Econometrics

Chapter 9: Heteroskedasticity

In Choi

Sogang University

What is Heteroskedasticity?

- The assumption of homoskedasticity means that $E(u_t^2 \mid all\ x's) = \sigma^2$ (σ^2 is a constant that does not change over i).
- If this is not true, that is if the variance of u_t is different for different values of the x's, then the errors are heteroskedastic.

What is Heteroskedasticity?

Example

(Cross-section consumption function)
y_i : household i's consumption
x_i : household i's income
Consider the regression equation

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

Then, $Var(y_i \mid x_i) = Var(u_i \mid x_i)$. It is highly likely that $Var(y_i \mid x_i)$ is large when x_i is large.

Why Worry About Heteroskedasticity?

- OLS is still unbiased and consistent, even if we do not assume homoskedasticity.

Example

Consider the model

$$y_t = \beta_0 + \beta_1 x_t + u_t, \; E(u_t) = 0 \text{ and } E(u_t^2) = \sigma_t^2.$$

Assume for simplicity that $\{x_t\}$ is a sequence of constants. Recall that

$$\hat{\beta}_1 - \beta_1 = \sum_{t=1}^{n} w_t u_t,$$

where $w_t = \frac{x_t - \bar{x}}{\sum_{t=1}^{n}(x_t - \bar{x})^2}$. Then, $E(\hat{\beta}_1) = \beta_1$ since

$$E(\hat{\beta}_1 - \beta_1) = E(\sum_{t=1}^{n} w_t u_t)$$

Why Worry About Heteroskedasticity?

- Furthermore, if we assume $\sigma_t^2 < M$ for all t,

$$
\begin{aligned}
Var(\hat{\beta}_1) &= E\left[\left(\hat{\beta}_1 - \beta_1\right)^2\right] \\
&= E\left[\left(\sum_{t=1}^{n} w_t u_t\right)^2\right] \\
&= \sum_{t=1}^{n} w_t^2 \left(E u_t^2\right) \\
&= \sum_{t=1}^{n} w_t^2 \sigma_t^2 \\
&= \frac{\sum_{t=1}^{n}(x_t - \bar{x})^2 \sigma_t^2}{\left(\sum_{t=1}^{n}(x_t - \bar{x})^2\right)^2} \leq \frac{M}{\sum_{t=1}^{n}(x_t - \bar{x})^2}.
\end{aligned}
$$

Thus, if $\sum_{t=1}^{n}(x_t - \bar{x})^2 \to \infty$ as n increases, the estimator is consistent.

Why Worry About Heteroskedasticity?

- The usual formula for standard errors should be different if we have heteroskedasticity. Thus, we can not use the usual t or F test for drawing inferences.

Robust Standard Errors

- Consider the simple linear regression model. Then, $Var(\hat{\beta}_1)$ can be estimated by

$$\frac{\sum_{t=1}^{n}(x_t - \bar{x})^2 \hat{u}_t^2}{\left(\sum_{t=1}^{n}(x_t - \bar{x})^2\right)^2}.$$

- Adjusted t-ratio is defined by

$$\frac{\hat{\beta}_1}{\sqrt{\frac{\sum_{t=1}^{n}(x_t - \bar{x})^2 \hat{u}_t^2}{\left(\sum_{t=1}^{n}(x_t - \bar{x})^2\right)^2}}}.$$

  This has a standard normal distribution in the limit. If this is used, hestroskedasticity is taken care of.

- Write

$$\sqrt{n}\left(\hat{\beta}_1 - \beta_1\right) = \left(\frac{1}{n}\sum_{t=1}^{n}(x_t - \bar{x})^2\right)^{-1} \frac{1}{\sqrt{n}}\sum_{t=1}^{n}(x_t - \bar{x})u_t.$$

The variance of $\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(x_t - \bar{x})u_t$ is $\frac{1}{n}\sum_{t=1}^{n}(x_t - \bar{x})^2\sigma_t^2$.

Robust Standard Errors

- By the WLLN,

$$\frac{1}{n}\sum_{t=1}^{n}(x_t-\bar{x})^2u_t^2-\frac{1}{n}\sum_{t=1}^{n}(x_t-\bar{x})^2\sigma_t^2\xrightarrow{p}0.$$

Since $u_t^2-\hat{u}_t^2\xrightarrow{p}0$, $\frac{1}{n}\sum_{t=1}^{n}(x_t-\bar{x})^2\sigma_t^2$ can be estimated by $\frac{1}{n}\sum_{t=1}^{n}(x_t-\bar{x})^2\hat{u}_t^2$.

- Thus, if $lim\frac{1}{n}\sum_{t=1}^{n}(x_t-\bar{x})^2=M$ and $lim\frac{1}{n}\sum_{t=1}^{n}(x_t-\bar{x})^2\sigma_t^2=L$,

$$\frac{\hat{\beta}_1-\beta_1}{\sqrt{\frac{\sum_{t=1}^{n}(x_t-\bar{x})^2\hat{u}_t^2}{\left(\sum_{t=1}^{n}(x_t-\bar{x})^2\right)^2}}}=\frac{\sqrt{n}\left(\hat{\beta}_1-\beta_1\right)}{\sqrt{\frac{\frac{1}{n}\sum_{t=1}^{n}(x_t-\bar{x})^2\hat{u}_t^2}{\left(\frac{1}{n}\sum_{t=1}^{n}(x_t-\bar{x})^2\right)^2}}}\xrightarrow{d}\frac{N(0,M^{-2}L)}{\sqrt{\frac{L}{M^2}}}=N(0,1)$$

Robust Standard Errors

- Important to remember that these robust standard errors only have asymptotic justification. In small samples, the t-statistic formed with the robust standard error may not have a distribution close a standard normal distribution, and inferences may not not be correct.

Generalized least squares (GLS) estimation

- Model

$$y_t = \beta_0 + \beta_1 x_{1t} + \cdots + \beta_k x_{kt} + u_t.$$

If $Var(u_t) = \sigma_t^2$, consider the transformed model

$$y_t/\sigma_t = \beta_0/\sigma_t + \beta_1 x_{1t}/\sigma_t + \cdots + \beta_k x_{kt}/\sigma_t + u_t/\sigma_t.$$

- The error term u_t/σ_t satisfy classical assumptions. Thus, the OLS estimator from this transformed model is BLUE. This estimator is called the generalized least squares (GLS) estimator.
- In practice, σ_t is not known.