Journal of Al Azhar University-Gaza (Natural Sciences), 2010, 12 : 35-54

A Stochastic Approximation – Iterative Least Squares Estimation Procedure

Shahnaz Ezaldeen Abu- Qamar

Department of Applied Statistics
Faculty of Economics and Administration Sciences
Al-Azhar University - Gaza

Received 11/5/2010 Accepted 2010

***Abstract**: Lawton and Sylvestre (1971) consider a fixed sample size estimation procedure for a nonlinear regression model. In this paper we propose a stochastic approximation – iterative least squares procedure. Our Procedure leads to a significant reduction in the sample size.*

1. INTRODUCTION AND SUMMARY

Consider the following nonlinear regression model:

$$Y\left(\underset{\sim}{X}\right)= g\left(\underset{\sim}{\theta};\underset{\sim}{X}\right)+\in \qquad (1.1)$$

Where g: $\Re^P \times \Re^r \to \Re,$ with $\Re^P$ and $\Re^r$ being Euclidean spaces, $\in$ is an unobservable random error, with E $(\in)$= 0, var $(\in)$= $\sigma^2;\sigma^2$ is a constant that may depend on $\underset{\sim}{X};Y\left(\underset{\sim}{X}\right)$ is an observable random response that can be observed at each level $\underset{\sim}{X} \in \Re^r$; and $\underset{\sim}{\theta} \in \Re^P$ is the parameters of interest. Based on the observations $Y_1, Y_2,....Y_n$. it has been known, [3], [4], [5], and [7], via classical procedures, how to estimate $\underset{\sim}{\theta} = (\theta_1,..., \theta_P)$. Our interest will be in the class of models which contain a component linear in some parameters but nonlinear in the remaining parameters. The objective

will be to estimate $\underset{\sim}{\theta}$ sequentially using a technique in which the optimal stochastic approximation method [1], is combined with the approach of eliminating linear parameters proposed by [4]. The sequential procedure is also compared with the fixed sample size procedure based fully on the Lawton and Sylvestre method.

Now to achieve our objective, i.e. to estimate $\underset{\sim}{\theta}$ sequentially, we can then use the following optimal stochastic approximation procedure [1]: Choose

$\underset{\sim}{\hat{\theta}}^{(1)}$ as an arbitrary initial estimate of $\underset{\sim}{\theta}$, then define the estimating sequence $\left(\underset{\sim}{\hat{\theta}}^{(n)}\right)$ by:

$$\underset{\sim}{\hat{\theta}}^{(n+1)} = \underset{\sim}{\hat{\theta}}^{(n)} - a_n h_o\left(Y_n\right) \quad , n = 1,2,.... \tag{1.2}$$

where

$$h_o\left(Y_n\right) = \left[-\left(\underset{\sim}{I}_n\left(\underset{\sim}{\theta}\right)\right)^{-1}\left[\underset{\underset{\sim}{\theta}}{grad}\, f\left(Y_n, \underset{\sim}{\theta}\right) / f\left(Y_n, \underset{\sim}{\theta}\right)\right]_{\underset{\sim}{\theta}=\underset{\sim}{\hat{\theta}}^{(n)}}\right] \quad , n = 1.2\ldots..$$

and a_n is sequence satisfying

$\sum_{n=1}^{\infty} a_n = \infty, \sum_{n=1}^{\infty} a_n^2 \prec \infty, \left(e.g. a_n = a/n, a \succ 0\right)$, and $\underset{\sim}{I}_n\left(\underset{\sim}{\theta}\right)$ is (P×P) – Fisher information matrix.

The Lawton and Sylvestre method (1971) of estimation may be applied when the nonlinear regression model (1.1) has the special form

$$Y\left(\underset{\sim}{X}\right)=\sum_{j=1}^{q}\theta_{1,j}g_j\left(\underset{\sim(2)}{\theta};\underset{\sim}{X}\right)+\in \qquad (1.3)$$

where $\theta_{1,1},\theta_{1,2},.....,\theta_{1,q}$ enter linearly into the model (1.3), $\underset{\sim(2)}{\theta}$ represents the vector of nonlinear parameters in (1.3), and the $g_j\left(\underset{\sim(2)}{\theta};\underset{\sim}{X}\right)$ are functions only of the nonlinear parameters and the predictor variables, i.e. $g_j:\Re^{(p-q)}\times\Re^r\rightarrow\Re,\ j=1,2,....,q.$ their method is used for the fixed sample size case when observations $Y_1,Y_2,....,Y_n$ are available.

Using their procedure, we take $\underset{\sim(2)}{\hat{\theta}}$ as an initial value $\underset{\sim(2)}{\theta}$ of and then determine the companion set of "best" values for $\underset{\sim(1)}{\theta}$ by the ordinary least squares procedure.

