Available online at www.isas.org.in/jisas

JOURNAL OF THE INDIAN SOCIETY OF AGRICULTURAL STATISTICS 70(1) 2016 51-61

Parameter Estimation in Non-linear Regression Models

Trijya Singh[1], S.K. Mandal[2] and Rajesh Kumar[3]

[1]*Department of Mathematics, Le Moyne College, Syracuse, NY13214, USA*

[2]*Sanjay Gandhi Post Graduate Institute of Medical Sciences, Lucknow*

[3]*ICAR-Indian Institute of Sugarcane Research, Lucknow*

Received 01 February 2016; 28 March 2016; Accepted 01 April 2016

SUMMARY

In many applications, the relationship between the dependent variable and an independent regressor is non-linear in parameters. In such situations, we do not get optimum estimates of parameters in closed form and various non-linear optimization algorithms are used to obtain the optimum estimates. These algorithms are iterative in nature and need good initial estimates of parameters as seed values for a faster and global convergence. This paper proposes various methods based on finite differences to estimate the parameters of non-linear models belonging to the asymptotic regression category. Some published data sets are used to illustrate the application of the proposed methods. It has been demonstrated that the proposed methods produce efficient initial estimates for optimization algorithms.

Keywords: Asymptotic regression, Non-linear regression, Finite differences, Growth curves, Optimization algorithms.

1. INTRODUCTION

In biological, medical, agricultural and economic sciences, quite often the relationship between the response variable and an associated regressor x_t is expressed as

$$y_t = g(x_t, \underset{\sim}{\theta}) + \varepsilon_t \qquad (1.1)$$

where $g(x_t, \underset{\sim}{\theta})$ is a smooth expectation or response function of x_t and a parameter vector $\underset{\sim}{\theta}$ and is assumed to be deterministic. We further assume that the model is exact. The mathematical form of $g(x_t, \underset{\sim}{\theta})$ is assumed to be known. The quantity ε_t is the true additive error with $E(\varepsilon_t) = 0$, for $t = 1,\ldots,n$. Normally, we assume that $\mathrm{Var}(\varepsilon_t) = \sigma^2$ for all t, but this assumption may be violated in some situations. For inference purposes, we also assume that ε_t is normally distributed. Given the pairs of observations (x_t, y_t), $t = 1,\ldots,n$, we have to estimate $\underset{\sim}{\theta}$. The values of x_i's are often equally spaced as they are under the researcher's control. This means that $(x_{t+1} - x_t)$ is the same for all t. The examples of x include dose of fertilizers or drugs, amount of a particular chemical, the time at which measurement is taken, etc. The linearity or non-linearity of model (1.1) depends on how the parameters occur in $g(x, \underset{\sim}{\theta})$ and not how the regressor x does. The response function

Corresponding author: Trijya Singh
E-mail address: singht@lemoyne.edu

$g(x, \underset{\sim}{\theta})$ may be a non-linear function in the variable, non-linear in parameters or non-linear in both. We do not consider the first case in the non-linear regression category. Only the last two cases fall in the category of non-linear regression. A simple test for non-linearity in parameters is that $\frac{\partial}{\partial \underset{\sim}{\theta}} g(x, \underset{\sim}{\theta})$ is independent of every coordinate of $\underset{\sim}{\theta}$. For examples of models falling in the categories above, one may refer to Ratkowsky (1983).

The models which are non-linear in parameters could be put into two categories, (1) intrinsically linear models and (2) intrinsically non-linear models. For the models in the first category the response function $g(x_t, \underset{\sim}{\theta})$ could be made linear in parameters by some suitable variable and parameter transformations. For example, for the model $y_t = At^b e^{-ct} + \varepsilon_t$ an alternative model could be written as $z_t = \alpha + bx_t - ct + e_t$ where $z_t = \log_e(y_t)$, $\alpha = \log_e(A)$ and $x_t = \log(t)$. The error e_t is the true error of the transformed model and does not contain any information about the exogenous variables. However, it is different from ε_t because the transformation of the model also transforms the error. But we still assume E(e_t) = 0. However, the assumption of constant error variance and normality of error terms may no longer be valid. The models in second category cannot be made linear in parameters by any transformation. An important model in this category is a two compartment model given by $C(t) = Ae^{-\alpha t} + Be^{-\beta t} + \varepsilon_t$ which is used to study the pharmaceutical behaviour of a drug in the bodies of humans or animals. Here, $C(t)$ is the concentration of drug in bloodstream at time t elapsed after drug administration. It is this category where the parameter estimation is a major problem.

