1. Consequences of Heteroskedasticity for OLS

- Assumption MLR. 5: Homoskedasticity $\text{var}(u_i \mid x) = \sigma^2$

- Now we relax this assumption and allow that the error variance depends on the independent variables, i.e., heteroskedasticity

**Consequences of heteroskedasticity**

- [IMPORTANT] OLS estimators are still unbiased and consistent under heteroskedasticity.
  - Recall that we prove unbiasedness under MLR.1 through MLR.4, but not MLR.5.
- Also, interpretation of R-squared is not changed in the presence of heteroskedasticity.
  - $R^2 \approx 1 - \frac{\sigma_u^2}{\sigma_y^2}$ where the unconditional error variance, $\sigma_u^2$, is unaffected by heteroskedasticity, (which refers to the conditional error variance)
- Heteroskedasticity invalidates variance formulas for OLS estimators
- The usual F-tests and t-test are not valid under heteroskedasticity because the variance formula for OLS estimator is wrong.
- Under heteroskedasticity, OLS is no longer the best linear unbiased estimator (BLUE); there might be more efficient linear estimator.
2. Heteroskedasticity-Robust Inference after OLS Estimation

- Formula for OLS standard errors and related statistics have been developed that are robust to heteroskedasticity of unknown form.
- To derive those formula, consider the followings:

  - Simple regression model case
    - Estimation model:  \( y_i = \beta_0 + \beta_1 x_i + u_i \)
    - We assumes that the first four Gauss-Markov assumptions hold, but not the last one.
    - The general form of contain heteroskedasticity:  \( \text{var}(u_i|x_i) = \sigma_i^2 \)
      - The subscript \( i \) indicates that the variance of the error depends upon the particular value of \( x_i \).

  - Then, following the same argument, we can show that  \( \text{var}(\hat{\beta}_1) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sigma_i^2}{SST_x^2} \).

  - Therefore, the sampling variance can be estimated by:  \( \hat{\text{var}}(\hat{\beta}_1) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2 \hat{u}_i^2}{SST_x^2} \).
Multiple regression model case

- Estimation model: \( y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik} + u_i \)

- The general form of contain heteroskedasticity: \( \text{var}(u_i|x) = \sigma_i^2 \)

- Formula for heteroskedasticity-robust OLS variance is:

\[
\hat{\text{var}}(\hat{\beta}_j) = \frac{\sum_{i=1}^{n} \hat{r}_{ij}^2 \hat{u}_i^2}{\text{SSR}_j}
\]

where \( \hat{r}_{ij} \) denotes the \( i \)th residual from regressing \( x_j \) on other independent variables, and \( \text{SSR}_j \) is the sum of squared residuals from this regression.

- The squared root of this formula provides heteroskedasticity-robust OLS standard error, and is also called White standard errors.

Discussion on heteroskedasticity-robust OLS standard error

- All formula are only valid in large samples.
- Using the formula, the usual t-test is valid asymptotically.
- The usual F-statistic does not work under heteroskedasticity, but heteroskedasticity robust versions are available in most software.
(Example) Hourly wage equation

- Estimated equation

\[
\widehat{\log(wage)} = -0.128 + 0.0904 \text{educ} + 0.0410 \text{exper} - 0.0007 \text{exper}^2
\]

\[
(0.105) (0.0075) (0.0052) (0.0001)
\]

\[
[0.107] [0.0078] [0.0050] [0.0001]
\]

- The usual OLS standard errors are presented in the parentheses and the heteroskedasticity-robust standard errors are presented in the squared brackets.

- The heteroskedasticity-robust standard errors may be larger or smaller than their nonrobust counterparts. The differences are often small in practice.

- The usual t-test said that all the independent variables are significant.

- \( H_0 : \beta_{\text{exper}} = \beta_{\text{exper}^2} = 0 \)

\[
F = 17.95 \quad \text{vs.} \quad F_{\text{robust}} = 17.99 \quad \Rightarrow \text{the null is highly rejected}
\]

- F-statistics are also often not too different.

- However, if there is strong heteroskedasticity, differences may be larger. To be on safe side, it would be better to always compute robust standard errors.
3. Testing for Heteroskedasticity

- It may still be interesting whether there is heteroskedasticity because then OLS may not be the most efficient linear estimator anymore.

