2.1 *Breusch-Pagan test (BP test)*

To test the above hypothesis, we follow these steps.

1) We need to find out whether u^2 is correlated with X. This can be done by regressing

population parameter b/c u = real error

$$u^2 = \delta_0 + \delta_1 X_1 + \ldots \delta_K X_K + \text{error}$$

Then perform a joint-hypothesis testing for

$H_0: \delta_1 = \delta_2 = \delta_3 = \ldots = \delta_K = 0$ (homosk.)

against H_a: H_0 is not true. So, if at least one of the $\delta \neq 0$, we already have heterosk.

2) We don't know the true value of u^2. So, we use $\hat{u}^2$ instead.

post-estimation command following "regress"

predict u_hat, residual

new variable name

an option to specify that we want to predict $\hat{u}$.

To perform the Breusch-Pagan Test in STATA BP

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

regress y x_1 x_2 x_3 x_4 ← estimate the main specification first.

predict u_hat, residual ← in order to predict $\hat{u}$.

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. → want to use BP-test for heterosk.

y X_1 X_2 X_3 X_4 X_5

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

Source	SS	df	MS
Model	226.077541	5	45.2155081
Residual	142.319996	674	.211157264
Total	368.397537	679	.542558964

Number of obs =	680
F(5, 674) =	214.13
Prob > F =	0.0000
R-squared =	0.6137
Adj R-squared =	0.6108
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
Model	8.22606613	5	1.64521323
Residual	76.2624962	674	.113149104
Total	84.4885623	679	.124430872

Number of obs =	680
F(5, 674) =	14.54
Prob > F =	0.0000
R-squared =	0.0974
Adj R-squared =	0.0907
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

test H_0 of $\delta_1=\delta_2=\delta_3=\delta_4=\delta_5=0$ ⇒ Test for joint sig.

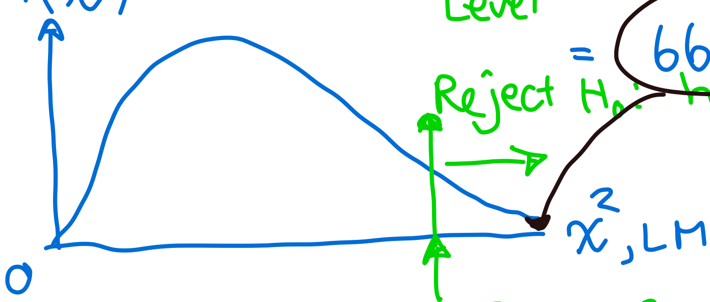
LM-test : $LM = n \cdot R^2_u$

$= 680 \cdot 0.0974$

$= 66.232$

5% significant Level

Reject H_0: homoske



$\chi^2_k = \chi^2_5 = 11.07$

* Since $LM > \chi^2_5$, we reject H_0 in favor of heteroskedasticity.

F-test : $F = \dfrac{R^2_{\hat{u}}/k}{(1-R^2_{\hat{u}^2})/n-k-1} = \dfrac{0.0974/5}{(1-0.0974)/(680-5-1)}$

$= 14.54$

Reject Ho: homost

5% significant level

$F_{k,n-k-1}$, $F_{5,674}$

2.21

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

① estimate the main regression (that we want to test for hetero)

② estat hettest

Y X1 X2 X3 X4 X5

```
. 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
------------------------------------------------------------------------------
```

Note: Heteroske. std. err. tend to be larger than homoske. std. err. So, we are ~~more~~ less likely to reject H₀: β = 0.

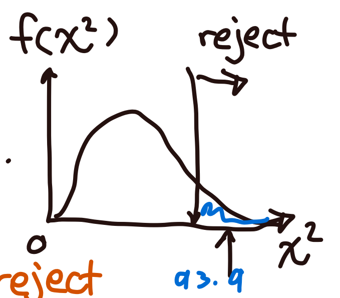
Robust X1 X2 X3 X4 X5

```
. 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
```

→ Homoskedasticity

χ^2 d.f. → the test statistic.

p-value



if p-value < 0.05, reject H₀ at 5% level of significance And conclude that we have the heteroskedasticity problem.

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.

From this example, we don't reject $H_0: \beta_{frosh} = 0$ if we use hetero-robust std. err., but we reject $H_0: \beta_{frosh} = 0$ if we use normal homoske. standard error.

Coef / Robust Std Err.

```
. 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
------------------------------------------------------------------------------
```

So, it is important to use the appropriate calculation for the std. err.! Otherwise ⇒ we can make a wrong conclusion.

This test is more difficult to pass than BP-test.

2.2 The White Test

(BP)

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

main regression: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + u$ (2)

1) like in the BP-test, we first have to obtain $\hat{u}_i$ from regression (2) $\Rightarrow$ generate $\hat{u}_i^2$ from $\hat{u}_i$

2) $\hat{u}^2 = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + ... + \theta_1 X_1 X_2 + \theta_2 X_1 X_3 + ... + \theta_{\#} X_k X_{k-1}$
$+ \gamma_1 X_1^2 + \gamma_2 X_2^2 + ... + \gamma_k X_k^2 + \text{error}$ (3)

3) Test for $H_0: \alpha_1 = \alpha_2 = ... = \theta_1 = \theta_2 = ... = \gamma_1 = \gamma_2 ... = 0$

Homoskedasticity

Ha: Ho is not true (we have heterosk.)