For the estimation of parameters of model (1.1), a commonly used approach is to minimize the sum of squares of errors (SSE) for a choice of parameters. The SSE is treated as a function of $\underset{\sim}{\theta}$ and is given by

$$SSE(\underset{\sim}{\theta}) = \sum_{t=1}^{n} \varepsilon_t^2 = \sum_{t=1}^{n} [y_t - g(x_t, \underset{\sim}{\theta})]^2 \qquad (1.2)$$

Analytically, by equating $\frac{\partial}{\partial \underset{\sim}{\theta}} SSE(\underset{\sim}{\theta})$ to zero, we get normal equations which are solved for $\underset{\sim}{\theta}$.

In case of linear models explicit solutions of these equations are available but in non-linear case it is not so. Therefore, one has to use iterative procedures like Fisher's method of scoring or the Gauss-Newton method. These methods need initial guess values to start the iterations and require extensive computation. An alternative method is to use non-linear optimization algorithms of computation like Levenberg-Marquardt or the package 'nls(stats)' in R to minimize (1.2) directly. This also needs good initial estimates.

The $SSE(\theta)$ at (1.2) is used as an objective function in algorithms but here we assume that V(ε_t) $= \sigma^2$, for all t = 1,…, n. Quite often, the assumption of constant error variance is violated. There are instances where the variance of y_t varies with x_t. That is, Var(y_t) $= h(x_t)\sigma^2$. For some such instances, one may refer to Chatterjee (1991, p. 48), Draper and Smith (1981, p. 112) and West (1980). In such situations, we use weighted least-square theory and minimize

$$SSE_W(\theta) = \sum_{t=1}^{n} w_t \{y_t - g(x_t, \underset{\sim}{\theta})\}^2 \qquad (1.3)$$

where w_t are inversely proportional to the variances V(ε_t)'s of error terms. Whether or not the variances vary could be ascertained by exploratory data analysis. The graph of y against x will reveal this. Alternatively, one may also use the plot of residuals of the assumed model against x-values for this purpose. If variance is not stable, as a rule of thumb, we use w_t is proportional to $\frac{1}{y_t^2}$ for t=1,2,…,n. In cases where it is known apriori that Var(ε_t) = $h(x_t)\sigma^2$, we take $w_t = \frac{1}{h(x_t)}$.

In earlier days, when access to computing facilities was not easily available researchers used the method of transformations to estimate the parameters of non-linear models which are intrinsically linear. For intrinsically non-linear models, graphical methods were used. The parameters were estimated in two stages. At the first stage, one parameter was estimated graphically and using this estimate, method of transformation was used to estimate the remaining parameters at the second stage. For the models which cannot be made linear in parameters by any transformation, some ingenious methods were also developed by researchers earlier for specific models. For the logistic growth model $y_t = \frac{k}{1+be^{-at}} + \varepsilon_t$, Nair (1954) discussed the application of various methods developed by Hotelling, Fisher, Yule, Rhodes and Hartley for estimating the parameters in detail. Later on, Cornell (1962) and Shah (1973) proposed methods based on partial sums for estimating the parameters of another such model given by $C(t) = Ae^{-\alpha t} + Be^{-\beta t} + \varepsilon_t$ where $C(t)$ is the concentration of a drug the blood stream at time t after its administration to humans or animals. Fresen and Juritz (1986) proposed a method of obtaining initial estimates of the parameters of the above model using the numerical integration technique. All these methods require that the observations are equally spaced. These methods are used in specific models but no method is available which could be used for a large class of non-linear regression models.

Nowadays, with the emergence of high speed computers and the availability of various optimization algorithms, use of computer software for minimizing $SSE(\theta)$ or $SSE_W(\theta)$ has become popular. With the least squares estimates obtained using software, researchers have realized that estimates obtained by transforming the model or by using graphical methods may be extremely poor. Even the ingenious methods mentioned above have been found to be far from satisfactory since they either underestimated or overestimated the parameters. In view of this, Wagner and Metzler (1967) asserted that such estimates could at best be used as initial seed values in optimization algorithms.

In the case of non-linear regression, the surface of objective function $SSE(\underset{\sim}{\theta})$ is quite often very rough and is full of spikes and troughs. This poses many problems. If initial seed values are poor then the algorithm may converge to a local minima or we may need a lot of iterations and computing time for convergence to the global minima. It is also possible that convergence may not occur at all (see Steyn and Van Wyk 1977). The residual sum of squares, $RSS = \sum_{t=1}^{n} [y_t - g(x_t, \underset{\sim}{\theta})]^2$, of fitted values for estimates provided by these methods may be very low, even then the estimates may be substantially different from the optimisation estimates, hence needing a large number of iterations for convergence to optimum values. For example, it has been noticed that the RSS of fitted values for estimates obtained by Cornell (1962) is quite low for the data in his paper but the estimates are quite different from the least squares estimates. In non-linear models, all or some parameters have some physical interpretation, and in such cases poor estimates may give misleading interpretations. In view of these considerations, it is desirable to develop a general methodology to find good initial estimates for optimization algorithms. Moreover, we also need to have several sets of good initial estimates to ensure a global minimum of $SSE(\theta)$.

