MGMG 522 : Session #8

Heteroskedasticity

(Ch. 10)

8-1

Heteroskedasticity

- Heteroskedasticity is a violation of the classical assumption #5 (observations of the error term are drawn from a distribution that has a constant variance). (See Figure 10.1 on p. 348)
- Heteroskedasticity can be categorized into 2 kinds:
  - Pure heteroskedasticity (Heteroskedasticity that exists in a correctly specified regression model).
  - Impure heteroskedasticity (Heteroskedasticity that is caused by specification errors: omitted variables or incorrect functional form).
- Heteroskedasticity mostly happens in a cross-sectional data set. This, however, doesn't mean that time-series data cannot have heteroskedasticity.

8-2

Pure Heteroskedasticity

- Classical assumption #5 implies that the observations of the error term are drawn from a distribution with a constant variance, or $VAR(\varepsilon_i) = \sigma^2$.
- In other words, the error term must be homoskedastic.
- For a heteroskedastic error term, $VAR(\varepsilon_i) = \sigma_i^2$.
- Heteroskedasticity can take on many forms but we will limit our discussion to one form of heteroskedasticity only.

Heteroskedasticity of the Form: $VAR(\varepsilon_i) = \sigma^2 Z_i^2$

- Z is called a proportionality factor because the variance of the error term changes proportionally to the Z factor (or variable). (See Figure 10.2 on p. 350 and Figure 10.3 on p. 351)
- The variance of the error term increases as Z increases.
- Z may or may not be one of the explanatory variables in the regression model.
- Heteroskedasticity can occur in at least 2 situations:

1. When there is significant change in the dependent variable of a <u>time-series</u> model.
2. When there are different amounts of measurement errors in the sample of different periods or different sub-samples.

Impure Heteroskedasticity

- Caused by specification errors, especially the omitted variables.
- Incorrect functional form is less likely to cause heteroskedasticity.
- Specification errors should be corrected first by way of investigating the independent variables and/or functional form.

8-5

How can omitted variables cause heteroskedasticity?

- Correct model: $Y = \beta_0+\beta_1X_1+\beta_2X_2+\varepsilon$
- If X_2 is omitted: $Y = \beta_0+\beta_1X_1+\varepsilon^*$
- Where, $\varepsilon^* = \beta_2X_2+\varepsilon$
- If ε is small compared to β_2X_2, and X_2 is heteroskedastic, the ε^* will also be heteroskedastic.
- Both the bias and impure heteroskedasticity will disappear once the model gets corrected.

8-6

Consequences of Heteroskedasticity

1. Pure heteroskedasticity does not cause bias in the coefficient estimates.
2. Pure heteroskedasticity increases variances of the coefficient estimates.
3. Pure heteroskedasticity causes OLS to underestimate the standard errors of the coefficients. (Hence, pure heteroskedasticity overestimates the t-values.)

8-7

Detection of Heteroskedasticity

- Because heteroskedasticity can take on many forms, therefore, there is no specific test to test for heteroskedasticity.
- However, we will discuss two tests as a tool for detecting heteroskedasticity of a certain form.
- The two tests are (1) Park Test, and (2) White Test.

8-8

Detection of Heteroskedasticity: Caveats

- Before you go on and conduct one or both of these tests, ask yourself these 3 questions:

1. Is your model already correctly specified? If the answer is "No", correct your model first.
2. Does the topic you are studying have a known heteroskedasticity problem as pointed out by other researchers?
3. When you plot a graph of the residuals against the Z factor, does it indicate a potential heteroskedasticity problem?

8-9

The Park Test

- The Park test tests for heteroskedasticity of the form, $VAR(\varepsilon_i) = \sigma^2 Z_i^2$.
- One difficulty with the Park test is the specification of the Z factor.
- The Z factor is usually one of the explanatory variables, but not always. The Z factor could very well be other variable not included in the regression model.

