Heteroskedasticity

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

$\text{Var}(u \mid \mathbf{x}) = \sigma_i^2$

Ch.8 Heteroskedasticity

1. Consequences of Heteroskedasticity for OLS
2. Heteroskedasticity-Robust Inference after OLS Estimation
3. Testing for Heteroskedasticity
4. Weighted Least Squares Estimation
5. The Linear Probability Model Revised*

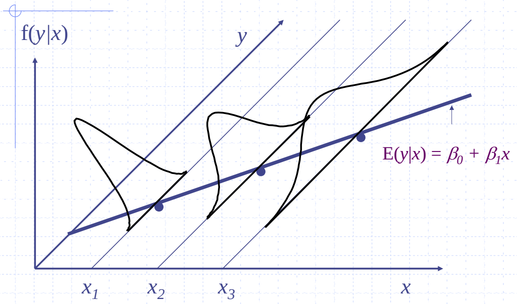
8.1 Consequences of Heteroskedasticity

What is Heteroskedasticity?

- The ***assumption of homoskedasticity*** implies that conditional on the explanatory variables, the variance of the unobserved error, u, was constant.

  MLR 5: $\text{Var}(u \mid x_1, x_2, \ldots, x_k) = \sigma^2$
- If this is not true, that is if the variance of u is different for different values of the x's, then the errors are heteroskedastic.

Example of Heteroskedasticity



OLSE with Heteroskedasticity

If we do not assume homoskedasticity…

- OLSE is still unbiased and consistent.
  - R^2 & adjusted R^2 is also unaffected by heteroskedasticity.
- The standard errors of the estimates are biased.
  - We can not use the usual t or F statistics.
  - OLSE is no longer BLUE.

8.2 Hetero.-Robust Inference for OLSE

- We have to find the unbiased standard errors of the estimates in heterosckedasticity.

  → ***Heterosckedasticity-Robust S.E.***
  - This *s.e.* is valid *at least asymptotically*, even if the form of unknown heterosckedasticity.
  - The t or F statistics calculated with this *s.e.* are valid asymptotically.

Variance with Heteroskedasticity 1

$$\hat{\beta}_1 = \beta_1 + \frac{\sum (x_i - \bar{x}) u_i}{SST_x}, \text{ where } SST_x = \sum (x_i - \bar{x})^2$$

$$Var(\hat{\beta}_1) = \frac{\sum (x_i - \bar{x})^2 \sigma_i^2}{SST_x^2} \quad (8.2),$$

A valid estimator for this when $\sigma_i^2 \neq \sigma^2$ is

$$Est.Var(\hat{\beta}_1) = \frac{\sum (x_i - \bar{x})^2 \hat{u}_i^2}{SST_x^2} \quad (8.3),$$

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

Econometrics 7

Variance with Heteroskedasticity 2

For the general multiple regression model, a valid estimator of $Var(\hat{\beta}_j)$ with heteroskedasticity is

$$Est.Var(\hat{\beta}_j) = \frac{\sum \hat{r}_{ij}^2 \hat{u}_i^2}{SSR_j^2}, \quad (8.4)$$

where $\hat{r}_{ij}$ is 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.

Econometrics 8

Robust Standard Errors

- Now that we have a consistent estimate of the variance, the square root can be used as a standard error for inference.

$$Est.s.e.(\hat{\beta}_j) = \sqrt{\frac{\sum \hat{r}_{ij}^2 \hat{u}_i^2}{SSR_j^2}}$$

- We call these ***Heterosckedasticity-robust standard errors***, or ***White's consistent s.e.***

Econometrics 9

Cont. Robust Standard Errors

- It is important to remember that these robust standard errors only have ***asymptotic*** justification.
  - With small sample sizes, t statistics formed with robust standard errors will not have a distribution close to the t, and inferences will not be correct.