Let $\underset{\sim(1)}{\hat{\theta}}\left(\underset{\sim(2)}{\theta}\right)=\left(\hat{\theta}_{1,1}(\underset{\sim}{\theta})_{(2)},\ \hat{\theta}_{1,2}\left(\underset{\sim(2)}{\theta}\right),.....,\hat{\theta}_{1,q}\left(\underset{\sim(2)}{\theta}\right)\right)'$ represent the vector of least squares estimates of the $\theta_{1,j}'s$ associated with a given set of $\underset{\sim(2)}{\theta}'s;$ namely, $\underset{\sim(2)}{\theta}=\left(\theta_{2,1},...,\theta_{2,P-q}\right)'$

Let $\underset{\sim}{Y}$ denote the (nx1) column vector of observed response values associated with the n observed values of the predictor vector, $\underset{\sim i}{x},i=1,2,...,n$.

Let $G_{\underset{\sim(2)}{\theta}}$ denote the (n×q) matrix with elements $g_j\left(\underset{\sim(2)}{\theta};\underset{\sim}{x_i}\right)$, $i=1,2,..,n$, $j=1,2,...,q.$ It then follows that, the vector $\underset{\sim}{\hat{\theta}}\left(\underset{\sim(2)}{\theta}\right)$, provided that $\left[G'_{\underset{\sim(2)}{\theta}}\ G_{\underset{\sim(2)}{\theta}}\right]^{-1}$ exists; is given by: $\underset{\sim(1)}{\hat{\theta}}\left(\underset{\sim(2)}{\theta}\right)=\left[G'_{\underset{\sim(2)}{\theta}}\ G_{\underset{\sim(2)}{\theta}}\right]^{-1}G'_{\underset{\sim(2)}{\theta}}\ \underset{\sim}{Y}$.

The reduced "model" associated with (1.3) is then given by:

$$Y\left(\underset{\sim}{X}\right)=\sum_{j=1}^{q}\hat{\theta}_{1,j}\left(\underset{\sim(2)}{\theta}\right)g_j\left(\underset{\sim(2)}{\theta};\underset{\sim}{X}\right)+\in^{*} \qquad (1.4)$$

Since $\hat{\theta}_{1.j}\left(\underset{\sim(2)}{\theta}\right)$ are strictly functions of $\underset{\sim(2)}{\theta}$'s, the model in (1.4) is a nonlinear regression model with only (p-q) parameters rather than the p parameters in the original model. Lawton and Sylvestre (1971) proposed an iterative method like the linearization method, and steepest descent method, [4] to estimate the remaining unknown nonlinear parameters. This procedure of Lawton and Sylvestre estimates the nonlinear parameters in an inconsequent fashion, that is the whole data must be used to find values of the estimators. If the data is drawn sequentially, then these procedures will not be suitable to use. However the stochastic approximation procedures have been shown to be "optimal" [1] in the sense that the estimating sequence $\left(\underset{\sim}{\theta}^{(n)}\right)$ is a consistent and asymptotically efficient estimator of $\underset{\sim}{\theta}$ such that, the variance of the asymptotic distribution of $n^{\frac{1}{2}}\left(\underset{\sim}{\theta}^{(n)}-\underset{\sim}{\theta}\right)$ achieves the Cramer-Rao lower bound for the variance of an unbiased estimator of $\underset{\sim}{\theta}$. The above results show that it is worthwhile to consider the use of stochastic approximation procedures to estimate

sequentially the nonlinear parameters in (1.4), instead of using any iterative classical method.

We will illustrate the fixed sample size procedure of [4] using the following example given by these authors. Let

$$y(x) = \theta_1 e^{(\theta_2 x)} + \in \tag{1.5}$$

Where θ_1 and θ_2 are two unknown parameters to be estimated, $\in$ is an unobservable random error and (x) is a response variable at the level $x.\theta_1$ appears linearly in the model (1.5). We seek the least squares estimators $\hat{\theta}_1, \hat{\theta}_2$ which minimize

$$Q(\theta_1, \theta_2) = \sum_{i=1}^{n} \left(y_i - \theta_1 e^{(\theta_2 X_i)}\right)^2 \tag{1.6}$$

It follows that the best value of θ_1 given θ_2 denoted by $\hat{\theta}_1(\theta_2)$

$$\left[\hat{\theta}_1(\theta_2) = \frac{\sum_{i=1}^{n} y_i e^{\theta_2 x_i}}{\sum_{i=1}^{n} e^{(2\theta_2 x_i)}}\right] \tag{1.7}$$

Now substitute (1.7) into (1.5). The linear parameter θ_1 is automatically replaced by its best companion value $\hat{\theta}_1(\theta_2)$ which is a function of θ_2 alone.

One then obtains the reduced "model", given by:

$$Y(X) = \hat{\theta}_1(\theta_2) e^{(\theta_2 X)} + \in^* \tag{1.8}$$

The parameter θ_2 will be estimated iteratively by using any of the iterative methods mentioned previously in section 1.

