Lab 11 - Heteroskedasticity

Spring 2018

Contents

1 Introduction

Today we study what happens when the assumption of homoskedasticity is violated and how to address such cases.

- One of the assumptions that we made in deriving the OLS estimator as the best linear unbiased estimator was **homoskedasticity**. This assumption says that the error u in the PRF has the same variance given any value of the explanatory variables. In other words,

$$Var(u|x) = \sigma^2$$

  This assumption implies that $Var(y|x) = \sigma^2$ i.e. the variance of y is *constant* and does not change with x.

- When this assumption is violated we say that the error term is **heteroskedastic**.

- Keep in mind that heteroskedasticity does not affect the estimate of the β's in your regression. The OLS estimator will still produce unbiased and consistent estimates of β.

- However, heteroskedasticity does affect the estimates of your standard errors. If s.e.($\hat{\beta}$) are derived under the assumption of homoskedasticity when in fact the error is heteroskedastic, s.e.($\hat{\beta}$) will be *biased*.

- This means that, if you do not correct for heteroskedasticity, the confidence intervals you construct as well as the test statistics you use to test hypotheses are no longer valid!

2 Heteroskedasticity

- If the error term is heteroskedastic then we have

$$Var(u_i|x_i) = \sigma_i^2$$

  where the subscript i on σ emphasizes the fact that the variance will depend on the particular value of x under consideration.

- In this case, even if the form of heteroskedasticity is not known, we can still produce a correct estimate of the $Var(\hat{\beta})$ and, hence its standard error. This standard errors are called **heteroskedasticity-robust standard errors** (or simply robust s.e.) and their formula is

$$\widehat{Var}(\hat{\beta}_j) = \frac{\sum_i^n \hat{r}_{ij}^2 \hat{u}_i^2}{SSR_j^2}$$

  where $\hat{r}_{ij}^2$ denotes the i^{th} residual from regressing x_j on all other regressors, and SSR_j is the sum of squared residuals from this regression.

- The standard errors estimated in this way are also known as White s.e., from the name of the person who first derived them.

- Once the robust standard errors are obtained, you can conduct correct hypothesis testing in the standard way and keep using the t and F statistics, now using the newly-derived standard errors.

3 Addressing heteroskedasticity in Stata

- Consider the following model of voting outcomes:

$$voteA = \beta_0 + \beta_1 prtystrA + \beta_2 democA + \beta_3 ln(expendA)_i + \beta_4 ln(expendB) + \epsilon_i \qquad (1)$$

  and try to estimate it with regular standard errors and with robust standard errors.

- You can estimate robust standard errors by simply adding the option `robust` at the end of your estimation command:

```
bcuse vote1, clear
reg voteA prtystrA democA lexpendA lexpendB
reg voteA prtystrA democA lexpendA lexpendB, robust
```

```
------------------------------------------------------------------------------
       voteA |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    prtystrA |   .2519175   .0712925     3.53   0.001     .1111729    .3926622
      democA |   3.792944    1.40652     2.70   0.008     1.016213    6.569674
    lexpendA |   5.779294   .3918197    14.75   0.000      5.00577    6.552819
    lexpendB |  -6.237836   .3974596   -15.69   0.000    -7.022495   -5.453178
       _cons |   37.66142   4.736036     7.95   0.000      28.3116    47.01123
------------------------------------------------------------------------------

------------------------------------------------------------------------------
             |               Robust
       voteA |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    prtystrA |   .2519175   .0660631     3.81   0.000     .1214968    .3823383
      democA |   3.792944   1.452168     2.61   0.010      .926096    6.659791
    lexpendA |   5.779294    .533142    10.84   0.000     4.726773    6.831816
    lexpendB |  -6.237836   .3561583   -17.51   0.000    -6.940959   -5.534714
       _cons |   37.66142   4.418941     8.52   0.000     28.93761    46.38522
------------------------------------------------------------------------------
```

- The top panel reports the result for the regular estimation, whereas the bottom panel reports the results for the heteroskedasticity-robust estimation (you can see this because the label for standard errors says “robust”). Note the following things:
- The estimates for the β’s do not change when you allow for heteroskedasticity.
- The estimated standard errors for the β’s change. In general, the standard errors could be larger or smaller, although empirically they tend to be larger. In our case, $s.e.(\hat{\beta}_1)$ is smaller but $s.e.(\hat{\beta}_2)$ is larger.

- By changing the s.e. we change the t-statistics and this potentially alters the result of your tests of hypotheses.
- Similarly, if you compute the same F test under the different estimators you will get different answers:

```
. test prtystrA democA

 ( 1)  prtystrA = 0
 ( 2)  democA = 0

       F(  2,   168) =    6.75
            Prob > F =    0.0015
```

```
. test prtystrA democA

 ( 1)  prtystrA = 0
 ( 2)  democA = 0

       F(  2,   168) =    7.55
            Prob > F =    0.0007
```

- Above we are testing the hypothesis that $\beta_1 = \beta_2 = 0$. The left panel reports the test under the assumption of homoskedasticity while the right panel corrects for heteroskedasticity. You can see that the p-values are different.

