MODIFIED TESTS FOR HETEROSCEDASTIC VARIANCES IN LINEAR STATISTICAL MODELS

Bandi Ramanjineyulu

Manager, Clinical Statistics, GlaxoSmithKline, Bangalore, Karnataka, India

Email id : ramanji.bandi@gmail.com

ABSTRACT

Heteroscedasticity is potentially a serious problem for research workers in statistics. The Literature on Inferential Problems for Linear Statistical, models under heteroscedasticity has grown enormously in the past three decades. The presence of heteroscedastic errors disturbs the optimal properties of OLS estimators of parameters of linear statistical model. In this case, the OLS estimators are no longer efficient estimators. There is a need of detecting the existence of the heteroscedastic errors in the linear model.

In the present study, an attempt has been made by proposing some test procedures for the linear statistical models with heteroscedastic errors.

I. INTRODUCTION:

One of the standard assumptions of the classical linear statistical model is that the disturbances are homoscedastic; i.e., all observations have equal variances. However, in many practical applications, it has been found that the disturbances are heteroscedastic. This is especially true for cross-sectional data, where significant variation exists in the size distribution of the units. Thus, Prais and Houthakker (1955) has found the problem of Heteroscedasticity in family budget studies utilizing observations on individuals with diverse incomes and family sizes.

The presence of heteroscedastic errors disturbs the optimal properties of ordinary least squares estimators of parameters of linear regression model. It is well known that if the errors are heteroscedastic, the ordinary least squares estimates of the regression coefficients are inefficient and the standard method of inference may produce misleading conclusions.

Thus, detecting the existence of heteroscedasticity and estimating the parameters of linear regression model under the problem of heteroscedasticity are of utmost importance.

A large spectrum of tests and estimation procedures for heteroscedasticity have been developed with a view to quickening of interest in the last few years.

The main contributions relating to the inference in linear regression models with heteroscedastic errors have been made by Glodfeld and Quandt (1965), Ramsey (1969), Glejser (1969), Lahiri and Egy (1981), Evans and King (1984), Conerly and Mansfield (1988), Greene (1997) and Nachane (2006) and others.

In the present research study, an attempt has been made by proposing some new testing procedures for the linear regression models with heteroscedastic disturbances.

II. TESTING EQUALITY OF REGRESSION COEFFICIENTS IN TWO LINEAR REGRESSION MODELS UNDER HETEROSCEDASTICITY USING STUDENTIZED RESIDUALS:

Consider two linear regression models for two regressions as

$$Y_1 = X_1\beta_1 + \epsilon_1,\ \epsilon_1 \sim N(0, \sigma_1^2 I_{n_1}) \tag{2.1}$$

$$Y_2 = X_2\beta_2 + \epsilon_2,\ \epsilon_2 \sim N(0, \sigma_2^2 I_{n_2}) \tag{2.2}$$

Where, Y_i is $(n_i \times 1)$ vector of observations on i^{th} dependent Variable;

X_i is $(n_i \times k)$ matrix of observations on k independent variables for i^{th} region;

β_i is $(k \times 1)$ vector parameters in i th region;

$\in_i$ is $(n_i \times 1)$ vector of disturbances for i th region; i = 1, 2.

By taking, $Y^1 = (Y_1^1, mY_2^1)^1, X = (X_1^1, mX_2^1)^1$,

$$Q = (0^1, mX_2^1)^1, D = (X, Q), \eta = (\beta_1^1, \gamma^1)^1,$$

$\gamma = \beta_2 - \beta_1, m = \dfrac{\sigma_1}{\sigma_2}$ and $\in = (\in_1^1, \in_2^1)^1$, the two linear regression models (2.1) and (2.2) can be compactly rewritten as

$$Y_{(n_1+n_2)\times 1} = D_{(n_1+n_2)\times 2k}\ \eta_{2k\times 1} + \in_{(n_1+n_2)\times 1} \tag{2.3}$$

To test for the equality of regression coefficients of two linear models (2.1) and (2.2) or to test for the constancy of two regressions of (2.1) and (2.2) under heteroscedasticity, one may state the null hypothesis as $H_0 : \gamma = 0$ against the alternative hypothesis $H_0 : \gamma \neq 0$.

