Multiple Regression Review

Antonio Pacifico
Federica Romei

October 22, 2012

1. The dataset for this exercise is hprice1.dta that contains 88 obsevations on the following data:

   - *price*: house price in \$ 1000s ;
   - *assess* : assessed value in \$ 1000s;
   - *sqrt*: size of house in square feet;
   - *bdrms*: number of bedrooms;
   - *colonial*: dummy variable
     - =1 if home is a colonial style;
     - =0 otherwise.

   Assume homoschedasticity and normality of the errors and run the regression

   $$price = \beta_0 + \beta_1 assess + \beta_2 sqrft + \beta_3 bdrms + \beta_4 colonial + u$$

   (a) Test at 5% the hypothesis that only colonial style is a significant variable in the house price determination.

   Solution: We have to test

   $$H_0 \quad \beta_1 = \beta_2 = \beta_3 = 0$$

   under homoschedasticity and normality of the errors.
   We can run the first regression with all the coefficient

   $$price = \beta_0 + \beta_1 assess + \beta_2 sqrt + \beta_3 bdrms + \beta_4 colonial + u$$

   and call the R^2 associated to this regression $R^2_{unrestricted}$. Then we can run another regression, where we set at zero β_1, β_2 and β_3, i.e. :

   $$price = \beta_{r,0} + \beta_{r,4} colonial + u$$

and we call the R^2 associated to this regression $R^2_{restricted}$.

We can easily compute

$$F = \frac{(R^2_{unrestricted} - R^2_{restricted})/q}{1 - R^2_{unrestricted}/(n - k_{unrestricted} - 1)}$$

where q is the number of restriction (3 in our case), n is the number of sample observation (88 in our case)and $k_unrestricted$ is the number of coefficient of the unrestricted regression excluded the constant (4 in our case). Once we have the F value we have to compare with $\tilde{F_{q,n-k-1}}$ at 5%. If $F > \tilde{F_{q,n-k-1}}$ we will reject H_0, we will not reject otherwise.

First of all, write on Stata

```
reg price assess sqrft bdrms colonial
```

and you should visualize :

```
    Source |       SS       df       MS              Number of obs =      88
-------------+------------------------------           F(  4,    83) =   99.80
       Model |  759860.557     4  189965.139           Prob > F      =  0.0000
    Residual |  157994.038    83  1903.54263           R-squared     =  0.8279
-------------+------------------------------           Adj R-squared =  0.8196
       Total |  917854.596    87  10550.0528           Root MSE      =   43.63

------------------------------------------------------------------------------
       price |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      assess |   .9471415   .0982252     9.64   0.000     .7517754    1.142508
       bdrms |   9.942738   6.930284     1.43   0.155    -3.841318    23.72679
       sqrft |  -.0033656   .0168439    -0.20   0.842    -.0368675    .0301363
    colonial |   9.186462   10.67217     0.86   0.392    -12.04005    30.41297
       _cons |  -40.56921   21.65287    -1.87   0.065    -83.63589    2.497472
------------------------------------------------------------------------------
```

where $R^2_{unrestricted} = 0.8279$.

Now run the restricted regression,i e.

```
reg price colonial
```

and you should visualize :

```
      Source |       SS       df       MS              Number of obs =      88
-------------+------------------------------           F(  1,    86) =    1.67
       Model |  17465.8965     1  17465.8965           Prob > F      =  0.2000
    Residual |  900388.699    86   10469.636           R-squared     =  0.0190
-------------+------------------------------           Adj R-squared =  0.0076
       Total |  917854.596    87  10550.0528           Root MSE      =  102.32

------------------------------------------------------------------------------
       price |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    colonial |   30.54851   23.65159     1.29   0.200    -16.4693    77.56631
       _cons |   272.3704   19.69173    13.83   0.000    233.2245    311.5162
------------------------------------------------------------------------------
```

where $R^2_{restricted} = 0.019$.

We can compute the $F - statistic$:

$$F_{3,83} = \frac{(0.8279 - 0.019)/3}{1 - 0.8279/(88 - 4 - 1)}$$

$$F = 132.78$$

The threshold is almost 2. Then you can reject the Null Hypothesis that all coefficients, except the constant and the one for colonial, are equal to zero at 5%. Indeed F>2.

Even if you don't have the threshold value, you can see that $F - statistic$ is quite large, then you have to reject H_0

(b) Now assume heteroschedasticity of the errors and test at 5% the hypothesis that only colonial style is a significant variable in the house price determination.

