Lecture 20

Heteroskedasticity
(sections 8.1 - 8.4 only)

I. Consequences of Heteroskedasticity for OLS

A. $\hat{\beta}_j$ are unbiased

B. $\hat{\beta}_j$ are ~~inefficient~~ consistent

C. $\hat{\beta}_j$ are inefficient

D. $se(\hat{\beta}_j)$ are biased and $t_{\hat{\beta}_j}$ are inappropriate for conducting statistical inference

E. ANOVA F-statistic for testing overall significance of regression is no longer appropriate.

II. Heteroskedasticity-Robust Inference After OLS Estimation

A. White's Standard Errors

B. Robust T-statistics

C. Robust F-statistics

III. Diagnosing and Testing for Heteroskedasticity

A. Residual Plots versus $X_1, X_2, \cdots, X_k$.

B. White's Test for Heteroskedasticity

C. Bruesch-Pagan Test for Heteroskedasticity

D. Special Case of White's Test

IV Weighted Least Squares

A. When Heteroskedasticity is known up to a multiplicative Constant

B. When Heteroskedasticity Function Must Be Estimated: Feasible GLS.

To determine the consequences of heteroskedasticity for OLS estimation let us consider the simple linear regression model

$$y_i = \beta_0 + \beta_1 x_i + u_i \qquad (1)$$

where $E(u_i) = 0 \;\; \forall i$ but $Var(u_i) = E(u_i^2) = \sigma_i^2$ and the variances of the random errors u_i are not equal. We know that

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_1^N (x_i - \bar{x}) u_i}{\sum_1^N (x_i - \bar{x})^2} . \qquad (2)$$

Therefore

$$E(\hat{\beta}_1) = \beta_1 + \frac{\sum_1^N (x_i - \bar{x}) E(u_i)}{SST_x}$$

$$= \beta_1 + \frac{\sum_1^N (x_i - \bar{x}) \cdot 0}{SST_x}$$

$$= \beta_1$$

and the OLS estimator of β_1 is still unbiased in the presence of heteroskedasity.

Moreover, it can be shown that the sampling distribution of $\hat{\beta}_1$ collapses on β_1 as the sample size N goes to infinity and thus

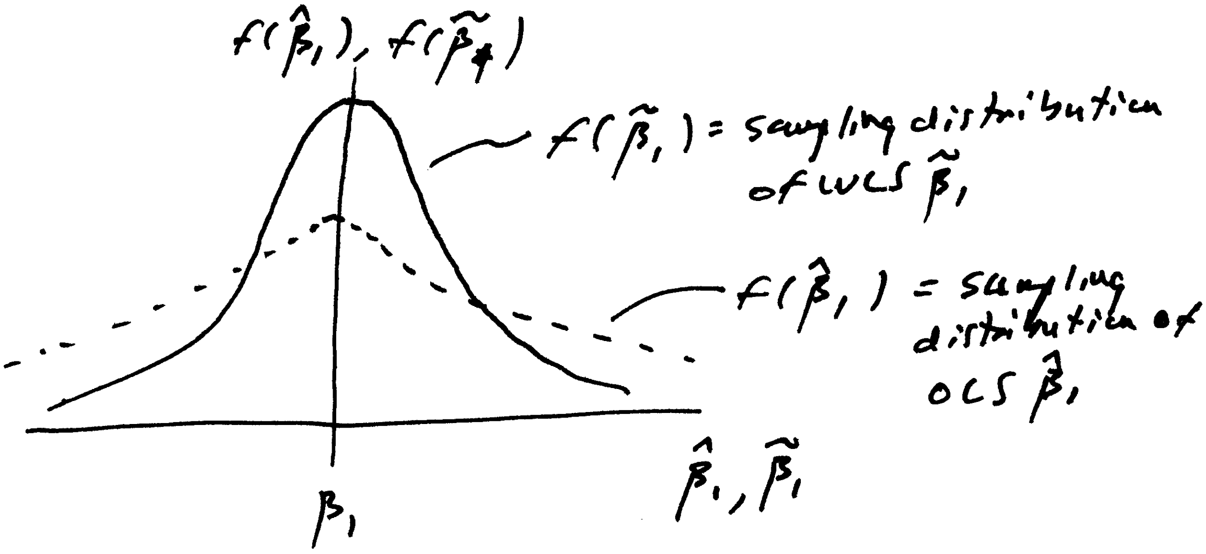
$$\text{plim } \hat{\beta}_1 = \beta_1$$

and, even in the presence of heteroskedasticity of the errors u_i, the OLS estimator of $\hat{\beta}_1$ is consistent.

