

Due in class Tuesday, November 17
(25 points)

The data for this assignment are contained in both the Excel file 'hprice1.xls' and the STATA file 'hprice1.dta' available at <http://u.arizona.edu/~rlo>. These data pertain to a cross-section sample of 88 homes in a certain locale. Be sure to attach the supporting computer print out to the completed assignment, show your work, and make clear where your answers are shown..

The variables of interest for this exercise are *assess* (assessed value \$1,000s), *sqrft* (size of house in square feet), *bdrms* (number of bedrooms), and *lotsize* (size of lot in square feet).

Some basic STATA commands that might be useful

To create the square of a variable named **Z** and give this new variable the name **ZSQRD**, type the command **generate ZSQRD = Z^2**

To estimate a regression model $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$ that generates robust standard errors, type the command **regress Y X1 X2, robust**

1. Use *OLS* to estimate the following model of assessed value

$$assess_i = \beta_0 + \beta_1 sqrft_i + \beta_2 bdrms_i + \beta_3 lotsize_i + \beta_4 (lotsize_i)^2 + u_i, i = 1, \dots, 88$$

Assume that the error term is an independently and normally distributed random variable with mean zero.

2. Test for heteroscedasticity of the error term at the 5% level of significance using the following tests.
 - a. An 'F' test.
 - b. The LM (Lagrange Multiplier) test.
 - c. Based on the above tests, what can you say about the statistical properties of *OLS* estimation of the model?
3. Estimate the assessment model with robust standard errors.
 - a. Explain how this procedure works and what it supposed to be doing.
 - b. How do the estimated coefficients and standard errors compare between the *OLS* procedure and the robust estimation procedure?

4. Suppose it is the case that $u_i \sim N(0, (lotsize_i)^2 \sigma^2)$.
 - a. Specify and estimate the appropriate Generalized Least Squares (*GLS*) aka Weighted Least Squares (*WLS*) transformation of the assessed value model.
 - b. Compare the *OLS* with the *GLS* estimates of the coefficients and standard errors.
5. Estimate the model $assess_i = \beta_0 + \beta_1 sqft_i + \beta_2 bdrms_i + \beta_3 lotsize_i + u_i$.
 - a. Compare your estimate of β_1 from this model with your estimate of β_1 from the model you estimated above that included the variable $(lotsize_i)^2$. What is the likely direction of the bias in estimating β_1 caused by omitting $(lotsize_i)^2$ from the regression?
 - b. Use the bias formula from class notes to obtain a theoretical expression for the bias in the *OLS* estimate of β_1 .