International Journal of Engineering & Technology, 7 (4.10) (2018) 532-535

International Journal of Engineering & Technology

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

Research paper

A Study on Misspecification and Predictive Accuracy of Stochastic Linear Regression Models

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 present study research article proposes a modified test for misspecification of the stochastic linear regression model and a new test for predictive accuracy of stochastic linear regression model. In addition to this modified Lagrange Multiplier (LM) test for misspecification of stochastic linear regression has been developed. In the derivation of the test statistics internally studentized residuals have been used. William A. Branch et.al [1] presented a stochastic non-linear self-referential model in which expectations are based on linear perceptions. I.sh. Torgovitski et.al [2] in this paper discussed the problem of raising the efficiency of the regression coefficients estimation as suggested an approach which allows as to reduce mathematical expectations of the square of deviation of the response prediction.

Keywords: *Stochastic linear regression model, OLS residuals vector, internally studentized residuals, test statistic, predictive accuracy BLUE, and OLS estimator.*

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 mis-specification 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. A Modified Test For Misspecification of the Stochastic Linear Regression Model based on Durbin-Watson Statistic by Using Internally Studentized Residuals

Consider the standard stochastic linear regression model

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

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

Assuming that the model (6.10.1) is misspecified because of omission of a relevant independent variable Z from this linear regression model. One may obtain the Ordinary Least Squares (OLS) residuals vector as

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

Also, Internally Studentized residuals are given by $q_i = \dfrac{e_i}{\hat{\sigma}\sqrt{1-h_{ii}}}, i = 1, 2, \ldots, n$

Where, $\hat{\sigma}^2 = \frac{e'e}{n-k}$ and h_{ii}: i^{th} diagonal element in the Hat matrix $H = \left(\left(h_{ij}\right)\right) = X\left(X'X\right)^{-1}X'$

Now, arrange Internally Studentized residuals according to values in ascending order of excluded independent variable Z. For testing the null hypothesis of mis-specification of the stochastic linear regression model because of exclusion of relevant independent variable Z, the modified Durbin-Watson test statistic is given by

$$d^* = \frac{\sum_{i=2}^{n}[q_i - q_{i-1}]^2}{\sum_{i=1}^{n} q_i^2}$$

Here, the subscript i denotes the index of observation and it does not necessarily means that the data is time series data. Compare the calculated value of d^* statistic with its critical value (From Durbin-Watson Table for critical values) and draw inference accordingly. If d^* is significant, then one may accept the hypothesis of mis-specification of the stochastic linear regression model due to exclusion of relevant variable Z. This proposed test may be applied by considering each of the excluded relevant independent variables.

3. A New Test for Predictive Accuracy of Stochastic Linear Regression Model Using Internally Studentized Residuals

In the stochastic model building, an important characteristic of specification of functional relationship is the predictive accuracy of stochastic model over various data sets. Consider the standard stochastic linear regression model as

$$Y_{n\times 1} = X_{n\times k}\beta_{k\times 1} + \in_{n\times 1} \text{ such that } \in \sim N\left(0, \sigma^2 In\right)$$

The Ordinary Least Squares (OLS) estimator of β is given by

$$\hat{\beta} = (X'X)^{-1}X'Y \;= (X'X)^{-1}X'(X\beta + \in)$$

$$\Rightarrow \hat{\beta} = \beta + (X'X)^{-1}X' \in$$

Suppose that there is an additional set of r observations $(r < k)$ are available on all Y and X's variables. For these new set of r observations, one may specify the same specification of stochastic linear regression model as

$$Y^*_{r\times 1} = X^*_{r\times k}\beta_{k\times 1} + \in^*_{r\times 1}, \text{ such that } \in^* \sim N\left(0, \sigma^2 In\right)$$

One may obtain the OLS forecast for Y^* as $\hat{Y}^* = X^*\hat{\beta}^*$, where $\hat{\beta}^* = \left(X^{*\prime}X^*\right)^{-1}X^{*\prime}Y^*$ or $\hat{\beta}^* = \beta + \left(X^{*\prime}X^*\right)^{-1}X^{*\prime} \in^*$