Unfortunately, although $\hat{\beta}_1$ is unbiased and consistent it is inefficient in the sense that for a given sample size N the weighted least squares estimator, say $\tilde{\beta}_1$, which we will define later, has a smaller sampling variance than the sampling variance of the OLS estimator $\hat{\beta}_1$.

(inefficiency)

This relationship is depicted below:



Obviously the weighted Least Squares estimator, $\tilde{\beta}_1$, is more efficient in estimating β_1 because it sampling distribution is "tighter" around β_1 than the sampling distribution of $\hat{\beta}_1$, the OLS estimator.

Using equation (2) above it follows that

$$Var(\hat{\beta}_1) = \frac{\sum_1^N (X_i - \bar{X})^2 Var(u_i)}{\left[\sum_1^N (X_i - \bar{X})^2\right]^2}$$

$$= \frac{\sum_1^N (X_i - \bar{X})^2 \sigma_i^2}{\left[\sum_1^N (X_i - \bar{X})^2\right]^2} \neq \frac{\sigma^2}{\sum_1^N (X_i - \bar{X})^2}$$

given that the u_i are assumed to be independent of each other. Thus, in the presence of heteroskedasticity, the variance of $\hat{\beta}_1$ is no longer equal to $\sigma^2 / \sum_1^N (X_i - \bar{X})^2$ and, thus, the usual OLS standard error

$$se(\hat{\beta}_1) = \sqrt{\frac{\hat{\sigma}^2}{\sum_1^N (X_i - \bar{X})^2}}$$

will be a biased and inconsistent estimator of the actual standard deviation of $\hat{\beta}_1$, namely,

$$sd(\hat{\beta}_1) = \sqrt{\frac{\sum_1^N (X_i - \bar{X})^2 \sigma_i^2}{\left[\sum_1^N (X_i - \bar{X})^2\right]^2}} \quad . \quad (3)$$

It would be nice to be able to get estimates of the σ_i, say $\hat{\sigma}_i^2$, that are unbiased and consistent. But, in general, that is impossible because with unique values of the explanatory variable X_i, there is only one OLS residual, $\hat{u}_i = y_i - \hat{y}_i$, available for estimating σ_i^2 and that is not enough.

However, in an ingenious paper Halbert White (1980) showed that although we can't estimate the σ_i^2 individually we can obtain a consistent estimate of $sd(\hat{\beta}_1)$ by the formula

$$se_w(\hat{\beta}_1) = \sqrt{\frac{\sum_1^N (X_i - \bar{X})^2 \hat{u}_i^2}{\left[\sum_1^N (X_i - \bar{X})^2\right]^2}} \quad . \qquad (4)$$

Thus, even in the presence of heteroskedasticity, a consistent (robust) t-statistic can be calculated as

$$t_{w,\hat{\beta}_1} = \frac{\hat{\beta}_1 - \beta_1^0}{se_w(\hat{\beta}_1)} \quad . \qquad (5)$$

The standard error (4) is called ~~calculate~~ ~~the~~ White's standard error for the OLS estimator $\hat{\beta}_1$ and (5) is called White's t-statistic. In a similar manner the overall F-statistic and subset F-statistics can be robustized to heteroskedasticity.

Although ~~That~~ we can no longer use the usual OLS standard errors, t-statistics, and F-statistics we can "adjust" these statistics, a la White (1980), so as to still be able to use OLS for statistical inference with appropriate modifications.

Diagnosing and Testing for Heteroskedasticity

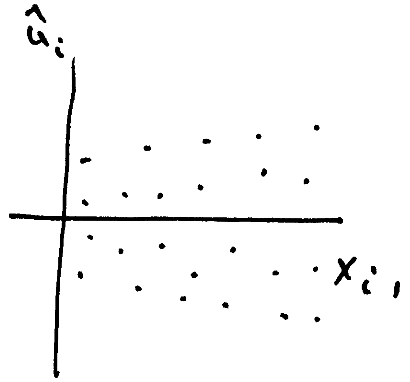
The best way to diagnose heteroskedasticity is to use residual plots.

