

Some Aspects of Nonlinear Methods in Model Selection

G.Y. Mythili¹, C.Narayana², J. Prabhakara Naik³, K.Vasu⁴, G. Mokesh Rayalu⁵

¹Assistant Professor, Department of Mathematics, School of Advanced sciences, VIT, Vellore

²Assistant professor, Department of Mathematics, Sriharsha Institute of P.G. Studies, Nellore

³Lecturer in Statistics, SSBN Degree & PG College(Autonomous), Anantapur

⁴Assistant Professor, Vidya Educational Institutions, Yadamari, Chittoor

⁵Assistant Professor, Department of Mathematics, School of Advanced sciences, VIT, Vellore

Abstract: Model means a set of relationships between two or more variables. These relationships can be expressed in terms of mathematical equations. A set of mathematical equations concerns with the economic variables is called a mathematical economic model. By introducing an error random variable, the mathematical economic model becomes an econometric model. Econometric model may be either in the form of a set of linear equations (linear regression model) or in the form of a set of nonlinear equations (Nonlinear regression model). If the curve of the regression is not a straight line, then the regression is said to the nonlinear regression. In this case, the nonlinear regression equation involves the terms as higher order of the type X^2 , X^3 and so on.

I. INTRODUCTION

A nonlinear regression model refers to a model having a regression function which is nonlinear either in the explanatory variables or in the unknown regression coefficients or in both explanatory variables and regression coefficients.

Nonlinear regression models can be broadly divided into two parts:

Nonlinear models which are linear in parameters but nonlinear in independent variables.

And Nonlinear models which are nonlinear in parameters.

A. Nonlinear Models which are Linear in Parameters

A general form of nonlinear regression model which is linear in parameters is given by

$$Y_i = \beta_0 + \beta_1 Z_{1i} + \beta_2 Z_{2i} + \dots + \beta_p Z_{pi} + \epsilon_i$$

$i = 1, 2, \dots, n$. (1) Where Z_i refers to any function of the basic independent variables X_1, X_2, \dots, X_k .

B. Nonlinear Models which are Nonlinear in Parameters

Any model which is not in the form (1) is called a nonlinear regression model, which is nonlinear in the parameters. These models are again two types :

Nonlinear regression models that are intrinsically linear;

and Nonlinear models that are intrinsically nonlinear.

A nonlinear model which can be expressed in the form (1) by suitable transformation of the variables is called nonlinear model that is intrinsically linear.

A nonlinear model which can not be expressed in the form (1) by any transformation, is called nonlinear model that is intrinsically nonlinear.

In the case of intrinsically linear models, the OLS estimation can be applied to the transformed models and the optimal estimators can be obtained for the parameters of the models.

In the case of intrinsically nonlinear models, OLS estimation fails to give the estimates of the parameters. However, the OLS estimation can be applied under iterative process to estimate the parameters of the models.

The inferential properties of estimators for the parameters in the nonlinear regression models are usually derived by using linear approximations in one form or another form. Therefore, in general, all the problems arising for linear models are also appear when dealing with nonlinear specifications. Moreover, for nonlinear regression models, the properties of estimators and test statistics can only be derived approximately or asymptotically. Little is known about the small sample properties.

II. NONLINEAR METHODS OF ESTIMATION

A few estimation methods are available for the estimation of parameters of the nonlinear statistical models.

A. Nonlinear Least Squares Estimation

Consider a nonlinear regression model is of the form

$$Y_i = f(X_{1i}, X_{2i}, \dots, X_{ki}; \beta_1, \beta_2, \dots, \beta_p) + \epsilon_i \quad \dots (2.1.1)$$

$i = 1, 2, \dots, n.$

Where Y is Dependent variable;

X_1, X_2, \dots, X_k are k-independent variables;

$\beta_1, \beta_2, \dots, \beta_p$ are p-parameters;

ϵ is error variable;

n is number of observations on each variable;

and $f(\cdot)$ is the known nonlinear functional form.

