EE325 Ch.7 Heteroscedasticity

Read Guajarati Ch. 11

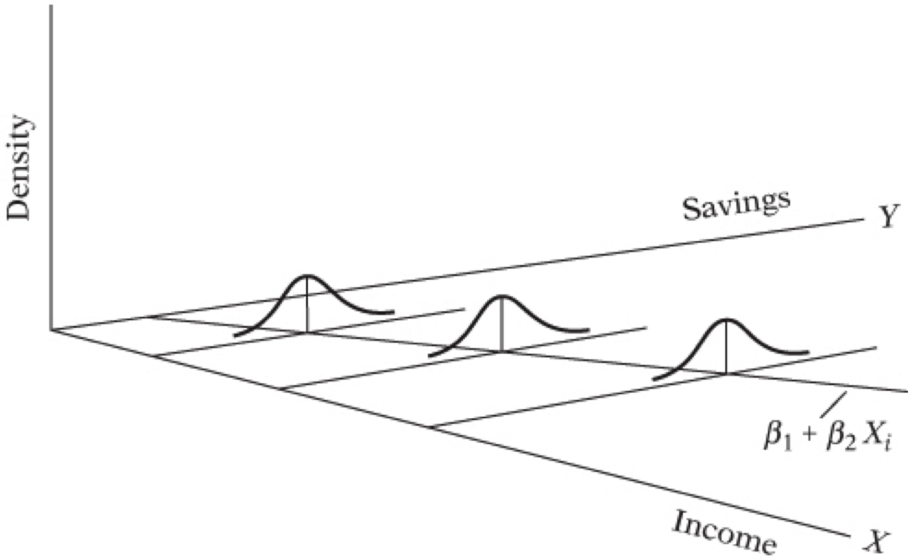
Outline

Nature of Heteroscedasticity

One of the important assumptions of CLRM is that the variance of each disturbance u_i term, conditional on the chosen values of the explanatory variables, is some constant number being equal to σ^2 (Homoscedasticity)

$$E(u_i^2) = \sigma^2$$

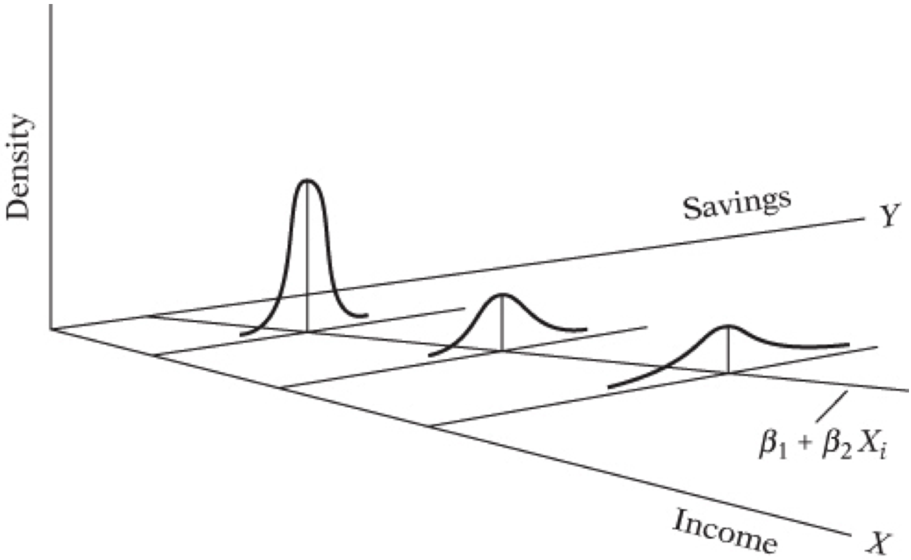
where i = 1, 2, ..., n


Density
Savings
Y
$\beta_1 + \beta_2 X_i$
Income
X

The conditional variance of Y_i increases as X increases. The variances of Y_i are then not the same. Here, there is heteroscedasticity.

