The Power of the Tests on the Multiple Regression Model

by Budi Pratikno

Submission date: 09-Feb-2021 02:32PM (UTC+0700)
Submission ID: 1505272524
File name: IJAST_Budi_P_14065-Article_2020.pdf (476.67K)
Word count: 2654
Character count: 13244

International Journal of Advanced Science and Technology
Vol. 29, No. 6, (2020), pp. 3274 – 3279

The Power of the Tests on the Multiple Regression Model

Budi Pratikno[1,*], Shahjahan Khan[2]

[1]*Department of Mathematics Jenderal Soedirman University, Purwokerto, 53122,Indonesia*
[2]*Department of Mathematics and Computing University of Southern Queensland Toowoomba, Queensland, Australia*

[1,*]*bpratikto@gmail.com,* [2]*Shajahan.Khan@usq.edu.au.*

Abstract

Many Authors have already studied estimation and testing using non-sample prior information (NSPI). They showed improvement of the estimation and the testing using NSPI. Here, we then study the power of the tests in testing hypothesis of the coefficient regression parameters on multiple regression model (MRM) when non-sample prior information (NSPI) is available on another subset of the regression parameters. Following the previous research of Pratikno [5], we used three different tests, namely unrestricted test (UT), restricted test (RT) and pre-test test (PTT). We then presented and graphically analyzed the power of the tests using ***R*** code. The result showed that the power of the PTT be a eligible choice among them.

Keywords: multiple regression, power of the tests, regression parameters

1. Introduction

Following the previous research such as Bancroft [32], Saleh [2], Yunus and Khan [18] and Pratikno [5], that inferences population parameters can be improved using non-sample prior information (NSPI). In term of the NSPI treatment, we represented again the classified of the NSPI as: (1) unknown if the NSPI is not available, (2) known if the exact value is available from the NSPI, and (3) uncertain if the suspected value is unsure (Pratikno, [5]). Many authors such as Khan and Saleh [27]; Khan et al. [28]; Saleh [2] and Yunus [17] also already studied the preliminary testing (pre-testing) on uncertain NSPI is to improve the quality of estimator. Moreover, Saleh [2], presented three estimators on the NSPI treatment, namely (i) the unrestricted estimator (UE), (2) (ii) the restricted estimator (RE), and (iii) the preliminary test estimator (PTE) (see Judge and Bock, [11]; Saleh, [2]). Khan [22], and Khan and Hoque [29] then provided the UE, RE and PTE for different linear models. Furthermore, for the testing purpose, Yunus [17] and Pratikno [5] described three statistical tests, namely the (i) unrestricted test (UT), (ii) restricted test (RT) and (iii) pre-test test (PTT).

Moreover, many authors have contributed to the estimation of parameter(s) areas in the presence of uncertain the NSPI Such as Bancroft ([32], [33]), Hand and Bancroft [6], and Judge and Bock [11]. Then, Khan ([23], [24]), Khan and Saleh ([25], [26], [30], [31]), Khan and Hoque [29], Saleh [2], and Yunus [17] also covered various work in the area of improved estimation using NSPI. Here, we noted some authors such as Tamura [16], Saleh and Sen ([3], [4]), Yunus and Khan ([18], [19], [20], [21]), and Yunus [17] have been started to study testing hypothesis using NSPI on nonparametric methods. Furthermore, Pratikno [5] started to use the NSPI in testing hypothesis on parametric

ISSN: 2005-4238 IJAST

models (i.e. regression models). Following Pratikno [5], we noted that the power and size of the tests are used to compare the performance of the UT, RT and PTT.

The paper presented the introduction in Section 1. Section 2 introduced theory of multiple regression model. The power function of the tests are obtained in Section 3. An illustrative example is given in Section 4. The conclusion is provided in Section 5.

2. Multiple Regression Model

For an n pair of observations on k independent variables ($X_1, \ldots, X_k$) and one dependent variable (Y), (X_{ij}, Y_i), for $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, k$, the multiple regression model (MRM) is defined as

$$Y_i = \beta_0 + \beta_1 X_{i1} + \ldots + \beta_k X_{ik} + e_i \qquad (1)$$

where Y_i is dependent (response) variable and X_{ik} is explanatory (predictor) variables. The equation (1) is then written in matrix form as

$$Y = \mathbf{X}\beta + e \qquad (2)$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_{r-1}, \beta_r, \ldots, \beta_k)'$ is a $(k + 1)$-dimensional column vector of unknown regression parameters, $\mathbf{Y} = (y_1, \ldots, y_n)'$ is $n\times 1$ vector of dependent variables, $\mathbf{X}$ is a $n\times(k + 1)$ matrix of the independent variables, and e is the error term as $N_n(\mathbf{0}, \sigma^2\mathbf{I}_n)$ with $\mathbf{I}_n$ is the identity matrix of order n and σ^2 is the variance of the error variables.