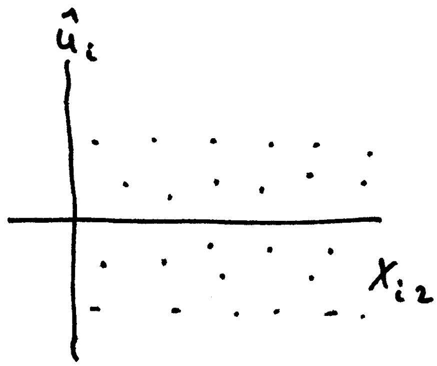
Consider the regression equation

$$y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + u_i .$$

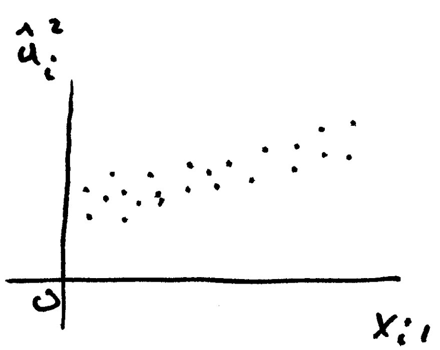
Let $\hat{u}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{i1} - \hat{\beta}_2 X_{i2}$ be the OLS residuals from this equation.

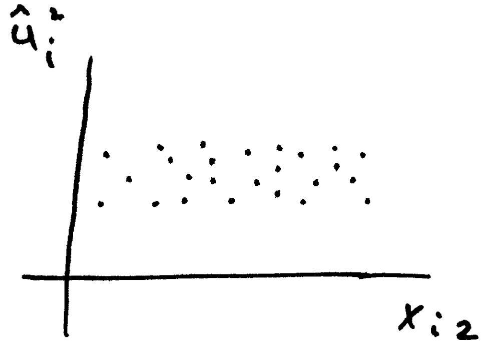
Consider the residual plots





or equivalently the squared residual plots





These plots indicate that the heteroskedasticity of the errors is related to the first explanatory variable X_{i1} and <u>not</u> to the second explanatory variable X_{i2}.

A reasonable parametrization for this heteroskedasticity is

$$\sigma_i^2 = \sigma^2 X_{i1} \qquad (6)$$

or, maybe, more generally

$$\sigma_i^2 = \sigma^2 X_{i1}^{\lambda} \qquad (7)$$

for unknown λ.

If we are confident that (6) is appropriate then we can use Weighted Least Squares (WLS) to estimate β_0, β_1, and β_2 more efficiently than using ~~weighted~~ Ordinary Least Squares.

If, instead, we adopt (7) we will have to estimate λ (as described below) and then apply Feasible WLS (or some call it

Feasible Generalized Least Squares (FGLS))

to estimate β_0, β_1, and β_2. Feasible WLS provides, asymptotically, more efficient estimates of β_0, β_1, and β_2 than OLS estimation.

Tests of Heteroskedasticity

Suppose that we suspect heteroskedasticity in our regression equation

$$y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik} + u_i$$

but we are not sure what is "causing" the heteroscedasticity.

Consider the following Heteroskedasticity Estimating equation

$$\hat{u}_i^2 = \delta_0 + \delta_1 X_{i1} + \delta_2 X_{i2} + \cdots + \delta_k X_{ik} + v_i \quad (8)$$

White's Heteroskedasticity Test (without cross product terms) uses equation (8) to test the following hypotheses

$$H_0: \delta_1 = \delta_2 = \cdots = \delta_k = 0 \quad \text{(no heteroskedasticity)}$$

$$H_1: \text{not } H_0 \quad \text{(heteroskedasticity is present)}$$

This test is carried out by computing the Overall F-statistic ($R^2_{\hat{u}^2}$ is the R^2 of eq. (8))

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

If the p-value of this F-statistic is greater than α (say .05) then we conclude that our regression equation does not have a heterskedasticity problem. Otherwise, we conclude that heteroskedasticity is a problem and we must take a remedial action before proceeding.

The LM version of the above test is called the Breusch-Pagan test for heteroskedasticity. Form the statistic

$$LM = N \cdot R^2_{\hat{u}^2} .$$

Under $H_0: \delta_1 = \delta_2 = \cdots = \delta_k = 0$ (no heteroskedasticity) this test statistic, in repeated samples, is distributed as a chi-square random variable with k degrees of freedom. Thus, if the p-value of this test statistic is greater than α (say .05), then H_0 is accepted. Otherwise, heteroskedasticity is indicated.

