Multiple Regression Review

Siria Angino
Federica Romei

October 27, 2013

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 $F_{q,n\tilde{-}k-1}$ at 5%. If $F > F_{q,n\tilde{-}k-1}$ we will reject H_0, we will not reject otherwise. The fastest way is to type on R

```
$R_u=summary(lm(price~ assess + sqrft +bdrms + colonial))\$r.squared$

$R_r=summary(lm(price~  colonial))\$r.squared$

a= (R_u - R_r)/3
b= (1 - R_u)/(length(price) - 4 -1)
Ftest=a/b
Ftest
```

and you should visualize :

```
    -> Ftest
[1] 130.0023
```

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 know the threshold value, $F - statistic$ is very 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. In this case, assuming heteroschedasticity, we use the option `linearHypothesis` on R (to use this command you need the package "car"). You can write

```
m1=lm(price~ assess + sqrft +bdrms + colonial)
```

```
rhs= c(0,0,0)
hm= rbind(c(0,1,0,0,0),c(0,0,1,0,0),c(0,0,0,1,0))
#HOMO
linearHypothesis(m1,hm,rhs)
#HETERO
linearHypothesis(m1,hm,rhs,vcov=vcovHC)
```

where m1 is the unrestricted regression. Then you build rhs and hm s.t.:

$$H_0: \;\; hm * \beta = rhs$$

$$H_1: \;\; hm * \beta \neq rhs.$$

Moreover if you want to perform a joint hypothesis test under heteroschedasticity you should write "vcov=vcovHC" in the command linearHypothesis. You will visualize

```
Linear hypothesis test

Hypothesis:
assess = 0
sqrft = 0
bdrms = 0

Model 1: restricted model
Model 2: price ~ assess + sqrft + bdrms + colonial

Note: Coefficient covariance matrix supplied.

  Res.Df Df     F    Pr(>F)
1     86
2     83  3 52.88 < 2.2e-16 ***
---
Signif. codes:  0 Ô***Õ 0.001 Ô**Õ 0.01 Ô*Õ 0.05 Ô.Õ 0.1 Ô Õ 1
```

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

2. The dataset we will use is CEOSAL1. It contains 209 observation for the year 1990 on the the following data:

   - *salary*: 1990 salary , thousands \$;
   - *sales* : 1990 firm sales, millions \$.

   We will not use the other variables.

   (a) Your employer wants to know what happens to the salary of CEO if the sales increases of 1%.

   Solution: If you want to estimate the level increase of the dependent variable when a regressor increases of 1% you have to estimate a **linear-log** model. In our case it will be:

   $$salary = \beta_0 + \beta_1 log(sales) + u.$$

   In this case if *sales* increases of 1%, then salary will change of $\frac{\beta_1}{100}$ thousands dollar. We can estimate the model using the command

   ```
   rob(lm(salary  lsales ))
   ```

   and obtain the following output:

   ```
   > rob(lm(salary~ lsales ))
                Estimate Std. Error   t value     Pr(>|t|) c.i. 2.5 %  c.i. 97.5%
   (Intercept) -898.9290  539.04891 -1.667620 9.690317e-02 -1955.4454    157.5874
   lsales       262.9015   62.29493  4.220271 3.650798e-05   140.8056    384.9973
   ```

   This means that if the sales increase of 1%, the salary of CEO will increase of $2,62$ thousand dollars.

   (b) Your boss wants to know how the percentage salary of CEO will change if the sales increases of 1 million.

   Solution: If you want to estimate the percentage change in the dependent variable when a regressor increases of 1 unit, you have to use a **log-linear** model. In our case it will be:

   $$lsalary = \beta_0 + \beta_1 sales + u.$$

   In this case if *sales* increases of 1 thousand the salary will increase of $\beta_1 * 100\%$. We can estimate the model using the command

   ```
   rob(lm(lsalary  sales ))
   ```

   and we obtain the following output:

```
> rob(lm(lsalary~ sales ))
               Estimate   Std. Error    t value      Pr(>|t|)   c.i. 2.5 %  c.i. 97.5 %
(Intercept) 6.84665e+00 5.991082e-02 114.280701 5.527125e-189 6.729227e+00 6.964073e+00
sales       1.49825e-05 7.483440e-06   2.002087  4.658052e-02 3.152282e-07 2.964977e-05
```

If *sales* increase of 1 million, the salary of Ceo will increases of 0.0015%.

(c) Your boss wants to know what is the elasticity between sales and salary.

Solution: If you want to estimate the percentage change in the dependent variable when a regressor increases of 1 percent (or the elasticity), you have to use a **log-log** model. In our case it will be:

$$lsalary = \beta_0 + \beta_1 lsales + u.$$

In this case if *sales* increases of 1 percent will increase of β_1%. We can estimate the model using the command

`rob(lm(lsalary  lsales ))`

and we obtain the following output:

```
rob(lm(lsalary~ lsales ))
             Estimate Std. Error   t value     Pr(>|t|) c.i. 2.5 % c.i. 97.5 %
(Intercept) 4.8219962 0.27603865 17.468555 1.363835e-42  4.2809704   5.3630220
lsales      0.2566717 0.03265347  7.860474 2.066079e-13  0.1926721   0.3206713
```

If *sales* increase of 1 percent, the salary of Ceo will increases of 0.25%.