By applying Least Squares Estimation to the models (2.1) and (2.2) the OLS residual vectors are given by

$$e_i = \left[I - X_i\ (X_i^1 X_i)\ X_i^1 \right] Y_i, i = 1, 2 \tag{2.4}$$

Consider the maximum likelihood estimators for σ_1^2 and σ_2^2 as

$$\hat{\sigma}_i^2 = \frac{e_i^1 e_i}{n_i}, i = 1, 2 \tag{2.5}$$

and hence, an estimate for m is given by

$$\hat{m}^2 = \left(\frac{\hat{\sigma}_1}{\hat{\sigma}_2} \right) \tag{2.6}$$

Define the sum of squares due to internally studentized residuals under H_0 as $e_1^{*1} e_1^*$ and unrestricted sum of squares due to internally studentized residuals as

$$e_{UR}^{*1}\ e_{UR}^{*} = \left[e_1^{*1} e_1^{*} + \hat{m}^2 e_2^{*1} e_2^{*} \right]$$

Here e_1^* and e_2^* are based on models (2.1) and (2.2) respectively.

To test, H_0 the proposed F-test statistic is given by

$$F = \left[\frac{\left[e_1^{*1} e_1^{*} - e_{UR}^{*1}\ e_{UR}^{*} \right] / k}{e_{UR}^{*1}\ e_{UR}^{*} / (n_1 + n_2 - 2k)} \right] \qquad (2.7)$$

This F- statistics follows F-distribution with $(k, n_1 + n_2 - 2k)$ degrees of freedom under H_0.

Remarks: The F- criterion given (2.7) can be obtained as the likelihood ratio test statistic $(-2 \log \lambda)$ as follows.

$$-2 \log \lambda = -\,2n_2 \log \left[\frac{\hat{m}_0}{\hat{m}_1} \right] + (n_1 + n_2) \log \left[\frac{e_1^{*1} e_1^{*}}{e_D^{*1} e_D^{*}} \right] \qquad (2.8)$$

Where $e_D^{*1} e_D$ is obtained by using the model (2.3) $\hat{m}_0$ *and* $\hat{m}_1$ are the maximum likelihood estimates of m under H_0 *and* H_1 respectively.

It can be shown that $-2\log\lambda$ follows χ^2- distribution with K degress of freedom as $n_2 \to \alpha$.

III. TESTING EQUALITY OF REGRESSION COEFFICIENTS IN SETS OF LINEAR REGRESSION MODELS UNDER HETEROSCEDASTICITY USING STUDENTIZED RESIDUALS:

Consider a set of p Linear models as

$$Y_i = X_i\ \beta_i + \in_i, \qquad i = 1, 2, ...p \qquad (3.1)$$

Y_i is $(n_i \times 1); \in_i = X_i$ is $(n_i \times 1)$; X_i is $(n_i \times k)$; β is $(k \times 1)$ matrices

$$\text{Also}, \in_i \sim N\left[0, \sigma_i^2\ In_i \right] \qquad i = 1, 2, ...p \qquad (3.2)$$

ISSN 2515-8260 Volume 07, Issue 02, 2020

It is assumed that $k < \underset{i}{\mathrm{Min}}\{n_i\}$ and Rank $(X_i) = k, \forall i = 1, 2, \ldots, p$

Under the proposed method, first estimate (3.1) and find the OLS residual vectors e_i's and hence obtain the internally studentized residual vectors as e_i^*'s. Now, consider

$$S_i^{*2} = \left[\frac{e_i^{*1} e_i^*}{X_i - k}\right] \tag{3.3}$$

as an independent estimate of σ_i^2 with $(n_i - k)$ degrees of freedom.

To test for the homoscedasticity one may state the null hypothesis as $H_0 = \sigma_i^2 = \sigma_2^2 = \ldots.. = \sigma_p^2$

A modified Bartlett's test for equality of variances may be used as

$$B_M = \left[\sum_{i=1}^{p} n_i - k\right]\left[\log \bar{S}^{*2} - \bar{C}\right] \tag{3.4}$$

where $$\bar{C} = \left[\sum_{i=1}^{p}(n_i - k)\, C_i \Big/ \sum_{i=1}^{p}(n_i - k)\right] \tag{3.5}$$

$$C_i = \log S_i^{*2} \tag{3.6}$$

$$\bar{S}^{*2} = \left[\frac{\sum_{i=1}^{p}(n_i - k)\, S_i^{*2}}{\sum_{i=1}^{p}(n_i - k)}\right] \tag{3.7}$$

Under H0, the test statistic B_M follows $\chi^2_{(p-1)}$. If H_0 is rejected then one may test the equality between the regression coefficients in linear regression models by using the following modified F-test based on internally studentized residual vectors.

