Economics 312 Spring 2020
Project #4 Assignment Due: 11:59pm, Monday, Feb. 24

Partner assignments

Shisham Adhikari	Roy Shannon
Blaise Albis-Burdige	Amrita Sawhney
Isabelle Caldwell	Lauren Rabe
Anne Cao	Grisha Post
Aditya Gadkari	Evian Oosthuizen
Catherine Gong	Max Nobel
Lirui Jiao	Paul Nguyen
Jonathan Li	Peter Mulgrew
Matt Wan	George Zhao
Charlie Lyu	

Hedonic price analysis is used to model the value of commodities with varying characteristics. Early applications of this analysis included cars and (in a recent daily problem) asparagus. 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. (This is one of Wooldridge's datasets and is used in some examples in the text.) 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 rate 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 are the advantages and disadvantages of performing this analysis at the neighborhood level rather than at the level of the individual house?

b. What effect (positive, negative, or none) would you expect each of the other variables in the dataset (neglecting *nnumber*, obviously) 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*.) Based on this assessment, would it make sense to include all of the potential variables in the regression or would you rule out certain ones having any effect? Do you foresee any issues with severe multicollinearity among the variables?

c. Consider the use of median house price as the dependent variable. Would you expect the variables of the model to affect all house prices within a neighborhood in the same way, so that the effect on the median would be a good proxy for the effect on the price of any given house? If not, how would this affect your interpretation of regression results? 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 log of price as the dependent variable in hedonic house-price models. Estimate a regression of the log of neighborhood house price on other neighborhood variables that might be associated with price. You may test alternative specifications of the right-hand side, both in terms of including and excluding variables and in terms of linear vs. non-linear specifications. Present a representative set of candidate regressions in a table created with `outreg2`, 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?

c. Based on your preferred regression, which factors are most important in determining house prices? Be sure to consider both the economic impact (*i.e.*, for which variables a one-standard-deviation change leads to the largest price changes) and statistical impact (*i.e.*, for which variables are we most confident that the impact is non-zero).

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 report 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?

b. What are the major factors that have large effects on values in neighborhood 130 (both positively and negatively) relative to other neighborhoods? To assess this, note that $y_{130} - \bar{y} = \beta_1 \left(x_{130,1} - \bar{x}_1 \right) + \beta_2 \left(x_{130,2} - \bar{x}_2 \right) + \ldots + \beta_k \left(x_{130,k} - \bar{x}_k \right) + u_{130}$. Each of the terms of this expression can be thought of as the contribution of one variable (or the error term) to the deviation of y_{130} from the mean of y. For example, the contribution of variable x_3 is $\beta_3 \left(x_{130,3} - \bar{x}_3 \right)$. It might be useful to construct a table with the variables down the rows and three columns next to them: one with the coefficient of that variable, one with the deviation of neighborhood 130's value from the mean of that variable, and one with the product. (Excel is really good at tasks like this.) The product column should sum to $y_{130} - \bar{y} - u_{130} = \hat{y}_{130} - \bar{y}$.

c. 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. Would this raise property values? How confident are you of your answer?

d. Board member Sarah Scardicat is worried about neighborhood crime. She has proposed an increase in police service that would require an increase in property tax rates from 43.7 to 47.7 and would lower crime from 0.881 to 0.6. Would this significantly increase property values?

e. Trevor Teacher has proposed to the board that they raise property taxes to hire additional teachers to lower the student-teacher ratio. Board members would be willing to do this only if it doesn't harm property values. Use your regression to compute point and 95% confidence-interval estimates of the amount of increase in the property-tax rate that could be imposed without hurting property values in order to lower the student-teacher ratio by one unit. (You might need to use `nlcom` for this part, even if your model is linear in these variables.)