# 7. Heteroscedasticity

(1) $$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u.$$

Assumption 5 (classical assumptions) states that the variance of $u$ (conditional on the explanatory variables) is constant. This is called the homoscedasticity assumption.

We use the term heteroscedasticity for the situation that this assumption fails, that is, the variance of the error terms depends in some way on the values of the regressors.

## 7.1 Consequences

In the presence of heteroscedasticity:

($i$) OLS estimators are not BLUE
($ii$) $\mathbb{V}\text{ar}[\widehat{\beta}_j]$ are biased, implying that $t$-, $F$-, and $LM$-statistics, and confidence intervals are no more reliable.
($iii$) OLS estimator are no more asymptotically efficient.

However,

($iv$) OLS estimators are still unbiased.
($v$) OLS estimators are still consistent

## 7.2 Heteroscedasticity-robust inference

Consider for the sake of simplicity

$$y_i = \beta_0 + \beta_1 x_i + u_i, \tag{2}$$

$i = 1, \ldots, n$, where

$$\mathbb{V}\text{ar}[u_i | x_i] = \sigma_i^2. \tag{3}$$

Then writing the OLS-estimator of $\beta_1$ in the form

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^n (x_i - \bar{x}) u_i}{\sum_{i=1}^n (x_i - \bar{x})^2}. \tag{4}$$

Because the error terms are uncorrelated,

$$\mathbb{V}\text{ar}[\hat{\beta}_1] = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 \sigma_i^2}{(\text{SST}_x)^2}, \tag{5}$$

where

$$\text{SST}_x = \sum_{i=1}^n (x_i - \bar{x})^2. \tag{6}$$

In the homoscedastic case, where $\sigma_i^2 = \sigma^2$ for all $i$ formula (5) reduces to the usual variance $\sigma_u^2 / \sum(x_i - \bar{x})^2$.

White (1980)* derives a robust estimator for (5) as

$$\widehat{\mathbb{V}\text{ar}[\hat{\beta}_1]} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2 \hat{u}_i^2}{(\text{SST}_x)^2}, \tag{7}$$

where $\hat{u}_i$ are the OLS residuals.

If we rewrite (1) in the matrix form

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{u}, \tag{8}$$

and write $\hat{\mathbf{b}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ as

$$\hat{\mathbf{b}} = \mathbf{b} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} \tag{9}$$

*White, H. (1980). A Heteroscedasticity-consistent covariance matrix estimator and direct test for heteroscedasticity. *Econometrica* 48, 817–838.

Given $\mathbf{X}$, the variance-covariance matrix of $\widehat{\mathbf{b}}$ is

$$\mathbb{C}\text{ov}[\widehat{\mathbf{b}}] = (\mathbf{X}'\mathbf{X})^{-1}\left(\sum_{i=1}^{n}\sigma_i^2\mathbf{x}_i\mathbf{x}_i'\right)(\mathbf{X}'\mathbf{X})^{-1}, \tag{10}$$

where $\mathbf{x}_i' = (1, x_{i1}, \ldots x_{ik})$ is the $i$th row of the data matrix $\mathbf{X}$ on $x$-variables.

Analogous to (7), an estimator of (10) is

$$\widehat{\mathbb{C}\text{ov}[\widehat{\mathbf{b}}]} = (\mathbf{X}'\mathbf{X})^{-1}\left(\sum_{i=1}^{n}\hat{u}_i^2\mathbf{x}_i\mathbf{x}_i'\right)(\mathbf{X}'\mathbf{X})^{-1}, \tag{11}$$

Heteroscedasticity robust standard error for estimate $\widehat{\beta}_j$ is the square root of the $j$th diagonal element of (11).

Remark 7.1: If the residual variances $\mathbb{V}\text{ar}[u_i] = \sigma_i^2 = \sigma_u^2$ are the same, then because

$$\mathbf{X}'\mathbf{X} = \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i',$$

(11) is

$$\mathbb{C}\text{ov}[\hat{\mathbf{b}}] = \sigma_u^2 (\mathbf{X}'\mathbf{X})^{-1} \left( \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i' \right) (\mathbf{X}'\mathbf{X})^{-1} = \sigma_u^2 (\mathbf{X}'\mathbf{X})^{-1},$$

i.e., the usual case.