$$E(u_i^2) = \sigma_i^2$$

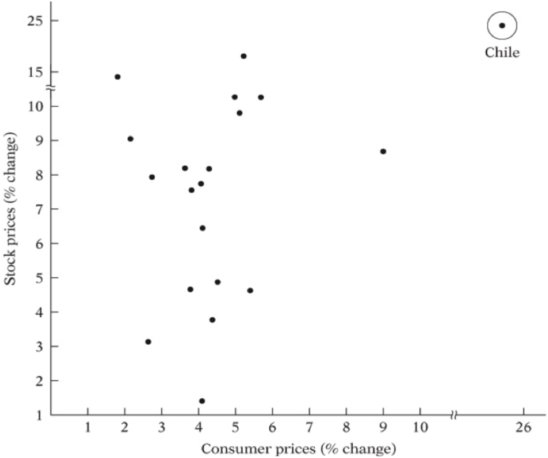
where i = 1, 2, . . ., n


Density
Savings
Y
$\beta_1 + \beta_2 X_i$
Income
X

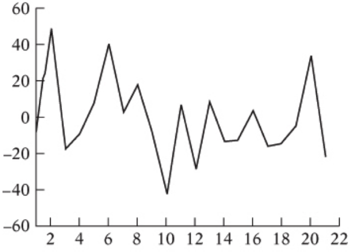
- Error learning
- As incomes grow, people have more discretionary income and more scope for choice about the disposition of their income. Hence, σ_i^2 is likely to increase with income
- As data collecting techniques improve, σ_i^2 is likely to decrease

- Error learning
- As incomes grow, people have more discretionary income and more scope for choice about the disposition of their income. Hence, σ_i^2 is likely to increase with income
- As data collecting techniques improve, σ_i^2 is likely to decrease

- Error learning
- As incomes grow, people have more discretionary income and more scope for choice about the disposition of their income. Hence, σ_i^2 is likely to increase with income
- As data collecting techniques improve, σ_i^2 is likely to decrease



- Heteroscedasticity can also arise as a result of the presence of outliers



(a) (b)

- Specification error – some important variables are omitted from the model.

- Incorrect data transformation
- Incorrect functional form

- Incorrect data transformation
- Incorrect functional form

- Prevail in cross-sectional data than in times series one because
  - Cross-sectional data compose of various samples of different entities
  - Time series cover same entity over time

Consequence of Heteroscedasticity

Normal OLS estimation with Homoscedasticity:

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

$$\hat{\beta}_2 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \frac{\sum x_i y_i}{\sum x_i^2}$$

$$var(\hat{\beta}_2) = \frac{\sum x_i^2 \sigma_i^2}{(\sum x_i^2)^2}$$

Since σ_i^2 is the same for all i : $\sigma_i^2 = \sigma_j^2, \forall i \neq j$

$$var(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_i^2}$$

$\hat{\beta}_2$ is best linear unbiased estimator (BLUE)

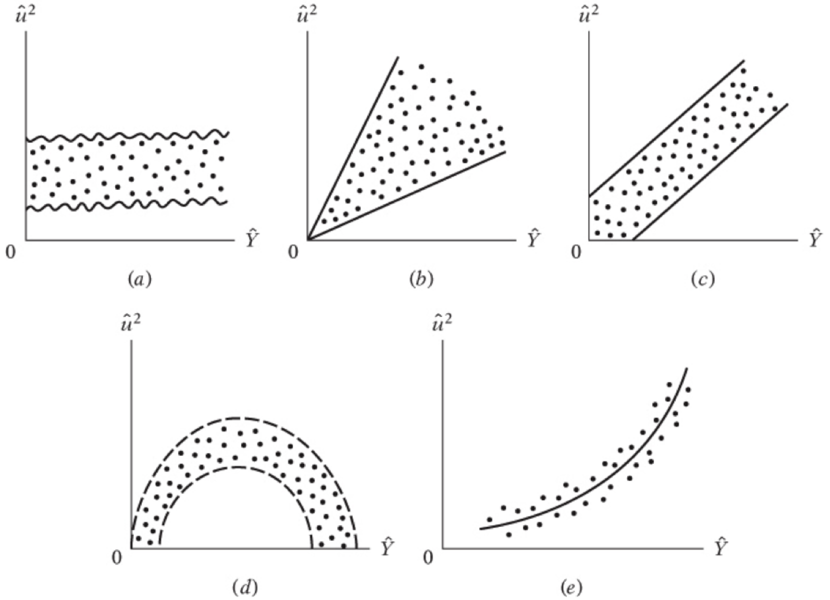
Normal OLS estimation with Heteroscedasticity:

