International Journal of Modern Mathematical Sciences, 2014, 11(3): 144-151

Modern Scientific Press

International Journal of Modern Mathematical Sciences

ISSN:2166-286X

Journal homepage:www.ModernScientificPress.com/Journals/ijmms.aspx

Florida, USA

Article

On the Performance of RESET and Durbin Watson Tests in Detecting Specification Error

Babatunde O.S[1], Oguntunde P.E[2], Ogunmola A. O[3] and Balogun O.S[4]

[1]*Department of Mathematics, Statistics & Computer Science, Kwararafa University, Wukari, Taraba State, Nigeria.*

[2]*Department of Mathematics, Covenant University, Ota, Ogun State, Nigeria.*

[3]*Department of Mathematics & Statistics, Federal University, Wukari, Taraba State, Nigeria.*

[4]*Department of Statistics and Operations Research, Modibbo Adama University, Yola, Adamawa State, Nigeria*

* Author to whom corresponding should be addressed; Email: olamee2323@gmail.com

Article history: Received 23 June 2014, Received in revised form 15 September 2014, Accepted 21 September 2014, Published 27 September 2014.

Abstract: When a model is created which correctly leaves out one or more important variables, one rarely know which test has the highest power for detecting the associated specification error. This research adopts the use of bootstrapping experiment. The models investigated consist of three omitted variables which have a coefficient that varies from 0.1 through 1 and 2. A bootstrap simulation approach was used to generate data for each of the models at different sample sizes (n) 20, 30, 50, and 80 respectively, each with 100 replications(r). For the models considered, the experiment reveals that the Ramsey Regression Equation Specification Error Test (RESET test) is more efficient than that of Durbin-Watson test in detecting the error of omitted variable in specification error.

Keywords: Bootstrap, Error, Regression, Simulation, Specification, RESET

Mathematical Subject Classification (2000): 62J05

---

1. Introduction

Knowing the consequence of specification error in linear regression is one thing but finding out whether one has committed such error is quite another since specification error cannot be committed deliberately. When a variable is inappropriately omitted from a model, the obvious remedy is to include that variable in the analysis (assuming the data for that variable is available). Specification error could occur because of the following; when an incorrect functional form is adopted, when we omit a relevant variable, when we include an irrelevant variable in the model.

The models considered in this work satisfy the assumptions of linear regression model (LRM) but some very important questions that arise when there is specification error are: what would happen if we use the LRM when the assumptions are not met, i.e. when it is not appropriate? What are the properties of the OLS estimators under a specification error?

According to Kelvin A. Clarke (2006), when a model is misspecified due to omitted variable, there is always the fear of omitted variable bias. He said a key underlying assumption is that the danger posed by omitted variable can be ameliorated by the inclusion of control variables. Also small amount of nonlinearity in control variables can also have a deleterious effect on the models considered (Achen 2005, Welch 1975).

Thursby(1979) compared the power of the test RESET to that of autocorrelation tests in detecting the errors of omitted variables in regression analysis. The experiments reveal that the RESET test is the most powerful test for detecting specification errors and is robust to autocorrelation. Research also suggest that RESET tests for GLMs have reasonable power properties in medium to large samples for testing functional and omitted variable in linear regression (Sapra 2005).

This research article is aimed at comparing the performance of the Ramsey's RESET test and the Durbin Watson test in detecting the error when variables are omitted in a regression model. We make use of the bootstrap simulation approach. In the next section, we explain how the data used was generated, followed by the procedure of the RESET test, procedure of the Durbin Watson test. This is followed by the discussion of our results, the conclusion and the references.

2. Methodology

Consider a standard linear regression model given as;

$$Y = X\beta + U \quad (1)$$

where;

Y is an $n \times 1$vector of dependent variables

X is an $n \times k$matrix of repressors

β is a $k \times 1$vector of parameters

U is an$n \times 1$ vector of disturbances and it is normally distributed with covariance matrix proportional to the identity matrix.

A three model of the form:

Model	Specification	Problem
1)	True: $y_t = 1.0 - 0.4x_{3t} + x_{4t} + 0.1x_{2t} + u_t$	
	Null: $y_t = \beta_0 + \beta_1 x_{4t} + \beta_2 x_{3t} u_t$	Omitted Variable
2).	True: $y_t = 1.0 - 0.4x_{3t} + x_{4t} + x_{2t} + u_t$	
	Null: $y_t = \beta_0 + \beta_1 x_{4t} + \beta_2 x_{3t} u_t$	Omitted Variable
3).	True: $y_t = 1.0 - 0.4x_{3t} + x_{4t} + 2.0x_{2t} + u_t$	
	Null:$y_t = \beta_0 + \beta_1 x_{4t} + \beta_2 x_{3t} u_t$	Omitted Variable

The true model is the model that has been specified correctly without any specification error and the null model is the model that contains the problem of omitted variable (for instance, x_{2t} is being omitted from all the three models). Observations on the dependent variables are generated according to one of the specifications labeled true.