Now, the OLS forecast error is given by

$$e^* = \left[Y^* - \hat{Y}^*\right] = X^*\beta + \in^* - X^*\hat{\beta}^*$$

$$= X^*\beta + \in^* - X^*\left(X^{*\prime}X^*\right)^{-1}X^{*\prime}\left(X^*\beta + \in^*\right)$$

$$\Rightarrow e^* = \in^* - X^*\left(X^{*\prime}X^*\right)^{-1}X^{*\prime} \in^*$$

One may have, $E\left[e^*\right] = 0$ and

$$\mathrm{Var}\left(e^*\right) = E\left[\in^* - X^*\left(X^{*\prime}X^*\right)^{-1}X^{*\prime} \in^*\right]\left[\in^* - X^*\left(X^{*\prime}X^*\right)^{-1}X^{*\prime} \in^*\right]'$$

$$= E\left[\in^* \in^{*\prime}\right] + X^*\left(X^{*\prime}X^*\right)^{-1}X^{*\prime}E\left[\in^* \in^{*\prime}\right]X^*\left(X^{*\prime}X^*\right)^{-1}X^{*\prime}$$

$$\Rightarrow \mathrm{Var}\left(e^*\right) = \sigma^2\left[I_r + X^*\left(X^{*\prime}X^*\right)^{-1}X^{*\prime}\right][\because \text{ Cross product terms vanish}]$$

Since, $\in^* \sim N\left(0, \sigma^2 I_r\right)$, e^* follows $N\left(0, \sigma^2 H^*\right)$, Where $H^* = \left(I_r + X^*\left(X^{*\prime}X^*\right)^{-1}X^{*\prime}\right)$

Suppose that forecast error vector is defined as

$$e^{**} = \left[Y^* - \hat{Y}^*\right] = \left[Y^* - X^*\hat{\beta}\right], \text{ where } \hat{\beta} = \beta + (X'X)^{-1}X' \in$$

$$= X^*\beta + \in^* - X^*\left[\beta + (X'X)^{-1}X' \in\right]$$

$$\Rightarrow e^{**} = \in^* - X^*(X'X)^{-1}X' \in$$

One may have, $E\left[e^{**}\right] = 0$ and

$$\mathrm{Var}\left[e^{**}\right] = E\left[\in^* - X^*(X'X)X' \in\right]\left[\in^* - X^*(X'X)X' \in\right]'$$

$$= E\left[\in^* \in^{*\prime}\right] + X^*(X'X)X'E[\in \in']X^*(X'X)X^{*\prime}$$

$$\mathrm{Var}\left[e^{**}\right] = \sigma^2\left[I_r + X^*(X'X)^{-1}X^{*\prime}\right][\because \text{ Cross product terms Vanish}]$$

Since, $\in^* \sim N\left(0, \sigma^2 I_r\right)$ and $\in \sim N\left(0, \sigma^2 I_n\right)$, the forecast error vector e^{**} follows $N\left(0, \sigma^2 H^{**}\right)$ where $H^{**} = I_r + X^*(X'X)^{-1}X^{*\prime}$

Thus, one may obtain internally studentized residuals vectors using the forecast error vectors e^* and e^{**} as q_* and q_{**} respectively, which are given by

$$q_{*i} = \frac{e_i^*}{\hat{\sigma}\sqrt{h_{ii}^*}} \text{ and } q_{**i} = \frac{e_i^{**}}{\hat{\sigma}\sqrt{h_{ii}^{**}}}$$

Where h_{ii}^* is the i^{th} diagonal element of $H^* = \left(\left(h_{ij}^*\right)\right) = \left[I_r + X^*\left(X^{*\prime}X^*\right)^{-1}X^{*\prime}\right]$

h_{ii}^{**} is the i^{th} diagonal element of $H^{**} = \left(\left(h_{ij}^{**}\right)\right) = \left[I_r + X^*\left(X^{*\prime}X^*\right)^{-1}X^{*\prime}\right]$

