Heteroscedasticity

In this case

$$Y_i = \alpha + \beta X_i + u_i$$

we assume

$$E(u_i) = 0$$
$$E(u_i^2) = \sigma_i^2$$
$$E(u_i u_j) = 0$$

So in this case the errors do not have a common variance.

An example of this is if regressing firm size on previous size across a sample of firms. Smaller firms tend to have larger variance of growth than larger firms, so the variance of size will decline as size increases.

The effect of this will be to leave the OLS estimator of β unbiased, but the estimated standard error will be biased.

Tests

1. Can check for heteroscedasticity by ordering the residuals by their x values. If the residuals increase as the x values increase then there is a problem.
2. Tests using residuals:

   Reset test: regress $\hat{u}_t$ on powers of Y i.e. $Y_t^2, Y_t^3, ..$ and test the significance of the coefficients.

   White test: regress $\hat{u}_t^2$ on all explanatory variables their squares and their cross products and test the significance of their coefficients

   Glejser test: regress $|\hat{u}_t|$ on X_t using different functional forms
3. Tests of differences between groups of observations: Goldfeld and Quandt test: split into two groups, one with large value of X and one with small, fit separate regressions and then do an F test on the equality of the error variances

   Can also leave some of the middle observations out to increase ability to discriminate. There is a more general likelihood ratio test for more than two groups when the number of observations is large.

Solutions

Use deflators:

- Transforming to logs often reduces heteroscedasticity
- Could deflate the variables by the variables that the variances vary with if you know it eg if in

$$Y_i = \alpha + \beta X_{1i} + \gamma X_{2i} + u_i$$

we know

$$var(u_i) = \sigma^2 X_{1i}$$

then we can deflate to get the unbiased standard errors and t tests

$$\frac{Y_i}{\sqrt{X_{1i}}} = \frac{\alpha}{\sqrt{X_{1i}}} + \beta + \gamma \frac{X_{2i}}{\sqrt{X_{1i}}} + v_i$$

- More generally if we knew the variances up to a multiplicative constant, such as $var(u_i) = \sigma^2 Z_i^2$ then we could we could divide the regression by the Z_i. So for

$$Y_i = \alpha_0 + \alpha_1 X_i + u_i$$

$$\frac{Y_i}{Z_i} = \frac{\alpha_0}{Z_i} + \alpha_1 \frac{X_i}{Z_i} + \frac{u_i}{Z_i}$$

then we estimate

$$\frac{Y_i}{Z_i} = \beta_0 \frac{1}{Z_i} + \beta_1 \frac{X_i}{Z_i} + w_i$$

then we will have $E(w_i) = 0$ and

$$E(w_i^2) = E\left(\left(\frac{u_i}{Z_i}\right)^2\right) = E\left(\frac{u_i^2}{Z_i^2}\right)$$

now as Z_i is non stochastic

$$E(w_i^2) = \frac{E(u_i^2)}{Z_i^2}$$

now as $E(u_i^2) = var(u_i) = \sigma^2 Z_i^2$

$$E(w_i^2) = \frac{\sigma^2 Z_i^2}{Z_i^2} = \sigma^2$$

so it is the unbiased estimator of the variance.

- Note that the purpose is to get efficient estimates of the variances by removing heteroscedasticity. Once you have these and have estimated the t ratios you should make all inferences on the original equation. Not the transformed one.
- If you are not sure then can at least try to correct the standard errors, using the White Heteroscedastic robust errors available in Microfit. This takes

$$Var(\hat{\beta}) = \frac{\sum x_i^2 \sigma_i^2}{(\sum x_i^2)^2}$$

and substitutes $\hat{u}_i^2$ for σ_i^2

$$Var(\hat{\beta}) = \frac{\sum x_i^2 \hat{u}_i^2}{(\sum x_i^2)^2}$$