By writing

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \cdot \\ \cdot \\ \cdot \\ Y_n \end{bmatrix}_{nx1}, X = \begin{bmatrix} X_{11} & X_{21} & \cdot & \cdot & \cdot & X_{k1} \\ X_{12} & X_{22} & \cdot & \cdot & \cdot & X_{k2} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ X_{1n} & X_{2n} & \cdot & \cdot & \cdot & X_{kn} \end{bmatrix}_{nxk}, \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \cdot \\ \cdot \\ \cdot \\ \beta_p \end{bmatrix}_{px1}$$

$$\epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \cdot \\ \cdot \\ \cdot \\ \epsilon_n \end{bmatrix}_{nx1} \quad \text{and} \quad f(X, \beta) = \begin{bmatrix} f(X_{j1}, \beta) \\ f(X_{j2}, \beta) \\ \cdot \\ \cdot \\ \cdot \\ f(X_{jn}, \beta) \end{bmatrix}_{nxk}, j = 1, 2, \dots, k$$

the nonlinear regression model can be expressed in the matrix form as

$$Y = f(X, \beta) + \epsilon \quad \dots (2.1.2)$$

Where, Y is (nx1); $f(X, \beta)$ is (nx1); X is (nxk);

β is (px1) and ϵ is (nx1) matrices.

Assume that the errors are i.i.d's with $\epsilon \sim (0, \sigma^2 I)$, but the exact form of the distribution is unknown.

Define the residual sum of squares as

$$R(\hat{\beta}) = [Y - f(X, \hat{\beta})]' [Y - f(X, \hat{\beta})] \quad \dots (2.1.3)$$

The nonlinear least squares estimator $\hat{\beta}$ of β can be obtained by minimizing $R(\hat{\beta})$ with respect to $\hat{\beta}$ and then solve the nonlinear normal equations for $\hat{\beta}$.

By writing $R(\hat{\beta})$ as

$$R(\hat{\beta}) = \sum_{i=1}^n [Y_i - f(X_i, \hat{\beta})]^2, \text{ where } X_i \text{ is } i^{\text{th}} \text{ k-dimensional observation on k-independent variables.}$$

The p-nonlinear normal equations are given by

$$\frac{\partial R(\hat{\beta})}{\partial \hat{\beta}} = 0 \Rightarrow \sum_{i=1}^n [Y_i - f(X_i, \hat{\beta})] \left[\frac{\partial f(X_i, \beta)}{\partial \beta} \right]_{\beta = \hat{\beta}} = 0 \quad \dots (2.1.4)$$

Since, $f(X_i, \beta)$ is nonlinear in β 's, the normal equations will be nonlinear both in X 's and β 's. In general, $\hat{\beta}$ will not be linear function of Y and optimality properties may not be derived for the nonlinear least squares estimators.

In practice, it is difficult to solve the nonlinear normal equations. In this case, generally one may use iterative methods to obtain the estimators for the parameters. Under certain regularity conditions the Least Squares Estimator $\hat{\beta}$ will be consistent and asymptotically normally distributed. A reasonable estimator for σ^2 is given by

$$\hat{\sigma}^2 = \frac{R(\hat{\beta})}{n - k}.$$

B. Taylor Series Expansion Method or Linear Approximation Method

Consider the nonlinear regression model in the matrix form as

$$Y_{n \times 1} = f_{n \times 1}(X_{n \times k}, \beta_{p \times 1}) + \epsilon_{n \times 1} \quad \dots (2.1.5)$$