Econometrics 10

A Robust *LM* Statistic

1. Run OLS on the restricted model and save the residuals $\breve{u}$.
2. Regress each of the excluded variables on all of the included variables (q different regressions) and save each set of residuals $\breve{r}_1, \breve{r}_2, \ldots, \breve{r}_q$
3. Regress a variable defined to be = 1 on $\breve{r}_1 \breve{u}$, $\breve{r}_2 \breve{u}$, …, $\breve{r}_q \breve{u}$, with ***no*** intercept.
4. The LM statistic is $n - SSR_1$, where SSR_1 is the sum of squared residuals from this final regression.

Econometrics 11

8.3 Testing for Heteroskedasticity

- Essentially we want to test

  H_0: $Var(u|x_1, x_2, \ldots, x_k) = \sigma^2$,

  which is equivalent to

  H_0: $E(u^2|x_1, x_2, \ldots, x_k) = E(u^2) = \sigma^2$
  - If we assume the relationship between u^2 and x_j will be linear, we can test as a linear restriction.
- So, for $u^2 = \delta_0 + \delta_1 x_1 + \ldots + \delta_k x_k + v$ (8.12) this means testing

  H_0: $\delta_1 = \delta_2 = \ldots = \delta_k = 0$ (8.13)

Econometrics 12

The Breusch-Pagan Test

- After regressing the residuals squared on all of the x's, we can use the R^2 to form an F or LM test.

$$F = \frac{R^2_{\hat{u}^2}/k}{(1-R^2_{\hat{u}^2})/(n-k-1)} \sim F_{k,n-k-1} \quad (8.15)$$

$$LM = n \cdot R^2_{\hat{u}^2} \sim \chi^2_k \quad (8.16)$$

Econometrics 13

The White Test

- The White test allows for nonlinearities by using squares and crossproducts of all the x's.

(residual)^2 = $\delta_0 + \delta_1 x_1 + \delta_2 x_2 + \delta_3 x_1^2 + \delta_4 x_2^2 + \delta_5 x_1 x_2 + v$ (8.19)'

  - Still just using an F or LM to test whether all the x_j, x_j^2, and $x_j x_h$ are jointly significant.
  - ⇔The Breusch-Pagan test will detect any linear forms of heteroskedasticity.

Econometrics 14

Alternate form of the White test

- The fitted values from OLS, $\hat{y}$, are a function of all the x's, so $\hat{y}^2$ will be a function of the squares and crossproducts and $\hat{y}$ and $\hat{y}^2$ can proxy for all of the x_j, x_j^2, and $x_j x_h$.

$$\hat{u}^2 = \delta_0 + \delta_1 \hat{y} + \delta_2 \hat{y}^2 + v \quad (8.20)$$

- Regress the residuals squared on $\hat{y}$ and $\hat{y}^2$ and use the R^2 to form an F or LM statistic.
  - Note only testing for 2 restrictions now.

Econometrics 15

8.4 Weighted Least Squares

- While it's always possible to estimate robust standard errors for OLS estimates, if we know something about the specific form of the heteroskedasticity, we can obtain more efficient estimates than OLS.
- The basic idea is going to be to transform the model into one that has homoskedastic errors – called ***weighted least squares***.

Econometrics 16

Case of form to a multiplicative constant

- Suppose the heteroskedasticity can be modeled as Var(u|$\mathbf{x}$) = $\sigma^2 h(\mathbf{x})$, where the trick is to figure out what $h(\mathbf{x}) \equiv h_i$ looks like
- Because we know Var(u|$\mathbf{x}$) = $\sigma^2 h_i$,

$$Var\left(u_i/\sqrt{h_i}\,|x\right) = \sigma^2$$

- So, if we divided our whole equation by $h(\mathbf{x})$, we would have a model where the error is homoskedastic. See (8.25) & (8.26).

Econometrics 17

Generalized Least Squares

- Estimating the transformed equation by OLS is an example of ***generalized least squares*** (GLS).
- GLS will be BLUE in this case.
- GLS is a weighted least squares (WLS) procedure where each squared residual is weighted by the inverse of Var(u_i|$\mathbf{x}_i$).