To test a subset of regression parameters when there is NSPI on another subset of the regression parameters, we then let $\boldsymbol{\beta}_1 = (\beta_0, \ldots, \beta_{r-1})$ be a subset of r regression parameters and $\boldsymbol{\beta}_2 = (\beta_r, \ldots, \beta_k)$ as another subset $(k+1-r)$ regression parameters with $\boldsymbol{\beta} = \boldsymbol{\beta}' = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2')$, where $\boldsymbol{\beta}_1'$ is a sub-vector of order r and $\boldsymbol{\beta}_2'$ is a sub-vector of dimension s = k + 1 − r. Furthermore, the matrix $\mathbf{X}_{(n\times(k+1))}$ is partitioned as $(\mathbf{X}_1, \mathbf{X}_2)$ with $\mathbf{X}_1 = (\mathbf{1}, \mathbf{x}_1, \ldots, \mathbf{x}_{r-1})$ and $\mathbf{X}_2 = (\mathbf{x}_r, \ldots, \mathbf{x}_k$. To test that $\mathbf{X}_1$ has no significant effect on the response, i.e. $H_0^1: (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2') = (\mathbf{0}, \boldsymbol{\beta}_2')$ for $r \le k$, we test a subset of the r regression parameters from the complete model (k variates). The F statistics is then used to test $H_0: \beta_r = \ldots = \beta_k = 0$ against H_a : not all $\beta^1 s$ are zero for the reduced model (Wackerly et al., [7]), that is

$$F = \left(\frac{SSE_r - SSE_c}{(k-(r-1))} \bigg/ \frac{SSE_c}{(n-(k+1))} \right) \approx F_{(k-(r-1)),n-(k+1)} \qquad (3)$$

where SSE_r is sum square error in reduced model and SSE_c is sum square error in complete model. The detail of the F test statistic for testing hypothesis is found on Ohtani and Toyoda [13], and Gurland and McCullough [12].

Following Pratikno (2012), we then test $\boldsymbol{\beta}_1$ when NSPI is available on the value of $\boldsymbol{\beta}_2$ for testing H_0 : $\mathbf{H}_{1(q\times r)}\boldsymbol{\beta}_{1(r\times 1)} = \mathbf{h}_{1(q\times 1)}$ against H_a : $\mathbf{H}_{1(q\times r)}\boldsymbol{\beta}_{1(r\times 1)} > \mathbf{h}_{1(q\times 1)}$. Following Saleh (2006) the test statistic for testing H_0 : $\mathbf{H}_1\boldsymbol{\beta}_1 = \mathbf{h}_1$ is given by

$$F^* = \frac{1}{qs_e^2}\left((H_1\tilde{\beta}_1 - h_1)'(H_1[X_1'X_1]^{-1}H_1')(\mathrm{H}_1\,\tilde{\beta}_1 - h_1)\right) \qquad (4)$$

ISSN: 2005-4238 IJAST

where the least square estimator (LSE) of $\boldsymbol{\beta}_1$ is

$$\tilde{\beta}_1 = (X_1'X_1)^{-1} X_1'Y = C_1^{-1} X_1'Y, \text{ where } C_1 = X_1'X_1,$$

and $s_e^1 = \frac{1}{n-r}(Y - X_1\tilde{\beta}_1)'(Y - X_1\tilde{\beta}_1)$ is an unrestricted unbiased estimator of σ^2 $(s_e^2 \to \sigma^2)$. Under H_a, F^* follows a noncentral F distribution with $(q, n-r)$ degrees of freedom (df) and noncentrality parameter $\frac{\Delta_1^2}{2}$, and under H_0, F^* follows a central F distribution with $(q, n-r)$ df, with

$$\Delta_1^2 = \frac{(H_1\beta_1 - h_1)'(H_1C_1^{-1}H_1')^{-1}(H_1\beta_1 - h_1)}{\sigma^2} \tag{5}$$

Here, we assumed that $P(F^* \le x) = G_{q,n-r}(x;\Delta_1^2)$ with $G_{q,n-r}(x;\Delta_1^2)$ is the cdf of a noncentral F distribution with $(q, n-r)$ df and noncentrality parameter $\frac{\Delta_1^2}{2}$. Detail of the noncentral F distribution is found Pratikno [5].

