Heteroscedasticity Problem

1 Nature and Consequences of heteroscedasticity for OLS

- Heteroskedasticity (broadly) -
- Heteroskedasticity (in econometrics) -

1.1 Nature of Heteroskedasticity

1.2 Consequences of Heteroskedasticity

1. Does not affect the biasedness of the OLS estimators

2. Does not affect the value of R^2 and $adj.R^2$

3. Make the estimated value of $Var(\widehat{\beta}_{OLS})$ wrong

4. Affect the correctness of our inference

1.3 How can the estimated value of $Var(\widehat{\beta}_{OLS})$ be wrong?

Suppose

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

Given that assumption 1 to 4 are true, but assumption 5 (homoskedasticity) is violated. Thus,

$$Var(u_i|x_i) =$$

And from the OLS estimation steps, we can write

$$\hat{\beta}_1 = \beta_1 +$$

1.4 Two types of remedies

1. Passive

2. Active

2 Testing for heteroskedasticity

- The main point -

Suppose

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + u$$

Assume that assumption 1 to 4 are true. Our hypotheses to test for heteroskeroskedasticity would be

We know that $Var(u|\mathbf{x}) = E(u^2|\mathbf{x}) - [E(u|\mathbf{x})]^2$. But ______________________ according to assumption 4. Thus, H_0 and H_a can be written as

2.1 Breusch-Pagan test (BP test)

To perform the Breusch-Pagan Test in STATA

STATA commands (in case $k = 4$):

regress y x_1 x_2 x_3 x_4

predict u_hat, residual

generate u_hat_sq = u_hat^2

regress u_hat_sq x_1 x_2 x_3 x_4

** Then, check the F-statistic on the top right-hand corner of the result table.

Example: Finding the determinants of GPA.

```
. regress  termgpa attend priGPA final frosh soph

      Source |       SS           df       MS      Number of obs   =       680
-------------+----------------------------------   F(  5,    674)  =    214.13
       Model |  226.077541         5  45.2155081   Prob > F        =    0.0000
    Residual |  142.319996       674  .211157264   R-squared       =    0.6137
-------------+----------------------------------   Adj R-squared   =    0.6108
       Total |  368.397537       679  .542558964   Root MSE        =    .45952

------------------------------------------------------------------------------
     termgpa |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      attend |    .046594   .0036082    12.91   0.000     .0395093    .0536787
      priGPA |   .5329307   .0403281    13.21   0.000     .4537468    .6121146
       final |   .0503197   .0040339    12.47   0.000     .0423992    .0582403
       frosh |   .0974307   .0560211     1.74   0.082    -.0125662    .2074276
        soph |   .0689273   .0467006     1.48   0.140    -.0227689    .1606236
       _cons |  -1.361077   .1316861   -10.34   0.000    -1.619642   -1.102513
------------------------------------------------------------------------------
```

```
. predict u_hat, residual

. generate u_hat_sq = u_hat^2

. regress u_hat_sq  attend priGPA final frosh soph

 regress u_hat_sq attend priGPA final frosh soph

      Source |       SS       df       MS              Number of obs =     680
-------------+------------------------------           F(  5,   674) =   14.54
       Model |  8.22606613     5  1.64521323           Prob > F      =  0.0000
    Residual |  76.2624962   674  .113149104           R-squared     =  0.0974
-------------+------------------------------           Adj R-squared =  0.0907
       Total |  84.4885623   679  .124430872           Root MSE      =  .33638

------------------------------------------------------------------------------
    u_hat_sq |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      attend |  -.0088079   .0026413    -3.33   0.001    -.0139941   -.0036218
      priGPA |  -.1454432    .029521    -4.93   0.000    -.2034074   -.0874791
       final |   .0061879   .0029529     2.10   0.036     .0003899    .0119859
       frosh |  -.1077493   .0410085    -2.63   0.009    -.1882692   -.0272294
        soph |  -.0975658   .0341858    -2.85   0.004    -.1646892   -.0304423
       _cons |   .7368933   .0963968     7.64   0.000     .5476191    .9261674
------------------------------------------------------------------------------
```

Alternatively, you can use the following set of STATA commands:
regress y x_1 x_2 x_3 x_4
estat hettest x_1 x_2 x_3 x_4

```
. regress termgpa attend priGPA final frosh soph

      Source |       SS       df       MS              Number of obs =     680
-------------+------------------------------           F(  5,   674) =  214.13
       Model |  226.077541     5  45.2155081           Prob > F      =  0.0000
    Residual |  142.319996   674  .211157264           R-squared     =  0.6137
-------------+------------------------------           Adj R-squared =  0.6108
       Total |  368.397537   679  .542558964           Root MSE      =  .45952

------------------------------------------------------------------------------
     termgpa |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      attend |    .046594   .0036082    12.91   0.000     .0395093    .0536787
      priGPA |   .5329307   .0403281    13.21   0.000     .4537468    .6121146
       final |   .0503197   .0040339    12.47   0.000     .0423992    .0582403
       frosh |   .0974307   .0560211     1.74   0.082    -.0125662    .2074276
        soph |   .0689273   .0467006     1.48   0.140    -.0227689    .1606236
       _cons |  -1.361077   .1316861   -10.34   0.000    -1.619642   -1.102513
------------------------------------------------------------------------------

. estat hettest attend priGPA final frosh soph

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: attend priGPA final frosh soph

         chi2(5)      =    93.90
         Prob > chi2  =   0.0000
```

If the null hypothesis is rejected (we have the heteroskedasticity problem), we can use the ", robust" option in STATA. This option gives us the correct standard error, or "heteroskedasticity-robust standard error". We can now use the t-statistics in this case.

```
. regress termgpa attend priGPA final frosh soph, robust

Linear regression                                      Number of obs =     680
                                                       F(  5,   674) =  234.28
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.6137
                                                       Root MSE      =  .45952

------------------------------------------------------------------------------
             |               Robust
     termgpa |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      attend |    .046594   .0044101    10.57   0.000     .0379348    .0552532
      priGPA |   .5329307   .0426426    12.50   0.000     .4492023     .616659
       final |   .0503197   .0041066    12.25   0.000     .0422564     .058383
       frosh |   .0974307   .0633543     1.54   0.125    -.0269648    .2218262
        soph |   .0689273   .0520495     1.32   0.186    -.0332713    .1711259
       _cons |  -1.361077   .1448208    -9.40   0.000    -1.645431   -1.076723
------------------------------------------------------------------------------
```

2.2 The White Test

Similar to the Breush-Pagan test, but is stricter because it does not allow $\widehat{u}^2$ to be correlated with x^2 or interactions among different x_s.

Suppose

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + u$$

The White Test (special case) (save degree of freedom)

1. Get $\hat{\beta}_1, \hat{\beta}_2, ..., \hat{\beta}_k$ by OLS.
2. Calculate $\hat{u}_i^2 = [y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_k x_k)]^2$
3. Calculate $\hat{y}_i = (\hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_k x_k)$
4. Calculate $\hat{y}_i^2$
5. Estimate $\hat{u}_i^2 = \gamma_0 + \gamma_1 \hat{y}_i + \gamma_2 \hat{y}^2 + error$ (keep R^2 of this regression)
6. $LM = nR^2$
7. If $p - value > significance\ level$, cannot reject H_o.