Such that $\epsilon \sim iid(0, \sigma^2 I_n)$

Using Taylor series expansion and neglecting the terms from second order derivatives, the nonlinear function $f(X, \beta)$ may be approximated at $\beta = \beta^*$ by a linear function as follows :

$$\begin{aligned} f(X, \beta) &\simeq f(X, \beta^*) + Z(\beta^*)(\beta - \beta^*) \\ \text{or} \end{aligned}$$

$$\text{Where } Z(\beta^*) = \left[\frac{\partial f(X, \beta)}{\partial \beta} \right]_{\beta = \beta^*}$$

$$\begin{aligned} \text{or } Y &= f(X, \beta^*) + Z(\beta^*)(\beta - \beta^*) + \epsilon \\ \Rightarrow Y^* &= Z(\beta^*)\beta + \epsilon \quad \dots (2.1.6) \end{aligned}$$

Where $Y^* \simeq Y - f(X, \beta^*) + Z(\beta^*)\beta^*$

Since, (2.1.5) and (2.1.6) are identical for $\beta = \beta^*$, a consistent estimator $\hat{\beta}_{NL}$, say for β^* in (2.1.5) will almost surely coincide with a corresponding consistent estimator $\hat{\beta}_{LS}$, say for β^* in (2.1.6) in the limit, and thus these properties of $\hat{\beta}_{NL}$ and $\hat{\beta}_{LS}$ will be similar at least in large samples.

Model (2.1.6) is called the 'Linear Pseudomodel' by Malinvaud (1970). The least squares estimator $\hat{\beta}_{LS}$ for β is given by

$$\hat{\beta}_{LS} = [Z(\beta^*)^T Z(\beta^*)]^{-1} [Z(\beta^*)^T Y^*] \quad \dots (2.1.7)$$

Assume that ϵ has the covariance matrix $\Sigma_{\epsilon} = \sigma^{*2} I_n$, then the variance covariance matrix of the least squares estimator $\hat{\beta}_{LS}$ as

$$\Sigma_{\beta_{LS}} = \sigma^{*2} [Z(\beta^*)^T Z(\beta^*)]^{-1} \quad \dots (2.1.8)$$

Consequently, the covariance of $\hat{\beta}_{NL}$ is approximately

$$\Sigma \hat{\beta}_{NL} = \sigma^{*2} [Z(\beta^*)' Z(\beta^*)]^{-1} \text{ in large samples.}$$

Also, an estimate of σ^{*2} is given by

$$\hat{\sigma}^{*2} = \frac{[Y^* - Z(\beta^*) \hat{\beta}_{NL}]' [Y^* - Z(\beta^*) \hat{\beta}_{NL}]}{n - p} \quad \dots (2.1.9)$$

Under certain regularity conditions,

$$\hat{\beta}_{NL} \stackrel{Asy}{\sim} N \left\{ \beta, \sigma^{*2} [Z(\beta^*)' Z(\beta^*)]^{-1} \right\} \quad \dots (2.1.10)$$

Since, $Z(\beta^*) = \left[\frac{\partial f(X, \beta)}{\partial \beta} \right]_{\beta = \beta^*}$ is unknown, it may be

$$\text{approximated by } \hat{Z}(\beta^*) = \left[\frac{\partial f(X, \beta_{NL})}{\partial \beta_{NL}} \right]_{\beta = \hat{\beta}_{NL}} \quad (2.1.11)$$

The matrix $[Z(\beta^*)' Z(\beta^*)]^{-1}$ is accordingly approximated

$$\text{by } \hat{C} = [\hat{Z}(\beta^*)' \hat{Z}(\beta^*)]^{-1} = ((\hat{C}_{ij})) \quad (2.1.12)$$

To test the hypothesis, $H_0 : \beta_i = \beta_{i0}$, the test statistic is given by

$$t = \frac{\hat{\beta}_{NLi} - \beta_{i0}}{\hat{\sigma}^* \sqrt{\hat{C}_{ii}}} \underset{\sim}{\text{Approximately}} t_{n-p} \quad (2.1.13)$$

C. Maximum Likelihood Method of Estimation

The method of maximum likelihood estimation can be applied to nonlinear models. If the observations are independent, under certain regularity conditions, the maximum likelihood estimators have some optimum properties.

