*International Journal of Engineering & Technology*, 7 (4.10) (2018) 539-542

**International Journal of Engineering & Technology**

*Website: www.sciencepubco.com/index.php/IJET*

***Research paper***

# A Treatise on Testing General Linear Hypothesis in Stochastic Linear Regression Model

**C. Narayana[1] , B. Mahaboob[2], B.Venkateswarlu[3]*, J. Ravi sankar[4] and P. Balasiddamuni[5]**

[1]*Department of Mathematics, Sri Harsha institute of P.G Studies, Nellore.*
[2]*Department of Mathematics, K.L.E.F(Deemed to be University), Vaddeshwaram, Vijayawada, Andhra Pradesh.*
[3,4]*Department of Mathematics, VIT University, Vellore, Tamilnadu.*
[5]*Rtd. Professor, Department of Statistics, S.V. University, Tirupati, Andhra Pradesh.*
**Corresponding author E-mail: venkatesh.reddy@vit.ac.in*

**Abstract**

The main objective of this research article is to propose test statistics for testing general linear hypothesis about parameters in stochastics linear regression model using studentized residuals, RLS estimates and unrestricted internally studentized residuals. In 1998, M. Celia Rodriguez -Campos et.al [1] introduced a new test statistics to test the hypothesis of a generalized linear model in a regression context with random design. Li Cai et.al [2] provide a new test statistic for testing linear hypothesis in an OLS regression model that not assume homoscedasticity. P. Balasiddamuni et.al [3] proposed some advanced tools for mathematical and stochastical modelling.

***Keywords***: *OLS estimator, OLS residuals, RLS estimator, linear hypothesis, RLS residual vector, General linear hypothesis, PRESS, GLS, BLUS, BAVS, RSURE, stochastic linear regression model, studentized residuals, RLS estimators.*

## 1. Introduction

In spite of the availability of highly innovative tools in Mathematics, the main tool of the Applied Mathematician remains the stochastic regression model in the form of either linear or nonlinear model. More importantly, mastery of the stochastic linear regression model is prerequisite to work with advanced mathematical and statistical tools because most advanced tools are generalizations of the stochastic linear regression model. The various inferential problems of stochastic modelling are considered to be essential to both theoretical and applied mathematicians and statisticians. The selection between alternative models is an important problem in stochastic modelling. Specification of the stochastic regression model is an important stage in any stochastic linear regression analysis. It includes specifying both the expectation function and the characteristics of the error. The various Misspecification tests and testing general linear hypothesis in the stochastic linear regression models were studied by many mathematicians and statisticians. Most of these people have proposed their tests in stochastic linear regression models by using some inferential criteria. A cursory glance at the recent literature on Model Building clearly suggests a significant shift in the level of mathematical and stochastic rigor brought at research efforts concerning Model Building. A more careful inspection shows that this trend has not been uniform across in the literature. In particular, while mathematical and stochastic modeling efforts in certain fields of science and technology have been appreciable, other research fields of science remain under developed. Successful Mathematical and Stochastic Model buildings are not a collection of simple mechanistic and routine techniques but more of an art requiring wide-ranging knowledge and judgement. In the stochastic model building, the most difficult problem is the specification of the stochastic model. Under the problem of misspecification of the stochastic regression model, first task is that what set of regressors have to be included in the model; and the second task is that in which mathematical form of the regressors are to be included in the model.

## 2. Special Types of Residuals

Residuals have an important role on inference in stochastic regression models. They are very useful in analyzing various problems of stochastic linear regression models such as Autocorrelation, Heteroscedasticity, Misspecification, variable selection, Model selection etc. To detect (i) the disagreements between data and an assume model, (ii) violations of assumptions of the stochastic regression model, (iii) outlines in the data one may frequently use different types of residuals. Several types of residuals exist in the literature are Ordinary Least Squares (OLS) residuals; Stepwise Least Squares residuals; Generalized Least Squares (GLS) residuals; Abrahams and Koerts residuals; Best Linear Unbiased Scalar (BLUS) residuals; Recursive residuals; Best Augmented Unbiased with Scalar Matrix (BAUS) residuals; Independent stepwise residuals; Restricted Seemingly Unrelated Regression Equation (RSURE) residuals; Studentized (Internal and External) residuals; predicted residuals etc.

