ETC2410/BEX2410: Introductory Econometrics

Group Assignment 1, Semester 1, 2022

Due: At or before 4:30 p.m. Australian Eastern Standard Time, on Friday April 29th. 2022

Instructions Regarding Assignment Submission and Presentation

1. Group Assignment 1 must be electronically submitted by 4:30 p.m. Australian Eastern Standard Time, on Friday April 29th.
2. The assignment must be submitted as a PDF document.
3. The file needs to be uploaded by only one member of each group.
4. 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**.
5. Any assignment received after 4:30 p.m. on Friday, April 29th will be deemed late.
6. **A penalty of 5% per day will be imposed on assignments that are submitted after the due date**. For example, if a group receives a mark of 65% for an assignment that has been submitted one day late, the mark will be reduced to 60%.
7. Any assignment received after 4:30 p.m. Australian Eastern Standard Time on Friday, May 6th will receive a mark of zero.
8. The answers to **Questions 2 and 3** may be neatly handwritten and then converted to a PDF document. If the images of the handwritten parts of your assignment are not clear, your tutor will assign them a mark of zero.
9. **Question 1 of your assignment must be typed**. Please use **Times New Roman font size 12**.
10. When answering the assignment questions please impose the same structure on your answers that I have used in composing the assignment. For example, Q1, (a), i).
11. Please attach a number to any equation or diagram that you refer to when answering the assignment questions.
12. 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 output are not acceptable**. For example, the estimate regression equation below is reported in equation form with standard errors reported in parentheses below the estimated coefficients:

$$\widehat{dwage} = \underset{(0.02)}{0.163} + \underset{(1.75)}{20.45}\, educ.$$

13. When instructed to do so, **you must place relevant Eviews output in an appropriately labelled appendix**. The material in the appendices will not be marked. It will only be used to check that the results reported in the main body of the assignment have been correctly reported.
14. Unless otherwise instructed, all hypothesis tests should be conducted at the **5% significance level**.
15. A penalty of up to 10% will be imposed for careless presentation and/or failure to comply with the instructions above.

Peer Evaluation Survey

Each group member will be required to complete an anonymous peer evaluation survey. The survey will be conducted via the TeamMates app which will email you a unique link to the survey (delivered to your Monash student email address).You will be asked to evaluate your group members' participation and effort. The aim of the survey is to identify and address any dysfunctional groups as early as possible.

The survey may also be used to adjust your assignment marks in the following manner:

1. Consider hypothetical student called Arsene:

   - Let n0 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 n1 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

     $$n0 + n1 \geq 2,$$

     then Arsene's mark for the assignment is

     $$\max\{0, (1 - 0.4(n0) - 0.15(n1)\} GM.$$

     If

     $$n0 + n1 < 2,$$

     then Arsene's mark for the assignment will be equal to the GM.

2. 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.

Data Description for Question 1

An Excel file containing the data for Question 1 has been uploaded to the Assignments folder on Moodle. The file is called DATA_Q1_ASS1.wf1. The data set contains observations on a variety of personal characteristics of 474 employees of an American commercial bank. The variables in the data set are defined in Table 1 below.

Table 1	
Variable	Description
$educ_i$	years of education of individual i
$gender_i$	1 if individual i male, 0 otherwise
$jobcat_i$	1 if individual i works in an administrative job 2 if individual i works in a custodial job 3 if individual i works in managerial job
$logsal_i$	the natural log of individual i's annual salary
$logssal_i$	the natural log of individual i's starting annual salary
$race_i$	1 if individual i belongs to an ethnic minority, 0 otherwise

The variable logssal is included as a proxy for personal characteristics other than education, gender and race, such as prior experience, scholastic achievement etc.

When you open the Eviews workfile, you should check that the sample means of the variables in your data set are the same as those reported in Table 2 below.

Table 2	
Variable	Mean
educ	13.49156
gender	0.544304
jobcat	1.411392
logsal	10.35679
logssal	9.669405
race	0.219409

Question 1 (86 marks)

(a) 3 marks

i) Report the histogram of logsal.

(1 mark)

ii) What relevant conclusions do you draw from the histogram?

(2 marks)

(b) 3 marks

i) Report the scatterplot for logsal and race.

(1 mark)

ii) What relevant conclusions do you draw from the scatterplot?

(2 marks)

(c) 21 marks

i) Estimate the following linear regression equation by OLS:

$$\text{logsal} = \beta_0 + \beta_1 race + u. \quad (1)$$

Report the estimated equation **in equation form** in the main body of your assignment. Report the estimated coefficients and standard errors to **three decimal places**. **Place the Eviews output in Appendix (c)**.

(2 marks)

ii) Test the **individual significance** of the regressor *race*. State the null and alternative hypotheses, the form and sampling distribution of your test statistic under the null, the sample and critical value of your test statistic, your decision rule and your conclusion.

(6 marks)

iii) Carefully interpret $\widehat{\beta}_1$.