Solution: In this case we can't use the previous formula. Then we can run the regression

$$price = \beta_0 + \beta_1 assess + \beta_2 sqrft + \beta_3 bdrms + \beta_4 colonial + u$$

on Stata as before.

```
reg price assess bdrms sqrft colonial,r

Linear regression                                      Number of obs =      88
                                                       F(  4,    83) =   52.75
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.8279
                                                       Root MSE      =   43.63

------------------------------------------------------------------------------
             |               Robust
       price |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      assess |   .9471415    .122009     7.76   0.000     .7044705    1.189812
       bdrms |   9.942738   5.580125     1.78   0.078    -1.155907    21.04138
       sqrft |  -.0033656   .0176428    -0.19   0.849    -.0384564    .0317252
    colonial |   9.186462   11.41456     0.80   0.423    -13.51664    31.88956
       _cons |  -40.56921   24.97071    -1.62   0.108    -90.23495    9.096532
------------------------------------------------------------------------------
```

In this case, assuming heteroschedasticity, we use the option `robust` on Stata. As you can see the coefficients are the same, while the standard errors changed with respect to the homoschedastic case. On Stata, to perform a jointly test you can use the command `test` putting on parenthesis the coefficients you want to test. In our case:

test (assess=0) (bdrms=0) (sqrft=0).

Once you have your output you should compare $p-value$ with α you fixed. As in the $t-statistc$ if $p-value < \alpha$ you will reject H_0 at α % significant level.

In our case the output is :

```
test (assess=0) (bdrms=0) (sqrft=0)

 ( 1)  assess = 0
 ( 2)  bdrms = 0
 ( 3)  sqrft = 0

       F(  3,    83) =   70.10
            Prob > F =    0.0000
```

Our $\alpha = 0.05$ while $p-value = 0$, then we will reject H_0 at 5% significant level.

(c) Run the regression:

$$price = \beta_0 + \beta_1 assess + \beta_2 bdrms + \beta_3 sqrft + \beta_4 lotsd + \beta_5 colonial + u$$

A friend of yours claims that the marginal effect of the size of lot in square feet on the house's price is the same as the marginal effect of size of house in square feet on the house's price. Test in at least 2 ways this sentence.

Solution: You can run a regression on Stata typing :

```
reg price assess bdrms sqrft lotsize colonial,r
```

You will visualize the output:

```
reg price assess bdrms sqrft lotsize colonial,r

Linear regression                                      Number of obs =      88
                                                       F(  5,    82) =   44.18
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.8309
                                                       Root MSE      =  43.511

------------------------------------------------------------------------------
             |               Robust
       price |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      assess |    .904078   .1223519     7.39   0.000     .6606812    1.147475
       bdrms |   9.630255   5.777185     1.67   0.099    -1.862407    21.12292
       sqrft |   .0010714   .0180266     0.06   0.953    -.0347892    .0369319
     lotsize |   .0005993   .0002328     2.57   0.012     .0001362    .0010623
    colonial |   9.547572   11.49432     0.83   0.409     -13.3183    32.41344
       _cons |  -40.44767   24.81617    -1.63   0.107    -89.81494    8.919605
------------------------------------------------------------------------------
```

Then you can test

$$H_0 : \beta_3 = \beta_4$$

typing

```
test (sqrft-lotsize=0)
```

and you will get the output:

```
r. test (lotsize - sqrft=0)

 ( 1) - sqrft + lotsize = 0

       F(  1,    82) =    0.00
            Prob > F =    0.9791
```

Since $p - value > 0.05$ you cannot reject H_0 at 5% significance level.

Notice that the test has only one restriction. We can then rewrite the equation as:

$$price = \beta_0 + \beta_1 assess + \beta_2 bdrms + \gamma sqrft + \eta(lotsd + sqrft) + \beta_5 colonial + u$$

where

$$\gamma = \beta_3 - \beta_4$$

and test that :

$$H_0 : \quad \gamma = 0$$

In Stata you should create a new variable

```
gen lotsd= sqrft + lotsize
```

then you should type:

```
reg price assess bdrms sqrft colonial lotsd,r
```

and you will get:

```
reg price assess bdrms sqrft  colonial lotsd,r

Linear regression                                      Number of obs =      88
                                                       F(  5,    82) =   44.18
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.8309
                                                       Root MSE      =  43.511

------------------------------------------------------------------------------
             |               Robust
       price |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      assess |    .904078   .1223519     7.39   0.000      .6606812    1.147475
       bdrms |   9.630255   5.777185     1.67   0.099     -1.862407    21.12292
       sqrft |   .0004721   .0179405     0.03   0.979     -.0352173    .0361615
    colonial |   9.547572   11.49432     0.83   0.409      -13.3183    32.41344
       lotsd |   .0005993   .0002328     2.57   0.012      .0001362    .0010623
       _cons |  -40.44767   24.81617    -1.63   0.107     -89.81494    8.919605
------------------------------------------------------------------------------
```

$\gamma = 0.03$ and it has a $p_value = 0.979 > 0.05$ which means that you cannot reject H_0 at 5% significance level.