In this paper, we have proposed various methods for obtaining efficient initial estimates. In section 2, we have developed several methods based on finite differences for the models which could be written in the asymptotic regression form (see, Stevens 1951) by performing suitable transformations. In Section 3, we have discussed the applications of the proposed methods using four sets of published data.

2. ESTIMATION BASED ON FINITE DIFFERENCES

We shall first discuss the procedure to obtain initial estimates of the parameters of the models which belong to the asymptotic regression category as discussed in Steven (1951). In an asymptotic regression model, the stimulus-response relationship could be written as,

$$y_t = \alpha + \beta\rho^{x_t} + \varepsilon_t \qquad (2.1)$$

where $0 < \rho < 1$. Note that $y_t \to \alpha$ as $x_t \to \infty$. Many important non-linear models could be put in this form. Some of the important examples are as follows.

2.1 Mitscherlish Model

This model describes the relationship between fertilizer dose levels and crop yields as follows,

$$y_t = A\{1-10^{-c(x_t+b)}\} + \varepsilon_t \qquad (2.2)$$

where c determines the efficiency of the fertilizer, b indicates soil content in the fertilizer and A is the maximum attainable yield. This model could be put in the form (2.1) by taking $\alpha = A$, $\beta = -A10^{-bc}$ and $\rho = 10^{-c}$.

2.2 Logistic Growth Model

This model is given by

$$y_t = \frac{k}{1+be^{-ax_t}} + \varepsilon_t \qquad (2.3)$$

The variable transformation $z_t = \frac{1}{y_t}$ and parameter transformations $\beta = \frac{b}{k}$, $\rho = e^{-a}$ transform the model (2.3) in asymptotic regression form.

2.3 Gompertz Model

This growth model is given by,

$$y_t = a\exp\{-be^{-cx_t}\} + \varepsilon_t \qquad (2.4)$$

The transformations $z_t = \log_e(y_t)$, $\alpha = \log_e(a)$, $\beta = -b$ and $\rho = e^{-c}$ enable us to write (2.4) in the asymptotic form.

2.4 One Compartment Drug Concentration Model

This model is given by

$$C(t) = \alpha + ve^{-\lambda t} + \varepsilon_t \qquad (2.5)$$

where, $C(t)$ is the concentration of drug in blood stream of a human subject or animal at time t. Here $v = \beta$ and $\rho = e^{-\lambda}$ we can transform (2.5) to the form (2.1).

2.5 Proposed Methods

There could be two approaches to estimate the parameters of model (2.1). In the first, we estimate α graphically. We first plot y against x and take the asymptotic value of y as $\hat{\alpha}$. Then, we use the logarithmic transformation for $y - \hat{\alpha} = \beta\rho^x$ and estimate β and ρ by the method of least squares. In the second approach we first estimate ρ and then estimate α and β by regressing y on $(\hat{\rho})^x$. If $\hat{\rho}$ falls outside the interval (0; 1), then it is a warning against an attempt to fit an asymptotic regression function. In this paper, we have used the second approach to develop the estimates. The reason for our choice is guided by the findings of Stevens (1951).

Stevens proposed an efficient method for estimating α, β and ρ using the least squares method. His method is essentially Fisher's general iterative procedure which needs initial estimates of the parameters and the estimated information matrix. Stevens showed that the information matrix which is found by differentiating the normal equations of model (2.1) with respect to α, β and ρ and replacing the parameters by their estimates and y by its expected value, turns out to be a function of $\hat{\rho} = r$ only. Therefore, Stevens' method requires a

reasonably accurate initial estimate of ρ. If the precise estimate of ρ is obtained at the first stage, better estimates of α and β would be obtained at the second stage. Stevens provided tables for calculating the elements of information matrix. In many experiments, most useful levels *x* are rather low, so he provided tables only for *n* = 5; 6 and 7. Gomes (1953) enlarged the set of tables by providing a table for *n* = 4 also. Realizing the necessity of an efficient initial estimate of ρ for the iterative procedure, Patterson (1958) proposed quadratic estimates of ρ for *n* = 4; 5; 6; and 7 which are, in fact, ratios of two contrasts in y_x's. Since the methods provided by Stevens and Patterson are complicated and are available only for small values of *n*, we have developed simple methods for which there is no limitation on *n*.

