Journal of the Arkansas Academy of Science

Volume 21 Article 18

1967

Monte Carlo Study of Estimation for Nonlinear Models Subjected to Linearizing Transformations

Jeff Kennington
University of Arkansas, Fayetteville

James E. Dunn
University of Arkansas, Fayetteville

Arkansas Academy of Science Proceedings, Vol. 21, 1967

MONTE CARLO STUDY OF ESTIMATION EFFICIENCY FOR NONLINEAR MODELS SUBJECTED TO LINEARIZING TRANSFORMATIONS

Jeff Kennington and James E. Dunn

University of Arkansas

INTRODUCTION

The purpose of this study was to make a quantitative evaluation of several estimation procedures commonly applied to nonlinear models. Though the set of methods considered here is by no means complete, for example we have omitted Bayesian procedures, we feel that we have included those which are most commonly used. The general approach was to start with a model involving known parameters, generate artificial data by adding random errors to the expected values given by the model, and proceed to estimate the parameters by each of the procedures. By using the inherent speed of a modern digital computer for all computations, we were able to generate many sets of data based on the same model and thus were able to study quantitatively the average behavior of the various estimation procedures.

ESTIMATION PROCEDURES

In what follows, we shall suppose that statistically independent responses $y_1, y_2, \cdots, y_n$ are measured at specified times $t_1, t_2, \cdots, t_n$ respectively. In order to have a specific model to work with, suppose it is also known that $E(y_i) = t_i^A \exp(-Bt_i) = f_i$ where E denotes the expected value operator. How shall we estimate the unknown parameters A and B? In the following development, assume that all summations are over all n observations.

Nonlinear Least Squares (Modified Gauss-Newton)

We may argue that the i'th observation y_i may be expressed as $y_i = f_i + e_i$ where e_i is a random error inherent in making the observation with $E(e_i) = 0$ and var $(e_i) = \sigma_i^2$. Formally, the least squares estimates of A and B are those values, say $\hat{A}$ and $\hat{B}$, at which

$$Q(A,B) = \sum e_i^2 = \sum (y_i - f_i)^2 \qquad (1)$$

attains its minimum. Finding the location of this minimum, of course, poses the major computational problem.

Typically we will argue that these estimates must solve

$$q_1 = \frac{\partial Q}{\partial A} = \sum (y_i - f_i) \frac{\partial f_i}{\partial A} = 0$$

$$q_2 = \frac{\partial Q}{\partial B} = \sum (y_i - f_i) \frac{\partial f_i}{\partial B} = 0 \qquad , \qquad (2)$$

hence, we proceed to their values by an iteration of the form

$$\begin{bmatrix} A_{k+1} \\ B_{k+1} \end{bmatrix} = \begin{bmatrix} A_k \\ B_k \end{bmatrix} - \nu H^{-1} \begin{bmatrix} q_1 \\ q_2 \end{bmatrix} \Bigg|_{\substack{A = A_k \\ B = B_k}} \tag{3}$$

where the elements of the matrix H are

$$h_{11} = \frac{\partial q_1}{\partial A} = \sum(y_i - f_i)\frac{\partial^2 f_i}{\partial A^2} - \sum\left(\frac{\partial f_i}{\partial A}\right)^2$$

$$h_{12} = h_{21} = \frac{\partial q_1}{\partial B} = \sum(y_i - f_i)\frac{\partial^2 f_i}{\partial B \partial A} - \sum\frac{\partial f_i}{\partial A}\frac{\partial f_i}{\partial B}$$

$$h_{22} = \frac{\partial q_2}{\partial B} = \sum(y_i - f_i)\frac{\partial^2 f_i}{\partial B^2} - \sum\left(\frac{\partial f_i}{\partial B}\right)^2 \tag{4}$$

and all summations are over i = 1, 2, . . ., n. The well known Newton-Raphson iteration takes the form of (3) with $\nu = 1$ and H as shown. Hartley (1961) has recently suggested a modified Gauss-Newton procedure in which all terms of the form $\sum(y_i - f_i)\frac{\partial^2 f_i}{\partial \cdot \partial}$ are dropped from (4) and ν is chosen to minimize Q(A,B) in the interval $0 \leq \nu \leq 1$. We have used this latter procedure in the work that follows.

