On misspecification tests for stochastic linear regression model

Cite as: AIP Conference Proceedings **2177**, 020039 (2019); https://doi.org/10.1063/1.5135214
Published Online: 04 December 2019

B. Mahaboob, S. Vijay Prasad, J. Peter Praveen, Ranadheer Donthi, and B. Venkateswarlu

On Misspecification Tests for Stochastic Linear Regression Model

B. Mahaboob[1], S. Vijay Prasad[1], J. Peter Praveen[1], Ranadheer Donthi[2], B.Venkateswarlu[3, a)]

[1]*Department of Mathematics, Koneru Lakshmaih Education Foundation, Vaddeswaram, Guntur, Dist.AP-522502, India*
[2]*Department of Mathematics, St. Martin's Engineering College, Dhulapally, Kompally, Hyderabad, Telangana, India.*
[3]*Department of Mathematics, Vellore Institute of Technology, Vellore 632014, India*

[a)] Corresponding Author: venkatesh.reddy@vit.ac.in

Abstract. This research article explores some misspecification tests for stochastic linear regression model viz. Durbin-Watson test, Ramsey's regression specification error test, Lagrange's multiplier test, and UTTS' rainbow test. Any type of error occurred in the set of underlying assumptions of a stochastic linear regression model and the associated inferences lead to 'specification errors'. These errors will present particularly in specifying the error vector ε and the data matrix X. Generally specification errors are caused by the inclusion of irrelevant independent variables or exclusion of relevant independent variables in the stochastic linear regression model. Ivan Krivy et.al, in 2000 in their research article depicted two stochastic algorithms which are useful in estimating the parameters of nonlinear regression models. Russell Davidson et al. in 1984 (see [2]), in their paper, developed a simple computational procedure for performing a wide variety of model specification tests. In1993, LudgerRuschendorf et.al in their research article constructed as nonlinear regression representations of general stochastic processes so that they got a particular special regression representations on Markov chains and of certain m- dependent sequences.

INTRODUCTION

The terms data and specification of linear regression model are inextricably linked. Specification of stochastic regression model has strong implications for the data required and the limitations of data may restrict the specification to be feasible.

Specification of the stochastic regression model is as follows:

I. The listing of independent variables including lagged values in each equation.
II. Formulating the functional form relating independent variables to the dependent variables.
III. The stochastic properties of the error term
Thus the first two things inevitably lead to a certain amount of interaction between theory and data. The third item consists of assumptions about the unobservable error variables under first two items each specification provided estimates of the errors.
In the stochastic model building if the stochastic regression model has not been correctly specified then the tests are strictly invalid and the various stochastic statements are not correct.

Recent Trends in Pure and Applied Mathematics
AIP Conf. Proc. 2177, 020039-1–020039-5; https://doi.org/10.1063/1.5135214
Published by AIP Publishing. 978-0-7354-1924-7/$30.00

MIS-SPECIFICATION TEST FOR STOCHASTIC LINEAR REGRESSION MODEL BASED ON DURBIN-WATSON TEST STATISTIC

Consider the standard stochastic linear regression model

$$Y = X\beta + \varepsilon$$

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

One may obtain the OLS estimator $\hat{\beta}$ as the BLUE for β which is given by

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

OLS residual vector emeans the difference of column vectors Y and $\hat{Y}$ i.e., Y and $X\hat{\beta}$

Suppose that X^* be an excluded relevant independent variable which is excluded from the stochastic linear regression model. Now arrange the OLS residual in an order according to the increasing values of excluded independent variables X^*.

Let the ordered OLS residual be denoted by $\left[e_{(1)}, e_{(2)}, \ldots e_{(n)}\right]$

By using Durbin –Watson statistic one may test the null hypothesis that the model is mis-specified.

The Durbin-Watson test statistic is given by

$$d^* = \frac{\sum_{i=2}^{n}\left[e_i - e_{i-1}\right]}{\sum_{i=1}^{n} e_i^2}$$

Using the Durbin –Watson table of critical values if the calculated value of d^* is significant then the null hypothesis of misspecification of the stochastic linear regression may be accepted at the given level of significance.

