International Journal of Innovation and Applied Studies
ISSN 2028-9324 Vol. 8 No. 1 Sep. 2014, pp. 341-345

http://www.ijias.issr-journals.org/

Comparative Appraisal of some Specification Error Test

O.S Babatunde[1], O. S Balogun[2], and T.J. Akingbade[3]

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

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

[3]Department of Statistics, Kogi State University, Ayingba, Kogi State, Nigeria

ABSTRACT: This research work compared the power of some specification error test using bootstrapping experiment to generate the data for each of the models at different sample sizes (n) 20, 30, 50, and 80 respectively, each with 100 replications(r). The models in this research consist of three omitted variables. For the models considered, the experiment reveals that the Ramsey Regression Equation Specification Error Test (RESET test) is more efficient than that of Q test in detecting the error of omitted variable in specification error.

KEYWORDS: Bootstrap, Error, Simulation, Specification, RESET, Q test.

1 INTRODUCTION

Specification is a process of developing a regression model. It involved the procedure of an appropriate functional form for the model and choosing which variable to include. Specification error occurs when an independent variable is correlated with the error term i.e. when a wrong model has been estimated. It implies that at least one of the key features or assumptions of the model is incorrect. And in consequence, estimation of the model may yield results that are incorrect or misleading.

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 mis-specified 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).

Olubusoye O. E. et el (2004) compared the power of RESET, White Test and Q- test in detecting specification errors arising from omitted variables, functional misspecification and contemporaneous correlation residuals. He concluded that RESET is the most powerful test for detecting incorrect functional form and the test is robust to autocorrelation and heteroscedastic disturbance. 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).

Bootstrap is a particular resampling scheme with replacement. In statistics and econometrics, bootstrapping has come to mean to resample repeatedly and randomly form an original initial sample using each bootstrapped sample to compute a statistic. The resulting empirical distribution of the statistic is then examined and interpreted as an approximation to the true sampling distribution.

It is one of the several econometric methods that have grown in popularity as a consequence of the great advances in computing technology.

This research article is aimed at comparing the performance of the Ramsey's RESET test and the Q 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 Q test. This is followed by the discussion of our results, the conclusion and the references.

2 METHODOLOGY

Consider the standard linear regression model given as

$$Y = X\beta + U$$

Where: Y is an n x 1 vector of dependent variables

X is an n x k matrix of regressors

β is a k x 1 vector of parameter

U is an n x 1 vector of disturbance and 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 contained the problem of omitted variable i.e x_{2t} is been omitted for all the three models. Observations on the dependent variables are generated according to one of the specification labeled true.

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

- Mean, $\hat{\beta} = \frac{\sum_{i=1}^{r}\beta}{r}$, where r = number of replications.
- Bias$(\hat{\beta}) = \hat{\beta} - \beta$
- Variance$(\hat{\beta}) = \frac{1}{r}\sum_{i=1}^{r}(\hat{\beta} - \beta)^2$
- MSE$(\hat{\beta}) = \frac{1}{n}E(\hat{\beta} - \beta)^2$
- RMSE$(\hat{\beta}) = \sqrt{\text{MSE}(\hat{\beta})}$

Which are used to check if the models satisfy the assumptions of linear regression model (LRM) and also check the effect of omitted variable 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 i$(\beta_0 = 1, \beta_2 = 0.1, \beta_3 = -0.4, \beta_4 = 1)$ned a numerical value on the basis of assumed σ^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'sand 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 β's for each bootstrap sample regression of y on fixedx. 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 Q TEST

From the assumed model, we obtained the OLS residual, we ordered the residuals in step in an increasing values of X and compute the Q- statistic from the residuals given as,

$$Q_n = \frac{\sum_{n=1}^{n} \sum_{n=1}^{n-1} (\bar{u}_t - \bar{u})(\bar{u}_{t-1} - \bar{u})}{\sum_{n=1}^{n} (\bar{u}_t - \bar{u})^2}$$

This follows a null chi square distribution at appropriate level of significance, using STATA package, the residual will be ordered using command tsset var (name) and subsequently compute the Q statistic using the command wntestq (variable name).

3 RESULT AND DISCUSSION

In this section, we make a critical comparison between the performance of RESET and Q 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;

$\mathrm{H_o}$: There is no specification error

Vs

$\mathrm{H_1}$: $\mathrm{H_o}$ is false

Take $\alpha = 0.05$

Percentage Rejection Region of the Hypothesis

Table 1: Result for Model 1

N	RESET	Q TEST
20	51.72	16.79
30	33.38	16.14
50	27.64	48.18
80	67.05	16.85

In model 1, the performance of RESET produced the best result at$n = 80$, except at $n = 50$ when the performance of Q test has substantial power.

Table 2: Result for Model 2

N	RESET	Q test
20	92.58	21.42
30	85.89	24.47
50	31.74	76.88
80	95.51	26.94

In model 2, RESET has the highest percentage based on the performance. It produced the best result at$n = 80$. The performance of Q test increased above seventy percent of the rejection region at $n = 50$

Table 3: Result for Model 3

N	RESET	Q test
20	62.00	47.50
30	28.96	15.77
50	35.79	93.30
80	78.72	54.47

For model 3, the performance of RESET produced the best result at $n = 80$ while the performance of Q test increased more than ninety three percent.

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 Q test.

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 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% compared to the Q test which has a higher percentage but the RESET still has a better result than the Q test as a whole and at $n = 80$, the result of RESET once again gets better especially in model 2 as compared to the Q test.

4 Conclusion

We have examined the performances of two powerful tests; the RESET and Q test 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 Q test.

References

[1] Achen, Christopher H. "Why Lagged Dependent Variables Can Supress the ExplanatoryPower of Other Independent Variables".*Annual Meetings of the Political Methodolody 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] Asteriou, Dimitrios, Hall, Stephen G. (2011). Misspecification, Wrong Regressors, Measurment Errors and Wrong Functional Forms. Applied Econometrics (Second ed.). London: Palgrave Macmillian. pp. 172-179.

[4] Babatunde O. S, Ikughur A. J, Ogunmola A.O, Oguntunde P. E. On the Effect of Autocorrelation in Regression Model due to Specification Error. International Journal of Modern Mathematical Sciences, 2014, 10(3): 239-246

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

[6] Hesterberg, T. (1998). "Simulation and Bootstrapping for Teaching Statistics" *American Statistical Association: Proceedings of the Section on Statistical Education*, 44-52.

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

[8] *Mackinnon, James G. (1992). Model specification Test and Artificial Regression. Journal of Economic Literature 30(1): 102-146*

[9] Thursby, J. G.(1979). "Alternative Specification Error Tests: A Comparative Study". *Journal of the American Statistical Association*. 74 (365): 222-225. DOI:10.1080/01621459.1979.10481641

[10] Thursby, Jerry G, Schmidt, P. (1977). "Some Properties of Tests for Specification Error in a Linear Regression Model".*Journal of the American Statistical Association.*Vol 72, No. 359