3. The Power of the Tests

Following Pratikno [5], we represented the power of the UT, the power of the RT and the power of the PTT for testing the above hypothesis on a subset of the regression parameters as follows.

3.1. The power of the UT

Following Pratikno [5], the formula of the UT is already derived as below, we then now used this formula to compute the power on generate simulation data

$$\begin{aligned} \pi^{UT}(\lambda) &= P(L^{UT} > F_{\alpha_1,q,n-r} \mid M_n) = 1 - P(L_1^{UT} \le F_{\alpha_1,q,n-r} - \Omega_{ut}) \\ &= 1 - P(L_1^{UT} \le F_{\alpha_1,q,n-r} - k_{ut}\zeta_1) \end{aligned} \tag{6}$$

where $\Omega_{ut} = \frac{\sigma}{qs_{ut}^2}(\lambda_1)'[\gamma_1]^{-1}(\lambda_1), \gamma_1 = H_1(X_1'X_1)^{-1}H_1', \zeta_1 = (\lambda_1)'[\gamma_1]^{-1}(\lambda_1)$ and $k_{ut} = \frac{\sigma}{qs_{ut}^2}$.

3.2. The power of the RT

Similarly, we follow Pratikno [5], the formula of the RT is presented in the equation (7).

$$\begin{aligned} \pi^{RT}(\lambda) &= P(L^{RT} > F_{\alpha_1,q,n-r} \mid M_n) = P(L_2^{RT} > F_{\alpha_2,q,n-r} - \Omega_{rt}) \\ &= 1 - P(L_2^{RT} \le F_{\alpha_2,q,n-r} - \Omega_{rt}) = 1 - P(L_2^{RT} \le F_{\alpha_2,q,n-r} - k_{rt}\zeta_1) \end{aligned} \tag{7}$$

where, $\Omega_{rt} = \frac{\sigma}{qs_{ut}^2}(\lambda_1)'[\gamma_1]^{-1}(\lambda_1), \gamma_1 = H_1(X_1'X_1)^{-1}H_1', \zeta_1 = (\lambda_1)'[\gamma_1]^{-1}(\lambda_1)$ and $k_{rt} = \frac{\sigma}{qs_{rt}^2}$.

3.3. The power of the PTT

In the similar way, following Pratikno [5], the formula of the PTT is then given in the equation (9).

$$\begin{aligned}\pi^{PTT}(\lambda) &= P(L^{PT} < F_{\alpha_3,q,n-s}, L^{RT} > F_{\alpha_2,q,n-r} \mid M_n) + P(L^{PT} \ge F_{\alpha_3,q,n-s}, L^{UT} > F_{\alpha_1,q,n-r} \mid M_n) \\ &= P\,[L^{PT} < F_{\alpha_3,q,n-s}]\; P\,[L^{RT} > F_{\alpha_2,q,n-r}] + d_{1r}(a,b) \\ &= [1 - P(L^{PT} > F_{\alpha_3,q,n-s})]\; P\,[L^{RT} > F_{\alpha_2,q,n-r}] + d_{1r}(a,b)\end{aligned} \tag{9}$$

where, $a = F_{\alpha_3,q,n-s} - \dfrac{\sigma}{q s_{pt}^2}(\lambda_2)'[\gamma_{pt}]^{-1}(\lambda_2) = F_{\alpha_3,q,n-s} - k_{pt}\zeta_2$, and $d_{1r}(a.b)$ is bivariate F probability integral and $d_{1r}(a,b)$ is defined as

$$d_{1r}(a,b) = \int_a^{\alpha}\int_b^{\alpha} f(F^{PT}, F^{UT})\, dF^{PT} dF^{UT} = 1 - \int_0^{b}\int_0^{a} f(F^{PT}, F^{UT})\, dF^{PT} dF^{UT} \tag{10}$$

with $b = F_{\alpha_1,q,n-r} - \Omega_{ut}$. The integral $\int_0^{b}\int_0^{a} f(F^{PT}, F^{UT})\, dF^{PT} dF^{UT}$ is the cdf of the correlated bivariate noncentral F distribution of the UT and PT (Amos and Bulgren, [8]), and it is computed using *R-code.*

4. A Simulation Study

Referring to Pratikno [5], a simulation was conducted using the generate random data using R. The independent variables (x_j, $j=1, 2, 3$) were generated from the uniform distribution ($n=100$). The error vector (e) was generated from the normal distribution ($n=100$). The MRM is then used by setting different β, and it is defined as $\beta_1 = (\beta_0, \beta_1)$, $\beta_2 = (\beta_2, \beta_3)$ with $\alpha = 0.05$. The graphs for the power are computed using the formulas in the equations (6), (7) and (9).