2. STOCHASTIC APPROXIMATION – ITERATIVE LEAST SQUARES PROCEDURE

In this section, we describe a new sequential procedure for estimating the parameters in the model given by (1.3), which combines the stochastic approximation technique, with the iterative least squares technique. For abbreviation this will be referred to, as the SA-ILS procedure. Clearly, the reduced "model" in (1.8) is a nonlinear model with a single parameter θ_2. In order to estimate θ_2 sequentially by using optimal stochastic approximation procedure (see Sec. 1), we shall consider certain probability models for $\in$ by using the reduced "model" in (1.8), and then find the probability density function for y, $f(y;\theta_2)$ by transformation. Thus by using optimal stochastic approximation procedure of the form (1.2), in order to estimate θ_2 sequentially, choose $\overset{\wedge}{\theta}_2^{(1)}$ as an arbitrary initial estimate of θ_2, and then define the estimating sequence $\left(\overset{\wedge}{\theta}_2^{(n)}\right)$ by:

$$\overset{\wedge}{\theta}_2^{(n+1)} = \overset{\wedge}{\theta}_2^{(n)} + a_n\left[I_n\left(\overset{\wedge}{\theta}_2^{(n)}\right)\right]^{-1}\left[\frac{df(y_n;\theta_2)/d\theta_2}{f(y_n;\theta_2)}\right]_{\theta_2=\overset{\wedge}{\theta}_2^{(n)}}, n=1,2,... \quad (2.1)$$

where $a_n = 1/n$ since the optimal value of (a) that minimizes the variance of the asymptotic distribution of $n^{\frac{1}{2}}\left(\overset{\wedge}{\theta}_2^{(n)} - \theta_2\right)$ [1] is given by 1. For notational convenience we set

$f_n\left(Y_n;\overset{(n)}{\theta_1},\theta_2\right) = f\left(y_n;\theta_2\right)$ and denote $\underset{\sim(1)}{\overset{\wedge}{\theta}^{(n)}}\left(\underset{\sim(2)}{\overset{\wedge}{\theta}^{(n)}}\right)$ by $\underset{\sim(1)}{\overset{\wedge}{\theta}^{(n)}}$.

Then (2.1) becomes

$$\hat{\theta}_2^{(n+1)} = \hat{\theta}_2^{(n)} + \frac{1}{n}\left[I_n \hat{\theta}_2^{(n)} \right]^{-1} \left[\frac{df(y_n;\theta_2)/d\theta_2}{f(y_n;\theta_2)} \right]_{\theta_2=\hat{\theta}_2^{(n)}}, \text{n=1,2,…} \quad (2.2)$$

Let $\hat{\underset{\sim}{\theta}}^{(n)}_{(1)}\left(\underset{\sim}{\theta}_{(2)}\right) = \left(\hat{\theta}_{1.1}^{(n)}(\underset{\sim}{\theta}_{(2)}),...,\hat{\theta}_{1,q}^{(n)}\left(\underset{\sim}{\theta}_{(2)}\right)\right)'$ represent the vector of iterative least squares estimates.

Let Y_1 be the first observation, $\underset{\sim}{Y}_2$ be the vector of the first two observations, and so on, $\underset{\sim}{Y}_n$ denoting the vector of the first n observed response values which have associated observed values of the predictor vector, $\underset{\sim}{X}_i, i = 1,2,....,n$. Let $G_{\hat{\underset{\sim}{\theta}}^{(n)}_{(2)}}$ denote the $(n \times q)$ matrix with elements $g_j\left(\hat{\underset{\sim}{\theta}}^{(n)}_{(2)}; \underset{\sim}{x}_i\right)$, $i = 1,2,....,n$, $j = 1,2,....,q$ It then follows that the sequence $\left(\hat{\underset{\sim}{\theta}}^{(n)}_{(1)}\left(\hat{\underset{\sim}{\theta}}^{(n)}_{(2)}\right)\right)$, provided that $\left[G'_{\hat{\underset{\sim}{\theta}}^{(n)}_{(2)}} \; G_{\hat{\underset{\sim}{\theta}}^{(n)}_{(2)}} \right]^{-1}$ exists, is given by:

$$\hat{\underset{\sim}{\theta}}^{(n)}_{(1)}\left(\hat{\underset{\sim}{\theta}}^{(n)}_{(2)}\right) = \hat{\underset{\sim}{\theta}}^{(n-1)}_{(1)}\left(\hat{\underset{\sim}{\theta}}^{(n-1)}_{(2)}\right) + \left[\left(G'^{(n)}_{\hat{\underset{\sim}{\theta}}_{(2)}} G^{(n)}_{\hat{\underset{\sim}{\theta}}_{(2)}} \right)^{-1} G'_{\hat{\underset{\sim}{\theta}}^{(n)}_{(2)}} \underset{\sim}{y}_n - \hat{\underset{\sim}{\theta}}^{(n-1)}_{(1)}\left(\hat{\underset{\sim}{\theta}}^{(n-1)}_{(2)}\right) \right]$$

, n=1,2,….

i.e.

$$\underset{\sim}{\overset{\wedge}{\theta}}_{(1)}^{(n)} = \underset{\sim}{\overset{\wedge}{\theta}}_{(1)}^{(n-1)} + \left[\left(G'_{\underset{\sim}{\overset{\wedge}{\theta}}_{(2)}^{(n)}} G_{\underset{\sim}{\overset{\wedge}{\theta}}_{(2)}^{(n)}} \right)^{-1} G'_{\underset{\sim}{\overset{\wedge}{\theta}}_{(2)}^{(n)}} Y_{\sim n} - \underset{\sim}{\overset{\wedge}{\theta}}_{(1)}^{(n-1)} \right] \quad n=1,2,\ldots \qquad (2.3)$$