4) We can use LM-test: LM-statistic $= n \cdot R^2_{\hat{u}^2}$

R^2 from eq. (3)

5) LM > critical value $\rightarrow$ reject H_0: homosk.

LM < ———— $\rightarrow$ don't reject Ho $\rightarrow$ implies ~~Hetero~~ Homo skedasticity

$\chi^2_{d.f.} \Rightarrow$ # of slope parameter in (3)

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

1. Get $\hat{\beta}_1, \hat{\beta}_2, ..., \hat{\beta}_k$ by OLS. $\rightarrow$ in the main regression (2)
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 $\hat{y}_i^2$

   $\rightarrow$ representative of x^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. $LM = nR^2$ from this regression
7. If $p - value > significance\ \ level$, cannot reject H_o.

p-value < sig. level $\rightarrow$ reject Ho in favor of heteroskedasticity.

3 Remedial measures

As mentioned before, there are 2 types of remedies – passive and active.

- The passive remedies just re-calculate the *std.err.* or $\hat{\beta}$ using the heteroskedasticity-robust standard error formula(s).
- The active remedies include the "weighted least squares (WLS) estimators", "generalized least squares (GLS) estimators", or "feasible GLS estimator".

3.1 *Weighted Least Squares (WLS)*

We assume that the heteroskedasticity may take the pattern

$$Var(u|\underbrace{X_1, X_2, \ldots X_K}_{\mathbf{X} \leftarrow \text{vector of } X_s}) = \sigma^2 \overbrace{h(X_1, X_2, \ldots X_K)}^{\text{a function which explains the}}$$

heteros. pattern

From

$$\begin{aligned} Var(u_i|\mathbf{x}) &= E(u_i^2|\mathbf{x}) - [\underbrace{E(u_i|\mathbf{x})}_{=0}]^2 \\ &= E(u_i^2|\mathbf{X}) \end{aligned}$$

We get

$$Var(u_i|\mathbf{X}) = E(u_i^2|\mathbf{X}) = \sigma^2 h_i(\mathbf{X})$$

$$\frac{E(u_i^2|\mathbf{X})}{h_i(\mathbf{X})} = \sigma^2$$

$$E\left(\frac{u_i^2}{\sqrt{h_i}^2}\middle|\mathbf{X}\right) = \sigma^2$$

$$E\left(\left(\frac{u_i}{\sqrt{h_i}}\right)^2\middle|\mathbf{X}\right) = \sigma^2$$

Therefore, even though $Var(u_i|\mathbf{X})$ is not constant,

↳ the variance of $\frac{u_i}{\sqrt{h_i}}$ is a constant !!

To make the error term become homoskedastic, we weight every term by $\sqrt{h_i}$.

$$\frac{Y_i}{\sqrt{h_i}} = \frac{\beta_0}{\sqrt{h_i}} + \beta_1\frac{X_{1i}}{\sqrt{h_i}} + \beta_2\frac{X_{2i}}{\sqrt{h_i}} + \ldots + \beta_K\frac{X_{Ki}}{\sqrt{h_i}} + \frac{u_i}{\sqrt{h_i}}$$

$$Y_i^* = \beta_0 X_{0i}^* + \beta_1 X_{1i}^* + \beta_2 X_{2i}^* + \ldots + \beta_K X_{Ki}^* + u_i^* \quad (1)$$

where $Y_i^* = \frac{Y_i}{\sqrt{h_i}}$, $X_{0i}^* = \frac{1}{\sqrt{h_i}}$, $X_{1i}^* = \frac{X_{1i}}{\sqrt{h_i}}$, etc.

estimating (1) would give unbiased $\hat{\beta}_{OLS}$ and would not require heteroskedasticity-robust std. err. because $\underbrace{Var(u_i^* | \mathbf{x}_i^*) = \sigma^2}_{\text{homoskedasticity!}}$

How do we find h_i *or* $\sqrt{h_i}$, the heteroskedasticity function?

1. If the heteroskedasticity is "known" to be caused by a multiplicative constant, we can adjust using that constant.

suppose $sav_i = \beta_0 + \beta_1 inc_i + u_i$

- We expect the heteroskedasticity pattern to vary with income (the more income the person earns the more variations in their savings.) $\rightarrow$ $Var(u|x) = \sigma^2 \underbrace{h(x)}$

$Var(u_i | income_i) = \sigma^2 income_i$

So, we transform the raw data an estimate $\frac{sav_i}{\sqrt{inc_i}} = \beta_0 \frac{1}{\sqrt{inc_i}} + \beta_1 \frac{inc_i}{\sqrt{inc_i}} + \frac{u_i}{\sqrt{inc_i}}$

$Var\left(\frac{u_i}{\sqrt{inc_i}} \middle| x\right) = E\left(\left(\frac{u_i}{\sqrt{inc_i}}\right)^2 \middle| inc_i\right) = \sigma^2$

2. If the heteroskedasticity pattern is not known, we can estimate it. This procedure is called "Feasible Generalized Least Squares" (also called Feasible GLS or FGLS)

3.2 Feasible GLS (not covered in this class.)

Wooldridge Ch. 8 part 4b

Since $Var(\hat{\beta}_j)$ would not be unbiased, we can make valid inferences about β_j, e.g. can use t-test, F-test, etc.