ISSN: 0974-5823 Vol. 7 (Special Issue 5, April-May 2022)

International Journal of Mechanical Engineering

TESTING OF MULTIVARIATE NONLINEAR REGRESSION HYPOTHESIS USING NONLINEAR LEAST SQUARE (NLS) ESTIMATION

Dr. Kesavulu Poola[1], J. Anil Kumar[2], V. Pavankumari[3], Putta Hemalatha[4], Prof. M. Bhupathi Naidu[5].

[1]Associate professor, CMS - Jain University, Bengaluru.
[2]Assistant professor, S V College of Engineering, Tirupati.
[3]Assistant professor, Malla reddy Engineering College, Hyderabad.
[4]Assistant professor, Institute of Aeronautical Engineering, Hyderabad.
[5]Professor, Department of *Statistics*, S. V. University, Tirupati.

ABSTRACT

The present study paper confers the different models on testing the nonlinear regression hypothesis using nonlinear least squares (NLS) estimator through nonlinear studentized residuals and nonlinear predicted residual. Moreover this research paper mentioned internally nonlinear studentized residuals, externally nonlinear studentized residuals to test the hypothesis of multivariate nonlinear regression models. This research article proposes a new way of parameter estimation using nonlinear least squares method. According to Pesaran, M. Hashem, and Angus s. Deaton. (1978) uses nonlinear least squares asymptotic properties. The key principle of this article is to carter extremely pioneering model of testing nonlinear regression hypothesis using NLS estimator through nonlinear (internally and externally) studentized residuals and predicted NLS Methods.

KEY WORDS: *Nonlinear models, nonlinear least square estimation, Studentized and Predicted residuals*

1. INTRODUCTION

Least squares method and maximum likelihood methods are the quite popular in parameter estimation of linear models. In this approach, the method has to satisfy all pre assumption of normality with mean zero mean and unknown population variance. But in the case of nonlinear models building of construction of inferential facets together with Parameter estimation and hypothesis testing concerning the parameters of the nonlinear regression models are quite difficult. In recent era, researchers focused on the erection of well-organized parameter estimation of the nonlinear regression models. Since three decades these nonlinear models have been studied. The parameter estimation procedures and testing of hypothesis for nonlinear regression models and error assumptions are common analogous to those made for linear regression models. In the present research study some methods of testing multivariate nonlinear hypotheses using nonlinear least square estimation, studentized and predicted residuals for multivariate nonlinear models has been proposed.

2. LEAST SQUARE ESTIMATION OF MULTIVARIATE NONLINEAR REGRESSION MODEL

Suppose, standard multivariate nonlinear regression model.

$$Y_{it} = f_i\left(X_t, \theta\right) + e_{it} \qquad i = 1, 2, \ldots M, \qquad t = 1, 2, \ldots n$$

Where $\hat{e}_t = \left(\hat{e}_{1t}, \hat{e}_{2t}, \ldots \hat{e}_{Mt}\right) \qquad t = 1, 2, \ldots, n$

Then the multivariate least square estimator minimizes

$$S_n\left(\theta, \hat{\Sigma}_n\right) = \frac{1}{n}\sum_{t=1}^{n}\left[Y_t - f\left(X_t, \theta\right)\right]'\left(\hat{\Sigma}_n\right)^{-1}\left[Y_t - f\left(X_t, \theta\right)\right] \qquad \ldots(2.1)$$

Here we shall tax $\hat{\Sigma}_n$ be any random variable that converges almost surely to Σ and has $\sqrt{n}\left(\hat{\Sigma}_n - \Sigma\right)$ bounded in probability.

i.e., given S>0, there is bound 'b' and sample size N, such that

$$P\left(\sqrt{n}\left|\hat{\sigma}_{\alpha\beta n} - \sigma_{\alpha\beta}\right| < b\right) > 1 - S \qquad \forall n > N$$

Here $\sigma_{\alpha\beta}$ is the typical element Σ.

3. TESTING OF MULTIVARIATE NONLINEAR REGRESSION HYPOTHESIS TESTING USING NLLS ESTIMATOR

The , multivariate nonlinear regression model

$$Y_{it} = f_i\left(X_t, \theta^0\right) + e_{it} \qquad t = 1, 2.,,, n \qquad \ldots(3.1)$$

Where $f_i\left(X_t, \theta^0\right)$ is a functional form and e_{it} is i.i.d with $N\left(o, \sigma_{In}\right)$

$\therefore$ Test procedure for multivariate nonlinear hypothesis is

$H_0 : h\left(\theta^0\right) = 0$ ….. $H_1 : h\left(\theta^0\right) \neq 0$

Here $n\left(\theta^0\right)$ is continuously first-order differential function; Map's ¡ ^{p} and ¡ ^{q} with Jocobian.

$$\therefore H\left(\theta^0\right) = \frac{\partial}{\partial\theta'}\left[h\left(\theta^0\right)\right] \qquad \ldots(3.2)$$

Here $H\left(\theta^0\right)$ is an order of $q \times p$

In order to estimate $H\left(\theta^0\right)$ at $\theta^0 = \hat{\theta}_n^0$, where $\hat{\theta}_n$ i.e. iterative multivariate NLLS estimator of θ.

