

Introductory Econometrics

Assignment 1

Instructions:

1. The assignment must be electronically submitted by 5:00pm on Friday, 23 April in PDF format.
2. The file needs to be uploaded 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:

$$\widehat{UNDER5MR} = 75.459 - 0.004 GDP\textit{PC}$$

(7.607) (0.001)

5. The answer to Question 1 may be handwritten and then converted to a pdf. If the images of the handwritten parts are not clear, the markers will mark it as zero. The rest of your 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. All special consideration applications are considered by a central body in the Faculty, not by the ETC2410 teaching team. However, 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 Farshid 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 (understanding unbiasedness and efficiency) (40 marks)

Consider the problem of measuring the comovement of a particular stock's excess return (Qantas) with the excess return of a market portfolio (AllOrds). Excess return is the return minus risk free rate of return (say the 3 month term deposit rate). We denote the stock's excess return by y and the market portfolio's excess return by x . We use data on the previous n months on y and x and we use the linear model:

$$y_t = \beta_0 + \beta_1 x_t + u_t \quad t = 1, \dots, n$$

which in matrix notation is

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}.$$

We use the index t because we are using time series observations. We assume that $E(\mathbf{u} | \mathbf{X}) = \mathbf{0}$ (this means that the part of the movement in Qantas shares not related to the current month's market events are completely related to events specific to Qantas, i.e. it is also not predictable using previous months or future months market events. Five estimators for the slope parameter β_1 have been proposed (in our sample, $x_n \neq 0$, $\bar{x} \neq 0$, $x_2 \neq x_1$, $x_n \neq \bar{x}$, and $\frac{x_n+x_1}{2} \neq \bar{x}$, so all of the proposed estimators are well defined):

1. $\hat{\beta}_1^{[1]} = \frac{y_n}{x_n}$, the ratio of the excess return of Qantas shares to excess return of the market portfolio for the most recent month (observation n)
2. $\hat{\beta}_1^{[2]} = \frac{\bar{y}}{\bar{x}} = \frac{\frac{1}{n} \sum_{t=1}^n y_t}{\frac{1}{n} \sum_{t=1}^n x_t}$, the ratio of the average excess return of Qantas shares to the average excess return of the market portfolio,
3. $\hat{\beta}_1^{[3]} = \frac{y_2 - y_1}{x_2 - x_1}$, the slope of the line connecting the first two observations
4. $\hat{\beta}_1^{[4]} = \frac{y_n - \bar{y}}{x_n - \bar{x}}$, the slope of the line connecting the last observation (x_n, y_n) to the sample average point (\bar{x}, \bar{y}) ,
5. $\hat{\beta}_1^{[5]} = \frac{(y_n + y_1)/2 - \bar{y}}{(x_n + x_1)/2 - \bar{x}}$, the slope of the line connecting the midpoint between the first and last observations $(\frac{x_n+x_1}{2}, \frac{y_n+y_1}{2})$ to the sample average point (\bar{x}, \bar{y}) .

- (a) Express as many as the above estimators as the number of people in your group¹ as β_1 plus some additional terms. Each group member should do one of the above estimators and in your submission you should specify which group member initially attempted each part (even if the group later revised the answer). Then, consider the expected value of each of these estimators, and using the implications of $E(\mathbf{u} \mid \mathbf{X}) = \mathbf{0}$, determine if each of the above is in general (i.e. without any additional assumptions) an unbiased estimator of β_1 [Hint: $E(\mathbf{u} \mid \mathbf{X}) = \mathbf{0}$ implies that the expected value of any u times any function of x_1, x_2, \dots, x_n is zero, for example $E\left(\frac{u_1}{x_1}\right) = 0$ and also $E\left(\frac{u_n}{\bar{x}}\right) = 0$]. Team members can compare their answers, decide which answer or answers are better, and then revise the other answers to make all answers be the most elegant and complete as they can be. Only the final answers need to be submitted, but you should state who did each part. (32 marks).
- (b) What additional assumption(s) is(are) needed before we can say that, conditional on \mathbf{X} , the variance of the OLS estimator $\hat{\beta}_1^{[OLS]} = \frac{\widehat{Cov}(y, x)}{\widehat{Var}(x)}$ is smaller than the variance of other unbiased estimators in the above list (**you do not need to derive the variance of any of the estimators to answer this question**). You need to explain these assumptions in the context of this question, i.e. where y is the excess return of Qantas shares and x is the excess return of the AllOrds index. (8 marks)

¹This means that if a group has 3 members, then that group needs to consider the first 3 estimators only. If a group has 4 members, then that group must consider the first 4 estimators only, and if a group has 5 members, then that group must consider all 5 estimators.

