

EE 325

Heteroscedasticity

EE325 2/2012 (Ajarn Kaewkwan Tangtipongkul)

The nature of Heteroscedasticity

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

$$E(u_i^2) = \sigma^2 \quad i = 1, \dots, n$$

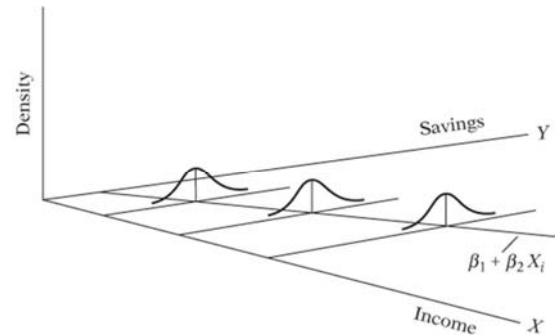
EE325 2/2012 (Ajarn Kaewkwan Tangtipongkul)

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

$$E(u_i^2) = \sigma_i^2 \quad i = 1, \dots, n$$

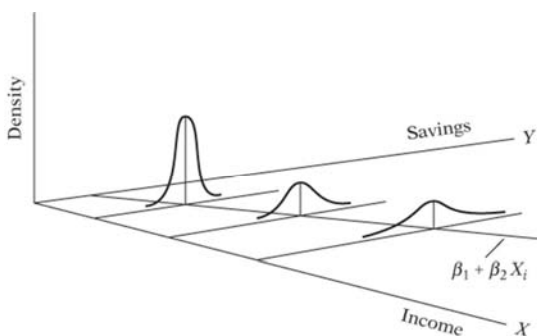
EE325 2/2012 (Ajarn Kaewkwan Tangtipongkul)

Homoscedastic disturbances



EE325 2/2012 (Ajarn Kaewkwan Tangtipongkul)

Heteroscedastic disturbances

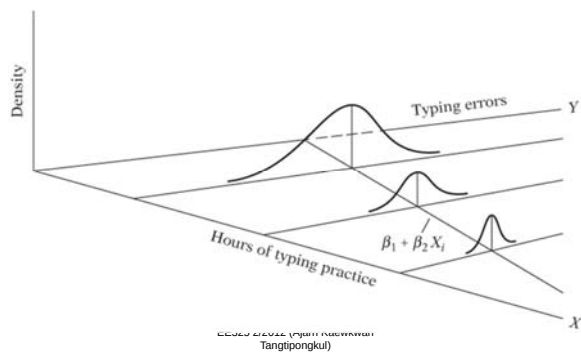


EE325 2/2012 (Ajarn Kaewkwan Tangtipongkul)

Several reasons why the variances of u_i may be variable

- Following the error-learning models, as people learn, their errors of behavior become smaller over time or the number of errors become more consistent

EE325 2/2012 (Ajarn Kaewkwan Tangtipongkul)



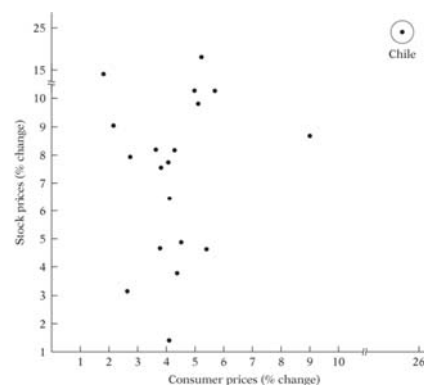
EE325 2/2012 (Ajarn Kaewkwan Tangtipongkul)

- As incomes grow, people have more discretionary income and hence 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

EE325 2/2012 (Ajarn Kaewkwan Tangtipongkul)

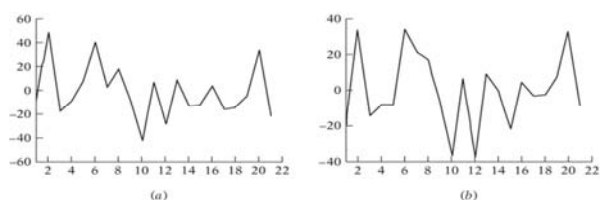
- Heteroscedasticity can also arise as a result of the presence of **outliers**
 - Percent rate of change of stock prices(Y) and consumer prices(X) for the post-World War II period through 1969 for 20 countries