RAMSEY'S REGRESSION SPECIFICATION ERROR TEST (RESET) FOR THE MISSPECIFICATION OF STOCHASTIC LINEAR REGRESSION MODEL

Consider the standard stochastic linear regression model in an alternative form

$$Y_i = X_i'\beta + \varepsilon_i \quad \text{, i=1, 2…n}$$

where X_i' 's are $(k \times 1)$ vectors; β is $(k \times 1)$ unknown parametric vector; ε_i 's are random error observations and Y_i is i^{th} entry on Y.

By augmenting model with a set of independent variables Z_i 's gives the augmented stochastic linear regression model as $Y_i = X_i'\beta + Z_i'\delta + \varepsilon_i \quad ; i = 1, 2, ..n$

where Z_i may be chosen independent variables which are responsible for misspecification of the model. Under mis-specification test one may use F-test to test $H_0 = \delta = 0$

Here the crucial problem is the choice of Z_i 's.Often Z_i 's may be approximated by higher powers in X_i 's or alternatively higher powers of estimates $\hat{Y}_i = X_i'\hat{\beta}$ where $\hat{\beta}$ is the OLS estimator of β .Thus the Ramsey RESET test consists of the following steps

I. Regress Y_i on X_i' and obtain $\hat{Y}_i'$.

II. Regress Y_i on $X_i', \hat{Y}_i^2, \hat{Y}_i^3, \hat{Y}_i^4$

III. Test the coefficients of all the powers of $\hat{Y}_i'$ are zero by using F –test.

Under null hypothesis the F-test statistic follows $F_{[q,(n-q)]}$ where q is the number of Z_i 's in the model.

The best choice for Z_i 's may be $Z_i = [X_i^2, X_i^3, X_i^4]$. The F-test statistic may be given by

$$F = \frac{[R_*^2 - R^2]/q}{[1-R_*^2]/(n-q)} \sim F_{[q,(n-q)]}$$

where R^2: The coefficient of multiple determinative without $Z_i's$

R_*^2 : The coefficient of multiple determinative including Z_i 's

q= Number of Z_i 's in the model.

LAGRANGE MULTIPLIER TEST FOR MISS-SPECIFICATION OF STOCHASTIC LINEAR REGRESSION MODEL

Consider the standard stochastic linear regression model

$$Y = X\beta + \varepsilon \tag{1}$$

where Y is $n \times 1$ matrix, X is $n \times k$ matrix, β is $k \times 1$ matrix and ε is $n \times 1$ matrix such that $\varepsilon \sim N(0, \sigma^2 I_n)$

The BLUE for β is the OLS estimator $\hat{\beta}$ which is given by $\hat{\beta} = (X'X)^{-1} X'Y$ (2)

Suppose that Z be $(n \times p)$ matrix of regressors which are included in the model and their coefficients to be tested for mis-specifiaction of the model

Write the augmented stochastic model as

$$Y = X\beta + z\delta + U \tag{3}$$

Where δ is $(p \times 1)$ vector of parameters associated with Z . U is $(n \times 1)$ vector of errors.

By using the OLS estimation one may estimate the restricted stochastic linear regression model (1) and obtain the OLS residual vector $e = [Y - \hat{Y}]$ where $\hat{Y} = X\hat{\beta}$

If one assumes that unrestricted regression model or augmented regression model (3) is the true regression model then the OLS residual vector *e* should be related to *Z*.

This suggests that one regress *e* on all the regressors including those in the restricted regression which gives

$$Y = X\beta + Z\delta + U \tag{4}$$

Here Uis an error term with usual properties.For large sample size

$$\chi^2 = nR^2 \overset{asy}{\sim} \chi_p^2 \tag{5}$$

Where p is the number of restrictions imposed by the stochastic restricted regression

R^2 is the coefficient of multiple of determination obtained from the augmented stochastic regression model.