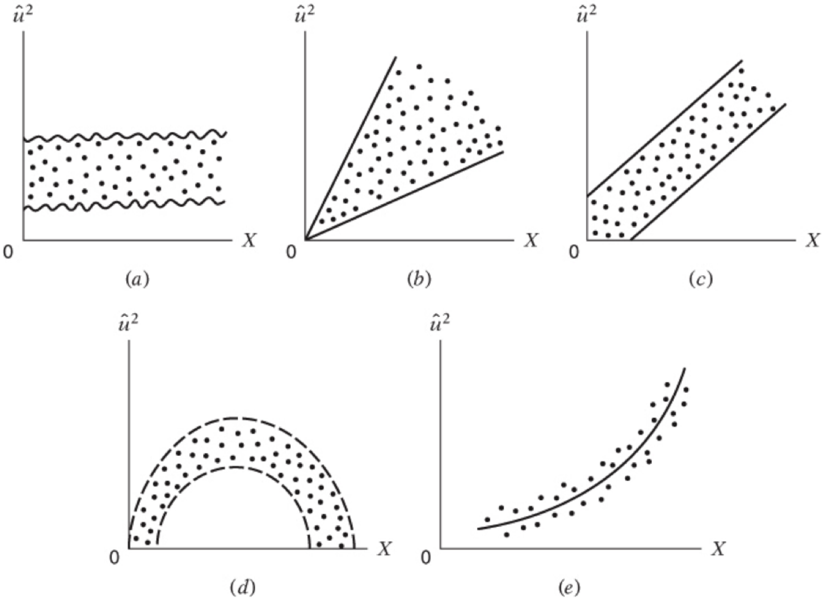
$$var(\hat{\beta}_2) = \frac{\sum x_i^2 \sigma_i^2}{(\sum x_i^2)^2}$$

$\hat{\beta}_2$ is no longer best linear unbiased estimator (BLUE)

Detection of Heteroscedasticity

- Informal method
  - graphical method
- Formal methods
  - Park test
  - Breusch-Pagan Test
  - White's General Heteroscedasticity Test


$\hat{u}^2$
$\hat{Y}$
0
(a)
(b)
(c)
(d)
(e)


$\hat{u}^2$
X
0
(a)
$\hat{u}^2$
X
0
(b)
$\hat{u}^2$
X
0
(c)
$\hat{u}^2$
X
0
(d)
$\hat{u}^2$
X
0
(e)

Park formalizes the graphical method by suggesting that σ_i^2 is a function of the explanatory variable (X_i). The functional form is:

$$\sigma_i^2 = \sigma^2 X_i^{\beta} e^{\nu_i}$$

or

$$\ln \sigma_i^2 = \ln \sigma^2 + \beta \ln X_i + \nu_i$$

where ν_i is the stochastic disturbance term.

Since σ_i^2 is generally not known. Park advises using $\hat{u}_i^2$ as a proxy and running the following regression:

$$\ln \hat{u}_i^2 = \ln \sigma^2 + \beta \ln X_i + \nu_i = \alpha + \beta \ln X_i + \nu_i$$

If β turns out to be statistically significant, it would suggest that heteroscedasticity is present in the data.

Example:
Table 11 Relationship between compensation and productivity

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

where
Y_i = average compensation in thousands of dollars
X_i = average productivity in thousands of dollars
i = i^{th} employment size of the establishment

Step 1:
Run the OLS regression disregarding the heteroscedasticity question

$$\begin{array}{rcccc} \hat{Y}_i & = & 1992.3452 & + & 0.2329X_i \\ se & & (936.4791) & & (0.0998) \\ t & & (2.1275) & & (2.333) \\ R^2 & = & 0.4375 & & \end{array}$$

and then obtain $\hat{u}_i^2$ from this equation

Step 2:
Once we obtain $\hat{u}_i^2$, we run the regression

$$\ln \hat{u}_i^2 = \ln \sigma^2 + \beta \ln X_i + \nu_i$$
$$\ln \hat{u}_i^2 = \alpha + \beta \ln X_i + \nu_i$$

$$\begin{aligned} \widehat{\ln \hat{u}_i^2} &= 35.817 && - && 2.8099 \ln X_i \\ se &= (38.319) && && (4.216) \\ t &= (0.934) && && (-0.667) \\ R^2 &= 0.0595 \end{aligned}$$

Step 3:
Examine the significance of β

Consider the k-variables linear regression model

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + u_i$$

We assume that

$$E(u|X_1, X_2, \ldots, X_k) = 0$$

So OLS estimators are unbiased and consistent

Step 1:
Estimate the equation, using OLS

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + u_i$$

Obtain the squared OLS residuals, $\hat{u}_i^2$

Step 2:
Run the regression

$$\hat{u}_i^2 = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + \cdots + \alpha_k X_{ki} + u_i$$

Keep the R-squared, $R^2_{\hat{u}_i^2}$, from this regression

Step 3:
Form either the F statistics:

$$F = \frac{R^2_{\hat{u}_i^2}/k}{(1 - R^2_{\hat{u}_i^2})/(n-k-1)}$$

or the LM statistics for heteroscedasticity:

$$LM = nR^2_{\hat{u}_i^2}$$

under the null hypothesis, LM is distributed asymptotically as χ^2_k

The Breusch-Pagan test is an asymptotic, or large sample, test and in the present example 30 observations may not constitute a large sample. It should also be pointed out that in small samples, the test is sensitive to the assumption that the disturbances, u_i, are normally distributed.

