Computer Homework 4

This homework examines the use of transformations in linear regression models. Use the data in the *Stata* data file mls.dta. Consider the following model:

$$\text{Model A:} \qquad SP(\lambda) = \alpha + \beta\, SQFT + \varepsilon$$

where $SP(\lambda)$ is a Box-Cox transformation of selling price. The observational subscript has been omitted to simplify notation.

1. Standardize selling price by its geometric mean. The geometric mean is simply the exponential transformation of the sample mean of the log of selling price.

2. Conditional on λ, the model is linear in the parameters. Use a grid search over $\lambda \varepsilon [0,1]$ in conjunction with OLS to obtain ML estimates of Model A. Make the grid search accurate to the first decimal place. (See the *Stata* program below.) The ML estimate of λ is the value that minimizes the sum-of-squared residuals, and the ML estimates of α and β are the OLS estimates obtained with the minimizing value of λ.

3. Compare the "goodness of fit" of the Box-Cox model with that of the linear model ($\lambda=1$) and the semi-log model ($\lambda=0$).

4. Test the hypothesis $\lambda=0$. What does this mean?

5. Estimate the model using the "canned" Box-Cox procedure in *Stata*. The command line is

   boxcox ssp sqft, model(lhs)

   where **model(lhs)** indicates that the BC transformation is applied only to the dependent variable. This procedure uses a gradient based method rather than a grid search.

***Stata* Program**

The grid search is easily accomplished using the looping capabilities of *Stata*. The program below standardizes SP by its geometric mean, then computes the Box-Cox transformed value of SP for lambda in the interval [0, 0.5] by increments of 0.1. The *Stata* **scalar** command is similar to **generate**, but works with scalars rather than vectors. The **while**, **if**, and **else** commands execute the lines between the open and close braces, { } if the condition in the statement is met. In this program, the **while** statement nests the **if** and **else** statements. Note the double equal in the if statement. My comments (between /* and */) may be omitted from your program.

```
gen lsp=log(sp)
sum lsp
gen ssp=sp/exp(r(mean))                    /*Standardize by Geometric Mean*/
scalar i=0                                 /*Set iteration counter*/
while i<=5 {
        scalar j=i*0.1
        if j==0 gen bcp=log(ssp)           /*BC trans. for zero lambda*/
        else gen bcp=((ssp^j)-1)/j         /*BC trans. for nonzero lambda*/
        display " "
        display "BOX-COX MODEL WITH PARAMETER " j
        reg bcp sqft                       /*OLS for given lambda*/
        drop bcp
        scalar i=i+1                       /*update counter*/
}
```