(1) Breusch-Pagan test for heteroskedasticity

■ Rationale

- $H_0: \text{var}(u|x) = \sigma^2$; Homoskedasticity

- $\text{var}(u|x) = E(u^2|x) - \left[ E(u|x) \right]^2 = E(u^2|x)$ under MLR.4: $E(u|x) = 0$

=> $E(u^2|x) = E(u^2) = \sigma^2$

- This implies that the mean of $u^2$ must not vary with $x_1, \ldots, x_k$. 
Heteroskedasticity

How to test

- Consider the following model:
  \[ \hat{u}^2 = \delta_0 + \delta_1 x_1 + \ldots + \delta_k x_k + \text{error} \]
  - Regress squared residuals on all explanatory variables

- Hypothesis testing: \( H_0 : \delta_1 = \ldots = \delta_k = 0 \)
  - The hypothesis means that all the explanatory variables do not explain the squared residuals, implying homoskedasticity.

- F-statistic:
  \[ F = \frac{R^2_{\hat{u}^2} / k}{(1 - R^2_{\hat{u}^2}) / (n - k - 1)} \sim F_{k, n-k-1} \]
  - where \( R^2_{\hat{u}^2} \) is the R-squared from the regression of \( \hat{u}^2 \) on all explanatory variables

- LM test (alternative test statistic)
  \[ LM = n \cdot R^2_{\hat{u}^2} \sim \chi^2_k \]
  - This test is only asymptotically valid.
(2) White test for heteroskedasticity

- How to test
  - Consider the following model with 3 independent variable case:
    \[ \hat{u}^2 = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + \delta_3 x_3 + \delta_4 x_1^2 + \delta_5 x_2^2 + \delta_6 x_3^2 + \delta_7 x_1 x_2 + \delta_8 x_1 x_3 + \delta_9 x_2 x_3 + \text{error} \]
    - Regress squared residuals on all explanatory variables, their squares, and interactions.
  - Hypothesis testing: \( H_0 : \delta_1 = \ldots = \delta_9 = 0 \)
    - The White test detects more general deviations from heteroskedasticity than the Breush-Pagan test
  - \( LM = n \cdot R_{\hat{u}^2}^2 \sim \chi^2_9 \) (we can also use an F-test for this hypothesis.)
  - Disadvantage of this form of the White test:
    - Including all squares and interactions leads to a large number of estimated parameters
      (e.g. \( k = 6 \) leads to 27 parameters to be estimated)

- Alternative form of the White test
  - \( \hat{u}^2 = \delta_0 + \delta_1 \hat{y} + \delta_2 \hat{y}^2 + \text{error} \)
  - Hypothesis testing: \( H_0 : \delta_1 = \delta_2 = 0, \quad LM = n \cdot R_{\hat{u}^2}^2 \sim \chi^2_2 \)
4. Weighted Least Squared Estimation

(1) The Heteroskedasticity is known up to a multiplicative constant (Weighted LS)

- Assume that $\text{var}(u_i | x_i) = \sigma^2 h(x_i)$, $h(x_i) = h_i > 0$
  
  ➢ The functional form of the heteroskedasticity, i.e., $h(x_i)$, is known.

- $y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik} + u_i$

  $\Rightarrow \begin{bmatrix} y_i \\ \sqrt{h_i} \end{bmatrix} = \beta_0 \begin{bmatrix} 1 \\ \sqrt{h_i} \end{bmatrix} + \beta_1 \begin{bmatrix} x_{i1} \\ \sqrt{h_i} \end{bmatrix} + \ldots + \beta_k \begin{bmatrix} x_{ik} \\ \sqrt{h_i} \end{bmatrix} + \begin{bmatrix} u_i \\ \sqrt{h_i} \end{bmatrix}$

  $\Leftrightarrow y_i^* = \beta_0 x_{i0}^* + \beta_1 x_{i1}^* + \ldots + \beta_k x_{ik}^* + u_i^*$; Transformed model

- Example: Savings and income
  
  ➢ $\text{sav}_i = \beta_0 + \beta_i \text{inc}_i + u_i$, \hspace{1em} $\text{var}(u_i | \text{inc}_i) = \sigma^2 \text{inc}_i$

  $\Rightarrow \begin{bmatrix} \text{sav}_i \\ \sqrt{\text{inc}_i} \end{bmatrix} = \beta_0 \begin{bmatrix} 1 \\ \sqrt{\text{inc}_i} \end{bmatrix} + \beta_1 \begin{bmatrix} \text{inc}_i \\ \sqrt{\text{inc}_i} \end{bmatrix} + u_i^*$

  ➢ Note that this regression model has no constant.
The transformed model is homoskedastic

\[
E(u_i^* \mid x_i) = E \left( \frac{u_i \left( \frac{1}{\sqrt{h_i}} \right)^2}{\sqrt{h_i}} \right) = \frac{E(u_i^2 \mid x_i)}{h_i} = \frac{\sigma_i^2 h_i}{h_i} = \sigma^2
\]

Provided that the other Gauss-Markov assumptions hold as well, OLS applied to the transformed model is the best linear unbiased estimator.