The test equation (8) can be generalized to include the cross product terms $X_1 \cdot X_2$, $X_1 \cdot X_3, \cdots, X_{k-1} \cdot X_k$ and squared terms

$X_1^2, X_2^2, \dots, X_k^2$, that is,

$$\hat{u}_i^2 = \delta_0 + \delta_1 X_{i1} + \cdots + \delta_k X_{ik} + \text{squared and cross product terms} + v_i \qquad (9)$$

Then we compute the overall F-statistic for this equation and proceed as before. This test procedure is call ~~the~~ White's Test for Heteroskedasticity (with crossproduct terms).

One additional test for heteroskedasticity is called a special case of the White Test for Heteroskedasticity. Consider the test equation

$$\hat{u}_i^2 = \delta_0 + \delta_1 \hat{y}_i + \delta_2 \hat{y}_i^2 + \text{error} \qquad (10)$$

where $\hat{y}_i$ are the OLS fitted values of the y_i.

The white's special case test consists of testing

$$H_0: \delta_1 = \delta_2 = 0 \quad \text{(no heteroskedasticity)}$$

vs.

$$H_1: \text{not } H_0 \quad \text{(heteroskedasticity is present)}.$$

This is an overall-F test with

$$F = \frac{R^2_{\hat{u}^2}/2}{(1-R^2_{\hat{u}^2})/(N-3)}.$$

If the p-value of this F-statistic is greater than α (say .05) then we conclude that our regression equation does not have a heteroskedasticity problem. Otherwise, we conclude that heteroskedasticity is present.

Weighted Least Squares

To understand the rationale for weighted Least Squares (WLS) consider the multiple regression

$$y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik} + u_i$$

and assume $E(u_i^2) = \sigma_i^2$. Now multiply the dependent variable and all of the explanatory variables including the intercept term (1) by the term $(1/\sigma_i)$ resulting in the transformed equation

$$\frac{y_i}{\sigma_i} = \beta_0 \cdot \frac{1}{\sigma_i} + \beta_1 \frac{X_{i1}}{\sigma_i} + \cdots + \beta_k \frac{X_{ik}}{\sigma_i} + \frac{u_i}{\sigma_i}$$

or

$$y_i^* = \beta_0 X_{i0}^* + \beta_1 X_{i1}^* + \cdots + \beta_k X_{ik}^* + u_i^* \quad (11)$$

where $y_i^* = (y_i/\sigma_i)$, $X_{i0}^* = (1/\sigma_i)$, $X_{i1}^* = (X_{i1}/\sigma_i)$, $\ldots$, $X_{ik}^* = (X_{ik}/\sigma_i)$ and $u_i^* = u_i/\sigma_i$.

Now equation (11) is a homoskedastic equation! $E(u_i^{*2}) = E(u_i^2/\sigma_i^2) = \sigma_i^2/\sigma_i^2 = 1$ for all i. The Ordinary Least Squares (in eq. (11)) estimators of $\beta_0, \beta_1, \cdots, \beta_k$, call them $\tilde{\beta}_0, \tilde{\beta}_1, \cdots, \tilde{\beta}_k$, are termed the <u>weighted Least Squares estimators</u>. Note that these estimators <u>do not</u> coincide with the OLS estimators of $\beta_0, \beta_1, \cdots, \beta_k$ in the original (untransformed) regression equation. The WLS estimators can also be thought of as the estimators that minimize the following weighted least squares criterion:

$$\sum_1^N (y_i - \tilde{\beta}_0 - \tilde{\beta}_1 X_{i1} - \cdots - \tilde{\beta}_k X_{ik})^2 \cdot \frac{1}{\sigma_i^2} \quad (12)$$

(12) is a weighted sum of squared residuals with each squared residual being weighted by the <u>reciprocal</u> of the variance of u_i.

Equivalently, each residual is weighted by the reciprocal of the standard deviation of the u_i (i.e. $\frac{1}{\sigma_i}$).

In weighted least squares the observations that exhibit the greatest variability around the sample regression function carry the least weight in determining the estimates of the regression coefficients while the observations that exhibit the least variability carry the greatest weight.