Econometrics 18

Feasible GLS

- The form of the heteroskedasticity is unknown in most cases, so we need to estimate $h(\mathbf{x}_i)$.
- Typically, we start with the assumption of a fairly flexible model, such as

  $\mathrm{Var}(u|\mathbf{x}) = \sigma^2\exp(\delta_0 + \delta_1 x_1 + \ldots + \delta_k x_k)$. (8.30)

  - Our assumption implies that

    $u^2 = \sigma^2\exp(\delta_0 + \delta_1 x_1 + \ldots + \delta_k x_k)v$

    where $E(v|\mathbf{x}) = 1$, then $E(v) = 1$

Econometrics 19

Cont. Feasible GLS

- From (8.30), we can write

  $\ln(u^2) = \alpha_0 + \delta_1 x_1 + \ldots + \delta_k x_k + e$. (8.31)

  where $E(e) = 1$ and e is independent of $\mathbf{x}$.

  - Now, we can use $\hat{u}$ as an estimate of u, and the inverse of exponential fitted values ($1/\hat{h} = 1/\exp(\hat{g})$) as the weight.
- Summing up,
  1. Run the original OLS model, save the residuals, $\hat{u}$, square them and take the log
  2. Regress $\ln(\hat{u}^2)$ on all of the independent variables and get the fitted values, $\hat{g}$
  3. Do WLS using $1/\exp(\hat{g})$ as the weight

Econometrics 20

WLS Wrapup

- When doing *F* tests with WLS, form the weights from the unrestricted model and use those weights to do WLS on the restricted model as well as the unrestricted model.
- Remember we are using WLS just for efficiency – OLS is still unbiased & consistent.
- Estimates will still be different due to sampling error, but if they are very different then it's likely that some other Gauss-Markov assumption is false.

Econometrics 21

Logarithm in Heteroskedasticity

- Under heteroskedasticity, we assume:
  - $\mathrm{Var}[y_i|x_i] = E[y_i - E[y_i|x_i]\ |x_i]^2 = \sigma_i^2$ (a.1)
  - $E[y_i|x_i] = \beta_0 + \beta_1 x_i \equiv \mu_i$. (a.2)
- By ***Taylor series expansion***, we can get
  - $f(y_i) \approx f(\mu_i) + f'(\mu_i)(y_i - \mu_i)$. (a.3)
- Here, since $f(\mu_i)$ can be treated as nonrandom variable under conditional $\mathbf{x}$,
  - $E[f(y_i)|x_i] = f(\mu_i)$. (a.4)

Econometrics 22

Cont. Logarithm in Heteroskedasticity

$$
\begin{aligned}
\mathrm{Var}[f(y_i)|x_i] &= E[f(y_i) - E[f(y_i)|x_i]\ |x_i]^2 && \leftarrow\text{(a.1)}\\
&= E[f(y_i) - f(\mu_i)\ |x_i]^2 && \leftarrow\text{(a.4)}\\
&= E[f'(\mu_i)(y_i - \mu_i)\ |x_i]^2 && \leftarrow\text{(a.3)}\\
&= \{f'(\mu_i)^2\}E[(y_i - \mu_i)\ |x_i]^2 &&\\
&= \{f'(\mu_i)\}^2\sigma_i^2 && \leftarrow\text{(a.1)}
\end{aligned}
$$

If we assume $f(y_i) = \ln(y_i)$ & $\sigma_i^2 = \sigma^2 E[y_i|x_i]^2$,

$$
\begin{aligned}
\{f'(\mu_i)\}^2\sigma_i^2 &= \{1/\mu_i\}^2\sigma^2\mu_i^2 && \leftarrow\text{(a.2)}\\
&= \sigma^2 && \text{(Homoskedasticity)}
\end{aligned}
$$

Econometrics 23