Now, one can obtain the sampling distributions as

(i) $\frac{q_*' H^{*^{-1}} q_*}{\sigma^2} \sim \chi_r^2$ [using $\hat{\beta}^* = \beta + \left(X^{*\prime}X^*\right)X^{*\prime} \in^*$]

(ii) $\frac{q_{**}' H^{**^{-1}} q_{**}}{\sigma^2} \sim \chi_r^2$ [using $\hat{\beta} = \beta + (X'X)X' \in$]

Since $\left[\frac{e'e}{\sigma^2}\right]$ has independent χ^2_{n-k} and hence $\left[\frac{q'q}{\sigma^2}\right]$ has independent χ^2_{n-k} distributions; for testing the predictive accuracy of stochastic linear regression model under null hypothesis, the following modified Chow test statistics have been proposed.

(i) $$F^* = \left[\frac{q_*^1\left[I_n x^*\left(x^{*1}x\right)^{-1}x^{*1}\right]^{-1}q_*/r}{q^1 q/(n-k)}\right] \square\, F_{[r,(n-k)]}$$ and

(ii) $$F^{**} = \left[\frac{q^1_{**}\left[I_n + x^*\left(x^{*1}x\right)^{-1}x^{*1}\right]^{-1}q_{**}/r}{q^1q/(n-k)} \right] \sim F_{[r,(n-k)]}$$

Here $q^1_*q_*$ = Internally Studentized Residual sum of squares based on OLS forecast error vector e^*

$q^1_{**}q_{**}$ = Internally studentized Residual sum of squares based on OLS forecast error vector e^{**}

$q^1q =$ Internally studentized Residual sum of squares based on OLS residual vector e.

Remarks:

1) Under null hypothesis of predictive accuracy of stochastic linear regression model, the modified chow test statistic F* may be suggested than F^{**}.

2) In the presence of stochastic regressors, under null hypothesis of predictive accuracy of the stochastic linear regression model, one may have

(i) $\frac{q^1_*q_*}{\sigma^2} \overset{asy}{\sim} \chi^2_r$, (ii) $\frac{q^1_{**}q_{**}}{\sigma^2} \overset{asy}{\sim} \chi^2_r$ and $\frac{q^1q}{\sigma^2}$ has independent χ^2_{n-k}. Thus, for testing the predictive accuracy of the stochastic linear regression model, the test statistic is given by

$$F^{***} = \left[\frac{q^1_*q_*/r}{q^1q/(n-k)} \right] \overset{asy}{\sim} F_{[r,(n-k)]}$$

4. Modified Lagrange Multiplier (Lm) Test for Misspecification of Stochastic Linear Regression Model Using Internally Studentized Residuals:

Consider the standard stochastic linear regression model

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

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

The best linear unbiased estimator (BLUE) for β is the OLS estimator $\hat{\beta}$, which in given by

$$\hat{\beta} = (X'X)^{-1}X'Y \tag{4.2}$$

One may obtain the restricted OLS residual vector $\in_R = (Y - X\hat{\beta})$ and hence the restricted Internally Studentized residuals q_{Ri} as

$$q_{Ri} = \frac{e_{Ri}}{\hat{\sigma}\sqrt{1-h_{ii}}}, \quad i=1, 2\ldots n \tag{4.3}$$

Where $\hat{\sigma}^2 = \frac{e'_R e_R}{n-k}$ and h_{ii} = i^{th} diagonal element of Hat matrix $H = X(X'X)^{-1}X' = ((h_{ij}))$.

Suppose that Z be $(n\times p)$ matrix of independent variables which are included in the model and their coefficients to be tested for mis-specification of the model. One may write the augmented or unrestricted stochastic linear regression model as

$$Y = X\beta + Z\delta + \in^* \tag{4.4}$$

Or $Y = X^*\Gamma + \in^*$, where, $X^* = [X \quad Z]$, $\Gamma = \begin{bmatrix}\beta \\ \delta\end{bmatrix}$

Such that $\in^* \sim N(0, \sigma^2_{\in^*} I_n)$, where δ is $(p\times1)$ vector of parameters associated with z.