Consider the following three-variable regression model

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$$

The White's test proceeds as following:

Step 1:
Given the data, we estimate the equation above and obtain the residuals, $\hat{u}_i^2$

Step 2:
We then run the following (auxiliary) regression:

$$\hat{u}_i^2 = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + \alpha_4 X_{2i}^2 + \alpha_5 X_{3i}^2 + \alpha_6 X_{2i} X_{3i} + \nu_i$$

Get the R-squared from this (auxiliary) regression

Step 3:
Under the null hypothesis that there is no heteroscedasticity, it can be shown that sample size (n) times the R-squared obtained from the auxiliary regression asymptotically follows the chi-square distribution with df equal to the number of regressors (excluding the constant term) in the auxiliary regression. That is:

$$nR^2 \sim \chi^2_{df}$$

Step 4:
If the chi-square value obtained in step 3 exceeds the critical chi-square value at the chosen level of significance, the conclusion is that there is heteroscedasticity. If not, there is no heteroscedasticity, which is to say that in the auxiliary regression

$$\hat{u}_i^2 = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + \alpha_4 X_{2i}^2 + \alpha_5 X_{3i}^2 + \alpha_6 X_{2i} X_{3i} + \nu_i$$

$$\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = \alpha_6 = 0$$

Example from page 387-388

From cross-sectional data on 41 countries,

$$\ln Y_i = \beta_1 + \beta_2 \ln X_{2i} + \beta_3 \ln X_{3i} + u_i$$

where
Y_i = Ratio of trade taxes to total government revenue
X_2 = Ratio of the sum of exports and imports to GNP
X_3 = GNP per capita

By applying White's heteroscedasticity test to the residuals obtained from regression, the following results were obtained:

$$\begin{aligned} \widehat{\hat{u}_i^2} &= -5.8417 + 2.5629 \ln X_{2i} + 0.6918 \ln X_{3i} - 0.4081(\ln X_{2i})^2 \\ &\quad -0.0491(\ln X_{3i})^2 + 0.0015 \ln X_{2i} \ln X_{3i} \end{aligned}$$

$$R^2 = 0.1148$$

$$nR^2 = 41(0.1148) = 4.7068$$

The 5 percent critical chi-square value for 5 df is 11.0705
Since the calculated χ^2 is less than the critical value, there is no heteroscedasticity.

Remedial Measures

1. When σ_i^2 is known: The method of weighted least squares
2. When σ_i^2 is unknown

If σ_i^2 is known, the most straightforward method of correcting heteroscedasticity is by means of weighted least squares, for the estimators thus obtained are BLUE

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$
$$Y_i = \beta_1 X_{0i} + \beta_2 X_i + u_i$$

where $X_{0i} = 1$ for each i

Now assume that the heteroscedastic variance, σ_i^2, are known and we divide the equation above with such variance for each i:

$$\frac{Y_i}{\sigma_i} = \beta_1 \frac{X_{0i}}{\sigma_i} + \beta_2 \frac{X_i}{\sigma_i} + \frac{u_i}{\sigma_i}$$

$$Y_i^* = \beta_1 X_{0i}^* + \beta_2 X_i^* + u_i^*$$

$$\begin{aligned} var(u_i^*) &= E(u_i^*) = E[(\tfrac{u_i}{\sigma_i})^2] \\ &= \tfrac{1}{\sigma_i^2} E(u_i^2); && \text{since } \sigma_i^2 \text{ is known} \\ &= \tfrac{1}{\sigma_i^2} \sigma_i^2; && \text{since } E(u_i^2) = \sigma_i^2 \\ &= 1 \end{aligned}$$

Since we are still retaining the other assumptions of the classical model, the finding that u_i^* is homoscedastic suggests that if we apply OLS to the transformed model:

$$\frac{Y_i}{\sigma_i} = \beta_1 \frac{X_{0i}}{\sigma_i} + \beta_2 \frac{X_i}{\sigma_i} + \frac{u_i}{\sigma_i}$$

If will produce estimators that are BLUE. In short, the estimated β_1^* and β_2^* are now BLUE and not the OLS estimators $\hat{\beta}_1$ and $\hat{\beta}_2$.

