Econ 410 Spring 2004 Test 3: Building a Hedonic Model of Housing Prices

****

Introduction

You will be asked to specify the independent/explanatory variables and functional form for an equation whose dependent variable is the price of a house in Southern California. Before making these choices, it's vital to review the housing price literature and to think through the theory behind such models. Such a review is especially important in this case because the model we'll be building will be *hedonic* in, nature. A hedonic model is an equation for the price of a house that is a function of the size of that house. The model uses measures of the quality of a product as independent variables instead of measures of the market for that product (like quantity demanded, income, etc.). Hedonic models are most useful when the product being analyzed is heterogeneous in nature because we need to analyze what causes products to be different and therefore to have different prices. With a homogeneous product, hedonic models are virtually useless.

Perhaps the most-cited early hedonic housing price study is that of G. Grether and P. Mieszkowski.[1] Grether and Mieszkowski collected a 7-year data set and built a number of linear models of housing price using different combinations of variables. They included square feet of space, the number of bathrooms, and the number of rooms, although the latter turned out to be insignificant. They also included lot size and the age of the house as variables, specifying a quadratic function for the age variable. Most innovatively, they used several slope dummies in order to capture the interaction effects of various combinations of variables (like a hardwood floors dummy times the size of the house). Peter Linneman estimated a housing price model on data from Los Angeles, Chicago, and the entire United States.[2] His goal was to create a model that worked for the two individual cities and then to apply it to the nation to test the hypothesis of a national housing market. Linneman did not include any lot characteristics, nor did he use any interaction variables. His only measures of the size of the living space were the number of bathrooms and the number of non-bathrooms. Except for an age variable, the rest of the independent variables were dummies describing quality characteristics of the house and neighborhood. Although many of the dummy variables were quite fickle, the coefficients of age, number of bathrooms, and the number of non-bathrooms were relatively stable and significant. Central air conditioning had a negative, insignificant coefficient for the Los Angeles regression. This illustrates the importance of not relying solely on this type of analysis!

K. Ihlanfeldt and J. Martinez-Vasquez investigated sample bias in various methods of obtaining house price data and concluded that the house's sales price is the least biased of all measures.[3] Unfortunately, they went on to estimate an equation by starting with a large number of variables and then dropping all those that had t-scores below one, almost surely introducing bias into their equation since some of the omitted variables could have arguably been relevant/significant in another sample. Finally, Allen Goodman added some innovative variables to an estimate on a national data set. He included measures of specific problems like rats, cracks in the plaster, holes in the floors, plumbing breakdowns, and the level of property taxes. Although the property tax variable showed the capitalization of low property taxes, as would be expected, **the**

---

[1] G. M. Grether and Peter Mieszkowski, 'Determinants of Real Estate Values,' *Journal of Urban Economics,* April 1974, pp. 127-146. Another classic article of the same era is J. Kain and J. Quigley, 'Measuring the Value of Housing Quality,' *Journal of American Statistical Association, June 1970.*

[2] Peter Linneman, 'Some Empirical Results on the Nature of Hedonic Price Functions for the Urban Housing Market,' *Journal of Urban Economics,* July 1980, pp. 47-68.

[3] Keith Ihlanfeldt and Jorge Martinez-Vasquez, 'Alternate Value Estimates of Owner-Occupied Housing: Evidence on Sample Selection Bias and Systematic Errors,' *Journal of Urban Economics,* November 1986, pp. 356-369. Also see Eric Cassel and Robert Mendelsohn, 'The Choice of Functional Forms for Hedonic Price Equations: Comment,' Journal *of Urban Economics,* September 1985, pp.435-142.

rats variable was insignificant, and the cracks variable's coefficient asserted that cracks significantly increase the value of a house! [4]

The Housing Price Interactive Exam

Now that we've reviewed at least a portion of the literature, it's time to build your own model. Here is a simple model of the price of a house as a function of the size of that house,

$$P_i = 40.0 + 0.138S_i \quad (1)$$

where

P_i = the price (in thousands of dollars) of the i^{th} house,

S_i = the size (in square feet) of the i^{th} house.