(2 marks)

iv) Do your results provide conclusive evidence of racial discrimination in the salaries paid by the bank? Briefly explain. (Another way of posing this question would be to ask if differences in annual salary are "caused" by differences in race).

(2 marks)

v) Briefly evaluate the explanatory power of your model.

(1 mark)

vi) Carefully interpret $\widehat{\beta}_0$.

(2 marks)

vii) Construct a 95% confidence interval for β_0.

(2 marks)

viii) Interpret the 95% confidence interval for β_0 constructed in part vii) above.

(2 marks)

ix) Use the confidence interval obtained in viii) to determine whether or not $\widehat{\beta}_0$ is statistically significant at the 5% significance level. Justify your answer.

(2 marks)

(d) 20 marks

Assume that

$$E(\text{logsal}|educ, \text{logssal, gender, race, jobcat}) = \beta_0 + \beta_1 educ + \beta_2 \text{logssal} + \beta_3 gender + \beta_4 race + \beta_5 jobcat. \quad (2)$$

The associated linear regression model is given by

$$\text{logsal} = \beta_0 + \beta_1 educ + \beta_2 \text{logssal} + \beta_3 gender + \beta_4 race + \beta_5 jobcat + u. \quad (3)$$

i) Estimate (3) OLS. Report the estimated equation **in equation form** in the main body of your assignment. Report the estimated coefficients and standard errors to **three decimal places**. **Place the Eviews output in Appendix (d)**.

(2 marks)

ii) Based on the reported p-values, are any of the regressors individually insignificant at the 1% significance level? Briefly explain.

(2 marks)

iii) Test the joint significance of the regressors. State the null and alternative hypotheses, the form and sampling distribution of your test statistic under the null, the sample and critical values of your test statistic, your decision rule and your conclusion.

(6 marks)

iv) Carefully interpret $\widehat{\beta}_4$.

(2 marks)

v) Carefully interpret

$$\widehat{\beta}_3 - \widehat{\beta}_4.$$

(2 marks)

vi) Does the estimated regression equation provide conclusive evidence of racial discrimination in the salaries paid by the bank? Briefly explain. (Another way of posing this question would be to ask if differences in annual salary are "caused" by differences in race).

(3 marks)

vii) Briefly compare the explanatory power of the models given by (1) and (3). What conclusions do you draw from this comparison?

(3 marks)

(e) 6 marks

Test the null hypothesis that once we control for education, starting salary and job category, gender and race have no effect on annual salary. Test at the **10% significance level**. State the null and alternative hypotheses, the unrestricted and restricted models, the form and sampling distribution of your test statistic under the null, the sample and critical value of your test statistic, your decision rule and your conclusion. **Place the Eviews output in Appendix** (**e**).

(6 marks)

(f) 19 marks

Consider once again the model given by

$$E(\text{logsal}|educ, \text{logssal, gender, race, jobcat}) = \beta_0 + \beta_1 educ + \beta_2 \text{logssal} + \beta_3 gender + \beta_4 race + \beta_5 jobcat. \qquad (2)$$

i) Define **population A** to be the population of female managers with 12 years of education who belong to a racial minority, who received a given starting salary. Derive a theoretical expression for the average logsal of population A.

(2 marks)

ii) Define **population B** to be the population of male managers with 11 years of education who are not members of a racial minority and who received the same starting salary as the individuals in population A. Derive a theoretical expression for the average logsal of population B.

(2 marks)

iii) Derive the restriction(s) on the parameter(s) of (2) implied by the null hypothesis that the average logsal of population A is the same as that of population B. Report all steps in the derivation. No marks will be awarded for an unsupported answer.

(2 marks)

iv) Derive an appropriate regression model to test the null hypothesis that the average logsal of population A is the same as that of population B, against the alternative that the average logsal of population B is **greater**.

(3 marks)

v) Estimate the model derived in part iv) above. Report the estimated equation **in equation form** in the main body of your assignment. Report the estimated coefficients and standard errors to **three decimal places**. **Place the Eviews output in Appendix** (**f**).

(2 marks)

vi) Test the null hypothesis that the average logsalary of population A is the same as that of population B against the alternative that the average logsalary of population B is greater. State the null and alternative hypotheses, the form and sampling distribution of your test statistic under the null, the sample and critical value of your test statistic, your decision rule and your conclusion.

(6 marks)

vii) State in words the implication of your test conclusion.

(2 marks)

(g) 2 marks

Define the **racial pay gap** to be the difference between the average salary of individuals with a given set of characteristics who are members of a racial minority, and the average income of individuals with the same set of characteristics who are not members of a racial minority.

i) Use equation (2) above to derive a theoretical expression for the racial pay gap for males.

(1 mark)

ii) Use equation (2) above to derive a theoretical expression for the racial pay gap for females.

