Introductory Econometrics
Nonlinear Specification and Dummy Variables
Suggested Solution

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.

Solution:

(a) Gretl menu: Add—Logs of selected variables.

To create a new dummy variable, go Add—Define new variable and insert the dummy condition: `d_bdrms=bdrms>3` (do not copy-paste from here).

Model 1: OLS, using observations 1–88
Dependent variable: l_price

	Coefficient	Std. Error	t-ratio	p-value
const	−0.104364	0.585402	−0.1783	0.8589
l_assess	1.04564	0.153677	6.8042	0.0000
l_sqrft	−0.0407485	0.140024	−0.2910	0.7718
l_lotsize	0.00805690	0.0392243	0.2054	0.8378
d_bdrms	−0.00967710	0.0355108	−0.2725	0.7859

Mean dependent var	5.633180	S.D. dependent var	0.303573
Sum squared resid	1.871316	S.E. of regression	0.150153
R^2	0.766599	Adjusted R^2	0.755351
$F(4, 83)$	68.15284	P-value(F)	1.96e–25
Log-likelihood	44.56399	Akaike criterion	−79.12798
Schwarz criterion	−66.74129	Hannan–Quinn	−74.13769

Interpretation: estimated coefficients $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\beta}_3$ represent estimated elasticities of price with respect to individual variables, ceteris paribus. Houses with more than 3 bedrooms lose almost 1% of the price compared to houses with less than 3 bedrooms (exact number will be 0.00967 * 100% = 0.967%), ceteris paribus.

(b) See the lecture #6 slides about how to interpret coefficients in different functional forms.

(c) See the lecture #6 slides about how to interpret coefficients in different functional forms.

(d) To test for rational valuation, we run the F-test of the given joint hypotheses/linear restrictions with:

$$H_0 : \begin{cases} \beta_1 = 1 \\ \beta_2 = 0 \\ \beta_3 = 0 \\ \beta_4 = 0 \end{cases} \quad \text{vs} \quad H_A : \begin{cases} \beta_1 \neq 1 \\ \text{or } \beta_2 \neq 0 \\ \text{or } \beta_3 \neq 0 \\ \text{or } \beta_4 \neq 0 \end{cases}$$

Unrestricted model: the original one, restricted model:

$$\ln(\text{price}) = \beta_0 + 1 \cdot \ln(\text{assess}) + \epsilon.$$

We define a new variable `delta = ln(price) - ln(assess)`: Add—Define new variable and insert: `delta=l_price-l_assess` (do not copy-paste from here) and use it as a new dependent variable in the restricted model:

Model 2: OLS, using observations 1–88
Dependent variable: delta

	Coefficient	Std. Error	t-ratio	p-value
const	−0.0848135	0.0156709	−5.4122	0.0000

Mean dependent var	−0.084814	S.D. dependent var	0.147006
Sum squared resid	1.880149	S.E. of regression	0.147006
R^2	0.000000	Adjusted R^2	0.000000
Log-likelihood	44.35678	Akaike criterion	−86.71357
Schwarz criterion	−84.23623	Hannan–Quinn	−85.71551

Finally, we perform a standard F-test:

$$F = \frac{(1.880 - 1.871)/4}{1.871/83} \approx 0.1 \sim F_{4,83,0.95} = 2.48$$

We do not reject H_0; thus, rational valuation is not rejected.

Test of joint linear restrictions directly in Gretl:

After estimating the model, go to the model output window menu: Tests—Linear restrictions and insert (separated by Enter if inserted individually):

```
b[l_assess]=1
b[l_lotsize]=0
b[l_sqrft]=0
b[d_bdrms]=0
```

to get the Gretl output as below:

```
Restriction set
 1: b[l_assess] = 1
 2: b[l_lotsize] = 0
 3: b[l_sqrft] = 0
 4: b[d_bdrms] = 0

Test statistic: F(4, 83) = 0.0979464, with p-value = 0.98284

Restricted estimates:

             coefficient   std. error   t-ratio    p-value
  ---------------------------------------------------------
  const       -0.0848135    0.0156709    -5.412     5.43e-07 ***
  l_assess     1.00000      0.00000      NA        NA
  l_lotsize    0.00000      0.00000      NA        NA
  l_sqrft      0.00000      0.00000      NA        NA
  d_bdrms      0.00000      0.00000      NA        NA

  Standard error of the regression = 0.147006
```

We can check in the upper part that the restrictions were inserted correctly, and we also directly see the F-statistic (the same as we computed manually above) and the p-value of the test.

Note: Be sure you can test hypotheses manually using the test statistics and the critical values found in (printed) statistical tables as, during the exam, you will not be allowed to use any software.

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.

Solution:

(a)

$$\hat{\text{cost}}_{\text{2days,DS+CS}} = \hat{\beta}_0 + \hat{\beta}_1 \cdot 2 + \hat{\beta}_2 \cdot 1 + \hat{\beta}_3 \cdot 1 = \hat{\beta}_0 + 2\hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_3;$$
$$\hat{\text{cost}}_{\text{3days,DS}} = \hat{\beta}_0 + \hat{\beta}_1 \cdot 3 + \hat{\beta}_2 \cdot 1 + \hat{\beta}_3 \cdot 0 = \hat{\beta}_0 + 3\hat{\beta}_1 + \hat{\beta}_2.$$

(b) We test: $\beta_0 + 2\beta_1 + \beta_2 + \beta_3 = \beta_0 + 3\beta_1 + \beta_2$; restriction: $\beta_3 = \beta_1$.
Unrestricted model: the original one, restricted model:

$$\text{cost} = \beta_0 + \beta_1 \text{days} + \beta_2 \text{DS} + \beta_1 \text{CS} + \epsilon = \beta_0 + \beta_1(\text{days} + \text{CS}) + \beta_2 \text{DS} + \epsilon.$$

After transforming the data to a new variable days + CS, we run both regressions and perform a standard F-test of the joint hypotheses with $J = 1$ and $k = 3$.

(c) We run the F-test for the joint significance of two coefficients with:

$$H_0 : \begin{cases} \beta_2 = 0 \\ \beta_3 = 0 \end{cases} \qquad \text{vs} \quad H_A : \begin{cases} \beta_2 \neq 0 \\ \text{or } \beta_3 \neq 0 \end{cases}$$

Unrestricted model: the original one, restricted model: cost $= \beta_0 + \beta_1 \text{days} + \epsilon$.
Since we keep the dependent variable the same ($TSS_U = TSS_R$), we can compute the F-test in terms of R^2, i.e.:

$$F = \frac{(R_U^2 - R_R^2)/J}{(1 - R_U^2)/(n - k - 1)} = \frac{(0.8 - 0.65)/2}{(1 - 0.8)/(25 - 4)} \approx 7.9 \sim F_{2,21,0.95} = 3.47.$$

We reject H_0; hence we conclude that the two dummies are jointly significant.