EE325 2/2012 (Ajarn Kaewkwan Tangtipongkul)



Specification error

- Some important variables are omitted from the model



EE325 2/2012 (Ajarn Kaewkwan Tangtipongkul)

- Incorrect data transformation
 - E.g. Ratio or first difference transformations
- Incorrect functional form
 - E.g. Linear versus log-linear models

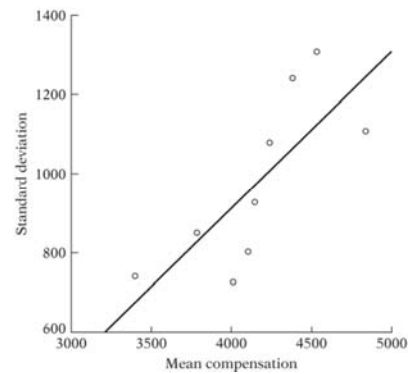
EE325 2/2012 (Ajarn Kaewkwan Tangtipongkul)

Example

TABLE 11.1 Compensation per Employee (\$) in Nondurable Manufacturing Industries According to Employment Size of Establishment, 1958

Industry	Employment Size (average number of employees)								
	1-4	5-9	10-19	20-49	50-99	100-249	250-499	500-999	1,000-2,499
Food and kindred products	2,994	3,295	3,565	3,907	4,189	4,486	4,676	4,968	5,342
Tobacco products	1,721	2,057	3,336	3,320	2,980	2,848	3,072	2,969	3,822
Textile mill products	3,600	3,657	3,674	3,437	3,340	3,334	3,225	3,163	3,168
Apparel and related products	3,494	3,787	3,533	3,215	3,030	2,834	2,750	2,967	3,453
Paper and allied products	3,498	3,847	3,913	4,135	4,445	4,885	5,132	5,342	5,326
Printing and publishing	3,611	4,206	4,695	5,083	5,301	5,269	5,182	5,395	5,552
Chemicals and allied products	3,875	4,660	4,930	5,005	5,114	5,248	5,630	5,870	5,876
Petroleum and coal products	4,616	5,181	5,317	5,337	5,421	5,710	6,316	6,455	6,347
Rubber and plastic products	3,538	3,984	4,014	4,287	4,221	4,539	4,721	4,905	5,481
Leather and leather products	3,016	3,196	3,149	3,317	3,414	3,254	3,177	3,346	4,067
Average compensation	3,396	3,787	4,013	4,104	4,146	4,241	4,388	4,538	4,843
Standard deviation	742.2	851.4	727.8	805.06	929.9	1,080.6	1,241.2	1,307.7	1,110.7
Average productivity	9,355	8,584	7,962	8,275	8,389	9,418	9,795	10,281	11,750

Source: The Census of Manufactures, U.S. Department of Commerce, 1958 (compiled by author).



EE325 2/2012 (Ajarn Kaewkwan Tangtipongkul)

OLS Estimation in the Presence of Heteroscedasticity

EE325 2/2012 (Ajarn Kaewkwan Tangtipongkul)

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

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

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

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

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

EE325 2/2012 (Ajarn Kaewkwan Tangtipongkul)

Is $\hat{\beta}_2$ still BLUE when we drop only homoscedasticity assumption and replace it with the assumption of heteroscedasticity?

$\hat{\beta}_2$ is **no longer best and the minimum variance** given by

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

EE325 2/2012 (Ajarn Kaewkwan Tangtipongkul)

The Method of Generalized Least Squares (GLS)

EE325 2/2012 (Ajarn Kaewkwan Tangtipongkul)

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

$$Y_i = \beta_1 X_{0i} + \beta_2 X_i + u_i \quad \text{where } X_{0i} = 1 \text{ for each } i$$

Now assume that the heteroscedastic variance σ_i^2 are known

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

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

EE325 2/2012 (Ajarn Kaewkwan Tangtipongkul)

$$\begin{aligned} \text{var}(u_i^*) &= E(u_i^*)^2 = E\left(\frac{u_i}{\sigma_i}\right)^2 \quad \text{Since } E(u_i^*) = 0 \\ &= \frac{1}{\sigma_i^2} E(u_i^2) \quad \text{Since } \sigma_i^2 \text{ is known} \\ &= \frac{1}{\sigma_i^2} (\sigma_i^2) \quad \text{Since } E(u_i^2) = \sigma_i^2 \\ &= 1 \end{aligned}$$

