Takashi Yamano
Fall Semester 2009

Lecture Notes on Advanced Econometrics

Lecture 9: Heteroskedasticity and Robust Estimators

In this lecture, we study heteroskedasticity and how to deal with it. Remember that we did not need the assumption of Homoskedasticity to show that OLS estimators are unbiased under the finite sample properties and consistency under the asymptotic properties. What matters is how to correct OLS standard errors.

Heteroskedasticity

In this section, we consider heteroskedasticity, while maintaining the assumption of no-autocorrelation. The variance of disturbance *i*, u_i, is not constant across observations but not correlated with u_j:

$$E(uu') = \begin{bmatrix} E(u_1u_1) & E(u_1u_2) & & E(u_1u_n) \\ E(u_2u_1) & E(u_2u_2) & & E(u_2u_n) \\ & & & \\ E(u_nu_1) & E(u_nu_2) & & E(u_nu_n) \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & 0 & & 0 \\ 0 & \sigma_2^2 & & 0 \\ & & & \\ 0 & 0 & & \sigma_n^2 \end{bmatrix} = \Sigma$$

or

$$E(uu') = \sigma^2 \begin{bmatrix} \sigma_1^2/\sigma^2 & 0 & & 0 \\ 0 & \sigma_2^2/\sigma^2 & & 0 \\ & & & \\ 0 & 0 & & \sigma_n^2/\sigma^2 \end{bmatrix} = \sigma^2\,\Omega$$

Notice that under homoskedasticity, $\Omega = I$.

Under heteroskedasticity, the sample variance of OLS estimator (under finite sample properties) is

$$\begin{aligned} Var(\hat{\beta}) &= Var[\beta + (X'X)^{-1}X'u] \\ &= E[(X'X)^{-1}X'uu'X(X'X)^{-1}] \\ &= (X'X)^{-1}X'E(uu')X(X'X)^{-1} \\ &= \sigma^2(X'X)^{-1}X'\Omega X(X'X)^{-1} \end{aligned} \qquad (1)$$

(See Theorem 10.1 in Greene (2003))

Unless you specify, however, econometric packages automatically assume homoskedasticity and will calculate the sample variance of OLS estimator based on the homoskedasticity assumption:

$$Var(\hat{\beta}) = \sigma^2 (X'X)^{-1}$$

Thus, in the presence of heteroskedasticity, the statistical inference based on $\sigma^2 (X'X)^{-1}$ would be biased, and *t*-statistics and *F*-statistics are inappropriate. Instead, we should use (1) to calculate standard errors and other statistics.

Finite Sample Properties of OLS Estimators
The OLS estimators are unbiased and have the sampling variance specified in (6-1). If u is normally distributed, then the OLS estimators are also normally distributed:

$$\hat{\mathrm{B}} \mid X \sim N[B, \sigma^2 (X'X)^{-1} (X'\Omega X)(X'X)^{-1}]$$

Asymptotic Properties of OLS Estimators
If $p\lim(X'X/n) = Q$ and $p\lim(X'\Omega X/n)$ are both finite positive definite matrices, then $Var(\hat{\beta})$ is consistent for $Var(\beta)$.

Robust Standard Errors
If Σ is known, we can obtain efficient least square estimators and appropriate statistics by using formulas identified above. However, as in many other problems, Σ is unknown. One common way to solve this problem is to estimate Σ empirically: First, estimate an OLS model, second, obtain residuals, and third, estimate Σ:

$$\hat{\Sigma} = \begin{bmatrix} \hat{u}_1^2 & 0 & 0 \\ 0 & \hat{u}_2^2 & 0 \\ & & \\ 0 & 0 & \hat{u}_n^2 \end{bmatrix}$$

(We may multiply this by (*n*/(*n*-*k*-1)) as a degree-of-freedom correction. But when the number of observations, *n*, is large, this adjustment does not make any difference.)