4 Testing for heteroskedasticity

- So far we have seen how to address heteroskedasticity but we have not said anything about how we test for it. There are two main tests for heteroskedasticity.
- The first one is the **Breusch-Pagan test**, which assumes a linear functional form between the variance of the error term and the independent variables in your regression.
- In both these test the null hypothesis H_0 is homoskedasticity.
- You can implement this test after running your regression by using the command `estat hettest`
- Testing for heteroskedasticity for the above regression yields:

```
Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of voteA

         chi2(1)      =     2.95
         Prob > chi2  =   0.0861
```

- The test statistic follows an χ^2 distribution. In this case we reject the null hypothesis of homoskedasticity at a 10% significance level (although not at a 5%) in favor of heteroskedasticity.
- A second test for heteroskedasticity is the **White test**. The White test is more general than the Breusch-Pagan test in that it allows for a non-linear relation between the variance of the error term and the x's. You can implement this test by typing
`estat imtest, white`

- The first part of the output is:

```
White's test for Ho: homoskedasticity
         against Ha: unrestricted heteroskedasticity

         chi2(13)     =     31.10
         Prob > chi2  =    0.0033
```

- This test statistic follows a χ^2 distribution and the p-value is 0.0033. In this case we can reject homoskedasticity in favor of heteroskedasticity even at a 1% confidence level.
- The tests that we conducted indicate that we should use the robust standard errors for our voting model above.

5 A simple example

- Consider the dataset `saving` (access using the `bcuse` command) on consumption, savings and income and the following consumption regression:

$$cons_i = \beta_0 + \beta_1 inc_i + \epsilon_i$$

- The estimate yields:

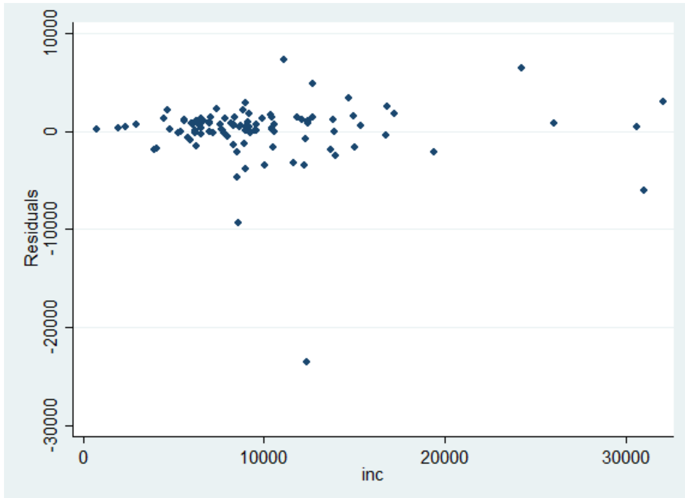
```
reg cons inc

      Source |       SS       df       MS              Number of obs =     100
-------------+------------------------------           F(  1,    98) =  219.89
       Model |  2.2480e+09     1  2.2480e+09           Prob > F      =  0.0000
    Residual |  1.0019e+09    98  10223460.8           R-squared     =  0.6917
-------------+------------------------------           Adj R-squared =  0.6886
       Total |  3.2499e+09    99    32827569           Root MSE      =  3197.4

------------------------------------------------------------------------------
        cons |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         inc |   .8533717   .0575488    14.83   0.000      .739168    .9675753
       _cons |  -124.8424   655.3931    -0.19   0.849    -1425.449    1175.764
------------------------------------------------------------------------------
```

- To visually check whether there is heteroskedasticity we can plot the residuals against our independent variable income:

```
predict u, resid
scatter u inc
```



From the picture, you can see that as income increases the variance of u increases. We can formally test for heteroskedasticity with the two tests discussed above:

```
. estat hettest
Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of cons

         chi2(1)      =    14.22
         Prob > chi2  =   0.0002

. estat imtest, white
White's test for Ho: homoskedasticity
         against Ha: unrestricted heteroskedasticity

         chi2(2)      =     1.85
         Prob > chi2  =    0.3967
```

- In this case only the first test rejects homoskedasticity. Reestimating the model with robust standard errors yields:

```
. reg cons inc, robust

Linear regression                                      Number of obs =     100
                                                       F(  1,    98) =  193.52
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.6917
                                                       Root MSE      =  3197.4

------------------------------------------------------------------------------
             |               Robust
        cons |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         inc |   .8533717   .0613441    13.91   0.000     .7316363     .975107
       _cons |  -124.8424   528.2192    -0.24   0.814    -1173.076    923.3915
------------------------------------------------------------------------------
```