Where $\underset{\sim}{\overset{\wedge}{\theta}}_{(2)}^{(1)}$ is an arbitrary initial value for the sequence $\left(\underset{\sim}{\overset{\wedge}{\theta}}_{(2)}^{(n)} \right)$, and

$\underset{\sim}{\overset{\wedge}{\theta}}_{(1)}^{(1)} = \left(G'_{\underset{\sim}{\overset{\wedge}{\theta}}_{(2)}^{(1)}} G_{\underset{\sim}{\overset{\wedge}{\theta}}_{(2)}^{(1)}} \right)^{-1} G'_{\underset{\sim}{\overset{\wedge}{\theta}}_{(2)}^{(1)}} Y_1$ is an initial value for the sequence $\left(\underset{\sim}{\overset{\wedge}{\theta}}_{(1)}^{(n)} \right)$

The reduced model associated with (1.3) is then given by:

$$Y\left(\underset{\sim}{X}\right) = \sum_{j=1}^{q} \overset{\wedge}{\theta}_{1,j}\left(\underset{\sim}{\theta}_{(2)}\right) g_j\left(\underset{\sim}{\theta}_{(2)}; \underset{\sim}{X}\right) + \in^{*} \qquad (2.4)$$

Since the $\overset{\wedge}{\theta}_{1,j}\left(\underset{\sim}{\theta}_{(2)}\right)$ are strictly functions of the $\underset{\sim}{\theta}_{(2)}$'s; the model (2.4) is a nonlinear regression model with only (p-q) parameters, and we will estimate them by using the optimal stochastic approximation procedure of the form (2.2). The main idea of SA- ILS procedure is to estimate the parameters which enter the model linearly, by using an iterative form of least squares estimator, sequentially, and then use a proper optimal stochastic approximation procedure to sequentially estimate nonlinear parameters. Therefore, we will use iterative least squares procedure in order to estimate $\theta_1(\theta_2)$ sequentially.

Now, we illustrate the use of the iterative least squares procedure using the example discussed by [4]. From (1.7) we have

$$\overset{\wedge}{\theta}_1(\theta_2) = \frac{\sum_{i=1}^{n} y_i e^{(\theta_2 x_i)}}{\sum_{i=1}^{n} e^{(2\theta_2 x_i)}}$$

Given an initial guess $\hat{\theta}_2^{(1)}$, we then have:

$$\hat{\theta}_1^{(1)}\left(\hat{\theta}_2^{(1)}\right)=\frac{y_1}{\left(e^{\hat{\theta}_2^{(1)} x_1}\right)}$$

Substitute $\hat{\theta}_1^{(1)}$ in (2.2), we will get $\hat{\theta}_2^{(2)}$ and so

$$\hat{\theta}_1^{(2)}\left(\hat{\theta}_2^{(2)}\right)=\frac{\sum_{i=1}^{2} y_i e^{\left(\hat{\theta}_2^{(2)} x_i\right)}}{\sum_{i=1}^{2} e^{\left(2\hat{\theta}_2^{(2)} x_i\right)}}$$

In general, at stage n, by substituting $\hat{\theta}_1^{(n-1)}$ in (2.2), we get $\hat{\theta}_2^{(n)}$ and then

$$\hat{\theta}_1^{(n)}=\hat{\theta}_1^{(n-1)}+\frac{1}{\sum_{i=1}^{n} e^{\left(2\hat{\theta}_2^{(n)} X_i\right)}}\left[\sum_{i=1}^{n} y_i e^{\hat{\theta}_2^{(n)} X_i}-\hat{\theta}_1^{(n-1)}\sum_{i=1}^{n} e^{(2\hat{\theta}_2^{(n)} X_i)}\right],\text{n=1,2,…,} \quad (2.5)$$

Therefore the SA_ILS procedure is given by the following two consecutive procedures:

$$\hat{\theta}_2^{(n+1)} = \hat{\theta}_2^{(n)} + \frac{1}{n}\left[I_n\left(\hat{\theta}_2^{(n)} \right) \right]^{-1} \left[\frac{df(y_n;\theta_2)/d\theta_2}{f(y_n;\theta_2)} \right]_{\theta_2 = \hat{\theta}_2^{(n)}}$$

, n=1,2,… and

$$\hat{\theta}_1^{(n)} = \hat{\theta}_1^{(n-1)} + \frac{1}{\sum_{i=1}^{n} e^{\left(2\hat{\theta}_2^{(n)} x_i \right)}} \left[\sum_{i=1}^{n} y_i e^{\hat{\theta}_2^{(n)} x_i} - \hat{\theta}_1^{(n-1)} \sum_{i=1}^{n} e^{\left(2\hat{\theta}_2^{(n)} x_i \right)} \right]$$

, n=1,2,…

Where $\hat{\theta}_2^{(1)}$ is an arbitrary initial value for the sequence $\left(\hat{\theta}_2^{(n)} \right)$ and

$\hat{\theta}_1^{(1)} = y_1 / e^{\left(\hat{\theta}_2^{(1)} X_1 \right)}$ is an initial estimate of $\left(\hat{\theta}_1^{(n)} \right)$ based on $\left(\hat{\theta}_2^{(1)} \right)$

The following algorithm illustrates the computation of the first three estimates for θ_1 and θ_2 in the previous example using the SA_ILS procedure:

Step 1: (Initialization): let $\hat{\theta}_2^{(1)}$ be an arbitrary initial estimate of θ_2.

