Review questions and exercises: Topic 4

Before you start

Make sure you are working in the correct directory: `pwd` . Change to another directory if necessary. Also, you may wish to open a command log and a result log:

```
cmdlog using logname.do
log using logname.log
```

You are using the `HPRICE2` data which contains observations on communities in the Boston area. We aim to relate median house price in the community (*price*, in USD) to various community characteristics. The variable *crime* measures the number of crimes per capita committed in the community, and *lowstat* is the percentage of residents of "lower socio-economic status" in the community. *nox* is the amount of nitrogen oxide in the air (parts per million). *stratio* is the average student-to-teacher ratio in the neighbourhood schools. *proptax* is the amount of property tax payable. *rooms* is the average number of rooms. *radial* is a highway access index.

1 Descriptive Statistics

Familiarise yourself with the data by looking at descriptive statistics and graphs of your choice. List any characteristics or properties of the data that you think are important.

2 Functional form

2.1 lin-lin

Estimate the following regression model:

$$\begin{aligned} price =& \beta_0 + \beta_1 crime + \beta_2 nox + \beta_3 dist + \beta_4 radial \\ &+ \beta_5 proptax + \beta_6 rooms + \beta_7 lowstat + \beta_8 stratio + u \end{aligned}$$

1. Interpret all coefficients.
2. Test for heteroskedasticity and misspecification. What do you find?
3. Produce a residual-versus-fitted plot and comment on it.
4. Modify your regression model to include an interaction effect between *dist* and *nox* (keep the main effects). Describe in words what the estimated effect of *nox* is on price. Do the same for the estimated effect of *dist* on *price*.
5. Calculate the marginal effect of *nox* on *price*, evaluated at *dist*= 1 and *dist*= 12. Are the marginal effects significant at the 5% level of signifcance?
6. Modify your regression model in 4.) to include *nox* squared. Keep everything else the same.
7. Repeat the marginal effects in 5.) for the new model.

2.2 log-lin and log-log

Estimate the following regression model:

$$\begin{aligned} log(price) =& \beta_0 + \beta_1 crime + \beta_2 log(nox) + \beta_3 log(dist) + \beta_4 radial \\ &+ \beta_5 log(proptax) + \beta_6 rooms + \beta_7 lowstat + \beta_8 stratio + u \end{aligned}$$

1. Interpret all coefficients.
2. Test for heteroskedasticity and misspecification. What do you find?
3. Produce a residual-versus-fitted plot and comment on it.
4. Modify your regression model to include an interaction effect between *log(nox)* and *radial* (keep the main effects). Describe in words what the estimated effect of *nox* is on log(price). Do the same for the estimated effect of *radial* on *price*.
5. Calculate the marginal effect of *radial* on *log(price)*, evaluated at *log(nox)*= 1.34 and *log(nox)*= 2.16. Are the marginal effects significant at the 5% level of signifcance?