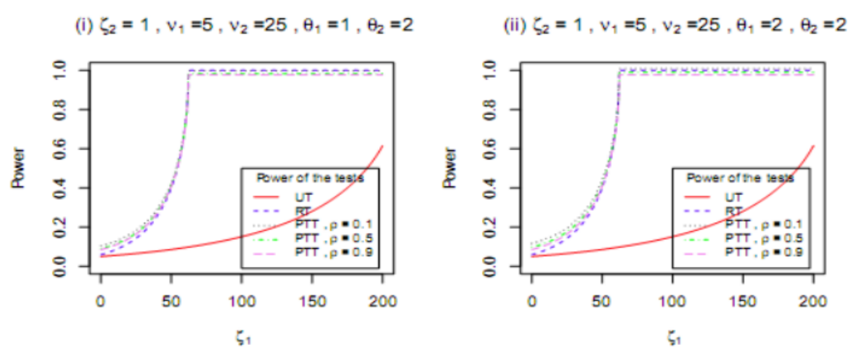

Figure 1: Power of the UT, RT and PTT against ζ_1 for some selected ρ, ζ_2, degrees of freedom and noncentrality parameters.

Figure 1. showed that the power of the UT is lower than both RT and PTT. It starts from a very small value (0.05 as minimum value) and slowly increases. The power of the RT reaches quickly to 1 for large ζ_1 . It is clear (see Figure 1.) that the power of the PTT depends on three parameters, namely ζ_1, ζ_2 and ρ , and the PTT is always larger than that

ISSN: 2005-4238 IJAST

of the UT and tends to be the same as of the RT. We then conclude that the PTT is more eligible test among them.

5. Conclusion

The result showed that the power of the RT and PTT are always greater than UT, but the power of the PTT tend to be larger or the same than RT. We then conclude that the PTT is more eligible test among them.

Acknowledgments

I thankfully to the LPPM UNSOED for providing me granting of research.

References

[1] A. H. El-Bassiouny and M. C. Jones. A bivariate *F* distribution with marginals on arbitrary numerator and denominator degrees of freedom, and related bivariate beta and *t* distributions. Statistical Methods and Applications, **18** (4), 465-481, 2009.
[2] A. K. Md. E. Saleh. Theory of preliminary test and Stein-type estimation with applications. John Wiley and Sons, Inc., New Jersey, 2006.
[3] A. K. Md. E. Saleh and P. K. Sen. Nonparametric estimation of location parameter after a preliminary test on regression. Annals of Statistics, **6** , 154-168, 1978
[4] A. K. Md. E. Saleh and P. K. Sen. Shrinkage least squares estima- tion in a general multivariate linear model. Procedings of the Fifth Pannonian Symposium on Mathematical Statistics, 307-325, 1982.
[5] B. Pratikno. Test of Hypothesis for Linear Models with Non-Sample Prior Information. Unpublished PhD Thesis, University of Southern Queensland, Australia, 2012.
[6] C. P. Han and T.A. Bancroft. On pooling means when variance is unknown. Journal of American Statistical Association, **63**, 1333-1342, 1968.
[7] D. D. Wackerly, W.Mendenhall III, and R. L.Scheaffer. Mathematical statistics with application, 7th Ed. Thomson Learning, Inc., Belmont, CA, USA, 2008.
[8] D. E. Amos and W. G. Bulgren. Computation of a multivariate *F* distribution. Journal of Mathematics of Computation, **26**, 255-264, 1972.
[9] D. G. Kleinbaum, L. L. Kupper, A. Nizam and K. E. Muller. Applied regression analysis and other multivariable methods. Duxbury, USA, 2008.
[10] F. J. Schuurmann, P. R. Krishnaiah, and A. K. Chattopadhyay. Table for a multivariate F distribution. The Indian Journal of Statistics **37**, 308-331, 1975.
[11] G. G. Judge and M. E. Bock. The Statistical Implications of Pre-test and Stein-rule Estimators in Econoetrics. North-Holland, NewYork,1978.
[12] J. Gurland and R. S. McCullough. Testing equality of means after a preliminary test of equality of variances. Journal of Biometrika, **49**(3-4), 403- 417, 1962.
[13] K. Ohtani and T. Toyoda. Testing linear hypothesis on regression coefficients after pre-test for disturbance variance. Journal of Economics Letters, **17**(1-2) (1985), 111-114, 1985.
[14] N. L. Johnson, S.Kotz and N. Balakrishnan. Continuous univariate distributions, Vol. 2, 2nd Edition. John Wiley and Sons, Inc., New York, 1995.
[15] P. R. Krishnaiah. On the simultaneous anova and manova tests. Part of PhD thesis, University of Minnesota, 1964.
[16] R. Tamura. Nonparametric inferences with a preliminary test.Bull. Math. Stat. **11**, 38-61, 1965.
[17] R. M. Yunus. Increasing power of M-test through pre-testing. Unpublished PhD Thesis, University of Southern Queensland, Australia, 2010.
[18] R. M. Yunus and S. Khan. Test for intercept after pre-testing on slope a robust method. In: 9th Islamic Countries Conference on Statistical Sciences (ICCS-IX): Statistics in the Contemporary World - Theories, Methods and Applications, 2007.
[19] R. M. Yunus and S. Khan. Increasing power of the test through pre-test a robust method. Communications in Statistics-Theory and Methods, **40**, 581-597, 2011a.
[20] R. M. Yunus and S. Khan. M-tests for multivariate regression model. Journal of Nonparamatric Statistics, **23**, 201-218, 2011b.
[21] R. M. Yunus and S. Khan. The bivariate noncentral chi-square distribution - Acompound distribution approach. Applied Mathematics and Computation, **217**, 6237-6247, 2011c.
[22] S. Khan. Estimation of the Parameters of two Parallel Regression Lines Under Uncertain Prior Information. Biometrical Journal, **44**, 73-90, 2003.