## Example 7.1: Wage example with heteroscedasticity-robust standard errors.

```
Dependent Variable: LOG(WAGE)
Method: Least Squares
Sample: 1 526
Included observations: 526
White Heteroscedasticity-Consistent Standard Errors & Covariance
================================================================
Variable        Coefficient    Std. Error    t-Statistic      Prob.
----------------------------------------------------------------
C                0.321378      0.109469      2.936         0.0035
MARRMALE         0.212676      0.057142      3.722         0.0002
MARRFEM         -0.198268      0.058770     -3.374         0.0008
SINGFEM         -0.110350      0.057116     -1.932         0.0539
EDUC             0.078910      0.007415     10.642         0.0000
EXPER            0.026801      0.005139      5.215         0.0000
TENURE           0.029088      0.006941      4.191         0.0000
EXPER^2         -0.000535      0.000106     -5.033         0.0000
TENURE^2        -0.000533      0.000244     -2.188         0.0291
================================================================
R-squared                0.461       Mean dependent var       1.623
Adjusted R-squared       0.453       S.D. dependent var       0.532
S.E. of regression       0.393       Akaike info criterion    0.988
Sum squared resid       79.968       Schwarz criterion        1.061
Log likelihood        -250.955       F-statistic             55.246
Durbin-Watson stat       1.785       Prob(F-statistic)        0.000
================================================================
```

Comparing to Example 6.3 the standard errors change slightly (usually little increase). However, conclusions do not change.

## 7.3 Testing for Heteroscedasticity

Start, as usual, with the linear model

(12) $$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u.$$

The null hypothesis of homoscedasticity is

(13) $$\mathbb{V}\mathrm{ar}[u|x_1,\ldots,x_k] = \mathbb{E}[u^2|x_1,\ldots,x_k] = \sigma^2,$$

since $\mathbb{E}[u|x_1,\ldots,x_k]^2 = 0$ by assumption 4.

Test therefore, whether $u^2$ is in some way related to the regressors $x_i$. The simplest approach is a linear function:

(14) $$u_i^2 = \delta_0 + \delta_1 x_1 + \cdots + \delta_k x_k + v_i.$$

The homoscadasticity hypothesis is then

(15) $$H_0 : \delta_1 = \cdots = \delta_k = 0,$$

i.e., $\sigma^2 = \delta_0$.

The error terms $u_i$ are unobservable. They must be replaced by the OLS-residuals $\widehat{u}_i$. Run therefore the regression

$$\widehat{u}_i^2 = \delta_0 + \delta_1 x_1 + \cdots + \delta_k x_k + v_i. \tag{16}$$

Estimating the parameters with OLS, the null hypothesis (15) can be tested with the overall $F$-statistic defined in (4.25), which can be written in terms of the $R$-square as

$$F = \frac{R^2_{\widehat{u}^2}/k}{(1 - R^2_{\widehat{u}^2})/(n-k-1)}, \tag{17}$$

where $R^2_{\widehat{u}^2}$ is the $R$-square of the regression (16).

The $F$-statistic is asymptotically $F$-distributed under the null hypothesis with $k$ and $n-k-1$ degrees for freedom.

Breuch-Bagan test:

Asymptotically (17) is equivalent to the Lagrange Multiplier (LM) test

(18) $$LM = nR^2_{\widehat{u}^2},$$

which is asymptotically $\chi^2$-distributed with $k$ degrees of freedom when the null hypothesis is true.

Remark 7.2: In regression (16) the explanatory variables can be also some external variables (not just $x$-variables).

White test:

Suppose, for the sake of simplicity, that in (1) $k = 3$, then the White-procedure is to estimate

(19)

$$\begin{aligned}\widehat{u}_i^2 &= \delta_0 + \delta_1 x_1 + \delta_2 x_2 + \delta_3 x_3 \\ &\quad +\delta_4 x_1^2 + \delta_5 x_2^2 + \delta_6 x_3^2 \\ &\quad +\delta_7 x_1 x_2 + \delta_8 x_1 x_3 + \delta_9 x_2 x_3 + v_i\end{aligned}$$

Estimate the model and use $LM$-statistic of the form (18) to test whether the coefficients $\delta_j$, $j = 1, \ldots, 9$, are zero.

Another option, which is more conserving on degrees of freedom, is to estimate

$$\widehat{u}_i^2 = \delta_0 + \delta_1 \widehat{y}_i + \delta_2 \widehat{y}_i^2 + v_i$$

and use the F or LM statistic for the null hypothesis $H_0: \ \delta_1 = 0, \delta_2 = 0$.

Remark 7.3: As is obvious, Breuch-Pagan (BP) test with $x$-variables is White test without the cross-terms.

