Introductory Econometrics
Assignment 2, Semester 1, 2022

Instructions:

1. The assignment must be electronically submitted by 4:30pm Australian Eastern Standard Time, on Friday, 27 May.
2. The file needs to be uploaded in PDF format by only one member of each group.
3. All members of the group must click the "Submit Assignment" button on Moodle and accept the University's submission statement. This step is essential, so please make sure that you do this.
4. When instructed to do so, you must report your results in equation form, with standard errors reported in parentheses below the parameter estimates. Screen shots of EViews (or any other statistical package) output are not acceptable. For example, the estimated regression equation below is reported in equation form with standard errors reported in parentheses below the estimated coefficients:
$$\widehat{UNDER5MR} = \underset{(7.607)}{75.459} - \underset{(0.001)}{0.004}\ GDPPC$$
5. The assignment must be typed. Please use Times New Roman font size 12.
6. Please attach a number to any equation or diagram that you refer to when answering the assignment questions.
7. Unless otherwise instructed, all hypothesis tests should be conducted at the 5% significance level.
8. If an assignment is submitted late the following penalty scheme applies: up to 24 hours late, assignment mark will be multiplied by 0.8; more than 24 hours but less than 48 hours late, assignment mark will be multiplied by 0.7; more than 48 hours but less than 72 hours late, assignment mark is multiplied by 0.6. Assignments that are more than 72 hours late are not accepted (i.e. they get a mark of zero).
9. If you are applying for special consideration for circumstances that may make you unable to engage in group work or to adhere to strict deadlines, please let Didier know as early as possible, so that alternative arrangements can be made in time.
10. A penalty of up to 10% will be imposed for failure to comply with the instructions above.

Peer Evaluation Surveys:

Each group member will be required to complete an anonymous peer evaluation survey. The survey will be done via the TeamMates app which will email you a unique link to the survey (to your Monash student email address). You will be asked to rate your group members' participation and effort (not their intelligence!). The aim of the survey is to identify and address any dysfunctional groups as early as possible. The survey will also be used to adjust your assignment marks in the following manner:

- Consider hypothetical student called Arsene:
  - Let n_0 equal the number of (D) votes that Arsene receives from his teammates. A (D) indicates that in the opinion of his teammates Arsene has contributed nothing to the completion of the assignment.
  - Let n_1 equal the number of (C) votes that Arsene receives from his teammates. A (C) vote means that in the opinion of his teammates Arsene has contributed less than it was agreed by the group that he would contribute.
  - Let GM equal Arsene's group submission mark. If $n_0 + n_1 \geq 2$, then Arsene's mark for the assignment is
    $$\max\{0, 1 - 0.4n_0 - 0.15n_1\} \times GM.$$
    If $n_0 + n_1 < 2$, then Arsene's mark for the assignment will be equal to the GM.
- If you fail to complete the survey by the deadline, we will assume that you have given everyone else in your group a (B) and that you have given yourself a (D).

Failure to complete the survey by the deadline will result in a loss of marks, so please complete the survey on time. It is important to communicate clearly with your group members and make sure that everyone understands what is expected from them.

Question 1 (33 marks)

The data for Question 1 is contained in the file 'a2q1.csv'. The file contains 54 observations on group averages of households on: total consumption expenditure (total), fraction of expenditure spent on food (fraction), household size (size).

Consider the linear regression model

$$fraction = \beta_0 + \beta_1 total + \beta_2 size + u.$$

1(a) Do you expect the error term to be homoskedastic? Briefly explain. (3 marks)

1(b) Is there any informal evidence to suggest that the error term may be heteroskedastic? (4 marks)

1(c) Perform the standard White test of the null hypothesis that the conditional variance of the error term in is homoskedastic against the alternative that it is a smooth function of the regressors. Specify any auxiliary regressions that you estimate in answering the question. State the null and alternative hypotheses in terms of restrictions on relevant parameters, specify the form and distribution of the test statistic under the null, the sample value and critical value of the test statistic, your decision rule and your conclusion. (8 marks)

1(d) Would it be a good idea to use the White test based on the fitted values in this case? (3 marks)

1(e) Show the line graph of the variable fraction. Is there any informal evidence to suggest that the error term may be serially correlated? (3 marks)

1(f) Perform a Breusch-Godfrey test of the null hypothesis that there is no first-order autocorrelation in the error term in. Specify any auxiliary regressions that you estimate in answering the question. State the null and alternative hypotheses in terms of restrictions on relevant parameters, specify the form and distribution of the test statistic under the null, the sample value and critical value of the test statistic, your decision rule and your conclusion. (8 marks)