Consider the nonlinear model as $Y = f(X, \beta) + \epsilon$ (2.1.14)

Where ϵ is assumed to follow $N[0, \phi]$

Here, ϕ is the variance - covariance matrix of ϵ .

Write the likelihood function of the observation vector Y as

$$L(\beta, \phi) = \frac{1}{(2\pi)^{n/2} |\phi|^{n/2}} e^{-\frac{1}{2} [Y - f(X, \beta)]' \phi^{-1} [Y - f(X, \beta)]} \quad (2.1.15)$$

$$\text{or } \ln L(\beta, \phi) = -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln |\phi| - \frac{1}{2} [Y - f(X, \beta)]' \phi^{-1} [Y - f(X, \beta)] \quad (2.1.16)$$

$$\text{or } -\ln L(\beta, \phi) = \frac{n}{2} \ln 2\pi + \frac{n}{2} \ln |\phi| + \frac{1}{2} [Y - f(X, \beta)]' \phi^{-1} [Y - f(X, \beta)] \quad (2.1.17)$$

The maximum likelihood estimators of β and ϕ can be obtained by maximising $\ln L(\beta, \phi)$ or minimising $-\ln L(\beta, \phi)$ with respect to β and ϕ .

Let $L^* = -\ln L(\hat{\beta}, \hat{\phi})$

Assuming that $\phi = \sigma^2 I_n$. Then

$$\ln L = \frac{n}{2} \ln 2\pi + \frac{n}{2} \ln \hat{\sigma}^2 + \frac{1}{2\hat{\sigma}^2} [Y - f(X, \hat{\beta})]^T [Y - f(X, \hat{\beta})] \tag{2.1.18}$$

Minimisation of L^* is equivalent to the minimisation of the residual sum of squares $[Y - f(X, \hat{\beta})]^T [Y - f(X, \hat{\beta})]$ with respect to $\hat{\beta}$.

Thus, the maximum likelihood estimator of β say $\hat{\beta}_{ML}$ equals to the OLS estimator of β say $\hat{\beta}_{OLS}$. i.e., $\hat{\beta}_{ML} = \hat{\beta}_{OLS}$

Also, the maximum likelihood estimator of σ^2 is given by

$$\hat{\sigma}_{ML}^2 = \frac{[Y - f(X, \hat{\beta}_{ML})]^T [Y - f(X, \hat{\beta}_{ML})]}{n} \tag{2.1.19}$$

Under certain regularity conditions, the maximum likelihood estimators are consistent, sufficient, asymptotically efficient and follow asymptotic normal distribution.

D. Newton – Raphson Method

Suppose $L(\beta)$ be the likelihood function which has continuous first and second derivatives with respect to the vector β of p -elements. To maximize $L(\beta)$, one may consider the first order condition as

$$\frac{\partial L(\beta)}{\partial \beta} \equiv L^1(\beta) = 0 \tag{2.1.20}$$

At any point β , the gradient or vector of first derivatives $L^1(\beta)$ provides much information concerning where one should move to reach a maximum. Many algorithms to maximize $L(\beta)$ are based on the gradient $L^1(\beta)$ and are called ‘Gradient Algorithms’.

An algorithm specifies how one should move from a point $\beta^{(0)}$ to the next point $\beta^{(1)}$. A gradient algorithm takes the form

$$\beta^{(1)} = \beta^{(0)} + K^0 H^0 L^1(\beta^{(0)}) \tag{2.1.21}$$

Where $L^1(\beta^{(0)})$ is the gradient evaluated at $\beta^{(0)}$;

K^0 is a scalar;

H^0 is a matrix to be specified;

K^0 gives the stepwise size;

and $H^0 L^1(\beta^{(0)})$ is the search direction.