$\in^*$ is an error vector with usual properties. By applying OLS estimation procedure, one may estimate the unrestricted stochastic linear regression model (4.4) and obtain the unrestricted OLS residual vector as

$$e_{UR} = \left[Y - X^*\hat{\Gamma}\right], \text{ where } \hat{\Gamma} = \left(X^{*\prime}X^*\right)^{-1}X^{*\prime}Y \tag{4.5}$$

The unrestricted internally studentized residuals are given by

$$q_{URi} = \frac{e_{URi}}{\hat{\sigma}_{\in^*}\sqrt{1-h^*_{ii}}}, \text{ i=1, 2… n} \tag{4.6}$$

Where $$\hat{\sigma}^2_{\in^*} = \frac{e'_{UR}e_{UR}}{n-k-p} \tag{4.7}$$

and h^*_{ii} = i^{th} diagonal element of Hat matrix $H^* = X^*\left(X^{*\prime}X^*\right)X^{*\prime} = \left(\left(h^*_{ij}\right)\right)$

If one assumes that unrestricted stochastic linear regression model (6.12.3) is the true model then the OLS residual vector e_R and hence q_R be related to Z, By regressing q_R on all regressors X and Z, One may obtain,

$$q_R = X\alpha + Z\eta + \in^{**} \tag{4.8}$$

Where $\in^{**}$ is an error term with usual properties.

In the case of large sample size, for testing null hypothesis of restricted stochastic linear regression model, one may use Lagrange Multiplier (LM) test statistic as $LM = nR^{*2} \overset{asy}{\sim} \chi^2_p$. Where p is the number of restrictions imposed by the stochastic restricted linear regression model; R^{*2} is the coefficient of multiple determination obtained from linear regression model (4.8). If χ^2 value is significant at the chosen level of significance, then one may reject the restricted stochastic linear regression model.

5. Conclusion

In the above monograph an attempt has been made by developing some new advanced tools to analyze inferential problems such as misspecification of the model, model predictive accuracy of the stochastic linear regression models. These ideas can be applied for analyzing inferential aspects of stochastic nonlinear regression models and random coefficients regression models by using different types of residuals other than studentized residuals. The research contribution made here could generate an immense interest in other research fellows to take up future research study in coming years.

References

[1] William A. Branch, Bruce Mc Gough, "Consistent expectations and misspecification in stochastic nonlinear economics", Journal of Economic dynamic and control, Vol. (29), issue 4, (2005), Pp: 659-676.

[2] I. Sh. Torgovitski, P.I. Baumstein, “Accuracy estimation of prediction on the stochastic regression model at the restricted sample”, Stochastic control, Proceeding of the second IFAC symposium, Vilnius, Lithuanian SSR, USSR, (1987),Pp: 19-23

[3] Gregory L. Light (2010), “Regression model misspecification and Causation with pedagogical demonstration”, Applied Mathematics Sciences, Vol. (4), (2010), Pp: 225-236.

[4] Byron J.T. Morgan, “Applied stochastic Modelling”, CRC Press, (2008), 978-1-58488-666-2.

[5] Berry L. Nelson, “Stochastic Modeling, Analysis and Simulation”, McGraw-Hill, (1995), 978-0070462137.

[6] Taylor, H.M. and Samuel karlin, “An Introduction to Stochastic Modeling”, Academic Press, London, (1998), 978-0-12-684887.

[7] Nafeez Umar, S. and Balasiddamuni, P. “Statistical Inference on Model Specification in Econometrics”, LAMBERT Academic Publishing, Germany, (2013).

[8] Nelson, B.L, “Stochastic Modeling”, McGraw-Hill, New York, (1995), 0-486-47770-3.