If χ^2 is the significant at the chosen level of significance then one may reject the restricted stochastic regression model. Otherwise one may not reject it and reject augmented stochastic regression model. It should be noted that Lagrange multiplier test is an alternative to Ramsey's RESET test.

UTTS' RAINBOW TEST FOR MISSPECIFICATION OF STOCHASTIC LINEAR REGRESSION MODEL.

If the true functional relation is not linear model then the Rainbow test is an appropriate test for misspecification of the model. This test rejects the null hypothesis of linear relationship under diagonization criterion.

Consider the standard stochastic linear regression model as

$$Y_{n\times 1} = X_{n\times k}\beta_{k\times 1} + \varepsilon_{n\times 1} \tag{6}$$

such that $\varepsilon \sim N\left(0, \sigma^2 I_n\right)$

Suppose that ee' be the ordinary least squares RSS based on sample

where $e = [Y - \hat{Y}] = [Y - X\hat{\beta}]$ (7)

and $\hat{\beta} = \left(X'X\right)^{-1} X'Y$

One may obtain $\tilde{e}'\tilde{e}$ as the OLS- RSS based on the center half of the sample.

Generally one may state the null and alternative hypothesis as

$$H_0 = E\left[\frac{e'e}{(n-k)}\right] = \sigma^2 = E\left[\frac{\tilde{e}'\tilde{e}}{\left(\frac{n}{2} - k\right)}\right]$$

$$H_1 = E\left[\frac{e'e}{(n-k)}\right] > E\left[\frac{\tilde{e}'\tilde{e}}{\left(\frac{n}{2}-k\right)}\right]$$

One may consider ee' as RRSS under H_0 and $\tilde{e}'\tilde{e}$ URRSS.

For testing H_0, the UTTS' Rainbow test statistic is given by

$$F = \frac{\left(e'e - \tilde{e}'\tilde{e}\right)/\left(n/2\right)}{\tilde{e}'\tilde{e}/\left(n/2 - k\right)} \sim F_{\left[\frac{n}{2},\left(\frac{n}{2}-k\right)\right]}$$

Compare the calculated value of F-statistic with its critical value and draw the inference accordingly. The crucial problemof thistest statistic is based on selecting suitable middle half of the sample.

The UTTSandChow test statistic behave alike except in choosing sub sample. More often Rainbow test places the data using their distances from mean andchooses the initial half of the observations in the case of simple regressor.

CONCLUSIONS

In the above research article some tests for misspecification in stochastic linear regression model namely Durbin-Watson test, Ramsey's regression specification error test (RESET), Lagrange multiplier test (LM test), UTTS' Rainbow test have been presented. Selection of the variables for the specification of stochastic linear regression model is one of the serious problems of model building by which one may select optimum number of variables which have more influence on dependent variables in the model. In the context of future research a number of methods can be developed as selection criteria of independent variables for specification of stochastic linear regression model.

REFERENCES

1. Raymond J. G. M. Florax and Peter Nijkamb, Tinbergen Institute Discussion Paper, 1-2003-081/3 (2003).
2. Russell Davidson and Tame Mackinnon, Inter. Economic Review, **25**, 485-502 (1984).
3. Jean-Marie Dufour, McGill University (2011).
4. George L. Light, Applied Mathematical Sciences, **4**, 225-236 (2010).
5. T.W. Anderson, Sankhya: The Ind. J. of Stat., **64**, 193-205 (2002).
6. Audley, R.J., Psychol. rev **67,** 1-15, (1963).
7. Byson J. T. Morgan, CRC Press, 978-1-58488-666-2 (2008).
8. Berry L. Nelson, McGraw-Hill, 978-0070462137 (1995).
9. Cohan .J.E, Harvard University Press (1971).
10. Pinsky, M.A and Samuel Kanlin, Academic Press, 978-0-12-381416-6 (2011).