Thus by using the estimated Σ, we have

$$X'\hat{\Sigma}X = X' \begin{bmatrix} \hat{u}_1^2 & 0 & 0 \\ 0 & \hat{u}_2^2 & 0 \\ & & \\ 0 & 0 & \hat{u}_n^2 \end{bmatrix} X .$$

Therefore, we can estimate the variances of OLS estimators (and standard errors) by using $\hat{\Sigma}$:

$$Var(\hat{\beta}) = (X'X)^{-1} X'\hat{\Sigma}X(X'X)^{-1}$$

Standard errors based on this procedure are called (**heteroskedasticity) robust standard errors** or **White-Huber standard errors**. Or it is also known as the **sandwich estimator of variance** (because of how the calculation formula looks like). This procedure is reliable but entirely empirical. We do not impose any assumptions on the structure of heteroskedasticity.

Sometimes, we may impose assumptions on the structure of the heteroskedasticity. For instance, if we suspect that the variance is homoskedastic within a group but not across groups, then we obtain residuals for all observations and calculate average residuals for each group. Then, we have $\hat{\Sigma}$ which has a constant $\hat{u}_j^2$ for group *j*. (In STATA, you can specify groups by using ***cluster***.)

In practice, we usually do not know the structure of heteroskedasticity. Thus, it is safe to use the robust standard errors (especially when you have a large sample size.) Even if there is no heteroskedasticity, the robust standard errors will become just conventional OLS standard errors. Thus, the robust standard errors are appropriate even under homoskedasticity.

A **heteroskedasticity-robust t statistic** can be obtained by dividing an OSL estimator by its robust standard error (for zero null hypotheses). The usual F-statistic, however, is invalid. Instead, we need to use the heteroskedasticity-robust Wald statistic. Suppose the hypotheses can be written as

$$H_0 : R\beta = r$$

Where R is a *q* x *(k+1)* matrix (q < (k+1)) and ***r*** is a *q* x 1 vector with zeros for this case. Thus,

$$R = \begin{bmatrix} 1\,0 & 0\ 0\ 0 \\ 0\,1 & 0\ 0\ 0 \\ & \\ 0 & 1\ 0\ 0 \end{bmatrix} = \left[I_q : 0_{q\times(k+1-q)} \right],\ r = \begin{bmatrix} 0 \\ 0 \\ \\ 0 \end{bmatrix}.$$

The **heteroskedasticity-robust Wald statistics** for testing the null hypothesis is

$$W = (R\hat{\beta} - r)'(R\hat{V}R')^{-1}(R\hat{\beta} - r)$$

where $\hat{V}$ is given in (7-2). The heteroskedasticity-robust Wald statistics is asymptotically distributed chi-squared with q degree of freedom. The Wald statistics can be turned into an appropriate F-statistics (q, q-k-1) by dividing it by q.

Tests for Heteroskedasticity

When should we use robust standard errors? My personal answer to this question is "almost always." As you will see in Example 7-1, it is very easy to estimate robust standard errors with STATA or other packages. Thus, at least I suggest that you estimate robust standard errors and see if there are any significant differences between conventional standard errors and robust standard errors. If results are robust, i.e., when you do not find any significant differences between two sets of standard errors, then you could be confident in your results based on homoskedasticity.

Statistically, you can use following two heteroskedasticity tests to decide if you have to use robust standard errors or not.

*The **Breusch-Pagan** Test for Heteroskedasticity*

If the homoskedasticity assumption is true, then the variance of error terms should be constant. We can make this assumption as a null hypothesis:

$$H_0\text{: } E(u|\,X) = \Phi^2$$

To test this null hypothesis, we estimate

$$\hat{u}^2 = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + \ldots + \delta_k x_k + e$$

Under the null hypothesis, independent variables should not be jointly significant. The F-statistics that test a joint significance of all independent variables is

$$F_{k,n-k-1} = \frac{R^2 / k}{(1 - R^2)/(n - k - 1)}$$

The LM test statistics is

$$LM = n\,R^2 \sim \Pi^2{}_k$$

*The **White** Test for Heteroskedasticity*

White proposed to add the squares and cross products of all independent variables:

$$\hat{u}^2 = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + ... + \delta_k x_k + \lambda_1 x_1^2 + \lambda_2 x_2^2 + ... + \lambda_k x_k^2 + \phi_1 x_1 x_2 + \phi_2 x_1 x_3 + ... + \phi\ x_{k-1} x_k + v$$

Because $\hat{y}$ includes all independent variables, this test is equivalent of conducting the following test:

$$\hat{u}^2 = \delta_0 + \delta_1 \hat{y} + \delta_2 \hat{y}^2 + v$$

We can use F-test or LM-test on H: $\delta_1 = 0\, and\ \delta_2 = 0$.

