*International Journal of Engineering & Technology*, 7 (4.10) (2018) 995-997

**International Journal of Engineering & Technology**

*Website: www.sciencepubco.com/index.php/IJET*

***Research paper***



# A Memoir on Nonlinear Regression Model and its Pseudo Model

**B. Mahaboob[1], J. PeterPraveen[2], J. Ravi Sankar[3], B.Venkateswarlu[4*] , C. Narayana[5]**

*[1,2]Department of Mathematics, K.L.E.F (Deemed to be University), Vaddeshwaram, Vijayawada, Andhra Pradesh.*
*[3,4]Department of Mathematics, VIT University, Vellore, Tamilnadu.*
*Department of Mathematics, Sri Harsha institute of P.G Studies, Nellore.*
**Corresponding author E-mail: venkatesh.reddy@vit.ac.in*

**Abstract**

The main objective of this article is to specify a nonlinear regression model, formulate the assumptions on them and aquire its linear pseudo model. A model may be considered a mathematical description of a physical, chemical or biological state or process. Many models used in applied mathematics and Mathematical statistics are nonlinear in nature one of the major topics in the literature of theoretical and applied mathematics is the estimation of parameters of nonlinear regression models. A perfect model may have to many parameters to be useful. Nonlinear regression models have been intensively studied in the last three decades. Junxiong Lin et.al [1] , in their paper, compared best –fit equations of linear and nonlinear forms of two widely used kinetic models, namely pseudo-first order and pseudo=second-order equations. K. Vasanth kumar [2], in his paper, proposed five distinct models of second order pseudo expression and examined a comparative study between method of least squares for linear regression models and a trial and error nonlinear regression procedures of deriving pseudo second order rare kinetic parameters. Michael G.B. Blum et.al [3] proposed a new method which fits a nonlinear conditional heteroscedastic regression of the parameter on the summary statistics and then adaptively improves estimation using importance sampling.

***Keywords***: *Nonlinear regression model, variance covariance matrix, linear pseudo model, least squares estimator, degrees of freedom.*

## 1. Introduction

Nonlinear regression analysis is currently the most fertile area of research in the modern theory of Statistical Science and Applied Statistics. It is a powerful technique for analysing data described by models which are nonlinear in the parameters. In the scattered diagram, if the plotted points cluster around a curve but not a straight line then the regression is said to be nonlinear regression. In this situation, the regression equation may contain the terms as higher orders of the type $X^2, X^3, X^4$ and so on. In many cases it may not be possible to transform a nonlinear regression model into a linear statistical model. Nonlinearities enter into statistical model in different forms. In the nonlinear regression models, if only the variables enter into nonlinearity, that is, nonlinear models which are linear in parameters, then these models can be handled in the linear model framework. Furthermore, if the nonlinearity enters into the parameters or into both the variables and parameters, and the nonlinear regression model can be expressed in the linear specification by means of a suitable transformation, that is, the nonlinear model which is intrinsically linear, then the model can be again handled in the linear model framework. In the case of nonlinear models that are intrinsically nonlinear, the inferential problem of estimating parameters will be more complicated.

## 2. General nonlinear regression model

An Suppose there exists a nonlinear relationship between a set of k free variables $X_i ; i = 1,2...k$ and a dependent variable Y p-parameters $\beta_1, \beta_2, ..., \beta_p$ and error variable $\varepsilon$ .let us chose a sample of n observations, one may specify a nonlinear regression model as

$$Y_i = f\left(X_{1i}, X_{2i}, ...., X_{ki}; \beta_1, \beta_2, ..., \beta_p\right) + \varepsilon_i \quad \text{values of i vary from 1 to n} \qquad (2.1)$$

Here f is the known nonlinear function which is the expectation function relating $E(Y)$ to the independent variables. By expressing,