Here are two commonly used estimators of Σ and that satisfy the condition under both null and alternative hypothesis.

$$S(\theta, \Sigma) = \sum_{t=1}^{n}\left[Y_t - f\left(X_t, \theta\right)\right]' \sum^{-1}\left[Y_t - f\left(X_t, \theta\right)\right] \qquad \ldots(3.3)$$

be the sum of square residual, evaluated under the restricted nonlinear least square estimate.

According to the asymptotic F-test, for $H_0 : h(\theta) = 0$, the standard form is

$$F = \frac{\left[S\left(\tilde{\theta}, \hat{\Sigma}\right) - S\left(\hat{\theta}, \hat{\Sigma}\right)\right] / q}{S\left(\hat{\theta}, \hat{\Sigma}\right) / (nM - P)} \sim \chi_q^2 \qquad \ldots(3.4)$$

Further $S^2 = \dfrac{S\left(\hat{\theta}, \hat{\Sigma}\right)}{nM - P}$ is independently distributed to $\hat{\theta}^0$ as the χ^2 distribution with $(n - p)$ degrees of freedom, Here S^2 is UBE of unknown variance σ^2.

$\therefore$ The nonlinear counterpart to the Wald Test staristic for testing $H_0 : h\left(\theta^0\right) = 0$ is given by

$$W = \frac{\hat{h}'\left(\hat{H}\hat{C}\hat{H}'\right)^{-1}\hat{h}}{qS^2} \sim \left(q, (n - p), \lambda\right) \qquad \ldots(3.5)$$

Here $\hat{h} = h\left(\hat{\theta}\right)$ and $\hat{H} = H\left(\hat{\theta}\right)$

4. STUDENTIZED RESIDUALS FOR MULTIVARIATE NONLINEAR MODELS

Consider the general multivariate nonlinear regression model

$$Y_{\alpha t} = f_\alpha\left(X_t, \theta_\alpha^0\right) + E_{\alpha t} \qquad \ldots(4.1)$$

Here θ_α^0 is p-dimensional vector matrix

Suppose $\hat{\theta}$ is the nonlinear least square estimator of θ for large sample, nonlinear least square residual vector.

$$e = \left(Y_i - \hat{Y}_t\right)$$

$$= \left(Y - f\left(\hat{\theta}\right)\right) \qquad \ldots(4.2)$$

Here $\hat{\theta}\ ;\ \theta + \left(F'F\right)^{-1} F'\varepsilon$ …(4.3)

And $F = F\left(\hat{\theta}\right) = \left[\dfrac{\partial}{\partial\theta_j} f\left(X_i, \theta\right)\right]_{n\times p}$ …(4.4)

Here $\dfrac{\partial}{\partial\theta_j} f\left(X_i, \theta\right)$ is the $(i, j)^{th}$ elements of $(n \times p)$ matrix $F(\theta)$ then general relationship between 'e' and 'ε' is

$e\ ;\ M\varepsilon$

Here $M = \left[I - F\left(F'F\right)^{-1} F'\right]$

or $e\ ;\ [I - H]\varepsilon$ where $M = \left(H_{ij}\right) = F\left(F'F\right)^{-1} F'$ is a symmetric idempotent matrix (or) HAT matrix

In scalar form

$$e_i\ ;\ \left[\varepsilon_i - \sum_{j=1}^{n} H_{ij}\varepsilon_j\right], \qquad j = 1, 2, \ldots, n \qquad \ldots(4.5)$$

Here H is HAT Matrix

Trace (H) = Rank (H) = P and $\sum_{i=j}^{n} H_{ij}^2 = H_{ij}$

Here ε follows $N_0\left(0, \sigma^2 I\right)$, so ε following normal distribution with zero mean and variance is $\sigma^2 I$. Here H controls the e.

