Economics 522A, Spring 2007

Lecture Note 15: Testing for Heteroskedasticity

1 White's Test

In the previous note, we obtained a simple estimator for the variance of $\hat{\beta}$, under the assumption that (y_i, x_i) were IID with $E[y_i|x_i] = x_i'\beta$. So we could form standard errors without assuming homoskedasticity, and compare them to the standard errors that assume homoskedasticity, and informally test whether there is heteroskedasticity. White (1980) derives a formal test for heteroskedasticity based on comparing the two estimates of the variance matrix of $\hat{\beta}$.

The robust standard errors are only valid asymptotically, and their performance for small samples may not be very good. Next, we discuss a few other tests for heteroskedasticity which make somewhat stronger assumptions on the *form* of the heteroskedasticity. In some cases, the tests are exact (do not require large samples).

2 Two-Sample Tests

Suppose we have a sample of observations y_i for $i = 1, \ldots, n$, and no covariates (other than a constant).

Also, suppose that the sample naturally splits into two groups. We order the observations so that

- Group 1: $i = 1, 2, \ldots, n_1$.
- Group 2: $i = n_1 + 1, n_1 + 2, \ldots, n_1 + n_2 = n$.

We will assume normality:

$$y_i \overset{\text{i.i.d.}}{\sim} N(\mu_1, \sigma_1^2), \quad i = 1, \ldots, n_1;$$

$$y_i \overset{\text{i.i.d.}}{\sim} N(\mu_2, \sigma_2^2), \quad i = n_1 + 1, \ldots, n_1 + n_2.$$

We would like to test whether the two variances are equal.

We could estimate the parameters separately for each group (say by least squares):

$$\hat{\mu}_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} y_i, \quad s_1^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (y_i - \hat{\mu}_1)^2;$$

$$\hat{\mu}_2 = \frac{1}{n_2} \sum_{i=n_1+1}^{n_1+n_2} y_i, \quad s_2^2 = \frac{1}{n_2 - 1} \sum_{i=n_1+1}^{n_1+n_2} (y_i - \hat{\mu}_2)^2;$$

By our earlier results,

$$\frac{(n_1 - 1)}{\sigma_1^2} \cdot s_1^2 \sim \chi^2_{n_1-1};$$

$$\frac{(n_2 - 1)}{\sigma_2^2} \cdot s_2^2 \sim \chi^2_{n_2-1}.$$

So, under $H_0 : \sigma_1^2 = \sigma_2^2$,

$$\frac{s_1^2}{s_2^2} \sim \frac{\chi^2_{n_1-1}/(n_1-1)}{\chi^2_{n_2-1}/(n_2-1)} \sim F_{n_1-1,n_2-1}.$$

The statistic s_1^2/s_2^2 is pivotal: its distribution does not depend on the value of σ_1^2 or σ_2^2 under the null hypothesis that the two variances are equal. So we can test the null hypothesis using the standard tables for the F distribution.

Suppose we have additional regressors x_i (a $k \times 1$ vector), but the observations still split into two groups as before. If we assume that

$$y_i|X \overset{\text{ind.}}{\sim} N(x_i'\beta_1, \sigma_1^2), \quad i = 1, \ldots, n_1,$$

and

$$y_i|X \overset{\text{ind.}}{\sim} N(x_i'\beta_2, \sigma_2^2), \quad i = n_1 + 1, \ldots, n_1 + n_2,$$

then we could run separate least squares regressions in the two subsamples to get

$$\hat{\beta}_1, s_1^2, \quad \hat{\beta}_2, s_2^2.$$

Now, we will have

$$\frac{(n_1 - k)}{\sigma_1^2} \cdot s_1^2 \sim \chi^2_{n_1-k};$$

$$\frac{(n_2 - k)}{\sigma_2^2} \cdot s_2^2 \sim \chi^2_{n_2-k}.$$

So, under $H_0 : \sigma_1^2 = \sigma_2^2$,

$$\frac{s_1^2}{s_2^2} \sim \frac{\chi^2_{n_1-k}/(n_1-k)}{\chi^2_{n_2-k}/(n_2-k)} \sim F_{n_1-k,n_2-k}.$$

3 Goldfeldt-Quandt Test

Suppose that:

$$y_i|X \overset{\text{ind.}}{\sim} N(x_i'\beta, \sigma^2(x_i)),$$

where the conditional variance $\sigma^2(x_i)$ is assumed to be monotonic in one of the elements of x_i, say x_{ij}.

Then, we could order observations by x_{ij}, and compare the variance estimated from the high-x_{ij} group to the variance estimated from the low-x_{ij} group, similar to the two-sample approach.

More formally, suppose that we take two subsamples from the whole sample:

- Group 1: individuals with $x_{ij} \leq M_1$.
- Group 2: individuals with $x_{ij} \geq M_2$.

In practice, it can be better to set $M_1 < M_2$, so that the "middle" observations are not used for the test. A common approach is to split the data into thirds based on x_{ij}, and compare the top third to the bottom third.

4 Parametrizing the Alternative: Breusch-Pagan Score Test

Suppose we assume

$$y_i|X \overset{\text{ind.}}{\sim} N(x_i'\beta, \sigma^2(x_i)),$$

and go further by assuming a particular functional form for $\sigma^2(x_i)$. Two common choices are:

$$\sigma^2(x_i) = x_i'\gamma,$$

and

$$\sigma^2(x_i) = \exp(x_i'\delta).$$

For example, let us suppose that

$$y_i|X \overset{\text{ind.}}{\sim} N(x_i'\beta, x_i'\gamma).$$

To make the notation clearer, suppose the first element of x_i is a constant, and let z_i be all the elements of x_i other than the constant. Then we could write

$$y_i|X \overset{\text{ind.}}{\sim} N(x_i'\beta, \gamma_1 + z_i'\gamma_2).$$

Now we have a fully specified conditional likelihood, and if we want we can estimate $\beta, \gamma_1, \gamma_2$ by conditional MLE. We could then test whether $\gamma_2 = 0$ to see whether there is conditional heteroskedasticity. We could use Wald, LR, or score/LM tests.

Breusch and Pagan (1979) and Godfrey (1978) suggested testing $H_0 : \gamma_2 = 0$ vs $H_1 : \gamma_2 \neq 0$ using a score/Lagrange multiplier test. The advantage is that we do not actually have to estimate γ_2, because the score test only requires that we calculated the restricted estimate. Here, the restricted model is:

$$y_i|X \overset{\text{ind.}}{\sim} N(x_i'\beta, \gamma_1),$$

which is just the homoskedastic conditionally normal model with $\gamma_1 = \sigma^2$. The Breusch-Pagan score test turns out to have a convenient representation:

1. Compute $\hat{\beta}$ and $\hat{\sigma}^2$ by MLE:

$$\hat{\beta} = (X'X)^{-1}X'y, \quad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} e_i^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - x_i'\hat{\beta})^2.$$

2. Regress $e_i^2/\hat{\sigma}^2$ on a constant and z_i, and form the explained sum of squares (the numerator term in R^2) from this second regression.

The score statistic can be shown to be equal to 1/2 the explained sum of squares from the second regression.

For a derivation of this result, see Ruud 18.3.2 and 18.7.