Research Paper
Impact Factor: 3.567
Peer Reviewed Journal

IJMDRR
E- ISSN –2395-1885
ISSN -2395-1877

HETEROSCADASTICITY – TESTS OF HETEROSCADASTICITY

Mr. P.Devendran
Lecturer, Vijayam Degree College, Chittor.

Recall that OLS makes the assumption that $V_j(\varepsilon) = \sigma^2$ for all j. That is, the variance of the error term is constant. (Homoskedasticity). If the error terms do not have constant variance, they are said to be heteroskedastic. [Tidbit from Wikipedia: The term means "differing variance" and comes from the Greek "hetero" ('different') and "skedasis" ('dispersion').] When heteroskedasticity might occur. Errors may increase as the value of an Independent Variable increases.

For example, consider a model in which annual family income is the Independent Variable and annual family expenditures on vacations is the Dependent Variable. Families with low incomes will spend relatively little on vacations, and the variations in expenditures across such families will be small. But for families with large incomes, the amount of discretionary income will be higher. The mean amount spent on vacations will be higher, and there will also be greater variability among such families, resulting in heteroskedasticity. Note that, in this example, a high family income is a necessary but not sufficient condition for large vacation expenditures. Any time a high value for an Independent variable is a necessary but not sufficient condition for an observation to have a high value on a Dependent Variable, heteroskedasticity is likely. Similar examples: Error terms associated with very large firms might have larger variances than error terms associated with smaller firms. Sales of larger firms might be more volatile than sales of smaller firms.

Heteroscedasticity

One of the key assumptions of regression is that the variance of the errors is constant across observations. If the errors have constant variance, the errors are called *homoscedastic*. Typically, residuals are plotted to assess this assumption. Standard estimation methods are inefficient when the errors are *heteroscedastic* or have non-constant variance.

Detecting Heteroskedasticity

Visual Inspection: Do a visual inspection of residuals plotted against fitted values; or, plot the Independent Variable suspected to be correlated with the variance of the error term. In Stata, after running a regression, you could use the rvfplot (residuals versus fitted values) or rvpplot command (residual versus predictor plot, e.g. plot the residuals versus one of the X variables included in the equation). In SPSS, plots could be specified as part of the Regression command.

- In a large sample, you'll ideally see an "envelope" of even width when residuals are plotted against the IV. In a small sample, residuals will be somewhat larger near the mean of the distribution than at the extremes. Thus, if it appears that residuals are roughly the same size for all values of X (or, with a small sample, slightly larger near the mean of X) it is generally safe to assume that heteroskedasticity is not severe enough to warrant concern.
- If the plot of residuals shows some uneven envelope of residuals, so that the width of the envelope is considerably larger for some values of X than for others, a more formal test for heteroskedasticity should be conducted.

Heteroscedasticity Tests

The MODEL procedure now provides two tests for heteroscedasticity of the errors: White's test and the modified Breusch-Pagan test.

Both White's test and the Breusch-Pagan are based on the residuals of the fitted model. For systems of equations, these tests are computed separately for the residuals of each equation.

The residuals of an estimation are used to investigate the heteroscedasticity of the true disturbances.
The WHITE option tests the null hypothesis

$$H_0 : \sigma_i^2 = \sigma^2 \quad \text{for all } i$$

White's test is general because it makes no assumptions about the form of the heteroscedasticity (White 1980). Because of its generality, White's test may identify specification errors other than heteroscedasticity. Thus White's test may be significant when the errors are homoscedastic but the model is misspecified in other ways.
White's test is equivalent to obtaining the error sum of squares for the regression of the squared residuals on a constant and all the unique variables in $\mathbf{J} \otimes \mathbf{J}$, where the matrix **J** is composed of the partial derivatives of the equation residual with respect to the estimated parameters.

IJMDRR
E- ISSN –2395-1885
ISSN -2395-1877

The modified Breusch-Pagan test assumes that the error variance varies with a set of regressors, which are listed in the BREUSCH= option.

Define the matrix *Z* to be composed of the values of the variables listed in the BREUSCH= option, such that $z_{i,j}$ is the value of the *j*th variable in the BREUSCH= option for the *i*th observation. The null hypothesis of the Breusch-Pagan test is

$$H_0 : \sigma_i^2 = \sigma^2(\alpha_0 + \alpha' \mathbf{z_i})$$

where σ_i^2 is the error variance for the *i*th observation, and α_0 and α are regression coefficients.
The test statistic for the Breusch-Pagan test is

$$b_p = \frac{1}{v}(\mathbf{u} - \bar{u}\mathbf{i})' Z(Z'Z)^{-1}Z'(\mathbf{u} - \bar{u}\mathbf{i})$$

where $\mathbf{u} = (e_1^2, e_2^2, \ldots, e_n^2)$, $\mathbf{i}$ is a $n \times 1$ vector of ones, and

$$v = \frac{1}{n}\sum_{i=1}^{n}\left(e_i^2 - \frac{\mathbf{e}'\mathbf{e}}{n}\right)^2$$

This is a modified version of the Breusch-Pagan test, which is less sensitive to the assumption of normality than the original test .