The Newton – Raphson method specifies

$$H^0 = - \left[\frac{\partial L^1(\beta)}{\partial \beta} \right]_{\beta = \beta^{(0)}}^{-1} = \text{Inverse of the Hessian matrix.} \tag{2.1.22}$$

Where,

$$L^1(\beta^{(0)}) = \left[\frac{\partial L(\beta)}{\partial \beta} \right]_{\beta = \beta^{(0)}} \quad (0)$$

If $L^1(\beta)$ is linear [or $L(\beta)$ is quadratic], the Newton – Raphson method, using $K=1$ converges in one iteration; i.e., starting from any $\beta^{(0)}$,

$$\beta^{(1)} = \beta^{(0)} - \left[\frac{\partial L^1(\beta)}{\partial \beta} \right]_{\beta = \beta^{(0)}}^{-1} L^1 \beta^{(0)} \quad (2.1.23)$$

gives the maximum of $L(\beta)$. If $L(\beta)$ is not quadratic, one can choose the step size K^0 different from 1. The method of choosing K in connection with the Newton – Raphson method has been suggested by Chow (1968) and Chow and Fair (1973).

E. Steepest Descent / Steepest Ascent Method

The method of steepest Descent or steepest Ascent specifies $H^0=I$, one way to choose the step size K^0 is to approximate $L(\beta)$ by a quadratic function and find K^0 to maximize $L(\beta^{(1)})$.

$$L(\beta^{(1)}) = L(\beta^{(0)}) + \frac{\partial L(\beta^{(0)})}{\partial \beta} (\beta^{(1)} - \beta^{(0)}) + \frac{1}{2} (\beta^{(1)} - \beta^{(0)})^T \frac{\partial^2 L(\beta^{(0)})}{\partial \beta \partial \beta^T} (\beta^{(1)} - \beta^{(0)}) \quad (2.1.24)$$

By substituting $K L^1(\beta^{(0)})$ for $(\beta^{(1)} - \beta^{(0)})$ one may write (2.1.24) as

$$L(\beta^{(1)}) = L(\beta^{(0)}) + K [L^1(\beta^{(0)})]^T [L^1(\beta^{(0)})] + \frac{1}{2} K^2 [L^1(\beta^{(0)})]^T \left[\frac{\partial L^1 \beta^{(0)}}{\partial \beta} \right] [L^1(\beta^{(0)})] \quad (2.1.25)$$

Minimization of (2.1.25) with respect to K gives

$$K^0 = - [L^1(\beta^{(0)})]^T [L^1(\beta^{(0)})] \left\{ [L^1(\beta^{(0)})]^T \left[\frac{\partial L^1(\beta^{(0)})}{\partial \beta} \right] [L^1(\beta^{(0)})] \right\}^{-1} \quad \dots (2.1.26)$$

This K^0 may be expensive to compute, since it requires the matrix

$$\frac{\partial L^1 \beta^{(0)}}{\partial \beta}$$

of second partial derivations of $L(\beta)$ with respect to β .

Although this method is very simple, it may not be used in many cases because, it may converge slowly. However this method can be valuable if it is combined with other algorithms such as Gauss-Newton methods etc.

F. Gauss – Newton Method

It is an approximation to the Newton – Raphson method in that when the matrix of second derivatives of L is computed, the second derivatives of the function $f(X, \beta)$ specifying the nonlinear model are ignored.

Consider the nonlinear model

$Y = f(X, \beta) + \epsilon$ and the objective function is of the form

$$R(\beta) = [Y - f(X, \beta)]^T \Omega [Y - f(X, \beta)] \quad (2.1.27)$$

With $\Omega = ((w_{ij}))$ for $i, j = 1, 2, \dots, n$.