2.1.1 Method 1

For equally spaced observations the model (2.1) can be written by a simple change of origin and scale of the regressor as

$$y_x = \alpha + \beta\rho^x + \varepsilon_t \tag{2.6}$$

where $x = 1,\ldots,n$. Let y_x be a function of an equally spaced design variable *x*. Then the first finite difference is defined as $\Delta y_x = y_{x+1} - y_x$, $x = 1, 2,.., n\text{-}1$, where Δ is called the forward difference operator. Δy_x represents the increment. For example, if we are interested in studying the growth of a plant, and y_x and y_{x+1} are its heights at two equally spaced time points, then Δy_x represents the increase in height which occurred during the period $(x, x + 1)$. Ignoring ε_x, we have

$$\Delta y_x = y_{x+1} - y_x = \beta(\rho - 1)\rho^x \tag{2.7}$$

Similarly,

$$\Delta y_{x+1} = \beta(\rho - 1)\rho^{x+1} \tag{2.8}$$

Moreover,

$$\Delta^2 y_x = \Delta y_{x+1} - \Delta y_x = y_{x+2} - 2y_{x+1} + y_x = \beta\rho^x(\rho - 1)^2 \tag{2.9}$$

From (2.7), we obtain

$$\sum_{x=1}^{n-2} \Delta y_x = \beta(\rho - 1)\sum_{x=1}^{n-2} \rho^x \tag{2.10}$$

and from (2.9), we have

$$\sum_{x=1}^{n-2} \Delta^2 y_x = \beta(\rho - 1)^2 \sum_{x=1}^{n-2} \rho^x \tag{2.11}$$

Therefore, (2.10) and (2.11) yield,

$$\hat{\rho}_1 = 1 + \frac{\sum_{x=1}^{n-2} \Delta^2 y_x}{\sum_{x=1}^{n-2} \Delta y_x} \tag{2.12}$$

2.1.2 Method 2

Let the backward difference operator ∇ be defined as

$$\nabla y_x = y_{x-1} - y_x \tag{2.13}$$

Ignoring the error term in (2.6), we have

$$\nabla y_x = y_{x-1} - y_x = \beta\rho^x\left(\frac{1}{\rho} - 1\right)$$

Similarly,

$$\nabla^2 y_x = y_{x-2} - 2y_{x-1} + y_x = \beta\rho^x\left(\frac{1}{\rho} - 1\right)^2$$

Now, we can easily obtain

$$\sum_{x=3}^{n} \nabla y_x = \beta\left(\frac{1}{\rho} - 1\right)\sum_{x=3}^{n} \rho^x \tag{2.14}$$

and

$$\sum_{x=3}^{n} \nabla^2 y_x = \beta\left(\frac{1}{\rho} - 1\right)^2 \sum_{x=3}^{n} \rho^x \tag{2.15}$$

From (2.14) ad (2.15), we obtain

$$\hat{\rho}_2 = \left(1 + \frac{\sum_{x=3}^{n} \nabla^2 y_x}{\sum_{x=3}^{n} \nabla y_x}\right)^{-1} \quad (2.16)$$

Now, we take the average of $\hat{\rho}_2$ and $\hat{\rho}_1$ to get the pooled estimate

$$\hat{\rho}_p = \frac{\hat{\rho}_1 + \hat{\rho}_2}{2}$$

2.1.3 Method 3

Squaring (2.7) on both sides and summing over *x*, we have

$$\sum_{x=1}^{n-2} (\Delta y_x)^2 = \beta^2 (\rho - 1)^2 \sum_{x=1}^{n-2} \rho^{2x} \quad (2.17)$$

Similarly, from (2.9) we obtain

$$\sum_{x=1}^{n-2} (\Delta^2 y_x)^2 = \beta^2 (\rho - 1)^4 \sum_{x=1}^{n-2} \rho^{2x} \quad (2.18)$$

From (2.17) and (2.18), we get

$$c = \frac{\sum_{x=1}^{n-2} (\nabla^2 y_x)^2}{\sum_{x=1}^{n-2} (\nabla y_x)^2} = (\rho - 1)^2$$

or $\rho = 1 \pm \sqrt{c}$. But $0 < \rho < 1$, therefore

$$\hat{\rho} = 1 - \sqrt{c} \quad (2.19)$$

2.1.4 Method 4: Method of Partial Sums

This method is useful when we have a large number of observations. We divide the observations into three equal parts. Let $n = 3r$. If the need arises, we drop one or two observations for the purpose of obtaining initial estimates, but all observations are utilized at the optimization stage.

Let $s_1 = \sum_{i=1}^{r} y_i$, $s_2 = \sum_{i=r+1}^{2r} y_i$ and $s_3 = \sum_{i=2r+1}^{3r} y_i$

We can assume x_i's to take values 1,2,…,*n* as they are equally spaced and we can use a suitable transformation of origin and scale. We can easily check that

$$s_1 = \sum_{i=1}^{r} y_i = \sum_{i=1}^{r} (\alpha + \beta \rho^i) = r\alpha + \beta \sum_{i=1}^{r} \rho^i$$

Similarly,

$$s_2 = \sum_{i=1}^{r} y_i = r\alpha + \beta \rho^r \sum_{i=1}^{r} \rho^i$$

and

$$s_3 = r\alpha + \beta \rho^{2r}) \sum_{i=1}^{r} \rho^i$$

Now,

$$s_1 - s_2 = \beta \sum_{i=1}^{r} \rho^i - \beta \rho^r \sum_{i=1}^{r} \rho^i = \beta (1 - \rho^r) \sum_{i=1}^{r} \rho^i \quad (2.20)$$