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ . \\ . \\ Y_n \end{bmatrix}_{n\times 1}, \quad X = \begin{bmatrix} X_{11} & X_{21} & . & . & X_{k1} \\ X_{12} & X_{22} & . & . & X_{k2} \\ . & . & . & . & . \\ . & . & . & . & . \\ X_{1n} & X_{2n} & . & . & X_{kn} \end{bmatrix}_{n\times k}$$

**Copyright © 2018 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.**

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ . \\ . \\ \beta_p \end{bmatrix}_{p\times 1}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ . \\ . \\ \varepsilon_n \end{bmatrix}_{n\times 1}$$

$$f(X,\beta) = \begin{bmatrix} f(X_{q1},\beta) \\ f(X_{q2},\beta) \\ . \\ . \\ f(X_{qn},\beta) \end{bmatrix}_{n\times 1}, q = 1,2,...,k$$

and

The most specific form of nonlinear statistical model is written as

$$y = F(x,\beta) + \varepsilon \tag{2.2}$$

The general nonlinear regression model (2.2) is of exactly the same form as the general linear regression model except that the $E(Y)$'s are nonlinear functions of the parameters.

**Assumptions on General Nonlinear Regression Models:**

(i) For nonlinear regression models, at least one of the derivatives of the expectation function with respect to the parameters depends on at least one of the parameters.

(ii. Here, we don't know the distribution which $\varepsilon$ follows

In other words, $\varepsilon \square (O, \sigma^2 I_n)$

That is, $E(\varepsilon) = O \quad \Rightarrow E(\varepsilon_i) = 0, \forall i = 1,2,...,n$

$$Var(\varepsilon) = E(\varepsilon\varepsilon') = \sigma^2 I_n \Rightarrow E(\varepsilon_i\varepsilon_j) = \sigma^2, \forall i = j = 1,2,...,n$$
$$= 0, \forall i \neq j$$

(iii) For nonlinear regression models with p parameters, the expectation surface $f(X,\beta)$ is a p-dimensional 'curved surface' in the n-dimensional dependent variable or response space; the parameter lines in the parametric space map to curves on the curved expectation surface; the 'Jacobian Determinant' which measures how large unit areas in $\beta$ become in Xinear regression models

Generally, three kinds of nonlinear regression models have been specified and found useful among other types.

**(i) Univariate Nonlinear Regression Model**

The first specification is a univariate nonlinear regression model with an additive error as

$$y = f(x,\beta) + \varepsilon \tag{3.1}$$

Here $X_i$ is a k-component vector denoting the $i^{th}$ observation on k independent variables; $\beta$ is a p-component vector of parameters; f(.) is a nonlinear function; and the errors $\varepsilon$ are distributed identically and independently with $\mu = 0$ *and* $\sigma^2$ is constant but not known.The nonlinear regression equation (3.1) is a generalization of the multiple linear regression model to the nonlinear case.

**(ii) g-variate Nonlinear Regression Model**

The second specification is a g-variate nonlinear regression model with additive errors. The $m^{th}$ nonlinear regression equation has been specified as

$$Y_{im} = f_m(X_i,\beta) + \varepsilon_{im}, i = 1,2,...,n \text{ , m=1, 2…g} \tag{3.2}$$

Here, g-of these nonlinear regression equations explain g dependent variables $Y_{i1}, Y_{i2,}..., Y_{ig}$.

The $m^{th}$ nonlinear regression function $f_m(.)$ may not include all independent variables in $X_i$ and all parameters in $\beta$ as arguments, but there is no loss in allowing all dependent variables and f denote the vector of g nonlinear regression functions. The nonlinear regression equation (3.2) is a generalization of a system of linear regression equations. One may rewrite (3.2) as

$$y = f(x,\beta) + \varepsilon \tag{3.3}$$

The error $\varepsilon_i$ is assumed to be independently and identically distributed with mean zero and variance – covariance matrix $\Omega$.