Step 2: (First approximation): For $\theta_2 = \hat{\theta}_2^{(1)}$ and data (x_1, y_1), the value of θ_1 which minimizes $\left(y_1 - \theta_1 e^{(\hat{\theta}_2^{(1)} x_1)} \right)^2$ is obtained as

$$\hat{\theta}_1^{(1)} = y_1 / e^{\hat{\theta}_2^{(1)} x_1}$$

Step 3: (Improvement): Treating $\overset{\wedge}{\theta}_1^{(1)}$ as if it was the known true value of θ_1, the second estimate of θ_2 is obtained from

$$\overset{\wedge}{\theta}_2^{(K+1)} = \overset{\wedge}{\theta}_2^{(K)} + \frac{1}{K}\left[I_K\left(\overset{\wedge}{\theta}_2^{(K)} \right) \right]^{-1} \left[\frac{df(y_K;\theta_2)/d\theta_2}{f(y_K;\theta_2)} \right]_{\theta_2 = \overset{\wedge}{\theta}_2^{K}},$$

, K=1,2,...

Step 4: For $\theta_2 = \overset{\wedge}{\theta}_2^{(2)}$ and data $(x_1, y_1), (x_2, y_2), \ldots, (y_K, x_K)$, the least squares estimate for θ_1 is easily seen to be in the form

$$\overset{\wedge}{\theta}_1^{(K)} = \overset{\wedge}{\theta}_1^{(K-1)} + \frac{1}{\sum_{n=1}^{K} e^{(2\overset{\wedge}{\theta}_2^{(K)} x_n)}} \left[\sum_{n=1}^{K} y_n e^{(\overset{\wedge}{\theta}_2^{(K)} x_n)} - \overset{\wedge}{\theta}_1^{(K-1)} \sum_{n=1}^{K} e^{\left(2\overset{\wedge}{\theta}_2^{(K)} x_n \right)} \right], \quad =1,2,\ldots$$

Step 5: Repeat the above steps until $\left| \frac{\left(\hat{\theta}_2^{(n+1)} - \hat{\theta}_2^{(n)} \right)}{\hat{\theta}_2^{(n)}} \right| \prec \delta$,

where δ is a small specified positive number**.**

3. EXAMPLE OF THE USE OF THE SA- ILS PROCEDURE UNDER DIFFERENT ERROR DISTRIBUTIONS

We shall consider the following nonlinear regression function;

3.1 Example: [2] Let

$$Y(X) = \theta_1 \sin(\theta_2 X) + \in$$

Also, assume the following two probability models for $\in$:

3.2 $\in$ is assumed to be normally distributed with mean zero and variance 1,

3.3 $\in$ is assumed to have a T- distribution with r degrees of freedom ($r \geq 1$)

Which includes the Cauchy distribution (r =1).

First of all we will explain in an analytical form the steps of the procedure for this sample.

Under 3.2 It follows that Y is also distributed as $N(\theta_1 Sin(\theta_2 X);1)$. Treating θ_1 as known initially; and differentiating the log of the density of Y with respect to θ_2 we get:

$$\frac{d \ln f\ (Y;\theta_1,\theta_2)}{d\theta_2} = X\ \theta_1 \cos(\theta_2 X)\left(Y - \theta_1 \sin(\theta_2 X)\right) \quad , -\infty \prec y \prec \infty$$

The Fisher information, $I(\theta_2)$, is seen to be equal to

$$I(\theta_2) = (x\theta_1 \cos(\theta_2 x))^2 \qquad (3.1)$$

From Section 2 the optimal transformation for the stochastic approximation procedure is

$$h_o(y_n) = -(I_n(\theta_2))^{-1}\left[\frac{d \ln f\left(y_n;\hat{\theta}_1,\theta_2\right)}{d\theta_2}\right],\ n = 1,2,....$$

$$= \frac{-\left(y_n - \hat{\theta}_1(\theta_2)\sin(\theta_2 x_n) \right)}{\left(\hat{\theta}_1(\theta_2) x_n \cos(\theta_2 x_n) \right)}$$

$$; x_n \notin \left\{ o \cup \frac{\pi}{2} + v\pi, v = 0,1,2,... \right\}$$

The optimal value of a that minimizes the variance of the asymptotic distribution of $n^{\frac{1}{2}}\left(\hat{\theta}_2^{(n)} - \theta_2 \right)$ [1] is given by 1, thus

$$a_n = 1/n$$

Now choose $\hat{\theta}_2^{(1)}$ as an arbitrary initial estimate of θ_2, then define the estimating sequence $\left(\hat{\theta}_2^{(n)} \right)$ by:

$$\hat{\theta}_2^{(n+1)} = \hat{\theta}_2^{(n)} + \frac{1}{n} \frac{\left(Y_n - \hat{\theta}_1^{(n)}\left(\hat{\theta}_2^{(n)} \right) \sin\left(\hat{\theta}_2^{(n)} X_n \right) \right)}{\left(\hat{\theta}_1^{(n)}\left(\hat{\theta}_2^{(n)} \right) X_n \cos\left(\hat{\theta}_2^{(n)} X_n \right) \right)}, n = 1,2,...$$