The Hessian of $R(\beta)$ is

$$H(\beta) = 2 Z(\beta)^T \Omega Z(\beta) - 2 \sum w_{ij} [Y_i - f(X_i, \beta)] \left[\frac{\partial^2 f(X_i, \beta)}{\partial \beta \partial \beta^T} \right] \quad (2.1.28)$$

However, since the mean of $\epsilon_i = Y_i - f(X_i, \beta^*)$ is assumed to be zero, these error terms should be small, at least close to the minimum of $R(\beta)$ if the variance is small. Thus, the first term on the R.H.S. of (2.1.28) is taken as an approximation of $H(\beta)$. Consequently,

$$[H(\beta)]^{-1} = [Z(\beta_n)^T \Omega Z(\beta_n)]^{-1} \quad (2.1.29)$$

By this method, one can have,

$$\beta_{n+1} \beta_n + [Z(\beta_n)^T Z(\beta_n)]^{-1} Z(\beta_n)^T [Y - f(X, \beta_n)] \quad (2.1.30)$$

$$\beta_{n+1} = [Z(\beta_n)^T Z(\beta_n)]^{-1} Z(\beta_n)^T [Y - f(X, \beta_n) + Z(\beta_n) \beta_n] \quad (2.1.31)$$

The least squares estimator for the model

$$Y - f(X, \beta_n) + Z(\beta_n) \beta_n = Z(\beta_n) \beta + \epsilon \quad (2.1.32)$$

Which is the linear Pseudo model at β_n . This shows that the Gauss algorithm can be viewed as a sequence of linear regressions. In each step, one may compute the least squares estimator for a linear approximation of the nonlinear model.

G. Method Of Scoring

It is another variation of the Newton-Raphson method applied to compute the maximum likelihood estimates for the parameters.

Consider the necessary condition that $\log L(\beta)$ is maximum at $\beta = \hat{\beta}$, as

$$\left[\frac{\partial \log L(\beta)}{\partial \beta} \right]_{\beta = \hat{\beta}} = 0 \quad (2.1.33)$$

Where $L(\beta)$ is the likelihood function. The quantity $\left[\frac{\partial \log L(\beta)}{\partial \beta} \right]$ is sometimes called the 'Efficient Score for β ' and is denoted by $S(\beta)$. Thus, the maximum likelihood estimator is the value of β for which the efficient score vanishes.

Sometimes the maximum likelihood equation $\left[\frac{\partial \log L(\beta)}{\partial \beta} \right]_{\beta = \hat{\beta}} = 0$ can be solved easily. But, often it is nonlinear and it has to be solved by using an Iterative method, for this purpose any one of the gradient methods can be used.

By the method of scoring, for single parameter case, one may expand $\left[\frac{\partial \log L(\beta)}{\partial \beta} \right]$ around $\beta = \beta^0$ by using Taylor series expansion. After neglecting terms involving from second derivatives, one may get

$$\frac{\partial \log L(\beta)}{\partial \beta} \simeq \left[\frac{\partial \log L(\beta)}{\partial \beta^0} \right] + (\beta - \beta^0) \left[\frac{\partial^2 \log L(\beta)}{\partial \beta^0^2} \right] \dots \quad (2.1.34)$$

In large samples, by substituting $I(\beta^0)$ for $\left[-\frac{\partial^2 \log L(\beta)}{\partial \beta^0^2} \right]$ and thus, one may get

$$\frac{\partial \log L(\beta)}{\partial \beta} \simeq S(\beta^0) - (\beta - \beta^0) I(\beta^0)$$

$$\therefore \frac{\partial \log L(\beta)}{\partial \beta} = 0 \Rightarrow S(\beta^0) - (\beta - \beta^0) I(\beta^0) = 0$$

$$\Rightarrow \beta = \beta^0 + \frac{S(\beta^0)}{I(\beta^0)} \tag{2.1.35}$$

Equation (2.1.35) is an important relation for Iterative process starting from

$$\beta^1 = \beta^0 + \frac{S(\beta^0)}{I(\beta^0)} \tag{2.1.36}$$

In the matrix notation, one can write

$$\beta^1 = \beta^0 + [I(\beta^0)]^{-1} S(\beta^0) \tag{2.1.37}$$

Where $I(\beta^0) = -E \left[\frac{\partial^2 \log \beta}{\partial \beta \partial \beta^1} \right]_{\beta = \beta^0}$, $S(\beta^0) = \left[\frac{\partial \log L(\beta)}{\partial \beta} \right]_{\beta = \beta^0}$

At the second iteration, one can have

$$\beta^2 = \beta^1 + [I(\beta^1)]^{-1} S(\beta^1) \tag{2.1.38}$$

Iteration process will be proceeded till the convergence.