- GLS is OLS on the transformed variables that satisfy the standard least-squared assumptions.
- The estimators thus obtained are known as GLS estimators, and these estimators are BLUE

$$\frac{Y_i}{\sigma_i} = \hat{\beta}_1^* \frac{X_{0i}}{\sigma_i} + \hat{\beta}_2^* \frac{X_i}{\sigma_i} + \frac{u_i}{\sigma_i}$$

$$Y_i^* = \hat{\beta}_1^* X_{0i}^* + \hat{\beta}_2^* X_i^* + u_i^*$$

$$\sum \hat{u^*}_i^2 = \sum (Y_i^* - \hat{\beta}_1^* X_{0i}^* - \hat{\beta}_2^* X_i^*)^2$$

$$\sum \frac{\hat{u}_i}{\sigma_i}^2 = \sum (\frac{Y_i}{\sigma_i} - \hat{\beta}_1^* \frac{X_{0i}}{\sigma_i} - \hat{\beta}_2^* \frac{X_i}{\sigma_i})^2$$

The GLS estimator of β_2^* is

$$\hat{\beta}_2^* = \frac{(\sum w_i)(\sum w_i X_i Y_i) - (\sum w_i X_i)(\sum w_i Y_i)}{(\sum w_i)(\sum w_i X_i^2) - (\sum w_i X_i)^2}$$

$$var(\hat{\beta}_2^*) = \frac{\sum w_i}{(\sum w_i)(\sum w_i X_i^2) - (\sum w_i X_i)^2}$$

where $w_i = \frac{1}{\sigma_i^2}$

Difference between GLS and OLS is:

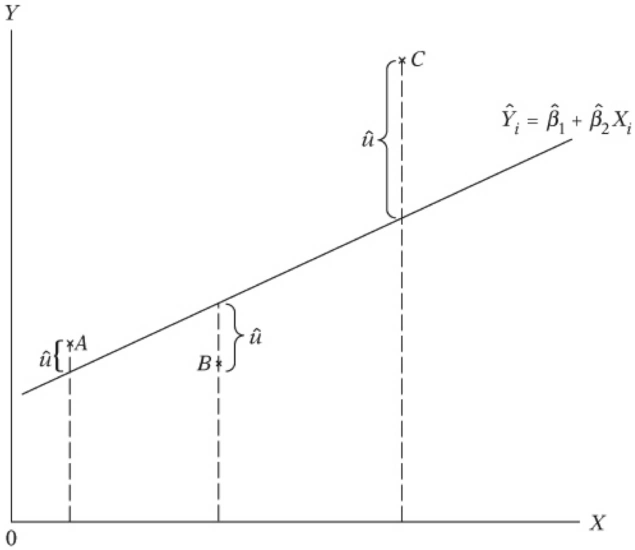
OLS minimizes

$$\sum \hat{u}_i^2 = \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)^2$$

GLS minimizes

$$\sum w_i \hat{u}_i^2 = \sum w_i (Y_i - \hat{\beta}_1^* - \hat{\beta}_2^* X_i)^2$$

where $w_i = \frac{1}{\sigma_i^2}$


Y
C
$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$
$\hat{u}$
$\hat{u}$
A
$\hat{u}$
B
0
X

Example:

TABLE 11.4
Illustration of Weighted Least-Squares Regression

Source: Data on Y and σ_i (standard deviation of compensation) are from Table 11.1. Employment size: 1 = 1–4 employees, 2 = 5–9 employees, etc. The latter data are also from Table 11.1.

Compensation, Y	Employment Size, X	σ_i	Y_i/σ_i	X_i/σ_i
3,396	1	742.2	4.5664	0.0013
3,787	2	851.4	4.4480	0.0023
4,013	3	727.8	5.5139	0.0041
4,104	4	805.06	5.0978	0.0050
4,146	5	929.9	4.4585	0.0054
4,241	6	1,080.6	3.9247	0.0055
4,387	7	1,241.2	3.5288	0.0056
4,538	8	1,307.7	3.4702	0.0061
4,843	9	1,110.7	4.3532	0.0081

Note: In regression (11.6.2), the dependent variable is (Y_i/σ_i) and the independent variables are $(1/\sigma_i)$ and (X_i/σ_i).