Equation 1 was estimated on a sample of 43 houses that were purchased in the same Southern California town (Monrovia) within a few weeks of each other. However, there are a number of additional independent variables we will incorporate:

N_i = the quality of the neighborhood of the i^{th} house (1 = best, 4 = worst) as rated by two local real estate agents;

A_i = the age of the i^{th} house in years;

BE_i = the number of bedrooms in the i^{th} house;

BA_i = the number of bathrooms in the i^{th} house;

CA_i = a dummy variable equal to 1 if the i^{th} house has central air conditioning, 0 otherwise;

SP_i = a dummy variable equal to 1 if the i^{th} house has a pool, 0 otherwise;

Y_i = the size of the yard around the i^{th} house (in square feet).

[4] Allen C. Goodman, "An Econometric Model of Housing Price, Permanent Income, Tenure Choice, and Housing Demand," *Journal of Urban Economics,* May 1988, pp. 327-353.

Test 3 Questions: Answer each in a clear, concise, yet comprehensive manner. Your work must correctly employ econometric theory and include all relevant supporting ANOVA printouts.

1) Read through the list of variables, making sure you understand the theory behind using each variable.
What are the expected signs of the coefficients?

2) Run your initial regression. Test for significance for each independent variable by using the appropriate t-scores. Test for joint significance using the F-statistic.

3) What is (are) the main problem(s) with including an irrelevant or superfluous independent variable? How might this problem affect your conclusions? Will the slope estimates likely be "BLUE?" Explain. Which variables seem potentially redundant? Test for superfluous variables. What are your conclusions?

4) What is (are) the main problem(s) with omitting a relevant independent variable? How might this problem affect your conclusions? Will the slope estimates likely be "BLUE?" Explain. Perform at least one "omitted variable" test. Which of the already defined variables do you think you must include in this hedonic price analysis?

There are a number of functional form modifications that could be made to improve the specification.
For example, you might consider:
(i) a log or semi log equation, or
(ii) a quadratic polynomial A_i^2 ***for age, as Grether and Mieszkowski did,***
(iii) combining variables by forming "slope dummies" such as $SP_i \cdot S_i$ or $CA_i \cdot S_i$ ***or***
(iv) combining variables by forming interactive variables that involve the neighborhood proxy variable such as $N_i \cdot S_i$ *or* $N_i \cdot BA_i$ ***.***

Develop your specification carefully. Think through each variable and/or functional form decision, and take the time to write out your expectations for the sign and size of each coefficient. Don't take the attitude that you should include every possible variable and functional form modification and then drop the insignificant ones. Instead, try to design the most logical hedonic model of housing prices you can from the beginning.

CONTINUED ON NEXT PAGE.

5) Find some variation (improvement) of the original form to test. Intuitively explain the variations you have adopted for your new functional form. Regress your updated functional form. Test your hypotheses for each coefficient with the t-test. Pay special attention to any functional form modifications. Test the overall significance of the equation with the F-test.

6) Describe the main problem(s) associated with multicollinearity. How might this problem affect your conclusions? Will the slope estimates likely be "BLUE?" Explain. Test for multicollinearity.

7) Describe the main problem(s) associated with heteroscedasticity. How might this problem affect your conclusions? Will the slope estimates likely be "BLUE?" Explain. Test for heteroscedasticity.

8) Describe the main problem(s) associated with serial correlation. How might this problem affect your conclusions? Will the slope estimates likely be "BLUE?" Explain. Test for serial correlation **(up to 5% bonus).**

9) Decide whether to accept your specification as the best one or to make a modification in your equation and estimate again. Remember your functional form is based on observations and theory. Unless you become absolutely convinced that your original form is sub-optimal, resist the temptation to estimate an additional specification just to see what it looks like. Do NOT force an issue. Remember that although your final specification is important, **your grade is primarily based on how well you justify your steps, while building a case for your final specification.**

Once you've decided to make no further changes, you're finished. Email me (and yourself) your answers in MSWord and your Excel worksheet (fraymond@bellarmine.edu). Be certain to make the subject "Econ 410 Test 3."