The criteria for evaluating the performance of the estimators in this research are the:

1) Mean: $\hat{\beta} = \frac{\sum_{i=1}^{r} \beta}{r}$, where r is the number of replications.
2) $Bias(\hat{\beta}) = \hat{\beta} - \beta$
3) $Var(\hat{\beta}) = \frac{1}{r}\sum_{i=1}^{r}(\hat{\beta} - \beta)^2$
4) $MSE(\hat{\beta}) = \frac{1}{n}E(\hat{\beta} - \beta)^2$
5) $RMSE(\hat{\beta}) = \sqrt{MSE(\hat{\beta})}$

Which are used to check if the models satisfy the assumptions of the linear regression model (LRM) and also to check the effect of omitted variables in the models before proceeding to test for specification error.

2.1. Data Generation

For the bootstrap experiment, the study considers the specification labeled 'true model' from the above models. Firstly, we considered the first model;$y_t = 1.0 - 0.4x_{3t} + x_{4t} + 0.1x_{2t} + u_t$, we assigned numerical values to all the parameters $(\beta_0 = 1, \beta_2 = 0.1, \beta_3 = -0.4, \beta_4 = 1)$in the model, the variance σ^2 is also assigned a numerical value on the basis of assumedσ^2, then the disturbance term U is generated.

The U generated was standardized. A random sample of size (n) of X was then selected from a pool of random numbers and numerical values for $y_t = 1.0 - 0.4x_{3t} + x_{4t} + 0.1x_{2t} + u_t$was computed for each of the sample sizes using Microsoft Excel software. The $x's$and $y's$ generated were copied from Microsoft Excel into STATA and then bootstrapped and replicated 100 times using a STATA command, each replication produces a bootstrap sample which gives distinct values of y that leads to different estimates of $\beta's$for each bootstrap sample regression of y on fixed$x$. The procedure above is then repeated for different sample sizes and was also performed on each of the three models.

2.2. Procedure for RESET (Regression Specification Error Test)

Using the equation;

$$Y_t = \beta_0 + \beta_1 X_{1t} + \cdots + \beta_k X_{kt} + U_t \qquad (2)$$

Also consider the model;

$$\hat{Y} = E\left[\frac{y}{x}\right] = \beta X \qquad (3)$$

Introducing $\hat{y}$ as a form of additional regressor(s);

We introduce $\hat{Y}^2, \dots, \hat{Y}^k$. The Ramsey's RESET test then tests whether $(\beta_1 X)^2, (\beta_1 X)^3, \dots, (\beta_1 X)^k$ has any power to explain Y. This is executed by estimating the following linear regression equation;

$$Y = \beta + \beta_1 \hat{Y}^2 + \cdots + \beta_k \hat{Y}^k \qquad (4)$$

Further test by a means of F- test whether β_1 through β_{k-1} are zero. If the null hypothesis states that all regression coefficients of the nonlinear terms are zero is rejected, then the model suffers from misspecification. That is,

$H_o: U \sim N(0, \sigma^2)$

Vs

$H_o: U \sim N(0, \sigma^2)$ where $U \neq 0$

The test is based on the argument regression

$Y = \beta_0 + X\beta + U$

The F-test procedure follows;

$$F = \frac{R^2_{new} - R^2_{old} / \text{number of parameters}}{1 - R^2_{new} / n - \text{number of parameters in the new model}} \qquad (5)$$

Let R^2obtained from (5) be R^2_{new} and that obtained from (2) be R^2_{old}.

If the computed F-value is significant at the chosen level of significance(α), we therefore accept the hypothesis that the model (2) is mis-specified.

Using STATA package, we subject the result of the bootstrap to analysis of the test RESET using the command ovtest which computes the RAMSEY RESET test.

2.3. Procedure for Durbin-Watson Test

From the assumed model, we obtained the Ordinary Least Squares (OLS) residuals. Since the assumed model is mis-specified, we ordered the residuals in steps, in increasing values of X, and then we compute the *d statistic* from the residuals as;

$$d = \frac{\sum_{t=2}^{n}(\hat{U}_t - \hat{U}_{t-1})^2}{\sum_{t=1}^{n}\hat{U}_t^2}$$

Where;

U_t is the usual OLS residual

U_{t-1} is the lagged residuals.

From the Durbin Watson table (*d* tables), if the residual '*d*' is significant, one would accept the hypothesis of the model specification, if it turns out to be the case, the remedial measures will naturally suggest themselves.

STATA was used to order the residuals using the command tssetvar(name) and subsequently to compute the Durbin-Watson statistic using the command estatdwatson.

NOTE: Durbin Watson statistic ranges in value from 0 – 4. A value near 2 indicates non autocorrelation, a value towards 0 indicates positive autocorrelation and a value towards 4 indicates negative autocorrelation.

2.4. Result and Discussion

In this section, we make a critical comparison between the performance of RESET and Durbin-Watson tests in detecting error when a variable is omitted in a regression model.