Source	SS	df	MS
Model	1327891.27	1	1327891.27
Residual	87312.7333	7	12473.2476
Total	1415204	8	176900.5

Number of obs	= 9
F(1, 7)	= 106.46
Prob > F	= 0.0000
R-squared	= 0.9383
Adj R-squared	= 0.9295
Root MSE	= 111.68

Y	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
X	148.7667	14.4183	10.32	0.000	114.6728	182.8605
_cons	3417.833	81.13632	42.12	0.000	3225.976	3609.69

$$\hat{Y}_i = 3,417.833 + 148.767X_i$$

Source	SS	df	MS
Model	175.811214	2	87.905607
Residual	.128115078	7	.018302154
Total	175.939329	9	19.5488143

Number of obs	=	9
F(2, 7)	=	4803.02
Prob > F	=	0.0000
R-squared	=	0.9993
Adj R-squared	=	0.9991
Root MSE	=	.13529

Ysigma	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
Xsigma	154.2118	16.95407	9.10	0.000	114.1218	194.3018
consigma	3406.277	80.96623	42.07	0.000	3214.822	3597.731

$$\widehat{\frac{\hat{Y}_i}{\sigma_i}} = 3,406.277\frac{1}{\sigma_i} + 154.212\frac{X_i}{\sigma_i}$$

- White's heteroscedaticity-consistent standard errors
- Several assumptions about the pattern of heteroscedaticity are required.

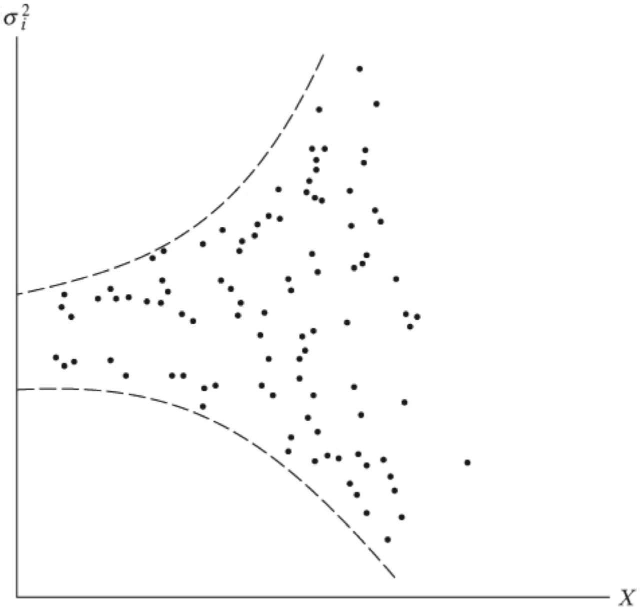
- White's heteroscedaticity-consistent standard errors
- Several assumptions about the pattern of heteroscedaticity are required.

Assumption 1: the error variance is proportional to X_i^2

$$E(u_i^2) = \sigma^2 X_i^2$$

$$\begin{aligned} Y_i &= \beta_1 + \beta_2 X_i + u_i \\ \frac{Y_i}{X_i} &= \frac{\beta_1}{X_i} + \beta_2 + \frac{u_i}{X_i} \\ &= \beta_1 \frac{1}{X_i} + \beta_2 + \nu_i \end{aligned}$$

$$\begin{aligned} E(\nu_i^2) &= E(\frac{u_i}{X_i})^2 = \frac{1}{X_i^2} E(u_i^2) \\ &= \sigma^2 \end{aligned}$$

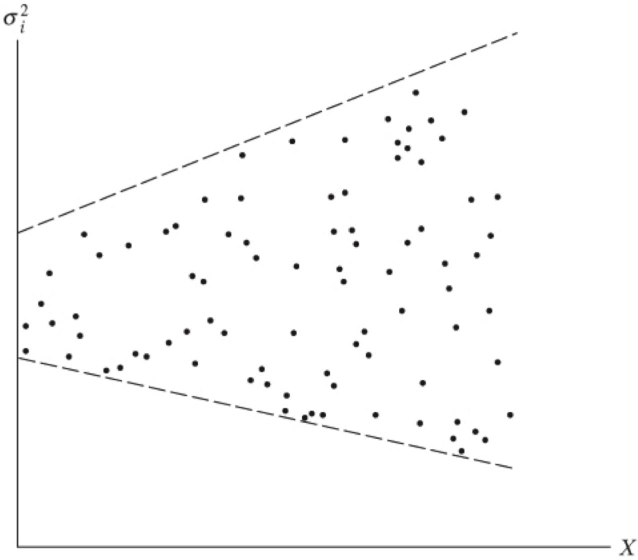

σ_i^2
X

Assumption 2: the error variance is proportional to the square root transformation of X_i