Breusch-Pagan TEST:
The Breusch-Pagan (BP) test is one of the most common tests for heteroskedasticity. It begins by allowing the heteroskedasticity process to be a function of one or more of your independent variables, and it’s usually applied by assuming that heteroskedasticity may be a linear function of all the independent variables in the model. This assumption can be expressed as

$$\varepsilon_i^2 = \alpha_0 + \alpha_1 X_{i1} + \ldots + \alpha_p X_{ip} + u_i$$

The values for ε_i^2 aren’t known in practice, so the $\hat{\varepsilon}_i^2$

are calculated from the residuals and used as proxies for ε_i^2 Generally, the BP test is based on the estimation of

$$\hat{\varepsilon}_i^2 = \alpha_0 + \alpha_1 X_{i1} + \ldots + \alpha_p X_{ip} + u_i$$

Alternatively, a BP test can be performed by estimating

$$\hat{\varepsilon}_i^2 = \delta_0 + \delta_1 \hat{Y}_i \text{ where } \hat{Y} \text{ represents the predicted values from}$$

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \ldots + \hat{\beta}_p X_{ip}$$

Here’s how to perform a BP test:

1. Estimate your model using OLS:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \ldots + \beta_p X_{ip} + \varepsilon_i$$

2. Obtain the predicted *Y* values after estimating the model.
3. Estimate the auxiliary regression using OLS:

$$\hat{\varepsilon}_i^2 = \delta_0 + \delta_1 \hat{Y}_i$$

4. From this auxiliary regression, retain the R-squared value:

$$R^2_{\hat{\varepsilon}^2}$$

IJMDRR
E- ISSN –2395-1885
ISSN -2395-1877

5. Calculate the *F*-statistic or the chi-squared statistic:

$$F = \frac{\frac{R^2_{\hat{\varepsilon}^2}}{1}}{\frac{\left(1-R^2_{\hat{\varepsilon}^2}\right)}{n-2}} \text{ or } \chi^2 = nR^2_{\hat{\varepsilon}^2}$$

The degrees of freedom for the *F*-test are equal to 1 in the numerator and $n – 2$ in the denominator. The degrees of freedom for the chi-squared test are equal to 1. If either of these test statistics is significant, then you have evidence of heteroskedasticity. If not, you fail to reject the null hypothesis of homoskedasticity.

White test:

In econometrics, an extremely common test for heteroskedasticity is the White test, which begins by allowing the heteroskedasticity process to be a function of one or more of your independent variables. It's similar to the Breusch-Pagan test, but the White test allows the independent variable to have a nonlinear and interactive effect on the error variance.

As in the Breusch-Pagan test, because the values for ε_i^2 aren't known in practice, the $\hat{\varepsilon}_i^2$ are calculated from the residuals and used as proxies for ε_i^2 The White test is based on the estimation of the following:

$$\hat{\varepsilon}_i^2 = \alpha_0 + \alpha_1 X_{i1} + \ldots + \alpha_p X_{ip} + \alpha_{p+1} X_{ip}^2 + \ldots \alpha_{2p} X_{ip}^2 + \alpha_{2p+1}\left(X_{i1}X_{i2}\right) + \ldots + u_i$$

Alternatively, a White test can be performed by estimating

$\hat{\varepsilon}_i^2 = \delta_0 + \delta_1 \hat{Y}_i + \delta_2 \hat{Y}_i^2$ where $\hat{Y}_i$ represents the predicted values from

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \ldots + \hat{\beta}_p X_{ip}$$

Follow these five steps to perform a White test:

1. Estimate your model using OLS:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \ldots + \beta_p X_{ip} + \varepsilon_i$$

2. Obtain the predicted *Y* values after estimating your model.
3. Estimate the model using OLS:

$$\hat{\varepsilon}_i^2 = \delta_0 + \delta_1 \hat{Y}_i + \delta_2 \hat{Y}_i^2$$

4. Retain the R-squared value from this regression: $R^2_{\hat{\varepsilon}^2}$
5. Calculate the F-statistic or the chi-squared statistic:

$$F = \frac{\frac{R^2_{\hat{\varepsilon}^2}}{1}}{\frac{\left(1-R^2_{\hat{\varepsilon}^2}\right)}{n-2}} \text{ or } \chi^2 = nR^2_{\hat{\varepsilon}^2}$$

The degrees of freedom for the *F*-test are equal to 2 in the numerator and $n – 3$ in the denominator. The degrees of freedom for the chi-squared test are 2. If either of these test statistics is significant, then you have evidence of heteroskedasticity. If not, you fail to reject the null hypothesis of homoskedasticity.