Consider the standard stochastic linear regression model

$$Y_{n\times 1} = X_{n\times k}\beta_{k\times 1} + \in_{n\times 1} \tag{2.1}$$

Such that $\in \sim N(0, \sigma^2 I_n)$

Where $\in$ is vector of unobservable errors which can be estimated with residual vector.

Define the OLS residual vector as $e = Y - \hat{Y} = Y - X\hat{\beta}$

$$= Y - X(X'X)^{-1}X'Y \quad \left[\because \hat{\beta} = (X'X)^{-1}X'Y\right]$$

$$= \left[I - X(X'X)^{-1}X'\right]Y$$

$$\Rightarrow e = [I - H]Y \tag{2.2}$$

Where $H = X(X'X)^{-1}X'$ is known as Hat matrix.

One may write, $H = \left((h_{ij})\right) = \left(\left(X_i'(X'X)^{-1}X_j\right)\right)$ (2.3)

One may obtain $E(\in) = 0,\ Var(e) = \sigma^2[I - H]^2$

or $Var(e) = \sigma^2[I - H]$ [$\because$ [I − H] is symmetric Idempotent Matrix]

$$\Rightarrow Var(e_i) = \sigma^2(1 - h_{ii}) \tag{2.4}$$

If the stochastic linear regression model contains constant or intercept term these,

$\sum_i h_{ij} = \sum_j h_{ij} = 1$ and $\sum_j h_{ij}^2 = h_{ii}$ also, each $h_{ii}$ falls in between $\frac{1}{n}$ and 1.

Studentized residuals are improved set of residuals, that can be obtained by scaling the residuals with large $h_{ii}$ get larger scaled residuals and residuals with small $h_{ii}$ set smaller scaled residuals. Scaling can done by dividing each of the residuals it's corresponding estimate of its standard deviation. These types of residual are known as studentized residuals. For a simple two- variable stochastic linear regression model, one may have

$$h_{ii} = \frac{1}{n} + \frac{(X_i - \bar{X})^2}{\sum_{i=1}^{n}\left[X_i - \bar{X}\right]^2} \tag{2.5}$$

Such that Trace (H) = K = No. of parameters. Here, $Rank(X) = K$.

Now, the studentized residuals can be defined as

**(i) Internally Studentized Residuals** $(q_i)$

The i[th] internally studentized residual is given by

$$q_i = \frac{e_i}{\hat{\sigma}\sqrt{(1 - h_{ii})}}, \quad i = 1, 2, ...., n \quad b \tag{2.6}$$

Where $\hat{\sigma}^2 = \left[e'e/n - k\right] = \frac{\sum_i e_i^2}{n-k}$ … (2.7)

Here $\left[q_i^2/n - k\right] \sim B\left[\frac{1}{2}, \frac{n-k-1}{2}\right]$.

Also, $E(q_i) = 0,\ Var(q_i) = 1$ and $\operatorname{cov}(q_i, q_j) = \frac{-h_{ij}}{\sqrt{(1-h_{ii})(1-h_{jj})}}$ $i \neq j$

**(ii) Externally studentized Residuals** $(q_i^*)$

The i[th] externally studentized residual is given by

$$q_i^* = \frac{e_i}{\hat{\sigma}_{(i)}\sqrt{1 - h_{ii}}},\ i = 1, 2, ..., n \tag{2.8}$$

Where $\hat{\sigma}_{(i)}^2 = \frac{(n-k)\hat{\sigma}^2 - \left[e_i^2/(1-h_{ii})\right]}{n-k-1}$ = Residual mean square without i[th] case.