$$E(u_i^2) = \sigma^2 X_i$$

$$\begin{aligned} Y_i &= \beta_1 + \beta_2 X_i + u_i \\ \frac{Y_i}{\sqrt{X_i}} &= \frac{\beta_1}{\sqrt{X_i}} + \beta_2\sqrt{X_i} + \frac{u_i}{\sqrt{X_i}} \\ &= \beta_1 \frac{1}{\sqrt{X_i}} + \beta_2\sqrt{X_i} + \nu_i \end{aligned}$$

$$\begin{aligned} E(\nu_i^2) &= E(\frac{u_i}{\sqrt{X_i}})^2 = \frac{1}{X_i} E(u_i^2) \\ &= \sigma^2 \end{aligned}$$


σ_i^2
X

Assumption 3: the error variance is proportional to the square of the mean of Y

$$E(u_i^2) = \sigma^2 [E(Y_i)]^2$$

$$\begin{aligned} Y_i &= \beta_1 + \beta_2 X_i + u_i \\ E(Y_i) &= \beta_1 + \beta_2 X_i \\ \frac{Y_i}{E(Y_i)} &= \frac{\beta_1}{E(Y_i)} + \beta_2 \frac{X_i}{E(Y_i)} + \frac{u_i}{E(Y_i)} \\ &= \beta_1 \frac{1}{E(Y_i)} + \beta_2 \frac{X_i}{E(Y_i)} + \nu_i \end{aligned}$$

Since $E(Y_i)$ is unknown, we then use $\hat{Y}$ instead.

$$\begin{aligned} E(\nu_i^2) &= E(\frac{u_i}{\hat{Y}_i})^2 = \frac{1}{\hat{Y_i}^2} E(u_i^2) \\ &= \sigma^2 \end{aligned}$$

Assumption 4: a log transformation such as

$$\ln Y_i = \beta_1 + \beta_2 lnX_i + u_i$$

very often reduces heteroscedasticity when compare with the regression

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

Example

TABLE 11.5
Sales and Employment for Companies Performing Industrial R&D in the United States, by Industry, 2005 (values are in millions of dollars)

Source: National Science Foundation, Division of Science Resources Statistics, Survey of Industrial Research and Development: 2005 and the U.S. Census Bureau Annual Survey of Manufacturers, 2005.

	Industry	Sales	R&D	Profits
1	Food	374,342	2,716	234,662
2	Textiles, apparel, and leather	51,639	816	53,510
3	Basic chemicals	109,899	2,277	75,168
4	Resin, synthetic rubber, fibers, and filament	132,934	2,294	34,645
5	Pharmaceuticals and medicines	273,377	34,839	127,639
6	Plastics and rubber products	90,176	1,760	96,162
7	Fabricated metal products	174,165	1,375	155,801
8	Machinery	230,941	8,531	143,472
9	Computers and peripheral equipment	91,010	4,955	34,004
10	Semiconductor and other electronic components	176,054	18,724	81,317
11	Navigational, measuring, electromedical, and control instruments	118,648	15,204	73,258
12	Electrical equipment, appliances, and components	101,398	2,424	54,742
13	Aerospace products and parts	227,271	15,005	72,090
14	Medical equipment and supplies	56,661	4,374	52,443

R&D expenditure, sales and profits in 14 industry groupings in the US are presented in 2005 (all figures in millions of dollars) and since the cross-sectional data presented in this table are quite heterogeneous, in regression of R&D on sales, heteroscedasticity is likely

Source	SS	df	MS
Model	208733442	1	208733442
Residual	1.0083e+09	12	84021567.1
Total	1.2170e+09	13	93614788.2

Number of obs	=	14
F(1, 12)	=	2.48
Prob > F	=	0.1410
R-squared	=	0.1715
Adj R-squared	=	0.1025
Root MSE	=	9166.3

rd	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
sales	.0437234	.0277404	1.58	0.141	-.0167178	.1041646
_cons	1337.874	5015.141	0.27	0.794	-9589.18	12264.93

$$\begin{array}{cccccc} \widehat{R\&D}_i & = & 1337.874 & + & 0.0437 Sale_i \\ se & & (5015) & & (0.0277) \\ t & & (0.27) & & (1.58) \\ R^2 & = & 0.1715 & & \end{array}$$

There is a positive relationship between R&D and sales, although it is not statistically significant at the traditional levels