As we know, variance of each e_i is a function of both σ^2 and $H_{ij}, i = 1, 2, \ldots n.$

The nonlinear least square residuals have a probability distribution that is scalar dependant. So, the nonlinear studentized residuals do not depend on either of these quantities and they have probability distribution and we have both internally nonlinear studentized residuals and externally nonlinear studentized residuals.

a) INTERNALLY NONLINEAR STUDENTIZED RESIDUALS

In nonlinear regression models, internally nonlinear studentized residuals are define by

$$e_i^* = \frac{\hat{e}_i}{\hat{\sigma}\sqrt{1 - h_{ij}}} \sim N(0,1) \qquad i = 1, 2, \ldots n \qquad \ldots(4.6)$$

Here $\hat{\sigma}^2 = \dfrac{e'e}{n - p}$

$$= \frac{\sum_{i=1}^{n} e_i^2}{n - p} \qquad \ldots(4.7)$$

Here $\left[\dfrac{e_i^{*^2}}{n - p}\right] \sim \beta$ - distribution with parameters $\dfrac{1}{2}$ and $(n - p - 1)\big/2$

It follows, $E\left(e_i^{*'}\right) = 0$ & $Var\left(e_i\right) = 1 \qquad \forall i = 1, 2, \ldots, n$

$$\text{Cov}\left(e_i^* e_j^*\right) = \frac{-h_{ij}}{\sqrt{\left(1-h_{ii}\right)\left(1-h_{jj}\right)}} \qquad \forall i \neq j = 1,2,...,n$$

Here h_{ij} add up to the trace of the hat matrix = P. Average 'h' is p/n which should be small, so usually $\sqrt{1-h_{ii}}$.

b) EXTERNALLY NONLINEAR STUDENTIZED RESIDUALS:

The externally nonlinear studentized residuals are define by

$$e_i^{**} = \frac{\hat{\varepsilon}_i / \left(1-h_{ij}\right)}{\sqrt{\text{MSE}_{(i)} / \left(1-h_{ij}\right)}} \qquad \text{...(4.8)}$$

$$= \frac{\hat{\varepsilon}_i}{\sqrt{\text{MSE}_{(i)}\left(1-h_{ij}\right)}} \qquad \text{...(4.9)}$$

Here $\text{MSE}_{(i)}$ is nestimate of σ^2 not baring data point 1.

i.e. $$e_i^{**} = \frac{\varepsilon_i}{\hat{\sigma}_{(i)}\sqrt{\left(1-h_{ij}\right)}} \qquad \forall i = 1,2,...,n \qquad \text{...(4.10)}$$

Based on the Normal distribution $\sigma_{(i)}^2$ and ε_i are

i.e., M.S.E (or) $\hat{\sigma}^2 = \dfrac{\left(n-p-1\right)\sigma_i^2 + \hat{\varepsilon}_i^2 / \left(1-h_{ii}\right)}{n-p}$

(or) $\sigma_{(i)}^2 = \hat{\sigma}^2 \left[\dfrac{n-p-e_i^*}{n-p-1}\right]$

So, the relationship between internally and externally nonlinear studentized residuals is given by

$$e_i^{**} = e_i^* \left[\frac{n-p-1}{n-p-e_i^*}\right], \quad i = 1,2,...,n \qquad \text{...(4.11)}$$

5. PREDICTED RESIDUALS FOR MULTIVARIATE NONLINEAR MODELS

Predicted residual sum of squares (PRESS) is also called Leave-one-out (LOO) stochastic, is regularly used in nonlinear regression analysis for cross-validation. In general, In non linear least squares, and studentized residuals fittiing is dependents on all the variables in the data. But in predicted residuals for nonlinear model i.e. i^{th} nonlinear predicted residual is depends on the fit to the data, where i^{th} care is excluded.

Suppose $\hat{\theta}$ is the nonlinear least square estimate of θ based on the full data, and $\hat{\theta}_{(i)}$ be the respective estimate where the i^{th} case is excluded.

Now the i^{th} nonlinear predicted residuals i.e.,

$$e_{(i)} = \left[Y_i - f_i\left(\hat{\theta}_{(i)}\right)\right] \qquad i = 1,2,...,n \qquad \text{...(5.1)}$$

Here $e_{(i)}$ is the prediction error

$\therefore$ the nonlinear PRESS defined by

$$\text{NLPRESS} = \sum_{i=1}^{N} e_{(i)}^2 \qquad \text{...(5.2)}$$

So, finally, the relationship between nonlinear predicted residuals and nonlinear studentized residual are given by

(i) $$e_i^* = \frac{e_{(i)}}{\hat{\sigma}\Big/\sqrt{\left(1-h_{ij}\right)}}$$ …(5.3)

(ii) $$e_i^{**} = \frac{e_{(i)}}{\hat{\sigma}_{(i)}\sqrt{\left(1-h_{ij}\right)}}$$ … (5.4)

and $$e_{(i)} = \frac{e_i}{\left(1-h_{ij}\right)} \quad \forall i = 1, 2, ..., n$$ …(5.5)

6. CONCLUSIONS

In the present research study, some inferential methods pertaining to testing of multivariate nonlinear regression models using NLLS estimator, studentized and predicted residuals for multivariate nonlinear models are proposed.