Example1: Step-by-Step Estimation for Robust Standard Errors

In the following do-file, I first estimate a wage model:

$$\log Wage = \beta_0 + \beta_1 female + \beta_2 educ + \beta_3 \exp er + \beta_4 \exp ersq + u$$

by using WAGE1.dta. Then, by using residuals from this conventional OLS, I estimate $\hat{\Sigma}$ and obtain robust standard errors by step-by-step with matrix. Finally, I verify what I get with robust standard errors provided by STATA. Of course, you do not need to use matrix to obtain robust standard errors. You just need to use STATA command, "robust," to get robust standard errors (e.g., **`reg y x1 x2 x3 x4, robust`**). But at least you know how robust standard errors are calculated by STATA.

```
. *** on WAGE1.dta
. *** This do-file estimates White-Huber robust standard errors

. set matsize 800
. clear
. use c:\docs\fasid\econometrics\homework\wage1.dta
.
. * Variable construction
. gen logwage=ln(wage)
. gen expsq=exper*exper
. gen x0=1
.
. * Obtain conventional OLS residuals
. reg logwage female educ exper expsq

  Source |       SS       df       MS              Number of obs =     526
---------+------------------------------           F(  4,   521) =   86.69
   Model |  59.2711314     4  14.8177829           Prob > F      =  0.0000
Residual |    89.05862   521   .17093785           R-squared     =  0.3996
---------+------------------------------           Adj R-squared =  0.3950
   Total |  148.329751   525   .28253286           Root MSE      =  .41345

------------------------------------------------------------------------------
 logwage |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  female |  -.3371868   .0363214     -9.283   0.000      -.4085411   -.2658324
    educ |   .0841361   .0069568     12.094   0.000       .0704692    .0978029
```

```
   exper |     .03891   .0048235      8.067   0.000        .029434    .0483859
   expsq |   -.000686   .0001074     -6.389   0.000       -.000897   -.0004751
   _cons |    .390483   .1022096      3.820   0.000       .1896894    .5912767
------------------------------------------------------------------------------
. predict e, residual
. gen esq=e*e
.
. * Create varaibles with squared residuals
. gen efemale=esq*female
. gen eeduc=esq*educ
. gen eexper=esq*exper
. gen eexpsq=esq*expsq
. gen ex0=esq*x0
.
. * Matrix construction
. mkmat logwage, matrix(y)
. mkmat efemale eeduc eexper eexpsq ex0, matrix(ex)
.
. * White-Huber robust standard errors
. matrix ixx=syminv(x'*x)
. matrix sigma=ex'*x
. matrix b_robust=(526/(526-5))*ixx*sigma*ixx

. * Here is the White-Huber robust var(b^)

. mat list b_robust
b_robust[5,5]
            female        educ       exper       expsq          x0
female   .00130927   .00003556  -1.277e-07   1.647e-09  -.00109773
  educ   .00003556   .00005914  -3.028e-06   1.570e-07  -.00077218
 exper  -1.277e-07  -3.028e-06   .00002186  -4.500e-07  -.00010401
 expsq   1.647e-09   1.570e-07  -4.500e-07   1.009e-08   5.805e-07
    x0  -.00109773  -.00077218  -.00010401   5.805e-07   .01179363

. * Take square root of the diagonal elements>> sd.error
. * Thus sd.er. for female is (0.0013)^.5=0.036, educ is
(0.000059)^.5=0.00768, . * and so on.
```