8-10

3 Steps of the Park Test

1. Run the OLS and get the residuals: $e_i = Y_i - b_0 - b_1 X_{1i} - b_2 X_{2i}$
2. Regress the $\ln(e)^2$ on the $\ln(Z)$: $\ln(e_i^2) = b_0 + b_1 \ln(Z_i)$
3. Check to see if b_1 is significant or not.
   – If b_1 is significant (meaning $b_1 \neq 0$), this implies heteroskedasticity.
   – If b_1 is not significant (meaning $b_1 = 0$), this implies that heteroskedasticity is unlikely. (However, it may still exist in other form or exist with other Z factor.)

8-11

The White Test

- In stead of having to specify the Z factor and heteroskedasticity of the form $VAR(\varepsilon_i) = \sigma^2 Z_i^2$, the White test takes a more general approach.

8-12

3 Steps in the White Test

1. Run the OLS and get the residuals:
   $e_i = Y_i - b_0 - b_1X_{1i} - b_2X_{2i} - b_3X_{3i}$
2. Regress e^2 on each X, square of each X, cross product of the Xs. For a 3-independent variable model, run OLS of

$$e_i^2 = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_1^2 + b_5X_2^2$$
$$+ b_6X_3^2 + b_7X_1X_2 + b_8X_1X_3 + b_9X_2X_3 + u_i$$

3. Test the overall significance of the model above with χ^2 test. The test statistic is nR^2.
   – If nR^2 > critical χ^2 value, this implies <u>hetero</u>skedasticity.
   – If nR^2 < critical χ^2 value, this implies that <u>hetero</u>skedasticity is unlikely.

◆ n = number of observations, R^2 = the value of the unadjusted R^2, the DF for the critical χ^2 value is the number of the independent variables in the model above.

8-13

Correcting Heteroskedasticity

◆ If heteroskedasticity still is present in your correctly specified regression model (pure heteroskedasticity), consider one of these three remedies.

1. Use Weighted Least Squares
2. Use Heteroskedasticity-Corrected Standard Errors
3. Redefine the variables (See an example on p. 366-369)

8-14

Weighted Least Squares

- Take an original model with a heteroskedastic error term,

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i \rightarrow (\text{Eq.}1)$$

- Suppose the variance is of the form,

$$\text{VAR}(\varepsilon_i) = \sigma_i^2 = \sigma^2 Z_i^2$$

- Eq.1 can be shown to be equal to,

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + Z_i u_i \rightarrow (\text{Eq.}2)$$

See 10-6 on p. 598 for proof.

- Divide Eq.2 with Z_i, obtaining

$$\frac{Y_i}{Z_i} = \frac{\beta_0}{Z_i} + \beta_1 \frac{X_{1i}}{Z_i} + \beta_2 \frac{X_{2i}}{Z_i} + u_i \rightarrow (\text{Eq.}3)$$

Weighted Least Squares (Con..)

- If Z happens to be one of the independent variables, use OLS with Eq.3 specification to obtain the coefficient estimates.
- If Z is not one of the independent variables, we must add an intercept term because Eq.3 has no intercept term. We will run OLS using Eq.4 specification.

$$\frac{Y_i}{Z_i} = \alpha_0 + \frac{\beta_0}{Z_i} + \beta_1 \frac{X_{1i}}{Z_i} + \beta_2 \frac{X_{2i}}{Z_i} + u_i \rightarrow (\text{Eq.}4)$$

- The interpretation of the slope coefficients becomes a bit tricky.
- There are two problems with WLS: (1) what is Z?, and (2) how Z relates to VAR(ε)?

White's Heteroskedasticity-Corrected Standard Errors (a.k.a. Robust Standard Errors)

- The idea of this remedy goes like this.
- Since, there is no bias in the coefficient estimates.
- But, the standard errors of the coefficient estimates are larger with heteroskedasticity than without it.
- Therefore, why not fix the standard errors and leave the coefficient estimates alone?
- This method is referred to as HCCM (heteroskedasticity-consistent covariance matrix).
- In EViews, you will choose "LS" and click on "Options", then select "Heteroskedasticity-Consistent Coefficient Covariance" and select "White".
- The standard errors of the coefficient estimates can be bigger or smaller than those from the OLS.
- This method works best in a large sample data.

Redefine the Variables

1. For example, take log of some variables.
2. Or, normalize some variables.