$x_n \notin \{0 \cup \frac{\pi}{2} + v\pi, v = 0,1,2,...\}$, and the estimating sequence of $\left(\hat{\theta}_1^{(n-1)} \right)$ is given by:

$$\hat{\theta}_1^{(n)} = \hat{\theta}_1^{(n-1)} + \frac{1}{\sum_{i=1}^{n}(\sin(\hat{\theta}_2^{(n)}X_i))^2}\left[\sum_{i=1}^{n} y_i \sin(\hat{\theta}_2^{(n)}X_i) - \hat{\theta}_1^{(n-1)}\sum_{i=1}^{n}(\sin(\hat{\theta}_2^{(n)}X_i))\right]^2,$$

n=1,2,…;

$x_1 \neq v\pi$, v=0,1,2,…, where $\hat{\theta}_1^{(1)} = y_1 / \sin(\hat{\theta}_2^{(1)}X_1),$ and $x_1 \neq v\pi$, v =0,1,2,…

Under 3.3 it follows that the density of y is

$$f(y;\theta_1,\theta_2) = \frac{\Gamma((r+1)/2)}{\Gamma(r/2)\sqrt{r\pi}}\left[1+(y-\theta_1\sin(\theta_2 X))^2/r\right]^{-\frac{(1+r)}{2}}$$
$$,-\infty \prec y \prec \infty.$$

Treating θ_1 as known initially, we obtain, after some tedious manipulations, that

$$I(\theta_2) = \frac{Rr^{3/2}}{r+3}\int_0^{\pi/2}\cos^{(r+1)}(\theta)d\theta,$$

where

$$R = \frac{2(\theta_1 X\cos(\theta_2 X)(1+r))^2\Gamma\left(\frac{r+1}{2}\right)}{\Gamma(r/2)r^2\sqrt{\pi r}}$$

It is straight forward to show that an alternative representation for I (θ_2) is

$$I(\theta_2) = \frac{r^{5/2}R}{(r+3)(r+1)}\int_0^{\pi/2}\cos^{(r-1)}(\theta)d\theta \qquad (3.2)$$

which has the advantage of a smaller power in the integrant. Using (3.2), we get the following: let $\left(\hat{\theta}_2^{(1)}\right)$ be an arbitrary initial estimate of θ_2, then it follows, after some simplifications, that the estimating

sequence $\left(\hat{\theta}_2^{(n)}\right)$ is given by:

$$\hat{\theta}_2^{(n+1)} = \hat{\theta}_2^{(n)} + \frac{(r+3)\sqrt{\pi}(y_n - \hat{\theta}_1^{(n)}(\hat{\theta}_2^{(n)})\sin(\hat{\theta}_2^{(n)}X_n))\left[\int_0^{\pi/2}\cos^{(r-1)}(\theta)d\theta\right]^{-1}}{2n\left[\hat{\theta}_1^{(n)}(\hat{\theta}_2^{(n)})X_n\cos(\hat{\theta}_2^{(n)}X_n\right]\Gamma\left[\frac{r+1}{2}\right]\left[r+\left(y_n-\hat{\theta}_1^{(n)}\left(\hat{\theta}_2^{(n)}\right)\sin(\hat{\theta}_2^{(n)}X_n)\right)^2\right]}; \quad (3.3)$$

$X_n \notin \left\{0 \cup \frac{\pi}{2} + v\pi, v = 0,1,2,...\right\}$; $n = 1,2,...$, $r \geq 1$, and, the estimating sequence

$\left(\hat{\theta}_1^{(n-1)}\right)$ is given by:

$$\hat{\theta}_1^{(n)} = \hat{\theta}_1^{(n-1)} + \frac{1}{\sum_{i=1}^{n}(\sin(\hat{\theta}_2^{(n)}x_i))^2}\left[\sum_{i=1}^{n} y_i \sin(\hat{\theta}_1^{(n)}x_i) - \hat{\theta}_1^{(n-1)}\sum_{i=1}^{n}(\sin(\hat{\theta}_2^{(n)}x_i))^2\right];$$

$n = 1,2,...$, $x_i \neq \pi v$ *for* i=1,2,…,n , v = 0,1,2,…, where

$$\hat{\theta}_1^{(1)} = y_1 / \sin(\hat{\theta}_2^{(1)}x_1).$$

Taking different cases of degrees of freedom r, the integral in (3.3) can be shown to be π/2, 1, 2/3, 384/945… for r=1,2,4,10, respectively. The estimating sequence $\left(\hat{\theta}_2^{(n)}\right)$ in each case has the form

$$\hat{\theta}_2^{(n+1)} = \hat{\theta}_2^{(n)} + \frac{\lambda_r\left(y_n - \hat{\theta}_1^{(n)}\left(\hat{\theta}_2^{(n)}\right)\sin\left(\hat{\theta}_2^{(n)}X_n\right)\right)}{n\left[\hat{\theta}_1^{(n)}\left(\hat{\theta}_2^{(n)}\right)X_n\cos\left(\hat{\theta}_2^{(n)}X_n\right)\left[r+\left(y_n-\hat{\theta}_1^{(n)}\left(\hat{\theta}_2^{(n)}\right)\sin\left(\hat{\theta}_2^{(n)}X_n\right)\right)^2\right]};$$