EE325 2/2012 (Ajarn Kaewkwan Tangtipongkul)

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

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

it 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 $\hat{\beta}_2$

EE325 2/2012 (Ajarn Kaewkwan Tangtipongkul)

GLS is OLS on the transformed variables that satisfy the standard least-squares assumptions

The estimators thus obtained are known as GLS estimators, and it is these estimators that are BLUE

EE325 2/2012 (Ajarn Kaewkwan Tangtipongkul)

$$\frac{Y_i}{\sigma_i} = \hat{\beta}_1^* \left(\frac{X_{0i}}{\sigma_i} \right) + \hat{\beta}_2^* \left(\frac{X_i}{\sigma_i} \right) + \left(\frac{\hat{u}_i}{\sigma_i} \right)$$

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

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

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

EE325 2/2012 (Ajarn Kaewkwan Tangtipongkul)

the GLS estimator of β_2^*

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

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

where $w_i = 1/\sigma_i^2$

EE325 2/2012 (Ajarn Kaewkwan Tangtipongkul)

Difference between OLS and GLS

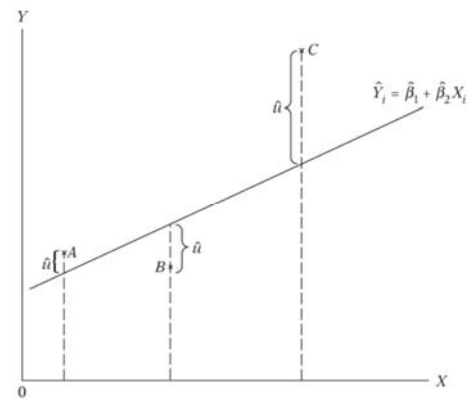
OLS we minimize

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

GLS we minimize

$$\sum w_i \hat{u}_i^2 = \sum w_i (Y_i - \hat{\beta}_1^* X_{0i} - \hat{\beta}_2^* X_i)^2, w_i = 1/\sigma_i^2$$

EE325 2/2012 (Ajarn Kaewkwan Tangtipongkul)



EE325 2/2012 (Ajarn Kaewkwan Tangtipongkul)

Example P 374-375

Value of α	Standard error of $\hat{\beta}_1$			Standard error of $\hat{\beta}_2$		
	OLS	OLS_{het}	GLS	OLS	OLS_{het}	GLS
0.5	0.164	0.134	0.110	0.285	0.277	0.243
1.0	0.142	0.101	0.048	0.246	0.247	0.173
2.0	0.116	0.074	0.0073	0.200	0.220	0.109
3.0	0.100	0.064	0.0013	0.173	0.206	0.056
4.0	0.089	0.059	0.0003	0.154	0.195	0.017

EE325 2/2012 (Ajarn Kaewkwan Tangtipongkul)

The most striking feature of these results is that OLS, with or without correction for heteroskedasticity, consistently overestimates the true standard error obtained by the (correct) GLS procedure, especially for large values of α , thus establishing the superiority of GLS

EE325 2/2012 (Ajarn Kaewkwan Tangtipongkul)

Detection of Heteroscedasticity

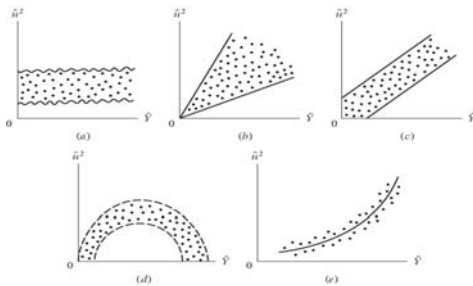
EE325 2/2012 (Ajarn Kaewkwan Tangtipongkul)

Detection of Heteroscedasticity

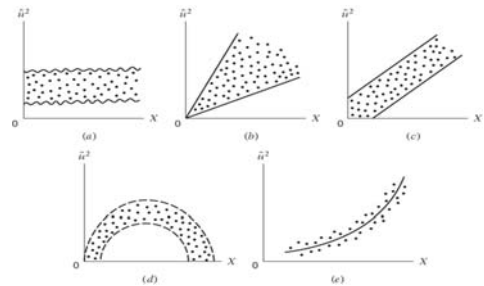
- Informal method
 - Graphical method
- Formal methods
 - Park Test
 - Breusch-Pagan Test
 - White's General Heteroscedasticity Test

