Heteroskedasticity

William Brasic

The University of Arizona

Homoskedasticity

Definition 1: Homoskedasticity

Under random sampling, the errors $\boldsymbol{u}$ are said to be homoskedastic if

$$\mathbb{V}\left[\boldsymbol{u}|X\right] = \sigma^2 I_n.$$

- Variance of errors is constant for each individual.
- Recall this is MLR Assumption 5.
    - Not needed for unbiased and consistent estimates, but needed for inference.

Heterskedasticity

Definition 2: Heteroskedasticity

Under random sampling, the errors $\boldsymbol{u}$ are said to be heteroskedastic if

$$\mathbb{V}\left[\boldsymbol{u}|X\right] = \text{diag}\left(\sigma_i^2\right).$$

- Variance of errors is individual specific.

Heteroskedasticity

Question 1

What are the implications of heteroskedasticity?

Heteroskedasticity

Question 1

What are the implications of heteroskedasticity?

Answer to Question 1

1. OLS is no longer BLUE (because it is no longer efficient).
2. $\widehat{\mathbb{V}}\left[\widehat{\beta}\right] = \widehat{\sigma}^2(X'X)^{-1}$ is a biased estimate of $\mathbb{V}\left[\widehat{\beta}\right] = \sigma^2(X'X)^{-1}$.
   - Thus, T-statistics and p-values will be incorrect, on average, meaning our inferences may very well be incorrect!

Heteroskedastic Consistent SE

Question 2

So our errors are heteroskedastic. Now the estimators of our standard errors are inconsistent so we can't do inference properly. What do we do?

Answer to Question 2

Adjust the standard errors to make them consistent again!

Variance of OLS Estimator

Definition 3: Variance of OLS Estimator

Under heteroskedasticity where $\mathbb{V}[\boldsymbol{u}] = \text{diag}\left(\sigma_i^2\right)$, the variance of the OLS estimator is

$$\begin{aligned}\mathbb{V}\left[\widehat{\boldsymbol{\beta}}\right] &= \mathbb{V}\left[\beta + (X'X)^{-1}X'\boldsymbol{u}\right] \\ &= \mathbb{V}\left[(X'X)^{-1}X'\boldsymbol{u}\right] \\ &= (X'X)^{-1}X'\mathbb{V}[\boldsymbol{u}]X(X'X)^{-1} \\ &= (X'X)^{-1}\left(X'\text{diag}\left(\sigma_i^2\right)X\right)(X'X)^{-1}.\end{aligned}$$

- We now need a consistent estimator of this matrix rather than $\widehat{\sigma}^2(X'X)^{-1}$.
  - We will discuss two of the four dominant heteroskedasticity corrections.

Variance of OLS Estimator

Theorem 1: HC0 is Consistent

White's (1980) heteroskedastic-consisent variance correction takes the form of

$$\widehat{\mathbb{V}}_{HC0}\left[\widehat{\boldsymbol{\beta}}\right] = \left(X'X\right)^{-1}\left(X'\text{diag}\left(\widehat{\sigma}_i^2\right)X\right)\left(X'X\right)^{-1}$$

and is consistent for $\mathbb{V}\left[\widehat{\boldsymbol{\beta}}\right]$.

- This is a large sample (asymptotic) correction.
- Can underestimate standard errors; need larger n for this bias to disappear asymptotically.

Variance of OLS Estimator

Theorem 2: HC1 is Consistent

The following variance-covariance estimator of

$$\widehat{\mathbb{V}}_{HC1}\left[\widehat{\boldsymbol{\beta}}\right] = \frac{n}{n-k}(X'X)^{-1}\left(X'\text{diag}\left(\widehat{\sigma}_i^2\right)X\right)(X'X)^{-1}$$
$$= \frac{n}{n-k}\widehat{\mathbb{V}}_{HC0}\left[\widehat{\boldsymbol{\beta}}\right]$$

and is consistent for $\mathbb{V}\left[\widehat{\boldsymbol{\beta}}\right]$.

- This is a large sample (asymptotic) correction.
- Corrects $HC0$'s underestimation of standard errors via the inflation factor $\frac{n}{n-k}$; does not need as large of a n for bias to disappear asymptotically.

HCSE

Question 3

Is heteroskedasticity a big deal?

HCSE

Question 3

Is heteroskedasticity a big deal?

Answer to Question 3

No! Sure, even with these corrections are estimator may no longer be BLUE. However, these variance-covariance estimators are consistent which means we can do inference like usual with large enough samples so nothing really changes. We will see these heteroskedasticity corrections are super easy to implement in R.

- Whenever estimating a regression, you should **always** use a correction so you don't have to worry about heteroskedasticity in large samples.

HCSE

Question 4

What if we don't think heteroskedasticity is present in our data set?

HCSE

Question 4

What if we don't think heteroskedasticity is present in our data set?

Answer to Question 4

Just use the heteroskedasticity correction anyway because you could be wrong. The gains in more reliable inference outweigh the potential gains in efficiency.

Thank You!