**(iii) System of Simultaneous Nonlinear Regression Equations Model**

The third specification of the nonlinear regression model is a system of simultaneous nonlinear equations with additive errors. The $m^{th}$ structural equation of the system has been specified as

$$\phi_m(Y_i, X_i; \beta_m) = \varepsilon_{im} \quad m = 1,2,...,g \quad i = 1,2,...,n \tag{3.4}$$

Generally, in this $m^{th}$ structural equation, the $m^{th}$ endogenous variable $Y_{im}$ is an explicit function of other endogenous variables in $Y_i$ and predetermined variables $X_i; \beta_m$ is a vector of parameters in $m^{th}$ structural equation. Denoting $\phi$ as the vector of g functions $\phi_1, \phi_2, ..., \phi_g$, the nonlinear regression equation (3.4) is a generalization of a system of linear simultaneous equations, one may write the structural nonlinear regression equation (3.4) as

$$\phi(y, x, \beta) = \varepsilon \tag{3.5}$$

## 3. Variance- covariance matrix and expectation of Best Linear Unbiased Estimator.

The general form of nonlinear statistical model is specified by

$$y = f(x,\beta) + \varepsilon \tag{4.1}$$

Such that $\varepsilon_i \overset{i.i.d}{\sim} (0, \sigma^2)$, i = 1,2,..,n or $\varepsilon$ independently follows $(O, \sigma^2 I_n)$.

One may approximate $f(X,\beta)$ using Taylor expansion up to $1^{st}$ degree, a linear approximation for the above is obtained as

$$f(X,\beta) \approx f(X,\beta^*) + (\beta - \beta^*)Z(\beta^*) \tag{4.2}$$

$$Z(\beta^*) = \begin{bmatrix} \frac{\partial f(X_1,\beta)}{\partial\beta} .... & \frac{\partial f(X_1,\beta)}{\partial\beta} \\ . \\ \frac{\partial f(X_n,\beta)}{\partial\beta} & \frac{\partial f(X_n,\beta)}{\partial\beta} \end{bmatrix}$$

Where

Here $f(X_i,\beta)$ is approximated by

$$f(X,\beta) = f(X,\beta^*) + \left[ \left.\frac{\partial f(X_i,\beta)}{\partial \beta}\right|_{\beta^1} .... \left.\frac{\partial f(X_i,\beta)}{\partial \beta}\right|_{\beta^1} \right](\beta - \beta^*) \quad (4.3)$$

Here, $\left[ \left.\frac{\partial F(X_i,\beta)}{\partial \beta}\right|_{\beta^*} .... \left.\frac{\partial F(X_i,\beta)}{\partial \beta}\right|_{\beta^*} \right]$

is the $i^{th}$ row of $Z(\beta^*)$

Owing to Malinward (1970) we can obtain the pseudo linear model for the above

$$y + \beta z(\beta^*) = f(x,\beta^*) + \varepsilon \quad (4.4)$$

This implies $y^* - \varepsilon = \beta z(\beta^*)$ (4.5)

Where $y^* + f(x,\beta^*) - z(\beta^*)\beta^* = y$ (4.6)

Since $\beta^*$ is not known. Here we should not apply pseudo linear model 4.5 directly for obtaining parameters.The LSE of $\beta$ is obtained by

$$\bar{\beta}_{LS} = \left[\gamma(\beta^*)\right]^{-1}\left[\gamma(\beta^*)'\right]^{-1}\left[\gamma(\beta^*)' y^*\right] \quad (4.7)$$

provided $y^*$ and $z(\beta^*)$ are known.