EE325 2/2012 (Ajarn Kaewkwan Tangtipongkul)

Graphical method



EE325 2/2012 (Ajarn Kaewkwan Tangtipongkul)



EE325 2/2012 (Ajarn Kaewkwan Tangtipongkul)

Park Test

Park formalizes the graphical method by suggesting that σ_i^2 is some function of the explanatory variable X_i . The functional form he suggests 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

EE325 2/2012 (Ajarn Kaewkwan Tangtipongkul)

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

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

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

EE325 2/2012 (Ajarn Kaewkwan Tangtipongkul)

Example

Table 11.1 Relationship between compensation and productivity

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

Y = average compensation in thousands of dollars
X = average productivity in thousands of dollars
i = ith employment size of the establishment

EE325 2/2012 (Ajarn Kaewkwan Tangtipongkul)

Step 1

Run the OLS regression disregarding the heteroscedasticity question

$$\hat{Y}_i = 1992.3452 + 0.2329 X_i$$

$$se = (936.4791) \quad (0.0998)$$

$$t = (2.1275) \quad (2.333)$$

$$R^2 = 0.4375$$

EE325 2/2012 (Ajarn Kaewkwan Tangtipongkul)

Step 2

We obtain $\hat{u}_i$ from this regression, and then in the second stage we run the regression

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

$$\ln \hat{u}_i^2 = 35.817 - 2.8099 \ln X_i$$

$$se = (38.319) \quad (4.216)$$

$$t = (0.934) \quad (-0.667)$$

$$R^2 = 0.0595$$

EE325 2/2012 (Ajarn Kaewkwan Tangtipongkul)

Breusch-Pagan Test (LM Test)

Consider the k-variables linear regression model

$$Y_i = \beta_1 + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + u_i$$

we assume that

$$E(u \mid x_1, x_2, \dots, x_k) = 0$$

, so that OLS is unbiased and consistent.

EE325 2/2012 (Ajarn Kaewkwan Tangtipongkul)

☺ Test procedure ☺

Step 1 Estimate Equation

$$Y_i = \beta_1 + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + u_i$$

by OLS and obtain the squared OLS residuals $\hat{u}^2$

Step 2 Run the regression

$$\hat{u}_i^2 = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + \alpha_4 X_{4i} + \dots + \alpha_k X_{ki} + \nu_i$$

Keep the R-Squared from this regression, $R_{\hat{u}^2}^2$

EE325 2/2012 (Ajarn Kaewkwan Tangtipongkul)

Step 3 Form either the F statistics or the LM statistic

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

Where k is the number of regressors in step 2

The LM statistic for Heteroscedasticity is

$$LM = n \cdot R_{\hat{u}^2}^2$$

Under the null hypothesis, LM is distributed asymptotically as χ_k^2

EE325 2/2012 (Ajarn Kaewkwan Tangtipongkul)

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.

EE325 2/2012 (Ajarn Kaewkwan Tangtipongkul)

White's General Heteroscedasticity Test

Consider the following three-variable regression model

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

The White test proceeds as follows:

Step 1 Given the data, we estimate

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

And obtain the residuals $\hat{u}_i$

EE325 2/2012 (Ajarn Kaewkwan Tangtipongkul)

Step 2 We then run the following (auxiliary) regression

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

Obtain the R-Squared from this (auxiliary) regression

EE325 2/2012 (Ajarn Kaewkwan Tangtipongkul)

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,

$$n \cdot R^2 \xrightarrow{asy} \chi_{df}^2$$

EE325 2/2012 (Ajarn Kaewkwan Tangtipongkul)

Step 4 If the chi-square value obtained in

$$n \cdot R^2 \xrightarrow{asy} \chi_{df}^2$$

Exceeds the critical chi-square value at the chosen level of significance, the conclusion is that there is heteroscedasticity. If it does not exceed the critical chi-square value, there is no heteroscedasticity, which is to say that in the auxiliary regression

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

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

EE325 2/2012 (Ajarn Kaewkwan Tangtipongkul)

Example P.387-388

From Cross-sectional data on 41 countries

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

Y = ratio of Trade taxes to total Government revenue

X_2 = ratio of the sum of Exports plus imports to GNP

X_3 = GNP per capita

EE325 2/2012 (Ajarn Kaewkwan Tangtipongkul)

