8 Heteroskedasticity

What does ie mean to have homoskedastic errors:

$$Var(u_i|x_i) = \sigma$$

What does it mean to have heteroskaedastic errors:

$$Var(u_i|x_i) = f(\sigma_i, x_i)$$

i.e. the conditional variance of the error term depends on the independent variables used in regression.

8.1 Consequence of Heteroskedasticity for OLS

1. Note that under MLR.1 to MLR.4 OLS estimates are unbiased and consistent. It has nothing to do with homoskedasticity assumption.

   Why: because both variances in population R-squared are unconditional variances, the population R-squared is unaffected by the presence of heteroskedasticity.

2. Where is the problem: estimator of variances of $\hat{\beta}$ are biased which makes the standard errors for confidence intervals incorrect.

   + Issues in inference of F or the LM statistic under heteroskedasticity.

8.2 Heteroskedasticity-Robust Inference after OLS Estimation

1. Problem: how do we adjust t, F and LM statistics to make the correct inference? This is under any form of Heteroskedasticity.

2. *Solution*: compute standard errors which take into account the possibility of heteroskedasticity.

3. Implementation in a single linear model:

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

Heteroskedasticity implies that u_i varies over i conditional on the x_i. This implies the following in the simplest of cases.

$$Var(u_i|x_i) = \sigma_i^2$$

4. OLS estimator (Recall 2.52 page 43, JW)

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^{n}(x_i - \bar{x})u_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \beta_1 + \frac{\sum_{i=0}^{n}(x_i - \bar{x})u_i}{SST_x}$$

5. Only using MLR.1 to MLR.4 we know the variance of the parameter estimate to be

$$Var(\hat{\beta}_1) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2\sigma_i^2}{SST_x^2}$$

When homoskedasticity implies $\sigma_i^2 = \sigma^2$, we can take the σ out of summation to simplify as:

$$Var(\hat{\beta}_1) = \sigma^2 \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{SST_x^2} = \sigma^2 \frac{1}{SST_x}$$

6. Under heteroskedasticity, White showed that an estimate of $\mathrm{Var}(\hat{\beta})$ is given by the following:

$$Var(\hat{\beta}) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2 \hat{u}_i^2}{SST_x^2}$$

where $\hat{u}_i$ is the OLS residual from initial regression of y on x.

White showed that

$$plim_{n \to \infty} \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2 \hat{u}_i^2}{SST_x^2} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sigma_i^2}{SST_x^2}$$

7. *In Multiple Regression Model*: this can be implemented by obtaining the $\hat{u}_i^2$ from the following regression:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + u$$

The square root of the unconditional variance is called the Heteroskedasticity-robust standard error:

$$\widehat{Var}(\hat{\beta}_j) = \frac{\sum_{i=1}^{n} \hat{r}_{ij}^2 \hat{u}_i^2}{SSR_j^2}$$

where $\hat{r}_{ij}$ denoted the i^{th} residual from regressing x_j on all other independent variables and SSR_j is the sum of squared residuals from this regression.

8. There are many adjustments to the standard error and method of computing it, but all are equivalent asymptotically or in large samples.

   We can use this to come up with correct t-statistics, F, LM and hence correct inference in the presence of heteroskedasticity

   ▷ Heteroskedasticity Robust t-Statistic is given as:

$$t = \frac{\text{estimate - hypothesized value}}{\text{standard error}}$$

9. Question: If Heteroskedasticity-robust standard errors (HRSE) are always correct, should we always use them?

   No, for small samples normal SE are better. Why? Because HRSE corrects for heteroskedasticity. If there is no hetero in the data in the first place, we are correcting for something which might induce an error in our inference.

10. Heteroskedasticity-robust F-statistic: is also called the Wald statistic can be obtained for *any* form of heteroskedasticity $\Longrightarrow$ derivation is involved and we wont need it for the moment.

   ▷ Similarly, remember Chow test of common coefficient across two groups? [Section 7.4c: TO DO computer exercise C14]

8.2.1 Computing Heteroskedasticity Robust LM Tests

1. To compute LM consistent with hetero, consider the following model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + u$$

2. We want to test whether $\beta_4 = 0, \beta_5 = 0$.

3. Estimate the restricted model and save the residuals $\tilde{u}$

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \tilde{u}$$

4. For each excluded variable x_{ei}, regress them on the remaining included variable x_{ii} and save the residuals as r_{ei}. This under (2) is as follows:

$$x_4 \text{ on } x_1, x_2 \text{ and } x_3 \text{ to obtain } r_4$$
$$x_4 \text{ on } x_1, x_2 \text{ and } x_3 \text{ to obtain } r_5$$

5. Regress the following and obtain SSR from the regression.

$$1 \text{ on } r_4\tilde{u} \text{ and } r_5\tilde{u} \text{ to obtain SSR}$$

6. LM statistic is $n - SSR$