Further, if the errors $\varepsilon_i \overset{i.i.d}{\sim} \left(0, \sigma^{*2}\right)$ The LSE of $\beta$ namely $\hat{\beta}_{LS}$ is consistent and follows asymptotically normal distribution. Hence the covariance variance matrix of $\hat{\beta}_{LS}$ is obtained by

$$Var(\hat{\beta}_{LS}) = \dot{\sigma}^2 \left[Z(\beta^*)\right]^{-1}\left[Z(\beta^*)\right]^{-1} \quad (4.8)$$

This is approximately correct as the model is approximately good. Now assume that NLLSE $\hat{\beta}_{LS}$ is sufficiently close to $\beta^*$. Then we can get an estimation for variance covariance matrix of $\hat{\beta}_{LS}$ by

$$\hat{\mathrm{Var}}\left(\hat{\beta}_{\mathrm{LS}}\right) = \hat{\sigma}^2_{\mathrm{LS}}\left[\mathrm{Z}\left(\hat{\beta}_{\mathrm{LS}}\right)' \mathrm{Z}\left(\hat{\beta}_{\mathrm{LS}}\right)\right]^{-1} \quad (4.9)$$

Here, $\hat{\sigma}^2_{\mathrm{LS}}$ is LSE of $\sigma^{*2}$.

For exxample,

$$\hat{\sigma}^2_{LS} = (n-p)^{-1}\left[f\left(X,\hat{\beta}_{LS}\right) - Y\right]\left[f\left(X,\hat{\beta}_{LS}\right) - Y\right] \quad (4.10)$$

If we put some assumptions that (4.5) is right and $\varepsilon$ follows Normal distribution with $O, \sigma^2 I_n$ then under particular regularity conditions of the asymptotic phenomenon we get

$$\left[\frac{(\mathrm{n}-\mathrm{p})\hat{\sigma}^2_{\mathrm{LS}}}{\dot{\sigma}^2}\right] = \frac{\varepsilon'\left\{\mathrm{I} - \mathrm{Z}(\beta^*)\left[\mathrm{Z}(\beta^*)'\mathrm{Z}(\beta^*)\right]^{-1}\mathrm{Z}(\beta^*)'\right\}\varepsilon}{\dot{\sigma}^2} \quad (4.11)$$

This follows Chi-Square distribution with d.f= $(\mathrm{n}-\mathrm{p})$.

## 4. Conclusion

In the above research article three kinds of nonlinear regression models have been specified. Mean vector and covariance of BLUE are estimated using principles of matrix calculus.

## References


[1] Michael G. B. Blum, Olivier Franwis (2010), "Nonlinear regression models for Appropriate Bayesian computation", Statistics and Computing, Springer, Vol. (20), Issue 1, Pp: 63-73.

[2] Shyamal D Peddada, Joseph K. Haseman, "Analysis of Nonlinear regression models. A Cautionary note", Dose-Response: An international journal Vol. (3), Issue 4, (2005).

[3] K. Vasanth Kumar, "Linear and Nonlinear regression analysis for the sorption kinetics methylene blue onto activated Carbon", Journal of Hazardous Materials B137, (2006), Pp: 1538-1544.

[4] Junxiong Lin, Lan Wang, "Comparison between linear and nonlinear forms of Pseudo-first order and pseudo-second order adoption kinetic models for the removal of methylene blue by activated carbon", Frontiers of Environmental Science and Engineering in China, Vol.(3), Issue 3, (2009), Pp: 320-324.

[5] Gordon K. Sonyth, "Nonlinear regression, Encyclopaedia of Environ metrics Vol. (3), (2002), Pp: 1405-1461.

[6] Gurleen K. Popli ,"A note on the instrumental variable estimators in the nonlinear models", Journal of Quantitative Economics Vol.16. no.2, (2000), Pp: 31-36.

[7] E. Grafarent and J. Awange, "Application of linear and nonlinear models", Springer Geophysics, (2012).

[8] Bates D.M and Walts D.G, "Nonlinear regression: Iterative Estimation and Linear Approximations in Nonlinear regression Analysis and its Applications", John Wiley and sons Inc Hobeken, NJ, USA (2008).

[9] Vasilyev D.M, "Theoretical and Practical Aspects of linear and nonlinear models order reduction Techniques", MIT, USA, (2008).

[10] Davidian M. and Giltinon D.M, "Nonlinear models for repeated measurement Data: An overview and update", Journal of Agricultural, Biological and Environmental statistics (JABES), Vol. (8), (2003), Pp: 387-419.