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

$$\hat{u}_i^2 = -5.8417 + 2.5629 \ln Trade_i + 0.6918 \ln GNP_i - 0.4081 (\ln Trade_i)^2 - 0.0491 (\ln GNP_i)^2 + 0.0015 (\ln Trade_i)(\ln GNP_i)$$

$$R^2 = 0.1148$$

EE325 2/2012 (Ajarn Kaewkwan Tangtipongkul)

$$n \cdot R^2 = 41(0.1148) = 4.7068$$

The 5 percent critical chi-square value for 5 df is 11.0705.

4.7068 < 11.0705 On the basis of the White test, that there is no heteroscedasticity

EE325 2/2012 (Ajarn Kaewkwan Tangtipongkul)

Remedial Measures

EE325 2/2012 (Ajarn Kaewkwan
Tangtipongkul)

Remedial Measures

- When σ_i^2 is known: The Method of Weighted Least Squares
- When σ_i^2 is not known

EE325 2/2012 (Ajarn Kaewkwan
Tangtipongkul)

When σ_i^2 is known

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.

EE325 2/2012 (Ajarn Kaewkwan
Tangtipongkul)

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/σ_i	X/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) .

EE325 2/2012 (Ajarn Kaewkwan
Tangtipongkul)

$$\hat{Y}_i = 3417.833 + 148.767 X_i$$

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 182.8665
_cons		3417.833	81.13632	42.12	0.000	3225.976 3609.69

EE325 2/2012 (Ajarn Kaewkwan
Tangtipongkul)

$$(\hat{Y}_i / \sigma_i) = 3406.639(1/\sigma_i) + 154.153(X_i / \sigma_i)$$

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) =	4803.82
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	16.95487	9.10	0.000	114.1218 194.3018
consigma		3406.277	80.96623	42.07	0.000	3214.822 3597.731

EE325 2/2012 (Ajarn Kaewkwan
Tangtipongkul)

When σ_i^2 is not known

Several assumptions about the pattern of X_i^2 heteroscedasticity

Assumption 1 The error variance is proportional to

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

EE325 2/2012 (Ajarn Kaewkwan Tangtipongkul)



EE325 2/2012 (Ajarn Kaewkwan Tangtipongkul)

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

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

EE325 2/2012 (Ajarn Kaewkwan Tangtipongkul)

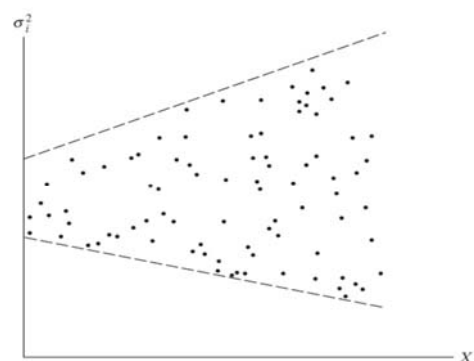
Assumption 2 The error variance is proportional to X_i
The **square root transformation**

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

EE325 2/2012 (Ajarn Kaewkwan Tangtipongkul)

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

EE325 2/2012 (Ajarn Kaewkwan Tangtipongkul)



EE325 2/2012 (Ajarn Kaewkwan Tangtipongkul)

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

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

EE325 2/2012 (Ajarn Kaewkwan Tangtipongkul)

$$E(Y_i) = \beta_1 + \beta_2 X_i$$

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

$$\begin{aligned} \hat{Y}_i &= \hat{\beta}_1 + \hat{\beta}_2 X_i \\ \frac{Y_i}{\hat{Y}_i} &= \beta_1 \left(\frac{1}{\hat{Y}_i} \right) + \beta_2 \left(\frac{X_i}{\hat{Y}_i} \right) + v_i \end{aligned}$$

EE325 2/2012 (Ajarn Kaewkwan Tangtipongkul)

Assumption 4 A log transformation such as

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

very often reduces heteroscedasticity when compared with the regression

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

EE325 2/2012 (Ajarn Kaewkwan Tangtipongkul)

Example

Heteroscedasticity

EE325 2/2012 (Ajarn Kaewkwan Tangtipongkul)

R&D Expenditure, Sales, and Profits in 14 Industry Groupings in the United States, 2005 (all figures in millions of dollars)

