Economics 311 Fall 2017
Problem Set #4 Due: Wednesday, October 4

This problem set again asks you to use Stata to perform some regressions and to interpret the results. Assemble your answers *along with the relevant Stata output* into a Word document and send it to me as an email attachment in .docx or .pdf format.

Hedonic price analysis is used to model the value of commodities with varying characteristics. The dependent variable is the price (or log of price) of a particular good. Regressors are characteristics of the good. Early applications of this analysis included cars (where the characteristics might be size, power, reliability, and fuel efficiency) and asparagus (where the characteristics were size and color). More recent applications have included computers and, very commonly, houses. Hedonic analyses of house prices are usually done with the individual house as the unit of analysis, but this question asks you to apply the analysis to a dataset in which the observations are neighborhoods.

The data set hprice.dta contains a set of variables for each of 506 neighborhoods within a metropolitan area. The variables in the data set are defined below:

nnumber = neighborhood number
price = median neighborhood house price ($)
crime = crimes committed in neighborhood per capita
nox = atmospheric concentration of nitrous oxides (p/100m)
rooms = average number of rooms in neighborhood houses
dist = weighted distance of neighborhood to 5 employment centers
radial = index of neighborhood access to "radial" highways
proptax = property tax per $1000 valuation in neighborhood
stratio = student-teacher ratio in neighborhood schools
lowstat = percentage of neighborhood people of "low status"

1. General considerations

a. What effect (positive, negative, or none) would you expect each of the other variables in the dataset to have on median house price? (Remember that these are "partial effects"—the effect of one variable on median house price *holding all the other explanatory variables at constant levels*.)

b. The creators of the data set chose to use the median rather than the mean house price in the neighborhood. What are the advantages and disadvantages of this choice? (Hint: Think about extreme values of house price within a neighborhood.)

2. What factors determine the median house price in these neighborhoods?

a. It usually works well to use the natural log of price as the dependent variable in hedonic house-price models rather than the level. As we will study soon, a change of 0.01 in the natural log of price is equivalent to a change of 1% in the level of price. So a coefficient of +0.05 on a particular regressor X_j would mean that an increase of one unit in X_j raises house price by 5%. Create a variable lprice for the log of price using `gen lprice = log(price)`. Estimate a regression of the log of neighborhood house price on neighborhood variables that might be associated with price. You may test alternative specifications of the right-hand side in terms of including and excluding variables. Present a small set of candidate regressions in a table created with `outreg`, then discuss which regression you are choosing for subsequent analysis and why you prefer it to the alternatives.

b. Interpret the coefficients of your preferred regression in terms of the effects of the regressor on the price of houses. (Remember that the dependent variable is in logs.) Does each coefficient have a plausible sign and magnitude?

3. Assessing the effects of proposed policies

Suppose that you have been hired by the board of directors of neighborhood number 130's homeowners' association. Members of the board believe that their neighborhood's property is undervalued and want to increase house values. They have asked you to advise them on the following propositions using your statistical analysis. For each, write a short essay supported by relevant tables showing your estimates and calculations.

a. Are board members justified in believing that property values for neighborhood 130 are too low, given the neighborhood's characteristics? (Hint: Calculate the residual for neighborhood 130. Residuals with a log dependent variable are also interpreted in percentages: a residual of +0.05 means that a neighborhood's actual price is 5% higher than the predicted price.)

b. Colin Commuter, one of the board members, has argued that improving a major road connecting the neighborhood to downtown would increase property values. The road improvement would increase the value of the radial access index from 4 to 8, but the increased traffic would increase nitrous oxide concentrations from 6.24 to 6.50. Based on your regression, would this be a good idea?

c. Board member Sarah Scardicat is worried about neighborhood crime. She has proposed an increase in police service that would require an increase in property taxes from 43.7 to 47.7 and lower crime from 0.881 to 0.6. Based on your regression, what effect would this have on property values?