Question 2 (A tribute to Hans Rosling) (60 marks)

Hans Rosling was a Swedish physician who passed away in 2017. See him in action in https://www.ted.com/talks/hans_rosling_shows_the_best_stats_you_ve_ever_seen

Suppose an NGO has employed your team as consultants. They have raised money for infrastructure projects aimed at reducing child mortality in developing countries. You have done some preliminary literature search for them by reading latest publications from the World Health Organisation (WHO)² and some recent research papers published in respected international journals³, and identified that the NGO can be effective by investing in the infrastructure that extends basic drinking water services and basic sanitation services to areas that do not have such services. The NGO has told you that they do not have sufficient resources to deal with two sets of infrastructure projects and asked you to advise them if they should concentrate on water or sanitation projects to be most effective in reducing child mortality. They have also told you that building infrastructure for basic drinking water or basic sanitation services for the same number of people cost the same.

Your team is asked to give a recommendation on what the NGO should do. Suppose you only have access to 2017 data for a sample of low income countries (the data set that you compiled in tutorial 2). Using that data set, answer the following questions (all results should be rounded to two decimal points and all hypothesis tests should be done at the 5% level of significance).

1. Is there any statistically significant evidence in the data to suggest that the expected value of the under-5 mortality rate in low income African countries is higher than that of the rest of low income countries? You are expected to do that by constructing a dummy variable named *Africa* that is equal to 1 if a country is in Africa, and is equal to 0 for all other countries, and then performing an appropriate regression that allows you to test $H_0 : E(\text{under5mr} \mid \text{Africa} = 1, \text{low income}) = E(\text{under5mr} \mid \text{Africa} = 0, \text{low income})$ against an appropriate alternative (since

²https://www.who.int/health-topics/children-environmental-health#tab=tab_2

³Lionel Kesztenbaum and Jean-Laurent Rosenthal (2017), “Sewers’ diffusion and the decline of mortality: The case of Paris, 1880–1914”, *Journal of Urban Economics*, 98, 174-186, <http://dx.doi.org/10.1016/j.jue.2016.03.001>

Joseph P. Ferrie, Werner Troesken (2008), “Water and Chicago’s mortality transition, 1850–1925”, *Explorations in Economic History*, 45, 1-16, <http://dx.doi.org/10.1016/j.eeh.2007.06.001>

all countries in our sample are low income countries, you can drop the explicit conditioning on “low income”).(14 marks)

2. With similar regressions as in the previous part, investigate if the expected value of percentages of population having access to basic drinking water and sanitation in African low income countries are the same as those in non-African low income countries. In this part, you do not need to write all steps of hypothesis testing formally. You only need to report your regression results (in equation form, not package output) and below each regression, write one sentence describing what you learn from the regression result. (6 marks)
3. Perform 2 simple regressions with under-5 mortality rate as the dependent variable in both, but one with SANITATION and the other with WATER as the independent variable. Both regressions should include an intercept. Interpret the estimate of the slope parameter in each of these regressions, and state if their signs make sense intuitively. Would you make a recommendation to the NGO based on these two regressions? Explain why or why not. (15 marks)
4. Perform a multiple regression with under-5 mortality rate as the dependent variable and SANITATION and WATER both as independent variables (include an intercept). Interpret the estimated slope coefficients, and comment on their statistical significance. What recommendation would you make based on this regression? Explain. (15 marks)
5. Do differences in access to water and sanitation services entirely explain the difference in under-5 mortality rates between African and non-African low income countries? Use your data set to explore this question. (10 marks)