Now let us carry the discussion a little further. Suppose we know the error variance is of the form

$$\sigma_i^2 = \sigma^2 h(X_{i1}, X_{i2}, \cdots, X_{ik}) = \sigma^2 h_i \quad (13)$$

where we know the function $h(X_{i1}, X_{i2}, \cdots, X_{ik}) = h_i$. Here we know the variance of each error up to a propotional constant, σ_i^2 We can then form the transformed regression equation

$$\frac{y_i}{\sqrt{h_i}} = \beta_0 \cdot \frac{1}{\sqrt{h_i}} + \beta_1 \frac{X_{i1}}{\sqrt{h_i}} + \cdots + \beta_k \frac{X_{ik}}{\sqrt{h_i}} + \frac{u_i}{\sqrt{h_i}}$$

$$y_i^* = \beta_0 X_{i0}^* + \beta_1 X_{i1}^* + \cdots + \beta_k X_{ik}^* + u_i^*.$$

This is a homoskedastic equation with the error variance being $E(u_i^{*2}) = \frac{1}{h_i} E(u_i^2) = \sigma^2$. Thus applying OLS to this transformed equation produces the Weighted Least Squares coefficient estimates $\tilde{\beta}_0, \tilde{\beta}_1, \cdots, \tilde{\beta}_k$ which are Best Linear Unbiased Estimators and thus efficient.

Now let's take another step. Suppose again that

$$E(u_i^2) = \sigma_i^2 = \sigma^2 h(X_{i1}, \cdots, X_{ik})$$

where this time let us assume that we know the form of the function $h(\cdot)$ but the function is dependent on some unknown parameters, say $\delta_0, \cdots, \delta_k$. For example, assume $h(\cdot)$ is of the form

$$h(X_1, \cdots, X_k) = \exp(\delta_0 + \delta_1 X_1 + \delta_2 X_2 + \cdots + \delta_k X_k)$$

and thus

$$\sigma_i^2 = \sigma^2 \exp(\delta_0 + \delta_1 X_1 + \delta_2 X_2 + \cdots + \delta_k X_k). \quad (14)$$

The heteroskedasticity specification (14) is called <u>multiplicative heteroskedasticity</u>.

Consider the <u>heteroskedasticity estimating equation</u> for multiplicative heteroskedasticity:

$$\log(\hat{u}_i^2) = \alpha_0 + \delta_1 X_{i1} + \delta_2 X_{i2} + \cdots + \delta_k X_{ik} + e. \quad (15)$$

Applying OLS to (15) we can obtain

$$\hat{h}_i = \exp[\hat{\alpha}_0 + \hat{\delta}_1 X_{i1} + \hat{\delta}_2 X_{i2} + \cdots + \hat{\delta}_k X_{ik}]. \quad (16)$$

using (16) we can form the transformed regression

$$\frac{y_i}{\sqrt{\hat{h}_i}} = \beta_0 \frac{1}{\sqrt{\hat{h}_i}} + \beta_1 \frac{X_{i1}}{\sqrt{\hat{h}_i}} + \cdots + \beta_k \frac{X_{ik}}{\sqrt{\hat{h}_i}} + \frac{u_i}{\sqrt{\hat{h}_i}}$$

$$y_i^{**} = \beta_0 X_{i0}^{**} + \beta_1 X_{i1}^{**} + \cdots + \beta_k X_{ik}^{**} + u_i^{**}. \quad (17)$$

Apply OLS to this transformed equation produces what are called <u>Feasible</u> weighted Least Squares estimators $\tilde{\tilde{\beta}}_0, \tilde{\tilde{\beta}}_1, \cdots, \tilde{\tilde{\beta}}_k$, where we have put the double tildas over the coefficients to distinguish them from the

the WLS estimators $\tilde{\beta}_0, \tilde{\beta}_1, \cdots, \tilde{\beta}_k$ which are derived assuming that the error variance σ_i^2 is known up to a proportional constant $\sigma_\cdot^2$ If the true specification of the equation's heteroskedasticity is (14) then the Feasible WLS estimates are consistent and asymptotically efficient.

In terms of statistical inference, the standard errors, t-statistics, and F-statistics associated with WLS or Feasible WLS (sometimes called Feasible Generalized Least Squares (FGLS)) are consistent and provide for more powerful inference (tests with higher power) than the tests based on robustified (White heteroskedasticity adjusted) test statistics.

So if an investigator knows the form of the heteroscedasticity of a regression equation, it is more efficient, both from an estimation as well as a statistical inference point of view, to use WLS or Feasible WLS. On the other hand, if the investigator doesn't know the form of the heteroskedasticity then he/she must rely on robustified t-statistics and F-statistics using the methods of White (1980).

See my SAS program Gpa3het.sas posted on the web for an example of WLS estimation and heteroskedasticity testing of the cumgpa equation that is discussed in Example 8.2, p. 262 in Wooldridge.

Also see my EVIEWS program Gpa3het.wf1 posted on the web that does the same work found in Gpa3het.sas. The only difference is that the weight for WLS in EVIEWS is $1/\sqrt{h_i}$ whereas in SAS the weight is defined as $1/h_i$.