsistency. The measurement error in *y* has no asymptotic effect on the limiting value. Hence the OLS estimator will be inconsistent.