The summary of the test of hypothesis of this research is given below:

H_o: *There is no specification error*

Vs

H_1: H_o *is false*

Take $\alpha = 0.05$

Percentage Rejection Region of the Hypothesis

Table 1: Result for Model 1

N	RESET	DWATSON
20	51.72	1.98
30	33.38	1.98
50	27.64	1.76
80	67.05	1.76

In model 1, the performance of RESET produced the best result at $n = 80$, the performance of Durbin-Watson falls within two percent of the rejection region.

Table 2: Result for Model 2

N	RESET	DWATSON
20	92.58	1.65
30	85.89	2.17
50	31.74	1.92
80	95.51	1.84

In model 2, RESET has the highest percentage based on the performance. It produced the best result at $n = 80$. The performance of Durbin-Watson test increased above two percent of the rejection region at $n = 20$

Table 3: Result for Model 3

N	RESET	DWATSON
20	62.00	1.48
30	28.96	1.96
50	35.79	1.99
80	78.72	1.92

For model 3, the performance of RESET produced the best result at $n = 80$ while the performance of Durbin-Watson test falls within two percent of the rejection region.

Comparing the percentage of rejection as a whole, for $n = 20$, the performance of RESET are quite good in all part but we obtained the best result when $n = 80$. This simply implies that as sample size increases, the percentage of rejection of RESET gets better than the Durbin Watson.

Comparing the percentage of rejection based on the same sample size, at $n = 20$, the performance of RESET is the best, when $n = 30$, for RESET, model 2 still perform better than other models but the effect tends to decrease as the sample size increases. At $n = 50$, the RESET does not perform well with all the result less than 50% but the RESET still has a better result than Durbin-

Watson and at $n = 80$, the result of RESET once again gets better especially in model 2 as compared to the Durbin Watson test.

3. Conclusion

We have examined the performances of two powerful tests; the RESET and Durbin Watson tests in detecting specification error in the presence of omitted variables in a regression model. The bootstrap simulation experiment indicates that the RESET test is more powerful and robust as compared to the Durbin Watson test.

References

[1] Achen, Christopher H..Why Lagged Dependent Variables Can Supress the Explanatory Power of Other Independent Variables. *Annual Meetings of the Political Methodology Section of the American Political Science Association, UCLA*, July 20-22, 2000.

[2] Anya McGuirk & Aris Spanos. The Linear Regression Model with Autocorrelated Errors: Just Say No to Error Autocorrelation. *Annual Meeting of the American Agricultural Economics Association*, July 28-31, 2002.

[3] Babatunde O.S, Ikughur A.J, Ogunmola A.O, Oguntunde P. E. Effect of omitted variable in misspecification error in regression analysis, *International Journal of Modern Mathematical Sciences*, 11(1)(2014): 49-57.

[4] Babatunde O. S, Ikughur A.J, Oguntunde P. E, Ogunmola A. O., On the effect of autocorrelation in regression model due to specification error, *International Journal of Modern Mathematical Sciences*. 10(3)(2014): 239-246. http://www.modernscientificpress.com/Journals/ViewArticle.aspx?XBq7Uu+HD/8eRjFUGMqlRdBaE41K3DNr06C/KfEk9T6KTMAZuAel3/IqG0I5hUUr

[5] Cochrane, D. and Orcutt, G. H., Application of Least Squares Regression to Relationships Containing Auto-Correlated Error Terms. *Journal of the American Statistical Association*, 44 (245)(1949): 32-61. doi:10.1080/01621459.1949.10483290

[6] Effron, B. and Tibshirani R.J., An Introduction to the Bootstrap, *Chapman & Hall/CRC*. New York.1993.

[7] Hesterberg, T., Simulation and Bootstrapping for Teaching Statistics. *American Statistical Association: Proceedings of the Section on Statistical Education*, (1998), p44-52.

[8] Kayode Ayinde, Emmanuel O. Apata, Oluwayemisi O. Alaba., Estimators of Linear Regression Model and Prediction under Some Assumption Violation. *Open Journal of Statistics*, 2(2012): 534-546.http://dx.doi.org/10.4236/ojs.2012.25069

[9] Kayode Ayinde and B.A Oyejola, A Comparative Study of the Performance of the OLS and Some GLS Estimators When Stochastic Regressors are Correlated with the Error Terms. *Research Journal of Applied Sciences,* 2(2007): 215-220.

[10] Thursby, J. G., Alternative Specification Error Tests: A Comparative Study. *Journal of the American Statistical Association*. 74(365)(1979): 222-225. DOI:10.1080/01621459.1979.1048164

[11] Thursby, J. G and Schmidt, P., Some Properties of Tests for Specification Error in a Linear Regression Model. *Journal of the American Statistical Association.* 72(359)(1977): 359

[12] Vaduva, I., Bootstrap Method in Density Estimation, *St. Cerc. Mat., Tom,* 46(3)(1994): 397-406