But, you do not need to go through this calculation yourself. STATA has a command called "robust."

```
. * Verify with STATA version of robust standard errors

. reg logwage female educ exper expsq, robust
Regression with robust standard errors                 Number of obs =     526
                                                       F(  4,   521) =   81.97
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.3996
                                                       Root MSE      =  .41345
------------------------------------------------------------------------------
         |               Robust
 logwage |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  female |  -.3371868   .0361838     -9.319   0.000      -.4082709   -.2661026
    educ |   .0841361     .00769     10.941   0.000        .069029    .0992432
   exper |     .03891   .0046752      8.323   0.000       .0297253    .0480946
   expsq |   -.000686   .0001005     -6.829   0.000      -.0008834   -.0004887
   _cons |    .390483   .1085985      3.596   0.000       .1771383    .6038278
------------------------------------------------------------------------------
end of do-file
```

End of Example 1

Example 2: the Breusch-Pagan *test*

```
. *** on WAGE1.dta
. *** This do-file conducts the Breusch-Pagan heteroskedasticity test
. set matsize 800
. clear
. use c:\docs\fasid\econometrics\homework\wage1.dta
. * Variable construction
. gen logwage=ln(wage)
. gen expsq=exper*exper

. * Obtain residuals from the level model
. reg wage female educ exper expsq
(Output is omitted)

. predict u, residual
. gen uu=u*u

. * the BP test
. reg uu female educ exper expsq
  Source |       SS       df       MS                  Number of obs =     526
---------+------------------------------               F(  4,   521) =   12.79
   Model |  24632.1712     4  6158.0428                Prob > F      =  0.0000
Residual |  250826.691   521  481.433189               R-squared     =  0.0894
---------+------------------------------               Adj R-squared =  0.0824
   Total |  275458.863   525  524.683548               Root MSE      =  21.942
------------------------------------------------------------------------------
      uu |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  female |  -5.159055   1.927575     -2.676   0.008      -8.945829   -1.372281
    educ |   1.615551   .3691974      4.376   0.000       .8902524    2.340849
   exper |   1.067343   .2559852      4.170   0.000       .5644535    1.570233
   expsq |  -.0189783   .0056986     -3.330   0.001      -.0301733   -.0077833
   _cons |  -18.15532   5.424263     -3.347   0.001      -28.81144   -7.499208
------------------------------------------------------------------------------
. test female educ exper expsq
 ( 1)  female = 0.0
 ( 2)  educ = 0.0
 ( 3)  exper = 0.0
 ( 4)  expsq = 0.0
       F(  4,   521) =   12.79
            Prob > F =    0.0000
```

The LM test statistics is 526 x 0.0894 = 47.02. This is significant at 1 percent level because the critical level is 13.28 for a chi-square distribution of four degree of freedom.

End of Example 2

Example 7-3: the White test

```
. * Obtain residuals from the log model
. reg logwage female educ exper expsq
(Output omitted)

. predict yhat
(option xb assumed; fitted values)
. predict v, residual
. gen yhatsq=yhat*yhat
. gen vsq=v*v
```

```
. * the White test
. reg vsq yhat yhatsq

  Source |       SS       df       MS                  Number of obs =     526
---------+------------------------------               F(  2,   523) =    3.96
   Model |  .605241058     2  .302620529               Prob > F      =  0.0197
Residual |   40.003265   523  .076488078               R-squared     =  0.0149
---------+------------------------------               Adj R-squared =  0.0111
   Total |   40.608506   525  .077349535               Root MSE      =  .27656
------------------------------------------------------------------------------
     vsq |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
    yhat |   -.187119   .2637874     -0.709   0.478      -.7053321     .331094
  yhatsq |   .0871914   .0812258      1.073   0.284      -.0723774    .2467603
   _cons |   .2334829   .2096599      1.114   0.266      -.1783961     .645362
------------------------------------------------------------------------------
. test yhat yhatsq

 ( 1)  yhat = 0.0
 ( 2)  yhatsq = 0.0

       F(  2,   523) =    3.96
            Prob > F =    0.0197
```

LM stat is 526 x 0.0149 = 7.84, which is significant at 5 percent level but not at 1 percent level.

End of Example 3