H. Quadratic Hill – Climbing Method

This method was proposed by Goldfeld, Quandt and Trotter (1966) which is another modification of the Newton – Raphson method.

When β^0 is far from the maximizing value, the matrix $\left[\frac{\partial L^1(\beta)}{\partial \beta^1} \right]_{\beta = \beta^0}$ of second partial derivatives may not be negative

definite. Taking a small step in the Newton – Raphson direction may lead one downhill rather than uphill. To ensure the negative definiteness of $-H^0$ the method of Quadratic hill climbing uses for $-H^0$ in the gradient algorithm.

$$\beta^1 = \beta^0 + K^0 H^0 L^1(\beta^0) \text{ as}$$

$$-H^0 = \left[\frac{\partial L^1(\beta^0)}{\partial \beta^1} - \alpha I \right]^{-1} \tag{2.1.39}$$

Where the scalar α is chosen to maximize $L(\beta)$ is a spherical region centered at β^0 , that is bounded by $(\beta - \beta^0)^T (\beta - \beta^0) = r_\alpha$, under the assumption that $L(\beta)$ is quadratic in that region. This method requires computing the characteristic roots of the matrix

$$\left[\frac{\partial L(\beta^0)}{\partial \beta^T} \right]$$

J. Conjugate Gradient Method

This algorithm was suggested by Fletcher and Reeves (1964). It does not require the use of first derivatives. It evaluates the function $L(\beta)$ to be maximized along mutually conjugate directions, beginning at a point β^0 . For a quadratic function,

$L(\beta) = \beta^T A \beta + \alpha^T \beta + c$, two direction vectors d_1 and d_2 are conjugate if $d_1^T A d_2 = 0$. When $A = I_m$, the 'm' columns of the identity matrix I_m are conjugate direction vectors. Let $d_1^0, d_2^0, \dots, d_m^0$ be m linearly independent direction vectors. Starting from β^0 , one can search along the directions d_i^0 , (i=1, 2, ..., m) sequentially, each time going along one direction d_i^0 . One begins by searching along d_1^0 , that is, by choosing a scalar λ_1 , to

$$\text{Max. } \lambda_1 L(\beta^0 + \lambda_1 d_1^0).$$

Having chosen $\hat{\lambda}_1$, one chooses a scalar λ_2 to

$$\text{Max. } \lambda_2 L(\beta^0 + \hat{\lambda}_1 d_1^0 + \lambda_2 d_2^0)$$

and so forth.

Having searched along all m directions,

$$\text{let, } \hat{\beta}^0 = \beta^0 + \sum_{i=1}^m \hat{\lambda}_i d_i^0 \equiv \beta^0 + \delta.$$

The following step is to Max. $L(\hat{\beta}^0 + K \delta)$ and set $\beta^1 = \hat{\beta}^0 + K \delta$.

To start with, let d_i^0 be the coordinate directions, ie., the m column vectors of the identity matrix I_m . The directions for the next iteration are $d_1^1 = d_2^0, d_2^1 = d_3^0, \dots, d_m^1 = \delta$. In the second iteration treat these d_i^1 (i = 1, 2, ..., m) as treated the d_i^0 (i = 1, 2, ..., m) in the first iteration and so forth. The above method was due to Powell (1964).

III. CONCLUSIONS

This paper contains various nonlinear methods of estimation based on some numerical methods besides the nonlinear least squares and maximum likelihood estimation procedures. A test for the specification of error in nonlinear regression model has been explained along with the estimation of a mixed general Cobb-Douglas type function with multiplicative and additive errors. The presence of the autocorrelated disturbances in the nonlinear regression models has been examined by considering first order autoregressive process. A seemingly unrelated nonlinear regression equations model has been discussed clearly in this paper.