$n = 1,2,...$, $x_{(n)} \notin \left\{0 \cup \frac{\pi}{2} + v\pi, v = 0,1,2,...\right\}$ where

$\lambda_r = 4,5,7,13$ *for* $\mathrm{r} = 1,2,4,10$ respectively. In each case the estimating sequence $\left(\hat{\theta}_1^{(n-1)}\right)$ is given by (3.4) .The above coefficient values for λ_r suggest that the general form is $\lambda_r = r + 3$, but we have not proved this analytically.

4. NUMERICAL SOLUTION USING THE LAWTON AND SYLVESTRE PROCEDURE

Lawton and sylvestre considered the special case when the model has a linear and nonlinear component (see equation (1.3)). They introduce a modification based on the idea of reducing the number of parameters that must be estimated by the iterative methods. For a sample Y_1, Y2,…,Y_n, the linear parameters are estimated at each stage by ordinary least squares and the estimates are then substituted into(1.3).

A general outline of the procedure was given is Section 1. We will discuss now the previous example in some detail.

Example:

We consider the model given in Section 3, that is

$$Y(X) = \theta_1 \sin(\theta_2 X) + \varepsilon$$

In Section 2 we have found that the least squares estimate of θ_1 given by: $\hat{\theta}_1(\theta_2) = \dfrac{\sum_{i=1}^{n} y_i \sin(\theta_2 x_i)}{\sum_{i=1}^{n} (\sin(\theta_2 x_i))^2}$,

and, the reduced " model " will be

$$Y(X) = \hat{\theta}_1(\theta_2)\sin(\theta_2 X) + \varepsilon^*,$$

Where this "model", is treated as a nonlinear model with a single parameter θ_2. We have used a linearization method as an iterative method for estimating the nonlinear parameter θ_2. This method has been explained previously [4]. We will apply it directly as follows: Let

$$\underline{Z}^{(j)} = \begin{bmatrix} X_1\hat{\theta}_1(\theta_2^{(j)})\cos(\theta_2^{(j)}X_1) \\ X_2\hat{\theta}_1(\theta_2^{(j)})\cos(\theta_2^{(j)}X_2) \\ \vdots \\ \vdots \\ X_n\hat{\theta}_1(\theta_2^{(j)})\cos(\theta_2^{(j)}X_n) \end{bmatrix};\quad \underline{Y} - \underline{g}^{(j)} = \begin{bmatrix} Y_1 - \hat{\theta}_1(\theta_2^{(j)})\sin(\theta_2^{(j)}X_1) \\ Y_2 - \hat{\theta}_1(\theta_2^{(j)})\sin(\theta_2^{(j)}X_2) \\ \vdots \\ \vdots \\ Y_n - \hat{\theta}_1(\theta_2^{(j)})\sin(\theta_2^{(j)}X_n) \end{bmatrix};$$

J=1,2,… .

Then, define an estimating sequence $\begin{pmatrix} \hat{\theta}_2^{(j)} \end{pmatrix}$ by:

$$\hat{\theta}_2^{(j+1)} = \hat{\theta}_2^{(j)} + \left(\underline{Z}'^{(j)}\,\underline{Z}^{(j)}\right)^{-1}\left(\underline{Z}'^{(j)}\left(\underline{Y} - \underline{g}^{(j)}\right)\right),\ j = 1,2,\ldots$$

Provided that $\left(\underline{Z}'^{(j)}\,\underline{Z}^{(j)}\right)^{-1}$ exists, i.e.

$$\hat{\theta}_2^{(j+1)} = \hat{\theta}_2^{(j)} + \left[\sum_{i=1}^{n}\left[X_i\hat{\theta}_1(\hat{\theta}_2^{(j)})\cos(\hat{\theta}_2^{(j)}X_i\right]^2\right]^{-1}\left[\sum_{i=1}^{n}\left[\left[X_i\hat{\theta}_1(\hat{\theta}_2^{(j)})\cos(\hat{\theta}_2^{(j)}X_i\right]\times\left[Y_i - \hat{\theta}_1(\hat{\theta}_2^{(j)})\sin(\hat{\theta}_2^{(j)}X_i\right]\right]\right]'$$

j=1,2,…, where $\hat{\theta}_2^{(1)}$ is an initial estimate of θ_2 .

The above estimates, $\hat{\theta}_2^{(j+1)}$, will be iteratively computed, in each iteration the "best" companion value of $\hat{\theta}_1\left(\hat{\theta}_2^{(j)}\right)$, will be computed by the least squares method.