Similarly,

$$s_2 - s_3 = \beta \rho^r (1 - \rho^r) \sum_{i=1}^{r} \rho^i \quad (2.21)$$

Now, from (2.20) and (2.21), we have

$$\frac{s_2 - s_3}{s_1 - s_2} = \rho^r$$

or $r \log \rho = \log_e \left(\frac{s_2 - s_3}{s_1 - s_2}\right)$

Thus, we get

$$\hat{\rho} = anti \log \left(\frac{1}{r} \log_e \frac{s_2 - s_3}{s_1 - s_2}\right) \quad (2.22)$$

2.1.5 Method 5

We can easily see that $\hat{\rho}_x = \frac{\Delta y_{x+1}}{\Delta y_x}$, x = 1,2,…,*n*-2. Thus we will have *n* - 2 estimates of ρ.

We take the geometric mean of these estimates as the final estimate of ρ. That is,

$$\hat{\rho} = \left(\prod_{x=1}^{n-2} \rho_x\right)^{\frac{1}{n-2}} \quad (2.23)$$

2.1.6 Method 6

From (2.7) and (2.8), we have

$$\sum_{i=1}^{n-2}(\Delta y_x) = \beta(\rho-1)\sum_{x=1}^{n-2}\rho^x$$

$$\sum_{i=1}^{n-2}(\Delta y_{x+1}) = \beta(\rho-1)\rho\sum_{x=1}^{n-2}\rho^x$$

From these equations we obtain

$$\hat{\rho} = \frac{\sum_{i=1}^{n-2}\Delta y_{x+1}}{\sum_{i=1}^{n-2}\Delta y_x} \qquad (2.24)$$

3. APPLICATIONS AND DISCUSSION

Before illustrating the applications of the proposed methods it is worthwhile to highlight some features of non-linear regression. In this case, the ideal way to estimate the parameters by the least squares method is to minimize the objective function $SSE(\theta)$ or $SSE_w(\theta)$ directly by applying non-linear optimization algorithms. These algorithms are iterative in nature and require good initial estimates of the parameters as seed values for convergence to the global minimum of the objective function. Moreover, to ensure a global minimum, we should also try several sets of initial estimates and see that the convergence occurs at the same values. These values are, in fact, the least squares estimates of the parameters.

Moreover, the exact distributional properties of the least squares estimates in the case of non-linear regression are not available in general. The estimates are biased. However, under certain regularity conditions, the estimates have been shown to be consistent with the finite asymptotic variances. For details one may refer to Chapter 5 of Seber and Wild (1989). The calculations for the estimates of asymptotic standard errors of the estimates are explained well in Ratkowsky (1983, pp. 15-17).

Another important aspect is concerned with the transformation of the model in order to facilitate the estimation of the parameters using transformed data. This practice has been in use for a long time. But as discussed earlier, estimates thus obtained are not the true least squares estimates but could be used as good initial estimates in the optimisation algorithms. The effect of the transformations on the parameter estimates for some commonly used non-linear regression models is explained in detail by Ratkowsky (1983, pp. 63-65). The findings assert that such estimates obtained using transformations provide good initial estimates.

Lastly, in order to compare the performance of various methods of finding initial estimates and to compare the performance of individual methods with that of a non-linear optimisation method, we could use several criteria. In addition to RSS, mean squared error given by $MSE = \frac{RSS}{n}$, number of iterations and computing time, one could also use the mean absolute percent error (MAPE), which is given by

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{\left| y_t - \hat{y}_t \right|}{y_t}$$

where $\hat{y}_t$ is the predicted value of y_t and n is the number of pairs in the data set. This measure indicates the predictive power of the method. After the convergence, we shall get final least squares estimates. The estimates of their standard errors can be calculated as mentioned in Ratkowsky (1983). For choosing the method of finding the initial estimates, the above criteria are good enough and standard errors of individual initial estimates have little practical significance.

To illustrate the application of the methods discussed in section 2, we have considered the data sets given in Stevens (1951) and Gomes (1953).

3.1 Data Set 1: Stevens (1951)

A thermometer, lowered into a refrigerated hold, gave the following six consecutive readings (°F) at half minute intervals. The data is displayed in the following table.

Time	0	1	2	3	4	5
Temperature Readings (°F)	57.7	45.7	38.7	35.3	33.1	32.2

3.2 Data set 2: Gomes (1953)

The mean yield of potatoes per plot of $\frac{1}{65}th$ of an acre in an experiment with 5 levels (0, 40, 80, 120 and 160 pounds per acre) of super phosphate is given in the following table.