Since the cross-sectional data presented in this table are quite heterogeneous, in a regression of R&D on sales, heteroscedasticity is likely

EE325 2/2012 (Ajarn Kaewkwan Tangtipongkul)

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

EE325 2/2012 (Ajarn Kaewkwan Tangtipongkul)

Source	SS	df	MS	
Model	208733442	1	208733442	Number of obs = 14
Residual	1.0083e+09	12	84021567.1	F(1, 12) = 2.48
Total	1.2170e+09	13	93614788.2	Prob > F = 0.1418
				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

EE325 2/2012 (Ajarn Kaewkwan
Tangtipongkul)

$$\hat{R} \& D_i = 1338 + 0.0437 Sales_i$$

$$se = (5015) (0.0277)$$

$$t = (0.27) (1.58)$$

$$r^2 = 0.172$$

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

EE325 2/2012 (Ajarn Kaewkwan
Tangtipongkul)

White Test

Source	SS	df	MS	
Model	9.2405e+16	2	4.6203e+16	Number of obs = 14
Residual	1.2022e+17	11	1.0929e+16	F(2, 11) = 4.23
Total	2.1263e+17	13	1.6356e+16	Prob > F = 0.0435
				R-squared = 0.4346
				Adj R-squared = 0.3318
				Root MSE = 1.00e+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

EE325 2/2012 (Ajarn Kaewkwan
Tangtipongkul)

$$\hat{u}_i^b = -46,746,325 + 578Sales_i + 0.000846Sales_i^2$$

$$se = (112,224,348) (1308) (0.003171)$$

$$t = (-0.42) (0.44) (0.27)$$

$$R^2 = 0.435$$

Using the R^2 value and $n=14$, we obtain $nR^2 = 6.090$

Under the null hypothesis of no heteroscedasticity, this should follow a chi-square distribution with 2 df (because there are two regressors). The p-value of obtaining a chi-square value of as much as 6.090 or greater is about 0.0476. Since this is a low value, the White test also suggests that there is heteroscedsticity.

EE325 2/2012 (Ajarn Kaewkwan
Tangtipongkul)

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 the error variance.

EE325 2/2012 (Ajarn Kaewkwan
Tangtipongkul)

White's heteroscedasticity-consistent standard errors

$$\hat{R} \& D_i = 1337.87 + 0.0437 \text{Sales}_i$$

$$se = (4892.447) \quad (0.0411)$$

$$t = (0.27) \quad (1.06)$$

$$r^2 = 0.172$$

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 the White procedure is strictly a large-sample procedure, where as we have only 14 observations

EE325 2/2012 (Ajarn Kaewkwan Tangtipongkul)

STATA

Source	SS	df	MS	Number of obs =	14
Model	208733442	1	208733442	F(1, 12) =	2.48
Residual	1.0083e+09	12	84021567.1	Prob > F =	0.1410
Total	1.2170e+09	13	93614788.2	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

. whitest

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

EE325 2/2012 (Ajarn Kaewkwan Tangtipongkul)

H_0 : Homoscedasticity

H_1 : Otherwise

White's general test statistic is 6.0842.

Degree of freedom =2

Critical value of Chi-square at 5 percent significance level is 5.99147

6.0842 > 5.99147 Reject the null hypothesis

The White test also suggests that there is heteroscedsticity.

EE325 2/2012 (Ajarn Kaewkwan Tangtipongkul)

Breusch-Pagan test

EE325 2/2012 (Ajarn Kaewkwan Tangtipongkul)

Source	SS	df	MS	Number of obs =	14
Model	208733442	1	208733442	F(1, 12) =	2.48
Residual	1.0083e+09	12	84021567.1	Prob > F =	0.1410
Total	1.2170e+09	13	93614788.2	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

EE325 2/2012 (Ajarn Kaewkwan Tangtipongkul)

. 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**

EE325 2/2012 (Ajarn Kaewkwan Tangtipongkul)



H_0 : *Homoscedasticity*

H_1 : *Otherwise*

Breusch-Pagan test's general test statistic is 8.83.

Degree of freedom =1

Critical value of Chi-square at 5 percent significance level is $8.83 > 3.84146$

Reject the null hypothesis

The Breusch-Pagan test also suggests that there is heteroscedsticity.



Source

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