Example 7.2: In the wage example Breusch-Pagan yields $R^2_{\hat{u}^2} = 0.025075$. With $n = 526$,

$$LM = nR^2_{\hat{u}^2} \approx 13.19$$

$df = 8$, producing $p$-value 0.1055. Thus there is not empirical evidence of heteroscedasticity.

White with cross-terms gives

$$R^2_{\hat{u}^2} = 0.086858$$

and

$$LM \approx 45.69$$

with $df = 36$ and $p$-value of 0.129. Again we do not reject the null hypothesis of homoscedasticity.

The alternative form of the White test gives

$$R^2_{\hat{u}^2} = 0.0079$$

and

$$LM \approx 4.165$$

with $df = 2$ and $p$-value of 0.125. Again we do not reject the null hypothesis of homoscedasticity.

Remark 7.4: When $x$-variables include dummy-variables, be aware of the dummy-variable trap due to $D^2 = D$! I.e., you can only include $D$s. Modern econometric packages, like EViews, avoid the trap automatically if the procedure is readily available in the program.

## 7.4 Weighted Least Squares (WLS)

Suppose the heteroscedasticity is of the form

$$\mathbb{V}\text{ar}[u_i|\mathbf{x}_i] = \sigma^2 h(\mathbf{x}_i), \tag{20}$$

where $h_i = h(\mathbf{x}_i) > 0$ is some (known) function of the explanatory (and possibly some other variables).

Dividing both sides of (1) by $\sqrt{h_i}$ and denoting the new variables as $\tilde{y}_i = y_i/\sqrt{h_i}$, $\tilde{x}_{ij} = x_{ij}/\sqrt{h_i}$, and $\tilde{u}_i = u_i/\sqrt{h_i}$, we get regression

$$\tilde{y}_i = \beta_0 \frac{1}{\sqrt{h_i}} + \beta_1 \tilde{x}_{i1} + \cdots + \beta_k \tilde{x}_{ik} + \tilde{u}_i, \tag{21}$$

where

$$\begin{aligned} \mathbb{V}\mathrm{ar}[\tilde{u}_i|\mathbf{x}_i] &= \frac{1}{h_i}\mathbb{V}\mathrm{ar}[u_i|\mathbf{x}_i] \\ &= \frac{1}{h_i} h_i \sigma^2 \\ &= \sigma^2, \end{aligned} \tag{22}$$

i.e., homoscedastic (satisfying the classical assumption 2).

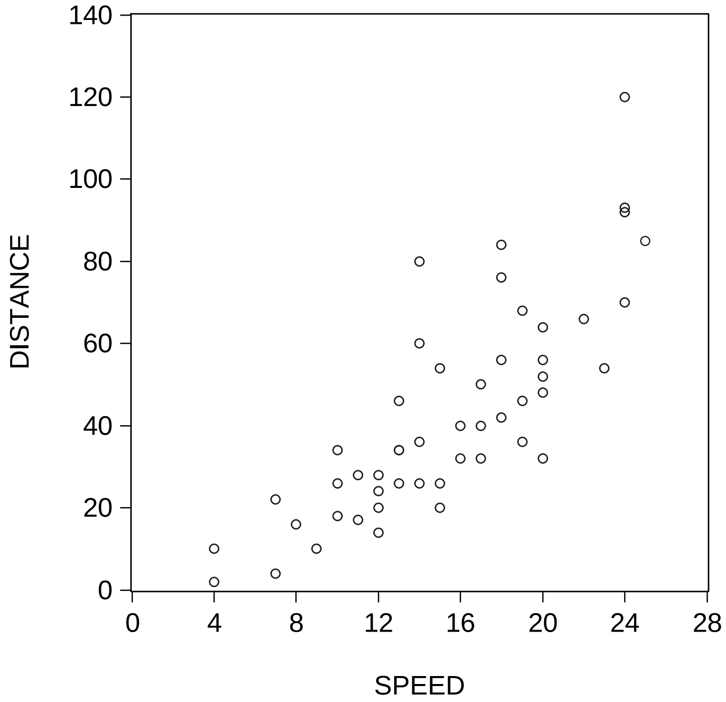
Applying OLS to (22) produces again BLUE for the parameters.

From estimation point of view the transformation leads, in fact, to the minimization of

$$(23) \quad \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_k x_{ik})^2 / h_i.$$

This is called Weighted Least Squares (WLS), where the observations are weighted by the inverse of $\sqrt{h_i}$.

Example 7.3: Speed and stopping distance for cars, $n = 50$ observations.




