Introductory Econometrics
Nonlinear Specification and Dummy Variables

by Hieu Nguyen

Fall 2024

1.

Consider the following model:

$$\ln(\text{price}) = \beta_0 + \beta_1 \ln(\text{assess}) + \beta_2 \ln(\text{sqrft}) + \beta_3 \ln(\text{lotsize}) + \beta_4 d\_bdrms + \epsilon,$$

where `price` is house sales price (in thousands GBP), `assess` is the assessed value (before the house was sold), `sqrft` is size of house (in square feet), `lotsize` is size of lot (in square feet), and `d_bdrms` is a dummy variable indicating if the house has more than 3 bedrooms.

(a) Use the data `housing.gdt` to estimate the model. First, transform the first four variables into logarithms, then construct the dummy variable as:

$$d\_bdrms = \begin{cases} 1 & \text{if bdrms} > 3 \\ 0 & \text{otherwise} \end{cases}$$

and run the regression. Interpret the estimated coefficients.

(b) Interpret coefficients in the model:

$$\ln(\text{price}) = \beta_0 + \beta_1 \text{assess} + \beta_2 \text{sqrft} + \beta_3 \text{lotsize} + \beta_4 d\_bdrms + \epsilon.$$

(c) Interpret coefficients in the model:

$$\text{price} = \beta_0 + \beta_1 \text{assess} + \beta_2 \ln(\text{sqrft}) + \beta_3 \ln(\text{lotsize}) + \beta_4 d\_bdrms + \epsilon.$$

(d) Suppose we would like to test in the model whether the assessed housing price is a rational valuation: if this is the case, then a 1% change in assess should be associated with a 1% change in price. In addition, `sqrft`, `lotsize`, and `d_bdrms` should not help to explain ln(price), once the assessed value has been controlled for. Define the hypotheses to be tested, the test statistic, and explain how you would conduct the test. Then test for rational valuation in Gretl.

2.

You have organized a ski trip to the mountains for a group of your friends and, as a true econometrician, you decided to estimate a model of the expenditures of each participant. You suppose that the cost of the trip for each person depends on how many days he or she spent there (some people arrived later and some left earlier) and on what type of skis he or she was going. Some people went on downhill skis, some went on cross-country skis and some people managed to go on both. Since you are friends only with people who like sports, there was nobody who was not skiing (i.e., everybody went on downhill or cross-country skis or both).

You specify the following model:

$$\text{cost} = \beta_0 + \beta_1 \text{days} + \beta_2 \text{DS} + \beta_3 \text{CS} + \epsilon,$$

where cost is the cost of the trip, days stands for the number of days the person stays in the mountains, DS is a dummy equal to 1 if the person goes on downhill skis, zero otherwise, and CS is a dummy equal to 1 if the person goes on cross-country skis, 0 otherwise.

(a) In terms of the parameters of your model, what is the expected cost of the trip for a person who spends two days in the mountains and goes both on cross-country and downhill skis? What is the expected cost for a person who spends three days in the mountains and goes on downhill skis only?

(b) How would you test that the two costs from the previous question are equal? Specify step by step how you would proceed: what regression(s) you would run, what results you would need for testing, what would be your test statistic, and what distribution it would have.

(c) You want to test if the two dummy variables are jointly significant in your model. Running the model with the dummies included leads to $R^2 = 0.8$, whereas running it without the dummies gives $R^2 = 0.65$. Knowing that you have 25 observations, test for the joint significance of the two dummies at 95% confidence level.