Fertilizer Level (x)	0	1	2	3	4
Yield (y)	229.1	231.8	254.2	250.6	249.6

For calculation of $\hat{\rho}$ using the proposed methods, we have written codes in R 3.2.2. and used the 'nls' package in R for obtaining the least squares estimates. Codes for calculating the predicted values, RSS, MSE and MAPE have also been written in R.

The initial estimates of parameters, RSS, MSE, MAPE, number of iterations needed for convergence to the values and computing time for Stevens' data and Gomes' data are given in Tables 1 and 2 respectively. The estimates obtained by the proposed methods are quite close to optimum least squares estimates and are simple to calculate. Moreover, in the proposed methods, there is no restriction on n.

Table 1. Parameter Estimates, RSS, MSE, MAPE (%), Numbers of Iterations (I) and Computing Time in Seconds (T) for the Different Methods for Stevens' Data

Method	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\rho}$	RSS	MSE	MAPE (%)	I	T
1	30.6891	26.8432	0.5533	0.0982	0.0164	0.3009	4	13
2	30.6891	26.8432	0.5533	0.0982	0.0164	0.3009	4	13
3	30.4552	26.1993	0.5627	1.0727	0.1788	0.6809	4	14
4	30.8665	26.7305	0.5458	0.1141	0.0190	0.3040	4	14
5	30.6517	26.5507	0.5534	0.2743	0.0457	0.4308	5	18
6	30.6891	26.8432	0.5534	0.0984	0.0164	0.3003	4	13
Patterson	30.6626	26.8601	0.5544	0.1003	0.0167	0.3004	5	15
Stevens	30.723	26.8210	0.5519	0.0973	0.0162	0.3013	3	8
NLS	30.7239 (0.2310)	26.8211 (0.2577)	0.5518 (0.0085)	0.0973	0.0161	0.3011	-	-

Figures within parentheses represent the standard errors of the estimates.

Table 2. Parameter Estimates, RSS, MSE, MAPE (%), Numbers of Iterations (I) and Computing Time in Seconds (T) for the Different Methods for Gomes' Data

Method	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\rho}$	RSS	MSE	MAPE (%)	I	T
1	276.4140	46.969	0.8279	155.7696	31.5390	1.7569	26	29
2	276.4140	46.969	0.8279	155.7696	31.5390	1.7569	26	29
3	252.0330	25.505	0.4409	140.8875	28.1775	1.7384	17	27
4	250.5219	24.1245	0.3571	156.7442	31.3488	1.7758	18	29
5	269.7175	40.6684	0.7870	149.3772	29.8754	1.6936	23	30
6	276.4141	46.9694	0.8279	155.7696	31.1539	1.7569	26	29
Gomes	255.5510	28.3230	0.5750	131.7861	26.3572	1.7014	11	68
NLS	255.5306 (17.2375)	28.3072 (17.0840)	0.5744 (0.4789)	131.7859	26.3571	1.7014	-	-

Figures within parentheses represent the standard errors of the estimates.

The estimates of Stevens and Gomes are closest to the optimum estimates because they have been obtained by iterative procedures aiming at the optimum values. But they involve tedious computing using tables and their use is limited to small sample sizes of n = 4; 5; 6 or 7. We still need some iterations to attain optimum estimates if their estimates are used as initial estimates in algorithms. The difference in number of iterations when compared with proposed estimates is not much. The main advantage of the methods proposed in this paper is that they are very simple to calculate. Moreover, they can also be used for large sample sizes and there is no need for any tabulated values. To demonstrate the application of the method for a moderately bigger sample sizes, we have considered the following data set reported by Ratkowasky (1983).

Growing time (x)	1	2	3	4	5	6	7
Weight of onion bulbs plus dry tops (y)	16.08	33.83	65.80	97.20	191.55	326.20	386.87

x	8	9	10	11	12	13	14	15
y	520.53	590.03	651.92	724.93	699.56	689.96	637.56	717.41

To fit logistic model of the form

$$y_t = \frac{\alpha}{(1+\exp(\beta - vx_t)} \qquad (3.1)$$

to Ratkowasky's data, we first use the transformations $z_t = \frac{1}{y_t}$, $a = \frac{1}{\alpha}$, $b = \frac{e^{\beta}}{\alpha}$ and $\rho = e^{-\nu}$ in order to put (3.1) in the form (2.1). Therefore,

$$z_t = a + b\rho^t + e_t,\ t = 1,\ldots,n. \qquad (3.2)$$

We can write the model in this form because x_t's are equally spaced. We applied the method of partial sums, that is, method 4 discussed earlier to obtain the estimate of ρ and then, regressing z_t on $\hat{\rho}^t$, we obtain estimates of the other parameters of (3.2). The estimates are obtained as $\hat{\alpha}$ = 0.0013549, $\hat{b}$ = 0.1189280 and $\hat{\rho}$ = 0.503567. From these estimates, we obtained estimates of parameters of (3.1) as $\hat{\alpha}$ = 738.0248, $\hat{\beta}$ = 4.7475 and $\hat{\nu}$ = 0.68803. The residual sum of squares of the fitted values is 17115.16. Using these estimates are initial estimates in a non-linear optimization algorithm, we obtain the optimum estimates in 7 iterations as $\hat{\alpha}$ = 702.871 (standard error = 13.9397), $\hat{\beta}$ = 4.4426 (standard error = 0.3508) and $\hat{\nu}$ = 0.6886 (standard error = 0.0574). The residual sum of squares of the fitted values is 8930.747. It is apparent that estimates obtained using the proposed method of partial sums are quite close to the optimum least squares estimates.