(1 mark)

(h) 12 marks

We wish to modify the model given by

$$E(\text{logsal}|educ, \text{logssal, gender, race, jobcat}) = \beta_0 + \beta_1 educ + \beta_2 \text{logssal} + \beta_3 gender + \beta_4 race + \beta_5 jobcat$$

to allow for the possibility that the racial pay gap may vary by gender. That is, the racial pay gap may be different for males and females.

i) Modify the model given by

$$E(\text{logsal}|educ, \text{logssal, gender, race, jobcat}) = \beta_0 + \beta_1 educ + \beta_2 \text{logssal} + \beta_3 gender + \beta_4 race + \beta_5 jobcat$$

to allow for the possibility that the racial pay gap may vary by gender. Report the extended model.

(2 marks)

ii) Use the extended model to derive an expression for the racial pay gap for males, controlling for education, starting salary and job category. (No marks will be awarded for an unsupported answer).

(2 marks)

iii) Use the extended model to derive an expression for the racial pay gap for females, controlling for education, starting salary and job category. (No marks will be awarded for an unsupported answer).

(2 marks)

iv) Test the null hypothesis that, controlling for education, starting salary and job category, the racial pay gap for males and females is the same, against an appropriate alternative. **Test at the 10% significance level**. State the null and alternative hypotheses, the form and sampling distribution of your test statistic under the null, the sample and critical value of your test statistic, your decision rule and your conclusion.

Report any equation that you estimate **in equation form** in the main body of your assignment. Report the estimated coefficients and standard errors to three decimal places. **Place the Eviews output in Appendix (h)**.

(6 marks)

Question 2 (5 marks)

(a) In the bivariate linear regression model

$$y_i = \beta_0 + \beta_1 x_i + u_i, i = 1,2,\ldots n, \tag{4}$$

the OLS estimators of the regression parameters are given by

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}, \tag{5}$$

$$\widehat{\beta}_0 = \bar{y} - \widehat{\beta}_1 \bar{x}. \tag{6}$$

By definition,

$$\widehat{u}_i = y_i - \widehat{y}_i, i = 1,2,\ldots n. \tag{7}$$

Prove that

$$\sum_{i=1}^{n} \widehat{u}_i = 0.$$

(3 marks)

(b) Recall from Topic 3 that the OLS estimators of β_0 and β_1 satisfy the equations

$$\left.\frac{\partial SSR(b_0, b_1)}{\partial b_0}\right|_{\widehat{\beta}_0, \widehat{\beta}_1} = -2\sum_{i=1}^{n}(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i) = 0,$$

$$\left.\frac{\partial SSR(b_0, b_1)}{\partial b_1}\right|_{\widehat{\beta}_0, \widehat{\beta}_1} = -2\sum_{i=1}^{n}x_i(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i) = 0.$$

Use this fact to prove that the vector of OLS residuals, $\widehat{u}$, is orthogonal to the vector x. That is,

$$x'\widehat{u} = 0.$$

(2 mark)

Question 3 (20 marks)

(a) Consider the bivariate linear regression model

$$y_i = \beta_0 + \beta_1 x_i + u_i . i = 1,2,\ldots n, \tag{4}$$

where

$$E(u_i|x_1,x_2,\ldots,x_n) = 0, i = 1,2,\ldots,n,$$

and

$$Var(y_i|x_1,x_2,\ldots,x_n) = \sigma^2, i = 1,2,\ldots,n. \tag{8}$$

Show that the OLS estimator of β_1,

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2},$$

may be written as

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})y_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2}. \tag{9}$$

(3 marks)

(b) The LIE for the mean states that given two random variables Z and W

$$E(Z) = E_W[E(Z|W)].$$

There is an analogous result for $Var(Z)$ which states that

$$Var(Z) = E[Var(Z|W)] + Var[E(Z|W)]. \tag{10}$$

Applying this result to $Var(\widehat{\beta}_1)$, and conditioning on $(x_1,x_2,\ldots,x_n)$, we obtain

$$Var(\widehat{\beta}_1) = E[Var(\widehat{\beta}_1|x_1,x_2,\ldots,x_n)] + Var[E(\widehat{\beta}_1|x_1,x_2,\ldots,x_n)]. \tag{11}$$

Note: When we condition on $(x_1,x_2,\ldots,x_n)$, we also condition on $\bar{x}$ since

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

i) Use (4), (9) and the properties of the expectation operator to prove that

$$[E(\widehat{\beta}_1|x_1,x_2,\ldots,x_n)] = ?$$

(8 marks)

ii) Use (8), (9) and the properties of the variance operator to prove that

$$Var(\widehat{\beta}_1|x_1,x_2,\ldots,x_n) = ?$$

(4 mark)

iii) Use your answers to i) and ii) together with (11) to derive an expression for $Var(\widehat{\beta}_1)$.

(2 marks)

Note: Your final expression for $Var(\widehat{\beta}_1)$ should involve σ^2, n and $E(\widehat{\sigma}_x^2)$, where $\widehat{\sigma}_x^2$ denotes the sample variance of x.

(c) In light of your answer to (b) above, what circumstances provide the best opportunity to precisely estimate $\widehat{\beta}_1$? Briefly explain.

(3 marks)