

UNIVERSITY OF ESSEX

Econometrics Exercise January 2022

INTRODUCTORY ECONOMETRICS

NOTE: The deadline for submitting this Econometric Exercise is 12.00 mid-day on Friday, 21 January 2022. The page limit is 15 printed sides, all inclusive, and the minimum font size is 12pt. It is acceptable to copy/paste computer code and output from Stata (or any other econometrics package), although you may need to reduce the font size to ensure the results are legible and clearly presented. In order to get full marks you are required to comment appropriately on the output.

Important: You must submit your answers online using FASER.

The times shown on your timetable are in UK time. Please check online for a conversion to your local time if you will be undertaking your assessment outside the United Kingdom.

You will be provided with an individualised data set. Your data set is unique to you. You **must not** use the data set of any other student for your Econometrics Exercise.

In this Exercise your task is to examine how different variables can be used to help explain household savings in a cross-section of households during the year 2020.

The variables in your data set are:

- *sav*: household savings from income during the year (in £10,000s);
- *age*: the age of the head of the household at the end of the year (in years);
- *nchild*: the number of children (aged 17 and under) in the household at the end of the year;
- *inc*: household income during the year (in £10,000s);
- *grad*: a dummy variable which equals 1 if the head of the household is a graduate and 0 otherwise; and,
- *hols*: the number of holidays taken by the household during the year.

1.
 - (a) For each of the variables, what is the average, the standard deviation, the minimum value, and the maximum value in your data set?
 - (b) What is the number of observations in your data set?
2.
 - (a) Calculate the correlation coefficient between *sav* and *inc*. Is there a strong relationship between these two variables?
 - (b) Construct a graph that plots savings (on the vertical axis) against income (on the horizontal axis). Does the visual evidence support your answer in part (a)?
 - (c) Generate a histogram for savings. Comment on the distribution of savings.
3.
 - (a) Fit a simple linear regression (SLR) of savings on income, and report the regression results.
 - (b) Is income a statistically significant determinant of savings at the 10% level?
 - (c) How much of the variation in savings can be attributed to the variation in income?
 - (d) What is the effect on savings of an extra £1,000 of income?
 - (e) What is your predicted value of savings at the mean income level?
4.
 - (a) Fit a multiple linear regression (MLR) of savings on the other five variables (*age*, *nchild*, *inc*, *grad* and *hols*), and report the regression results.
 - (b) Which of the regressors are individually significant at the 10% level? And at the 5% level?
 - (c) How much of the variation in savings can be explained by the estimated model?
 - (d) What is the effect on savings of an extra £1,000 of income? What is your predicted value of savings at the mean income level? Compare your results with those in the SLR model.
 - (e) Carry out a test of the joint significance of the variables *age*, *nchild*, *grad* and *hols*. Which model do you prefer – the SLR model in the previous question, or the MLR model in this question?
 - (f) Carry out a RESET test in the MLR model – is there evidence of functional form misspecification?
 - (g) Carry out a test for heteroskedasticity in the MLR model – is there evidence of heteroskedasticity?
5.
 - (a) Generate two new variables, *lsav*, the logarithm of savings, and *linc*, the logarithm of income. Calculate the average, the standard deviation, the minimum value, and the maximum value of these new variables.
 - (b) Calculate the correlation coefficient between *lsav* and *linc*. Is there a strong relationship between these two variables?
 - (c) Construct a graph that plots *lsav* (on the vertical axis) against *linc* (on the horizontal axis). Does the visual evidence support your answer in part (b)?

6.
 - (a) Fit a simple linear regression (SLR) of *lsav* on *linc*, and report the regression results.
 - (b) Is *linc* a statistically significant determinant of *lsav* at the 10% level?
 - (c) How much of the variation in *lsav* can be attributed to the variation in *linc*?
 - (d) What is the effect on savings of an increase of 1% in income?
 - (e) What is the predicted value of savings at the mean income level? Compare your answer with that in Question 3(e).
7.
 - (a) Fit a multiple linear regression (MLR) of *lsav* on *age*, *nchild*, *linc*, *grad* and *hols*, and report the regression results.
 - (b) Which of the regressors are individually significant at the 10% level? And at the 5% level?
 - (c) Carry out a test of the joint significance of the variables *age*, *nchild*, *grad* and *hols*. Which model do you prefer – the SLR model in the previous question, or the MLR model in this question?
 - (d) Carry out a RESET test in the MLR model – is there evidence of functional form misspecification?
 - (e) Carry out a test for heteroskedasticity in the MLR model – is there evidence of heteroskedasticity?
8.
 - (a) Generate a dummy (proxy) variable, *age50*, which is equal to one if the head of the household is aged 50 or over and is equal to zero otherwise. In what proportion of households is the head aged 50 or over?
 - (b) For your preferred model specification from previous questions, investigate whether there are significant differences in savings behaviour according to age group (i.e. under or over 50). In particular, test whether there are significant differences in the intercept of your preferred regression, and in the coefficient on *inc* or *linc*.
9.
 - (a) Prepare a table reporting the estimates (and corresponding standard errors) from the regressions in questions 3(a), 4(a), 6(a), 7(a) and 8(b). In this table, each column should report the results of one regression including the coefficient estimates, the standard errors (robust if necessary), the number of observations used and the R^2 .
 - (b) After running all these regressions, summarise the main findings you have obtained.

End of Paper