Write,

$$X = [Z\ U];\ X_i = [Z_i\ U_i];$$

$$\beta^1 = (\gamma^1, \delta^1); \quad \beta_i^1 = (\gamma_i^1, \delta_p^1)$$

$$\gamma^1 = (\gamma_1^1, \gamma_2^1, \ldots\ldots.., \gamma_p^1) \quad \delta^1 = (\delta_1^1, \delta_2^1, \ldots., \delta_p^1)$$

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ . \\ . \\ Y_p \end{bmatrix}, Z = \begin{bmatrix} Z_1 & 0 & . & . & 0 \\ 0 & Z_2 & . & . & 0 \\ . & . & . & . & . \\ . & . & . & . & . \\ 0 & 0 & . & . & Z_p \end{bmatrix} U = \begin{bmatrix} U_1 & 0 & . & . & 0 \\ 0 & U_2 & . & . & 0 \\ . & . & . & . & . \\ . & . & . & . & . \\ 0 & 0 & . & . & U_p \end{bmatrix} \text{and} \in = \begin{bmatrix} \in_1 \\ \in_2 \\ . \\ . \\ \in_p \end{bmatrix}$$

Here, all Z_i 's have q columns.

Now, the unrestricted linear regression model can be written as

$$Y = X\beta + \in \qquad (3.8)$$

By applying OLS estimation to (3.8), first obtain the OLS residual vector and hence the unrestricted internally studentized residual sum of squares as

$$e_{UR}^{*1}\ e_{UR}^{*} = e_1^{*1}\ e_1^{*} + e_2^{*1}\ e_2^{*} + ... + e_p^{*1}\ e_p^{*} \qquad (3.9)$$

Here, $e_i^{*1}\ e_i^{*}$ is the internally residual sum of squares which is obtained by regressing Y_i on X_i $i = 1, 2, ..., p$

Consider the hypothesis H_0: $\gamma_1 = \gamma_2 = ..., = \gamma_p = \gamma$ say under the H_0, the restricted linear model is given by

$$Y = X^* \beta^* + \in \qquad (3.10)$$

Where $\beta^* = (\gamma^1, \delta^1)$

$X^* = (Z^*, V^*)$

and $Z^{*1} = (Z_1^1, Z_2^1, ..., Z_p^1)$

By applying the OLS to the model (3.10) and hence the restricted internally studentized residual sum of squares can be obtained as $e_R^{*1} e_R^{*}$.

Now, the modified F-test statistic for testing $H_0 : \gamma_1 = \gamma_2 = ... = \gamma_p$ based on studentized residuals is given by

$$F^{*} = \frac{\left[e_R^{*1} e_R^{*} - e_{UR}^{*1} e_{UR}^{*}\right] / q}{e_{UR}^{*1} e_{UR}^{*} \Big/ \left(\sum_{i=1}^{p} n_i - pk\right)} \tag{3.11}$$

where q is the number of restrictions under H_0. Under the assumption of homoscedasticity,

F^* follows F distribution with $\left[q, \sum_{i=1}^{p} n_i - pk\right]$ degrees of freedom

IV. TWO MODIFIED TESTS FOR EQUALITY OF HETEROSCEDASTIC VARIANCES IN THE LINEAR MODEL USING STUDENTIZED RESIDUALS:

Consider a heteroscedastic linear regression model as

$$Y_{n\times1} = X_{n\times k}\,\beta_{k\times1} + \in_{n\times1} \tag{4.1}$$

Such that $\in_i \overset{iid}{\sim} N\left(0, \sigma_i^2\right)$, i =1, 2,……, n

Suppose that $\Sigma = \text{diag}\left\{1, \frac{\sigma_2^2}{\sigma_1^2}, \frac{\sigma_3^2}{\sigma_1^2}, ..., \frac{\sigma_n^2}{\sigma_1^2}\right\}$ (4.2)

Then one may have $\text{Var}(\in) = \sigma_1^2\,\Sigma$ (4.3)

Consider the null hypothesis

$Ho: \Sigma = I_n$ against $H_1: \Sigma \neq I_n$

Define the OLS residual vector as

$e = \left[Y - X\hat{\beta}\right]$, where $\hat{\beta} = \left(X^1X\right)^{-1} X^1Y$.