7. REFERENCES

8. Anselin, L., 1984a, Specification tests and model selection for aggregate spatial interaction: An empirical comparison, Journal of Regional Science 24, l-15..
9. Bai, J., A.J. Jakeman, and M. McAleer, 1992, Estimation and discrimination of alternative air pollution models, Ecological Modelling 64, 89-124.
10. Battese, G.E. and B.P. Bonyhady, 1981, Estimation of household expenditure functions: An application of a class of heteroscedastic regression models, Economic Record 57, 80-85.
11. Bera, A. and M. McAleer, 1989, Nested and non-nested procedures for testing linear and log-linear regression models, Sankhya B 51, 212-224.
12. Chai, T. and Draxler, R.R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? - Arguments against avoiding RMSE in the literature, Geoscientific Model Development, 7, 1247–1250.
13. CoxD R1961 Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability1 Berkeley University of California Press
14. David Pollard and Peter Radchenko 2006 Journal of Multivariate Analysis 97 548-562
15. Dorugade, A.V., 2014. New ridge parameters for ridge regression. Journal of the Association of Arab Universities for Basic and Applied Sciences, 15, pp.94-99.
16. Dr. Kesavulu Poola , Prof. M. Bhupathi Naidu(2008) . Importance of studentized and press residuals for nonlinear multivariate regression models. Journal of Emerging Technologies and Innovative Research, Volume 5, Issue 7, 353-356
17. Dr. Kesavulu Poola, P Hema Sekhar, and Dr. M Bhupathi Naidu (2020), Combined multiple forecasting model using regression, International Journal of Statistics and Applied Mathematics 2020; 5(6): 147-150
18. Edmond Malinvaud A tribute to his contributions in Econometrics Peter C B PhilipsYale University Frank Nielson 2013 Camel University
19. Engle and Robert F 1983Wald likelihood ratio and Lagrange multipliers tests in Econometrics
20. Jain, S., Chourse, S., Dubey, S., Jain, S., Kamakoty, J., and Jain, D (2016). Regression Analysis – Its Formulation and Execution in Dentistry. Journal of Applied Dental and Medical Science. Vol. 2, No. 1, pp. 199-208.
21. Kesavulu Poola, Vasu K, M Bhupathi Naidu, R Abbaiah and P Balasiddamuni (2016). The effect of multicollinearity in nonlinear regression models in International Journal of Applied Research., 2(12): 506-509.
22. Kloppers, P.H., Kikawa, C.R., and Shatalov, M.Y., (2012). A new method for least squares identification of parameters of the transcendental equations, International Journal of the Physical Sciences, 7, 5218–5223.
23. Kuchibhotla, A.K. and Patra, R.K., 2022. On least squares estimation under heteroscedastic and heavy-tailed errors. The Annals of Statistics, 50(1), pp.277-302.
24. Mahaboob B etal2016 Indian Streams Research Journal 6(4)
25. Mu B, Bai E W, Zheng W X and Zhu Q 2017 Automatic a 77322-335
26. Nagaraj, D., Wu, X., Bresler, G., Jain, P. and Netrapalli, P., 2020. Least Squares Regression with Markovian Data: Fundamental Limits and Algorithms. Advances in Neural Information Processing Systems, 33, pp.16666-16676.
27. Permai, S. D. and Tanty, H. (2018). Linear Regression Model using Bayesian Approach for Energy Performance of Residential Building. Procedia Computer Science. Vol. 135, pp 671-677.
28. Pesaran, M. Hashem, and Angus S. Deaton. "Testing non-nested nonlinear regression models." Econometrica: Journal of the Econometric Society (1978): 677-694.
29. Prabitha, J. and Archana, S. (2015). Application of Regression Analysis in Numeroustimes. International Journal of Science, Enginering and Technology Research. Vol. 4, Issue 4, pp. 1002-1005.
30. Singh, A. S. and Masuku, M. B. (2013). Application of Modeling and Statistical Regression Techniques in Research. Research Journal of Mathematical and Statistical Sciences. Vol. 1, No. 16, pp. 14-20
31. Vuong, Q.H., 1989, Likelihood ratio tests for model selection and non-nested hypotheses, Econometrica 57, 307-333.

32. Wang, S., Qian, L. and Carroll, R.J., 2010. Generalized empirical likelihood methods for analyzing longitudinal data. Biometrika, 97(1), pp.79-93.
33. White, H., 1982, Regularity conditions for Cox's test of non-nested hypotheses, Journal of Econometrics 19, 301-318.
34. Yang, G., Zhang, B. and Zhang, M., 2021. Estimation of Knots in Linear Spline Models. Journal of the American Statistical Association, pp.1-12.
35. Zhang, T., 2009. Some sharp performance bounds for least square regression with l1 regularization. The Annals of Statistics, 37(5A), pp.2109-2144.
36. Zheng, B. and Agresti, A. (2000). Summarizing the predictive power of a generalized linear model, Statist. Med., 19, 1771–1781