ISSN: 2005-4238 IJAST

[23] S. Khan. Estimation of parameters of the multivariate regression model with uncertain prior information and Student-*t* errors. Journal of Statistical Research, **39**(2) (2005), 79-94.

[24] S. Khan. Shrinkage estimators of intercept parameters of two simple regression models with suspected equal slopes. Communications in Statistics - Theory and Methods, **37**, 247-260, 2008.

[25] S. Khan. and A. K. Md. E. Saleh. Preliminary test estimators of the mean based on p-samples from multivariate Student-*t* populations. Bulletin of the International Statistical Institute. 50th Session of ISI, Beijing, 599-600, 1995.

[26] S. Khan. and A. K. Md. E. Saleh. Shrinkage pre-test estimator of the intercept parameter for a regression model with multivariate Student-t errors. Biometrical Journal, **39**, 1-17, 1997.

[27] S. Khan. and A. K. Md. E. Saleh. On the comparison of the pre-test and shrinkage estimators for the univariate normal mean. Statistical Papers, **42**(4), 451-473, 2001.

[28] S. Khan. , Z. Hoque and A. K. Md. E. Saleh. Improved estimation of the slope parameter for linear regression model with normal errors and uncertain prior information. Journal of Statistical Research, **31** (1), 51-72, 2002.

[29] S. Khan. and Z. Hoque. Preliminary test estimators for the multivariate normal mean based on the modified W, LR and LM tests. Journal of Statistical Research, Vol 37, 43-55, 2003.

[30] S. Khan. and A. K. Md. E. Saleh. Estimation of intercept parameter for linear regression with uncertain non-sample prior information. Statistical Papers. **46**, 379-394, 2005.

[31] S. Khan. and A. K. Md. E. Saleh. Estimation of slope for linear regression model with uncertain prior information and Student-*t* error. Communications in Statistics - Theory and Methods, **37**(16), 2564-258, 2008.

[32] T.A. Bancroft. On biases in estimation due to the use of the preliminary tests of singnificance. Annals of Mathematical Statistics, **15**, 190-204, 1944.

[33] T.A. Bancroft. Analysis and inference for incompletely specified models involving the use of the preliminary test(s) of singnificance. Biometrics, **20** (3), 427-442, 1964.

The Power of the Tests on the Multiple Regression Model

ORIGINALITY REPORT

8%	%	%	8%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Visvesvaraya Technological University, Belagavi Student Paper	5%
2	Submitted to Higher Education Commission Pakistan Student Paper	1%
3	Submitted to Mahidol University Student Paper	1%
4	Submitted to Virginia Community College System Student Paper	1%
5	Submitted to University of Southampton Student Paper	<1%

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		