REFERENCES

- [1] Andrews, D.W.K. and Fair, R.C. (1988), "Inference in Nonlinear Econometric Models with Structural Change", Review of Economic Studies, 55, 615-640.
- [2] Arrow, K., Chenery, H.B., Minhas, B. and Solow, R.M. (1961), "Capital-Labour Substitution and Economic Efficiency", Review of Economics and Statistics, 43, 225-250.
- [3] Bard, Y. (1974), "Nonlinear Parameter Estimation", Academic Press, New York.



- [4] Chambers, J.M. (1973), "Fitting Nonlinear Models : Numerical Techniques", *Biometrika*, 60, 1, 1-13.
- [5] Chow, G.C. (1983), "Econometrics", Mc Graw-Hill Book Company, Inc., Singapore.
- [6] Cook, R.D., Tsai, C.L. and Wei, B.C. (1986), "Bias in Nonlinear Regression", *Biometrika*, 73, 615-623.
- [7] Dadksah, K.M. (1984), "Confidence Interval for Predictions from a Logarithmic Model", *Review of Economics and Statistics*, 66, 527-528.
- [8] Davidson, R. and MacKinnon, J.G. (1981), "Several Tests for Model Specification in the Presence of Alternative Hypothesis", *Econometrica*, 49, 781-793.
- [9] De-Min Wu (1975), "Estimation of the Cobb-Douglas Production Function", *Econometrica*, 43, 739-744.
- [10] Gompertz, M.L. and Pantula, S.G. (1992), "Nonlinear Regression with Variance Components", *Journal of the American Statistical Association*, 87, 201-209.
- [11] Granger, C.W.J. and Terasvirta, T. (1993), "Modeling Nonlinear Economic Time Series", Oxford University Press, Oxford.
- [12] Gujarati, D.N. (1995), "Basic Econometrics", Third Edition, Mc Graw-Hill Inc., New York.
- [13] Guttman, I., Pereyya, V. and Scolink, H.D. (1973), "Least Squares Estimation for a Class of Non-Linear Models", *Technometrics*, 15, 209-218.
- [14] Johnston, J. (1984), "Econometric Methods", 3rd Edition, Mc Graw Hill, New York.
- [15] Judge, G.G. and et al., (1988), "The Theory and Practice of Econometrics", Second Edition, John Wiley and Sons, New York, p. 394.
- [16] Kmenta, J. (1964), "Some properties of Alternative Estimates of the Cobb-Douglas Production Function", *Econometrica*, 32, 183-188.
- [17] Kobyashi, M. (1991), "Testing for Autocorrelated Disturbances in Nonlinear Regression Analysis", *Econometrica*, 59, 1153-1159.
- [18] Markatou, M. and Manos, G. (1996), "Robust Tests in Nonlinear Regression Models", *Journal of Statistical Planning and Inference*, 55, 205-217.
- [19] Mc Cullagh, P. and Cox, D.R. (1986), "Invariants and Likelihood Ratio Statistics", *Ann. Statist.* 14, 1419-1430.
- [20] Owyong, D.T. and Bhanogi Rao, v.v. (1998), "Total Factor Productivity Growth in the Singapore Economy : Some Econometric Systems", *The Singapore Economic Review*, 43, 74-81.
- [21] Stato, R. (1967), "A Two Level Constant – Elasticity of Substitution Production Function", *Review of Economic Studies*, 34, 201-218.
- [22] Stephen, M.G. and Quardt, R.e. (1970), "The Estimation of Cobb-Douglas Type Functions with Multiplicative and Additive Errors", *International Economic Review*, 11, 251-257.
- [23] Thurby, J. (1980), "Alternatives CES Estimation Techniques", *Review of Economics and Statistics*, 62, 295-299..