Now, define the internally standentized residual vector as e^* based on OLS residuals.

Suppose that under H_1 the elements of Σ are known and hence Σ is known fixed matrix say Σ_o. To test H_o, a modified King test is given by

$$M_k = \frac{e^{*1} \sum_0^{-1} e^*}{e^{*1} e^*} \qquad (4.4)$$

For smaller values of M_k, the H_0 of heteroscedastic disturbances may be rejected.

Remarks: When the elements of $\sum$ are unknown a modified Bartlett's test for the equality of heteroscadastic variances can be applied by using internally studentized residuals.

By assuming the $\sigma_i^2 = g\left(z_i^1 r\right)$, i =1,2… n where g is a twice differential function independent of i and z_i, a modified Godfrey test based on internally studentized residuals for testing the homoscedasticity of disturbances

$H_o : \sigma_1^2 = \sigma_2^2 = ...\sigma_n^2$ is given by

$$M_G = \frac{\sum_{i=1}^{n} \left(\frac{i-1}{n}\right) e_i^{*2}}{\sum_{i=1}^{n} e_i^{*2}} \qquad (4.5)$$

Where e_i^*'s are the internally studentized residuals. For smaller of values M_G, the H_0 may be rejected.

V. CONCLUSIONS:

In the present study testing procedures for testing the equality between the regression coefficients in two/sets of linear regression models under heteroscedasticity have been suggested by using the studentized residuals. Here, a test for detecting the presence of heteroscedastic errors in the linear regression model has also been proposed.

This kind of research study can be further extended by considering various structures for heteroscedasticity with reference to linear and nonlinear regression models.

Further, the research contribution made in the present study could generate an immense interest in other researchers to take up an extension research work by developing testing procedures for Random Coefficient Regression models/ Sets of linear regression models or Seemingly Unrelated Regression Equations (SURE) models in the context of the various specifications of heteroscedastic or/ and autocorrelated errors.

References

1. Carlos, M. Jarque. (1981), "A Test for Heteroscedasticity in a Limited Dependent variable Model" Austral. J. Statist., 23, 159-163.

2. Conerly, M.D., and Mansfield, E.R. (1988), An Approximate Test for Comparing Heteroscedastic Regression Models, Journal of the American Statistical Association, 83, 81-817.

3. Davidson, R. and J.G. Mackinnon (1981), "Several tests for Model Specification in the presence of Alternative Hypotheses", Econometrica, 49, 781-793.

4. Evans, M.A., and M.L. King (1984), A Further class of tests for Heteroscedasticity", Mimeo (Monash University, Melbourne).

5. Glejser, H. (1969), A New Test for Heteroscedasticity, Journal of American Statistical Association, 64, 316-323.

ISSN 2515-8260 Volume 07, Issue 02, 2020

6. Goldfeld, S.M. and R.E. Quandt (1965), Some Tests for Homoscedasticity, Journal of the American Statistical Association, 60, 539-547.

7. Greene, W.H. (1997), “Econometric Analysis” (3rd Edition) Prentice Hall, New Jersery.

8. Harrison, M.J. and B.P.M. MC Cabe, (1979), “A Test of Heteroscedasticity Based on Ordinary Least Squares Residuals”, JASA, 74, 494-499.

9. Harvey, A.C. and Phillips, G.D.A (1974), “A Comparison of Some Tests for Heteroscedasty in the General Linear Model”; Journal of Econometrics, 2, 307-316.

10. Koenkar, K. (1981), “A Note on Studentizing a test for Heteroscedasticity”, Journal of Econometrics, 17, 107-112.

11. Lahari, K. and Egy. D. (1981), “Joint Estimator and testing for functional form and Heteroscedasticity”, Journal of Econometrics, 15, 299-307.

12. Nachane, D.M., (2006), “Econometrics: Theoretical Foundations and Empirical perspectives”, Oxford university press, New Delhi.

13. Ramsey, J.B. (1969), Tests for Specification error in Classical Linear Least Squares Regression Analysis, Journal of the Royal statistical society, B, 31, 250-271.