ECO 530 - Introduction to Econometrics

Fall 2023

Exercise 4

50 points

Due Date: Friday, October 27

Instructions

Complete the exercises below using the Nicaragua Rural Business Development Data (NicaRBD.RData).

Be sure to show all of your work. For this assignment, you should submit:

- A PDF containing your written answers, tables, and figures along with the R script
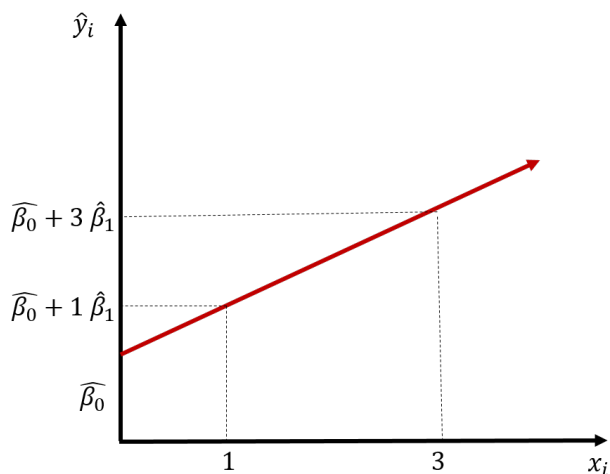
  that generates them.

Example

Consider the following equation depicting the relationship between an outcome y_i and an explanatory variable x_i

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

Imagine that you obtain estimates of β_0 and β_1 via OLS. You can use those estimates $(\hat{\beta}_0, \hat{\beta}_1)$ to predict y for any level of x. Remember, this is our guess at the value of $E[y|x]$

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

We can sketch our estimated relationship as:

That is, a line with its intercept at $\hat{\beta}_0$ and a slope of $\hat{\beta}_1$.

Q1 - Prepare the data

In these exercises, we are going to be working with the variables listed below. Food expenditure will serve as the outcome variable in all of your regressions. Create a data frame without missing values on these variables. Present a table of summary statistics and briefly describe the data in a paragraph.

- food.expend - Expenditure on food measured in purchasing power parity (PPP) adjusted dollars

- land - A respondent's landholding, measured in *manzanas.*

- hadloan - Equal to one if a respondent had a loan and zero otherwise

- crop - A factor variable with values 1-4 and labels "Beans", "Livestock", "Sesame", and "Casava"

Q2 - A single continuous regressor

Estimate the relationship between food expenditure and land holding via Ordinary Least Squares. Explicitly discuss your decision regarding whether or not to use heteroskedasticity robust standard errors. Report and discuss your results. Then, as in the example, plot or sketch the relationship you estimated.

Q3 - Add a binary control variable

Estimate the relationship between food expenditure and land holding again, this time conditioning on whether the respondent had a loan. Report and discuss your results. How do they differ from your estimates in Q2? Plot or sketch the relationship you estimated again. (Hint: You'll need to draw two lines this time.).

1 Q4 - Add a categorical control variable with more levels

Estimate the relationship between food expenditure and land holding again, this time conditioning on the crop a respondent produced. Report and discuss your results. Plot or sketch the relationship you estimated again.

Q5 - Interaction terms

Estimate the specification below:

$$food.expend_i = \beta_0 + \beta_1 land_i + \beta_2 hadloan_i + \beta_3 land_i \times hadloan_i + \epsilon_i$$

There are various ways to create the interaction term that you'll need to estimate β_3. You could:

1. Create it by hand: use *mutate()* to define a new variable equal to $land_i \times hadloan_i$

2. Write the equation in your *lm_robust()* function as *food.expend land*hadloan*. R will estimate the saturated model for you.

I prefer Option 1, as I feel it gives me better control. Report and discuss your results. Plot or sketch the relationship you estimated. Explicitly test the hypothesis that the effect of landholding of food expenditure is the same for respondents with and without a loan. Finally, make a coefficient plot showing the estimated effect of landholding on food expenditure for respondents with and without a loan.