OLS in the transformed model is called as weighted least squares (WLS)

\[
\min \sum_{i=1}^{n} \left( \frac{y_i}{\sqrt{h_i}} - b_0 \frac{1}{\sqrt{h_i}} - b_1 \frac{x_{i1}}{\sqrt{h_i}} - \ldots - b_k \frac{x_{ik}}{\sqrt{h_i}} \right)^2
\]

\[
\iff \min \sum_{i=1}^{n} \left( y_i - b_0 - b_1 x_{i1} - \ldots - b_k x_{ik} \right)^2 / h_i
\]

Observation with a large variance gets a smaller weight in the optimization problem

Why is WLS more efficient than OLS in the original model?

Observation with a large variance are less informative than observation with small variance and therefore should get less weight.

WLS is a special case of generalized least squares (GLS)
(2) Unknown heteroskedasticity function (feasible GLS)

- \( \text{var}(u|x) = \sigma^2 \exp(\delta_0 + \delta_1 x_1 + \ldots + \delta_k x_k) = \sigma^2 h(x) \)

  ➢ Assumed general form of heteroskedasticity. That is, \( h(x) \) is unknown and therefore should be estimated before running the estimation model.

  ➢ exp-function is used to ensure positivity.

- \( u^2 = \sigma^2 \exp(\delta_0 + \delta_1 x_1 + \ldots + \delta_k x_k) \cdot v \) where \( v \) is multiplicative error

\[
\Rightarrow \log(u^2) = \alpha_0 + \delta_1 x_1 + \ldots + \delta_k x_k + \epsilon \quad \text{where} \quad \alpha_0 = \log(\sigma^2) + \delta_0.
\]

- Steps for estimation of FGLS

  i. Regress \( y \) on all independent variables, \( x_1, \ldots, x_k \), and obtain the residuals, \( \hat{u}_i \)

  ii. Estimates the model, \( \log(\hat{u}^2) = \hat{\alpha}_0 + \hat{\delta}_1 x_1 + \ldots + \hat{\delta}_k x_k + \text{error} \), in order to obtain \( \hat{h}_i \).

\[
\Rightarrow \hat{h}_i = \exp(\hat{\alpha}_0 + \hat{\delta}_1 x_1 + \ldots + \hat{\delta}_k x_k)
\]

  iii. Apply WLS using \( \hat{h}_i \) obtained in the previous step, i.e., estimate the following model:

\[
y_i^* = \beta_0 x_{i0}^* + \beta_1 x_{i1}^* + \ldots + \beta_k x_{ik}^* + u_i^* \quad \text{where} \quad y_i^* = \frac{y_i}{\sqrt{\hat{h}_i}}, \quad x_{ij}^* = \frac{x_{ij}}{\sqrt{\hat{h}_i}} \quad \text{for} \quad j = 0, 1, \ldots, k, \quad \text{and} \quad u_i^* = \frac{u_i}{\sqrt{\hat{h}_i}}
\]
(Example) Demand for cigarettes

- Estimation by OLS

\[
\hat{\text{cigs}} = -3.64 + 0.880 \log(\text{income}) - 0.751 \log(\text{cigpric}) - 0.501 \text{educ} \\
(24.08) (0.728) (5.773) (0.167) \\
- 0.771 \text{age} - 0.0090 \text{age}^2 - 2.83 \text{restaurn} \\
(0.160) (0.0017) (1.11)
\]

\[n = 807, \quad R^2 = 0.0526\]

- \text{cigs}: cigarettes smoked per day; \quad \text{cigpric}: cigarettes price
- \text{restaurn}: smoking restriction in restaurants

- \text{p-value}_{\text{Breusch–Pagan}} = 0.000 \quad \Rightarrow \text{reject homoskedasticity}
Estimation by FGLS

\[ \hat{cigs} = -5.64 + 1.30 \log(income) - 2.94 \log(cigprice) - 0.463 \text{educ} \]

\[ (17.80) \quad (0.44) \quad (4.46) \quad (0.120) \]

\[ -0.482 \text{age} - 0.0056 \text{age}^2 - 3.46 \text{restauran} \]

\[ (0.097) \quad (0.0009) \quad (0.80) \]

\[ n = 807, \; R^2 = 0.1134 \]

Discussion

- The income elasticity is now statistically significant; other coefficients are also more precisely estimated (without changing qualitative results)
**What if the assumed heteroskedasticity function is wrong?**

- If the heteroskedasticity function is misspecified, WLS is still unbiased and consistent under MLR.1 through MLR.4, but not efficient.
- However, if we estimate $\hat{h}(x_i)$ with a misspecified functional form, then FGLS is biased and inconsistent.
- In contrast, OLS is always unbiased in the presence of heteroskedasticity.
- Practically, OLS estimation with robust standard error adjustment would be a safe way.