or $\hat{\sigma}_{(i)}^2 = \sigma^2\left[\frac{n-k-q_i^2}{n-k-1}\right]$ (2.9)

A relationship between $q_i$ and $q_i^*$ is given by

$$q_i^* = q_i\left[\frac{n-k-1}{n-k-q_i^2}\right]^{1/2}, \quad i = 1, 2, ...., n \tag{2.10}$$

The i[th] predicted residual is defined as

$$q_{(i)} = \left[Y_i - X_i'\hat{\beta}_{(i)}\right], \quad i = 1, 2, ..., n \tag{2.11}$$

Where $\hat{\beta}_{(i)}$ the OLS estimator of is $\beta$ based on a fit to the data with the i[th] case excluded. Further, the i[th] Predicted Residual Sum of Squares (PRESS) is defined as

$$PRESS = \sum_{i=1}^{n} q_{(i)}^2 \tag{2.12}$$

PRESS can be considered as a criterion for best stochastic model selection. Small value of PRESS reveals the better performance of the model. A relationships between $q_{(i)}$ and $e_i, q_i, q_i^*$ are given by

(i) $$q_{(i)} = \frac{e_i}{1 - h_{ii}} \tag{2.13}$$

(ii) $$q_i = \frac{e_{(i)}}{\hat{\sigma}/\left[(1-h_{ii})\right]^{\frac{1}{2}}} \tag{2.14}$$

and (iii) $$q_i^* = \frac{e_{(i)}}{\hat{\sigma}_{(i)}/(1-h_{ii})},\ i = 1, 2, .., n \tag{2.15}$$

Here, $e_{(i)}$, $\hat{\sigma}_{(i)}$ are the OLS estimators which are computed based on fit to the data without i[th] case.

## 3. Testing Linear Hypothesis about Parameters of Stochastic Linear Regression Model Using Studentized Residuals.

Consider the standard stochastic linear regression model

$$\underset{n\times 1}{Y} = \underset{n\times k}{X}\ \underset{k\times 1}{\beta} + \underset{n\times 1}{\in} \tag{3.1}$$

Such that $\in \sim N\left[0, \sigma^2 I_n\right]$. Where N refers to multivariate normal distribution and 0 is null mean vector. Suppose that $\beta$ obey the set of $m^{(\leq k)}$ linear restrictions in the form of general linear hypothesis as $H_0 : R_{m\times k}\beta_{k\times 1} = r_{m\times 1}$. where R is (mxk) known matrix and r is (mx1) known vector further, assume that R is having full row rank, which indicates that there are no linear dependencies among the hypotheses, One may replace unknown parametric vector $\beta$ by the OLS estimator $\hat{\beta}$ as $(X^1X)^{-1}X^1X$.

One may obtain the sampling distribution of $R\hat{\beta}$ as follows:

(i) $$E\left(R\hat{\beta}\right) = RE\left(\hat{\beta}\right) = R\beta \tag{3.2}$$

(ii) Var $$\left(R\hat{\beta}\right) = E\left[R\hat{\beta} - E\left(R\hat{\beta}\right)\right]\left[R\hat{\beta} - E\left(R\hat{\beta}\right)\right]^1$$

$$= RE\left[\hat{\beta} - \beta\right]\left[\hat{\beta} - \beta\right]^1 R^1$$

$$= R\,Var\left(\hat{\beta}\right)R^1$$

$$\Rightarrow Var\left(R\hat{\beta}\right) = \sigma^2\left[R\left(X^1X\right)^{-1}R^1\right]$$

$$\left[\because Var\left(\hat{\beta}\right) = \sigma^2\left(X^1X\right)^{-1}\right] \tag{3.3}$$