Visual inspection suggests somewhat increasing variability as a function of speed. From the linear model

$$\text{dist} = \beta_0 + \beta_1 \text{speed} + u$$

White test gives $LM = 3.22$ with $df = 2$ and $p$-val 0.20, which is not statistically significant.

Physics: stopping distance proportional to square of speed, i.e., $\beta_1(\text{speed})^2$.

Thus instead of a linear model a better alternative should be

$$\text{dist}_i = \beta_1(\text{speed}_i)^2 + \text{error}_i, \tag{24}$$

Human factor: reaction time $v_i = \beta_0 + u_i$, where $\beta_0$ is the average reaction time and the error term $u_i \sim N(0, \sigma_u^2)$.

During the reaction time the car moves a distance

$$v_i \times \text{speed}_i = \beta_0\text{speed}_i + u_i\text{speed}_i. \tag{25}$$

Thus modeling the error term in (24) as (25), gives

$$\text{dist}_i = \beta_0\text{speed}_i + \beta_1(\text{speed}_i)^2 + e_i, \tag{26}$$

where

$$e_i = u_i \times \text{speed}_i. \tag{27}$$

Because

$$\begin{aligned} \mathbb{V}\text{ar}[e_i|\text{speed}_i] &= (\text{speed}_i)^2\mathbb{V}\text{ar}[u_i] \\ &= (\text{speed})^2\sigma_u^2, \end{aligned} \tag{28}$$

the heteroscedasticity is of the form (20) with

$$h_i = (\text{speed}_i)^2. \tag{29}$$

Estimating (26) by ignoring the inherent heteroscedasticity yields

```
Dependent Variable: DISTANCE
Method: Least Squares
Included observations: 50
==============================================================
Variable         Coefficient    Std. Error   t-Statistic   Prob.
--------------------------------------------------------------
SPEED               1.239          0.560         2.213       0.032
SPEED^2             0.090          0.029         3.067       0.004
==============================================================
R-squared                0.667   Mean dependent var       42.980
Adjusted R-squared       0.660   S.D. dependent var       25.769
S.E. of regression      15.022   Akaike info criterion     8.296
Sum squared resid    10831.117   Schwarz criterion         8.373
Log likelihood        -205.401   Durbin-Watson stat        1.763
==============================================================
```

Accounting for the heteroscedasticity and estimating the coefficients from

$$\frac{\text{dist}_i}{\text{speed}_i} = \beta_0 + \beta_1 \text{speed}_i + u_i \tag{30}$$

gives

```
==============================================================
Variable         Coefficient    Std. Error   t-Statistic   Prob.
--------------------------------------------------------------
SPEED              1.261          0.426          2.963     0.00472
SPEED^2            0.089          0.026          3.402     0.00136
==============================================================
```

The results are not materially different. Thus the heteroscedasticity is not a big problem here.

Remark 7.5: The $R$-squares from (26) and (30) are not comparable. Comparable $R$-squares can be obtained by computing $\widehat{\text{dist}}$ using the coefficient estimates of (30) and squaring the correlation

$$R = \mathbb{C}\text{orr}(\text{dist}_i, \widehat{\text{dist}}_i). \tag{31}$$

The $R$-square for (30) is 0.194 while for (26) 0.667. A comparable $R$-square, however, is obtained by squaring (31), which gives 0.667, i.e., the same in this case (usually it is slightly smaller; why?).

## Feasible generalized Least Squares (GLS)

In practice the $h(\mathbf{x})$ function is rarely known. In order to guarantee strict positivity, a common practice is to model it as

$$h(\mathbf{x}_i) = \exp(\delta_0 + \delta_1 x_1 + \cdots + \delta_k x_k). \tag{32}$$

In such a case we can write

$$\log(u^2) = \alpha_0 + \delta_1 x_1 + \cdots + \delta_k x_k + e, \tag{33}$$

where $\alpha_0 = \log \sigma^2 + \delta_0$ and $e$ is an error term.

In order to estimate the unknown parameters the procedure is:

(i) Obtain OLS residuals $\widehat{u}$ from regression equation (1)
(ii) Run regression (33) for $\log(\widehat{u}^2)$, and generate the fitted values, $\widehat{g}_i$.
(iii) Re-estimate (1) by WLS using $1/\widehat{h}_i$, where $\widehat{h}_i = \exp(\widehat{g}_i)$.

This is called a *feasible GLS*.

Another possibility is to obtain the $\widehat{g}_i$ by regressing $\log(\widehat{u}^2)$ on $\widehat{y}$ and $\widehat{y}^2$.