Table 3 gives the observed values and fitted values for method 4 and non-linear algorithm estimates together with their RSS, MSE and MAPE values. The values predicted by method 4 are quite close to observed values in the beginning but slightly over predicting towards the end for a few points. Fig. 1 gives the curves for the observed and fitted values. Thus, the performance of the proposed method appears to be quite satisfactory for obtaining the initial estimates. Looking at the performance of the various methods in different data sets we conclude that methods 1 to 4 could be recommended for obtaining good initial estimates.

Table 3. Fitted values, RSS, MSE and MAPE (%) for method 4 and non-linear algorithm estimates for ratkowsky's data

Growing Time	Observed Values	Method 4	Non-linear Algorithm
1	16.08	16.36	16.088
2	33.83	31.8541	31.3317
3	65.8	60.7863	59.7623
4	97.2	111.8349	109.774
5	191.55	193.5051	189.9181
6	326.2	305.7005	297.5832
7	386.87	431.4088	417.4393
8	520.53	543.7892	523.2509
9	590.03	625.7042	599.5534
10	651.92	676.9529	646.9175
11	724.93	706.0146	673.6363
12	699.56	721.5828	687.9011
13	689.96	729.669	695.2934
14	637.56	733.8017	699.0651
15	717.41	735.8964	700.9743
RSS	-	17115.16	8929.883
MAPE (%)	-	6.1713	4.7852
MSE	-	1141.011	595.3255

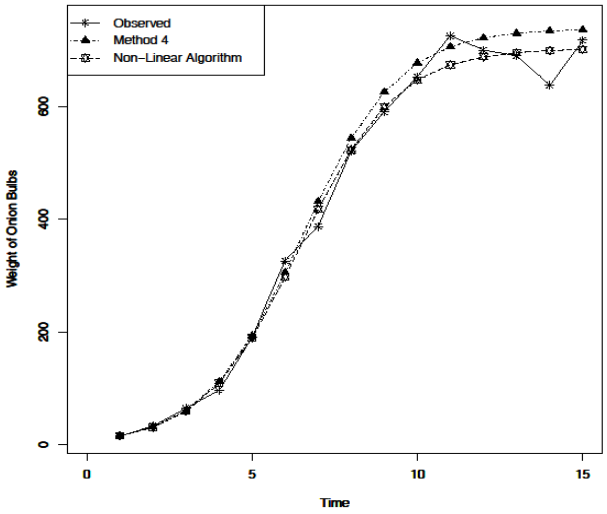

Fig. 1. Plot of observed and fitted values using Method 4 and non-linear algorithm estimates for ratkowasky's data

ACKNOWLEDGEMENTS

The authors are thankful to Prof. Ram Karan Singh, University of Lucknow, Lucknow, India, for suggesting the problem and giving useful comments during the preparation of the paper. The authors would also like to thank the referee for his/her constructive suggestions, which have improved the presentation of the paper.

REFERENCES

Chatterjee, S. (1991). *Regression Analysis by Examples*. John Wiley & Sons, New York.

Cornell, R.G. (1962). A method of fitting a linear combination of two exponentials. *Biometrics*, **18**, 104-113.

Draper, N.R. and Smith, H. (1981). *Applied Regression Analysis*. John Wiley & Sons, New York.

Fisher, R.A. (1950). *Statistical Methods for Research Workers*. 11th edition, Oliver and Boyd, London.

Fresen, J.L. and Juritz, J.M. (1986). A note on Foss's method of obtaining initial estimates for exponential curve fitting by numerical integration. *Biometrics*, **42,** 621-627.

Gomes, F.P. (1953). The use of Mitscherlich's regression law in the analysis of experiments with fertilizers. *Biometrics*, **9,** 498-516.

Nair, K.R. (1954). The fitting of growth curves. *In: Statistics and Mathematics in Biology*. Edited by Kempthorne, Lush and Bancroft.

Patterson, H.D. (1956). The use of autoregression in fitting exponential curves. *Biometrika*, **45,** 389-400.

Ratkowsky, D.A. (1983). *Nonlinear Regression Modeling-A Unified Practical Approach*. Marcel Dekker Inc., New York.

Rhodes, E.C. (1940). Population Mathematics, *JRSS*, **103,** 61-89.