5- A SIMULATION STUDY

In this section we report the findings of a simulation study to compare the properties of the SA-ILS procedure and Lawton and Sylvestre fixed sample size procedure. The model used that we discussed in Section 3 where $Y = \theta_1 \sin(\theta_2 X) + \varepsilon$

Where, the error term is assumed to have ϵ ~ N(0,1) and T-distribution with

r =1,10. Different Dseeds are taken as 1234.0D0, 123457.0D0, 17.0D0, 12.0D0 generate four samples each of size 100 from the T-distribution mentioned above, using the IMSL routine name GGAMR. Values of (θ_1, θ_2) are taken as (0.25,0.65), (0.45,0.85),(0.55,0.95) and (0.45,0.95) were used to give different pattern for the estimates and number of observations needed for convergence. For the fixed sample size n-1000 was used, we get the same results as if the sample size n=100.

The residuals ϵ were generated using the random normal deviate generator available in the IMSL routine name GGNML. We are interested in comparing the SA-ILS procedure and the fixed sample size Lawton and Sylvestre procedure from the point of view of the number of observations needed for convergence.

The following tables give the sample numbers of observations needed for convergence for the Lawton and sylvestre, and SA-ILS procedure, where δ is taken to be 10^{-2}.

Table 5.1: Model: $Y_i = \theta_1 \sin(\theta_2 X_i) + \varepsilon_i$, where $\{\varepsilon_i\} \sim N(0,1)$

			SA-ILS Procedure	Lawton and Sylvestre Procedure
Dseed	(θ_1, θ_2)	Initial $\hat{\theta}_2^{(1)}$	no. of observations	No. of iteration s
123457.0D0	(0.25,0.65)	0.75	8	51
12.0D0	(0.45,0.95)	0.75	6	nc.*
12.0D0	(0.45,0.95)	0.65	4	nc.*

- nc. Means no convergence

Table 5.2: Model: $Y_i = \theta_1 \sin(\theta_2 X_i) + \varepsilon_i$, where $\{\varepsilon_i\} \sim$ T-distribution with d.f.=1

			SA-ILS Procedure	Lawton and Sylvestre Procedure
Dseed	(θ_1, θ_2)	Initial $\hat{\theta}_2^{(1)}$	no. of observations	No. of iteration s
1234.0D0	(0.25,0.65)	0.55	9	nc.*
1234.0D0	(0.25,0.65)	0.65	9	nc.*
123457.0D0	(0.45,0.85)	0.75	11	108
123457.0D0	(0.45,0.95)	0.05	11	nc.*

Table 5.3: Model: $Y_i = \theta_1 \sin(\theta_2 X_i) + \varepsilon_i$, where $\{\varepsilon_i\} \sim$ T-distribution with d.f.=10

			SA-ILS Procedure	Lawton and Sylvestre Procedure
Dseed	(θ_1, θ_2)	Initial $\hat{\theta}_2^{(1)}$	no. of observations	no. of observations
17.0D0	(0.55,0.95)	0.05	11	266
17.0D0	(0.55,0.95)	0.15	6	201
17.0D0	(0.55,0.95)	0.25	7	nc.*
17.0D0	(0.55,0.95)	0.02	11	nc.*
17.0D0	(0.25,0.65)	0.55	9	323
17.0D0	(0.25,0.65)	0.75	7	160
17.0D0	(0.25,0.65)	0.45	6	333
17.0D0	(0.25,0.65)	0.60	6	174

6. Discussion and conclusions

We observe from the above tables that the SA-ILS procedure is more superior than the Lawton and sylvestre procedure, in the sense that the SA-ILS procedure requires numbers of observations ranging from 4 to 11, while the Lawton and Sylvestre procedure requires many iterations each using all available 100 observations, the number of the iterations is very large in most cases. Thus the SA-I L S procedure leads to a significant reduction in the amount of observations and calculations required.

References

1. Abdelhamid, Sami N. On estimation via optimal stochastic approximation procedures, *Statistical Theory, and Data Analysis*, K. Matusita (Editor) Elsevier Science Publishers B.V. (North Holland). (**1985**).
2. Albert, A. E. and Gardner, L.A. *Stochastic approximation and nonlinear regression*, Research Monograph No.42, M. I.T. Press, (**1967**).
3. Cognez P.; Page D.; Burgot G.; Allain H.; Burgot J.-L (1995) Elimination of linear parameters in non-linear regression: a fast and effective method for the determination of binding parameters. *International Journal of Pharmaceuties*, 123(1), 137-142.
4. Draper, N.R. and Smith, H. *Applied Regression Analysis*, John Wiley and Sons, Inc., New York. (**1966**).
5. Huang, M. N. L. and Huang, M. K, (1991). A Parameter-Elimination Method for Nonlinear Regression with Linear Parameters and Autocorrelation Errors *Biometrical Journal* 33(8), 937-950.
6. Lawton, W.H. and Sylvestre, E.A. Elimination of linear parameters in nonlinear regression, *Technometrics, 13,* **(1971)**. 461-467; Discussion, 477-481.
7. McCullagh, P. and Nelder, J.A. *Generalized Linear Models*, chapman-Hall London. **(1983).**