Since, $\hat{\beta}$ has multivariate normal distribution with mean vector $\beta$ and covariance matrix $\sigma^2\left(X^1X\right)^{-1}$ the sampling distribution of $R\hat{\beta}$ is given by

$$R\hat{\beta} \sim N\left[RB, \sigma^2\left\{R\left(X^1X\right)^{-1}R^1\right\}\right] \tag{3.4}$$

$$\Rightarrow \left(R\hat{\beta} - R\beta\right) \sim N\left[0, \sigma^2\left\{R\left(X^1X\right)^{-1}R^1\right\}\right]$$

or $$\left(R\hat{\beta} - r\right) \sim N\left[0, \sigma^2\left\{R\left(X^1X\right)^{-1}R^1\right\}\right] \tag{3.5}$$

From the distribution of quadratic forms, one may obtain

$$\left(R\hat{\beta} - r\right)^1\left[\sigma^2\left\{R\left(X^1X\right)^{-1}R^1\right\}\right]^{-1}\left(R\hat{\beta} - r\right) \sim \chi^2_m$$

It can be easily seen that $\left[R\left(X^1X\right)^{-1}R^1\right]$ is positive definite matrix. In general, error variance parameter $\sigma^2$ is unknown and it can be estimated by using OLS residual sum of squares (from Gauss-Marksoff Theorem) as

$$\hat{\sigma}^2 = \frac{e^1e}{n-k} = \frac{\sum_{i=1}^{n} e_i^2}{n-k} \tag{3.6}$$

Since, the OLS residuals have a distribution that is scale dependent and studentized residuals are scale independent, one may replace OLS residuals with studentized residuals. By replacing OLS residuals with internally studentized residuals ($q_i$), an estimate of $\sigma^2$ is given by

$$\tilde{\sigma}^2 = \frac{q^1q}{n-k} = \frac{\sum_i q_i^2}{n-k} \tag{3.7}$$

Also, $\tilde{\sigma}^2 \sim \chi^2_{(n-k)}$. Now, the test statistic for testing general linear hypothesis $H_o : R\beta = r,$ about the parameters of stochastic linear regression model is given by

$$F = \left[\frac{\left(R\hat{\beta} - r\right)^1\left[R\left(X^1X\right)^{-1}R^1\right]^{-1}\left(R\hat{\beta} - r\right)/m}{q^1q/n-k}\right] \sim F_{[m,(n-k)]} \tag{3.8}$$

One may compare the calculated value of F-statistic with its critical value (Table value) for [m, (n-k)] degrees freedom at chosen level of significance and draw the inference accordingly.

## 4. Testing General Linear Hypothesis in Stochastic Linear Regression Model Using RLS Estimators

Under General Linear Hypothesis $H_o : R\beta = r$ consists of a set of $m(\le k)$ linear restrictions, the Restricted Least Squares estimator of $\beta$ in the standard stochastic linear regression model $Y = X\beta + \in$ is given by

$$\beta^*_{RLS} = \hat{\beta} + \left(X^1X\right)^{-1}R^1\left[R\left(X^1X\right)^{-1}R^1\right]^{-1}\left(r - R\hat{\beta}\right) \tag{4.1}$$

Where $\hat{\beta}$ is the unrestricted OLS estimator of $\beta$. By defining Restricted Least Squares (RLS) residual vector

$$e^* = Y - X\beta^*_{RLS} = Y - X\hat{\beta} - X\left(\beta^*_{RLS} - \hat{\beta}\right) = e - X\left(\beta^*_{RLS} - \hat{\beta}\right)$$

(or)

$$e^{*1}e^* = e^1e + \left[\left(\beta^*_{RLS} - \hat{\beta}\right)^1\left(X^1X\right)\left(\beta^*_{RLS} - \hat{\beta}\right)\right] \tag{4.2}$$

Here, the cross product term variables, since $X^1e=0$.