1(g) Estimate the linear regression model $fraction = \beta_0 + \beta_1 total + \beta_2 size + u$ by OLS. Correct the standard errors based on the results from the tests in Question 1(c) and 1(f). Compare the results obtained with the usual OLS standard errors to the results with the corrected standard errors. (4 marks)

Question 2 (31 marks)

The construction of new housing is usually related to additional consumer spending on appliances and furniture, among other things, and therefore considered as a key economic indicator. The construction of new housing is monitored by, for instance, central banks by the variable 'housing starts'. This variable refers to the start of the construction on a new residential housing unit. The data for this question is contained in the file 'HOUSTNSA.csv'. The file contains monthly housing starts in the United States since 1959, and is downloaded from `https://fred.stlouisfed.org/series/HOUSTNSA`.

2(a) Estimate a linear regression model for the housing starts with an intercept and seasonal dummies with December as the base. Interpret the results. (4 marks)

2(b) Use the results from Question 2a to write down the estimated equation (with only the estimated coefficients and without the standard errors) for the linear regression model with an intercept and seasonal dummies with January as the base, without estimating the equation with the new base. Show all steps in the process. (4 marks)

2(c) Estimate a linear regression model for the housing starts with only an intercept and include the correlogram of the residuals in this model. Also include the correlogram of the residuals in a model with an intercept and the seasonal dummies included. Compare the correlograms and briefly discuss your findings. (6 marks)

2(d) Estimate an AR(2) model for the housing starts with the seasonal dummies included. Test the joint significance of the seasonal dummies. Identify the restricted and unrestricted models. State the null and alternative hypotheses, the form and distribution of the test statistic under the null, the sample value and critical value of the test statistic, your decision rule and your conclusion. (9 marks)

2(e) Test the hypothesis that there is only quarterly seasonality in the AR(2) model for housing starts, against the alternative that there is monthly seasonality in the AR(2) model for housing starts. Identify the restricted and unrestricted models. State the null and alternative hypotheses, the form and distribution of the test statistic under the null, the sample value and critical value of the test statistic, your decision rule and your conclusion. (8 marks)

Question 3 (36 marks)

We have the model with one fixed regressor

$$y_i = \beta x_i + u_i \tag{1}$$

for individuals $i = 1, \ldots, n$ with uncorrelated homoskedastic error terms $u_i \sim N(0, \sigma^2)$. Suppose that the value for σ^2 is known.

3(a) Show that the OLS estimator $\hat{\beta} = (X'X)^{-1}X'y = \frac{\sum_i x_i y_i}{\sum_i x_i^2}$ for β is unbiased. (2 marks)

3(b) Derive an expression for the standard error of the OLS estimator for β in terms of x_i and σ. (5 marks)

Suppose that the individuals are divided into groups $j = 1, \ldots, J$ each with n_j observations respectively, and we only observe the reported group means $\bar{y}_j$ and $\bar{x}_j$. The model becomes

$$\bar{y}_j = \tilde{\beta}\bar{x}_j + \bar{u}_j, \tag{2}$$

with error terms $\bar{u}_j = \frac{1}{n_j}\sum_{i=1}^{n_j} u_{ij}$, where u_{ij} indicates error term u_i of individual i belonging to group j.

3(c) Show that the error terms $\bar{u}_j$ are heteroskedastic. (4 marks)

3(d) Show that the OLS estimator for $\tilde{\beta}$ is unbiased. (3 marks)

3(e) Derive an expression for the standard error of the OLS estimator for $\tilde{\beta}$ in terms of x_{ij} and σ, where x_{ij} indicates x_i of individual i belonging to group j. (6 marks)

3(f) What are the consequences of heteroskedasticity in the errors for the OLS estimator of the parameters, their usual OLS standard errors reported by statistical packages, and the standard t-test and F-test for these parameters? (4 marks)

3(g) Test the null-hypothesis that $H_0 : \mathrm{E}[\bar{u}_j^2|\bar{x}_j] = \sigma^2$ for $j = 1, \ldots, J$, against the alternative that the variance is a smooth unknown function of $\bar{x}_j$. Explicitly state which regression(s) you use, the null and the alternative, and the test statistic with its distribution under the null. (5 marks)

3(h) Suppose that we observe the group size n_j for $j = 1, \ldots, J$. Regress $\bar{y}_j\sqrt{n_j}$ on $\bar{x}_j\sqrt{n_j}$. Show that the error terms of this regression are homoskedastic. (4 marks)

3(i) Apart from the problem of heteroskedasticity in the errors, what would be another reason to prefer the regression with individual data over the regression with grouped data? (3 marks)