Seber, G.A.F. and Wild, C.J. (1989). *Non-linear Regression*. John Wiley & Sons, New York.

Shah, B.K. (1973). obtaining preliminary estimates to the two term exponential model to blood concentration data. *J. Pharma. Sci*., **62,** 1208-1209.

Steyn, H.S. and Van Wyk, J.W.J, (1977). Some Methods for Fitting Compartment Models to Data, Weteukskaplike by trees. Vandie, PU VIY CHO Reeks B. Natuur Wetenskappe. Nr 91, Potchefstroom, South Africa.

Stevens, W.L. (1951). Asymptotic Regression. *Biometrics*, **7,** 247-267.

Wagner, J.G. and Metzler, C.M. (1967). Estimation of rate constants for absorption and elimination from blood concentration data. *J. Pharma. Sci.*, **56,** 658-659.

West, P.W. (1980). Use of diameter increment in basal area increment in tree growth studies. *Can. J. Forest. Res.*, **10**, 71-77.

APPENDIX

R Codes

Method-1

```
x<-c(0,1,2,3,4,5)
y<-c(57.5,45.7,38.7,35.3,33.1,32.2)
n<-length(x)-2
d1<-0;d2<-0
for(i in 1:n){d1=d1+(y[i+1]-y[i])
d2=d2+(y[i+2]-2*y[i+1]+y[i])}
ro1<-1+d2/d1
xt<-ro1^x
lm.result<-lm(y~xt)
summary(lm.result)
```

Codes for Non Linear Optimization Algorithm

```
x<-c(0,1,2,3,4,5)
y<-c(57.5,45.7,38.7,35.3,33.1,32.2)
a<-30.6891;b<-26.8432;r<-0.55328
nls.result<-
nls(y~Alpha+Beta*Ro^x,start=list(Alpha=a,Beta=b
,Ro=r),algorithm="port",trace=TRUE,control=nls.c
ontrol(maxiter = 100, tol = 1e-05, minFactor =
1/1024,
printEval = FALSE, warnOnly = FALSE))
summary(nls.result)
```

Note: For logistic model withinnls we will have to write

```
y~Alpha/(1+exp(Beta-Ro*x))
```

R codes for Method-2

```
x<-c(0,1,2,3,4,5)
y<-c(57.5,45.7,38.7,35.3,33.1,32.2)
n<-length(x)-2
d1<-0;d2<-0
for(i in 1:n){d1=d1+(y[i+1]-y[i])
d2=d2+(y[i+2]-2*y[i+1]+y[i])}
ro1<-1+d2/d1
m<-length(x)
D1<-0;D2<-0
for(i in 3:m){D1=D1+(y[i-1]-y[i])
D2=D2+(y[i-2]-2*y[i-1]+y[i])}
ro2<- 1/(1+D2/D1)
rop<-(ro1+ro2)/2
xt<-rop^x
lm.result<-lm(y~xt)
summary(lm.result)
```

R Codes for Method-3

```
x<-c(0,1,2,3,4,5)
y<-c(57.5,45.7,38.7,35.3,33.1,32.2)
n<-length(x)-2
d1<-0;d2<-0
for(i in 1:n){d1=d1+(y[i+1]-y[i])^2
d2=d2+(y[i+2]-2*y[i+1]+y[i])^2}
c<-d2/d1
ro<-1-sqrt(c)
xt<-ro^x
lm.result<-lm(y~xt)
summary(lm.result)
```

R Codes for Method-4

```
r<-5
x11<-c(16.08,33.83,65.8,97.2,191.55)
x22<-c(326.2,386.87,520.53,590.03,651.92)
x33<-c(724.93,699.56,689.96,637.56,717.41)
x1<-1/x11;x2<-1/x22;x3<-1/x33
x<-c(1,2,3,4,5,6,7,8,9,10,11,12,13,14,15)
y<-c(x1,x2,x3)
s1<-sum(x1);s2<-sum(x2);s3<-sum(x3)
y1<-s1-s2;y2<-s2-s3
z<-y2/y1
w<-log(z)
u<-w/r
ro<-exp(u)
xt<-ro^x
lm.result<-lm(y~xt)
summary(lm.result)
```

R Codes for calculating predicted values, RSS, MSE and MAPE

```
x<-c(0,1,2,3,4,5)
y<-c(57.5,45.7,38.7,35.3,33.1,32.2)
a0<-30.6891;b0<-26.8432;r0<-0.5533
yp<-function(x,a,b,r){
z<-a+b*r^x
return(z)
}
ypred<-yp(x,a0,b0,r0)
w1<-y-ypred
w2<-w1^2
RSS<-sum(w2)
MSE<-RSS/length(x)
w<-abs(w1)/y
u<-sum(w)/length(x)
MAPE<-u*100
```

Note: For logistic model within function we will have to write

```
z<-a/(1+exp(b-r*x))
```