$$\Rightarrow \left[e^{*1}e^* - e^1e\right] = \left[\left(\beta^*_{RLS} - \hat{\beta}\right)^1\left(X^1X\right)\left(\beta^*_{RLS} - \hat{\beta}\right)\right]$$ Using (4.1), one may express

$$\left[e^{*1}e^* - e^1e\right] = \left(r - r\hat{\beta}\right)^1\left[R\left(X^1X\right)^{-1}R^1\right]^{-1}\left(r - R\hat{\beta}\right)$$

or $$\left[e^{*1}e^* - e^1e\right] = \left(\beta^*_{RLS} - \hat{\beta}\right)\left(X^1X\right)\left(\beta^*_{RLS} - \hat{\beta}\right) \tag{4.3}$$

Using RLS estimator $\beta^*_{RLS}$ and Internal studentized residuals the t-test statistic for testing general linear hypothesis $H_o : R\beta = r$ in stochastic linear regression model is given by

$$F_R = \left[\frac{\left(\beta^*_{RLS} - \hat{\beta}\right)X^1X\left(\beta^*_{RLS} - \beta\right)/m}{q^1q/n-k}\right] \sim F_{(m,n-k)} \tag{4.4}$$

## 5. Testing General Linear Hypothesis in Stochastic Linear Regression Model Using Restricted and Unrestricted Internally Studentized Residuals.

Suppose the $e^1e$ be the unrestricted OLS residual sum of squares obtained by using OLS estimators and $e^{*1}e^*$ be the restricted OLS residual sum of squares under general linear hypothesis obtained by using restricted least squares estimators of parameters of the standard stochastic linear regression model. By replacing these residual sum of squares with their corresponding Internally Studentized residual sum of squares, one may obtain the test statistic for testing general linear hypothesis $H_0 : R\beta = r$ as

$$F_{IS} = \frac{\left[q^1_Rq_R - q^1q\right]/m}{q^1q/(n-k)} \sim F_{[m,(n-k)]} \tag{5.1}$$

Where $q^1_Rq_R$ = Restricted Internally Studentized Residual sum of squares.

$q^1q$ = Unrestricted Internally Studentized Residual sum of squares.
m = Number of Linear Restrictions about k parameters.

## 6. Conclusion

In the above research study test statistics for testing general linear hypothesis about parameters in stochastic linear regression model have been developed by using studentized residuals RLS estimators and unrestricted internally studentized residuals. These ideas can be extended to develop advanced tools for analyzing inferecutial aspects of stochastic nonlinear regression models and random coefficients regression models by using different types of residuals other than studentized residuals.

## References

[1] M. Celia Rodrignez-Campos Wenceslao 'Manteiga and Ricardo Cao, "Testing the hypothesis of a generalized linear regression model using nonparametric regression estimation", Journal of statistical planning and inference, (1998), Pp: 99-122.
[2] Li Cai, Andraw F. Hayes, "A new test of linear hypothesis in OLS regression under heteroscedasticity of unknown form", Journal of Education and behavioral statistics", Vol. (33), (2008), Pp: 21-40.
[3] Balasiddamuni, P. et.al., "Advanced Tools for Mathematical and Stochastic Modeling", Proceedings of the International Conference on Stochastic Modeling and Simulation, Allied Publishers, (2011).
[4] Byron J.T. Morgan, "Applied stochastic Modelling", CRC Press, (2008), 978-1-58488-666-2.
[5] Berry L. Nelson, (1995), "Stochastic Modeling, Analysis and Simulation", McGraw-Hill, (1995), 978-0070462137.
[6] Nelson, B.L. "Stochastic Modelling", McGraw-Hill, New York, (1995), 0-486-47770-3.
[7] Taylor, H.M. and Samuel karlin, "An Introduction to Stochastic Modeling", Academic Press, London, (1998), 978-0-12-684887-87
[8] Nafeez Umar, S. and Balasiddamuni, P, "Statistical Inference on Model Specification in Econometrics", LAMBERT Academic Publishing, Germany, (2013).