Source	SS	df	MS
Model	9.2405e+16	2	4.6203e+16
Residual	1.2022e+17	11	1.0929e+16
Total	2.1263e+17	13	1.6356e+16

Number of obs =	14
F(2, 11) =	4.23
Prob > F =	0.0435
R-squared =	0.4346
Adj R-squared =	0.3318
Root MSE =	1.0e+08

muhat2	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
sales	577.6563	1307.934	0.44	0.667	-2301.087	3456.4
sales2	.0008456	.0031711	0.27	0.795	-.006134	.0078253
_cons	-4.67e+07	1.12e+08	-0.42	0.685	-2.94e+08	2.00e+08

$$\begin{array}{lcccccc}
\widehat{u_i^2} & = & -46,746,325 & + & 578 Sale_i & + & 0.000846 Sale_i^2 \\
se & & (112,224,348) & & (1,308) & & (0.003171) \\
t & & (-0.42) & & (0.44) & & (0.27) \\
R^2 & = & 0.435 & & & &
\end{array}$$

Using R^2 value and n=14, we get $nR^2 = 6.090$. Under the null hypothesis of no heteroscedasticity, this should follow χ^2_2 (two regressors): 0.0476.The White test then suggests there is heteroscedasticity.

For remedial measures,

- the true error variance is unknown, we cannot use the method of weighted least squares to obtain heteroscedasticity-corrected standard errors and t-values.
- Therefore, we would have to make some educated guesses about the nature of error variance

```
Linear regression                                      Number of obs =      14
                                                       F(  1,    12) =    1.13
                                                       Prob > F      =  0.3083
                                                       R-squared     =  0.1715
                                                       Root MSE      =  9166.3
```

rd	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
sales	.0437234	.0410918	1.06	0.308	-.0458079	.1332547
_cons	1337.874	4892.447	0.27	0.789	-9321.852	11997.6

$$
\begin{array}{cccccc}
\widehat{R\&D}_i & = & 1337.874 & + & 0.0437 Sale_i \\
se & & (4892.447) & & (0.0411) \\
t & & (0.27) & & (1.06) \\
R^2 & = & 0.172 & &
\end{array}
$$

We see that the parameter estimates have not changed, the standard error of the intercept coefficient has decreased slightly, and the standard error of the slope coefficient has increased slightly. But remember that White procedure is strictly a large-sample one, where we have only 14 observations for this case.

Source	SS	df	MS
Model	208733442	1	208733442
Residual	1.0083e+09	12	84021567.1
Total	1.2170e+09	13	93614788.2

Number of obs	=	14
F(1, 12)	=	2.48
Prob > F	=	0.1410
R-squared	=	0.1715
Adj R-squared	=	0.1025
Root MSE	=	9166.3

RD	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
Sales	.0437234	.0277404	1.58	0.141	-.0167178	.1041646
_cons	1337.874	5015.141	0.27	0.794	-9589.18	12264.93

```
. whitetst

White's general test statistic :    6.0842  Chi-sq( 2)  P-value =  .0477
```

H_o : *Homoscedasticity*
H_a : *Otherwise*

White's general test statistics is 6.0842.

With degree of freedom = 2, then critical value of χ^2_2 at 5 percent significant level is 5.9915.

The calculated $\chi^2 > \chi^2_2$. We reject null hypothesis: the White's test suggests that there is heteroscedasticity

```
      Source |       SS       df       MS              Number of obs =      14
-------------+------------------------------           F(  1,    12) =    2.48
       Model |  208733442      1   208733442           Prob > F      =  0.1410
    Residual | 1.0083e+09     12  84021567.1           R-squared     =  0.1715
-------------+------------------------------           Adj R-squared =  0.1025
       Total | 1.2170e+09     13  93614788.2           Root MSE      =  9166.3

------------------------------------------------------------------------------
          RD |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       Sales |   .0437234   .0277404     1.58   0.141    -.0167178    .1041646
       _cons |   1337.874   5015.141     0.27   0.794     -9589.18    12264.93
------------------------------------------------------------------------------

. estat hettest

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

         chi2(1)      =     8.83
         Prob > chi2  =   0.0030
```

H_o : *Homoscedasticity*
H_a : *Otherwise*

Breusch-Pagan's general test statistics is 8.83.

With degree of freedom = 1, then critical value of χ_2^2 at 5 percent significant level is 3.8414.

The calculated $\chi^2 > \chi_2^2$. We reject null hypothesis: the Breusch-Pagan's test suggests that there is heteroscedasticity

Gujarati, D.N. (2009) Basic Econometrics. 5th ed. Singapore, McGraw-Hill.