Linearized — Unweighted Least Squares

Many references would suggest linearizing the model by a logarithmic transformation to obtain $E(\ln y_i) = A \ln t_i - Bt_i$ approximately. Assuming an additive error e_i^* with this model, then estimates A*, B* are obtained by minimizing $Q^*(A,B) = \sum(e_i^*)^2 = \sum(\ln y_i - A \ln t_i + Bt_i)^2$, i.e. solve the normal equations

$$A^*\sum \ln^2 t_i - B^*\sum t_i \ln t_i = \sum(\ln t_i)(\ln y_i)$$

$$-A^*\sum t_i \ln t_i + B^*\sum t_i^2 = -\sum t_i \ln y_i \tag{5}$$

Solving a set of linear equations of this type is an elementary exercise and thus leads to the popularity of this approach.

Linearized — Weighted Least Squares

Is there an intermediate procedure? The Gauss-Markov theorem suggests that for the linearized model $\ln y_i = A \ln t_i - Bt_i + e_i^*$, the best linear estimates of A and B, say A^+ and B^+, are obtained by minimizing $Q^+(A,B) = \sum w_i(e_i^*)^2 = \sum w_i(\ln y_i - A \ln t_i + Bt_i)^2$, i.e. solve

$$A^+\sum w_i \ln^2 t_i - B^+\sum w_i t_i \ln t_i = \sum w_i (\ln t_i)(\ln y_i)$$

$$-A^+\sum w_i t_i \ln t_i + B^+\sum w_i t_i^2 = -\sum w_i t_i \ln y_i \tag{6}$$

where the weights w_i are proportional to $[\mathrm{var}(\ln y_i)]^{-1}$. Since $\ln y_i$ is a nonlinear transformation, we can only approximate its variance. If $E(y) = C$ and we subject y to the transformation $z = T(y)$, then a commonly used approximation is $\mathrm{var}(z) = \mathrm{var}[T(y)] = \mathrm{var}(y)[T'(C)]^2$, cf. Rao (1952). In our case, $E(y_i) = f_i$, $\mathrm{var}(y_i) = \sigma_i^2$, and $T(y_i) = \ln y_i$ so that $\mathrm{var}(\ln y_i) = \sigma_i^2/f_i^2$ approximately. It follows that

$$w_i = [\mathrm{var}(\ln y_i)]^{-1} = f_i^2/\sigma_i^2 = [t_i^A \exp(-Bt_i)]^2/\sigma_i^2 . \tag{7}$$

The obvious problem now in trying to solve (6) is that the weights given in (7) involve the unknowns A and B. Hence, we will need to iterate to the solution of (6), i.e. set $v_i = 1$ or $v_i = y_i^2$ and solve (6), re-evaluate the v_i using these solutions and resolve (6), etc. until convergence.

ERROR DISTRIBUTIONS

Even though the above estimation procedures are distribution free in the sense that they may be used whatever the functional form of the error distribution, it is not evident what effect varying error distributions will have on the properties of the estimates of A and B. To examine this, we generated random errors from the following three probability distributions, each properly scaled so that

$$E(e) = 0 \text{ and } var(e) = 1 \qquad (8)$$

In each case, we assumed the existence of a sequence $U_1, U_2, \ldots$ of independent random numbers arising from a rectangular distribution on the (0, 1) interval. The power residue method, cf. I.B.M. (1959), which we used, provides an easy means for generating such a sequence on the digital computer.

Rectangular

If $U_1, U_2, \ldots$ is as defined above and $e_j = 2\sqrt{3}\,(U_j - 1/2)$, then $e_1, e_2, \ldots$ is a sequence of independent rectangular random numbers satisfying conditions (8) and with density function $g(e) = (2\sqrt{3})^{-1}$ on the interval $(-\sqrt{3}, \sqrt{3})$.

Normal

If $U_1, U_2, \ldots$ is as defined above and we set

$$e_j = (-2 \ln U_j)^{1/2} \cos 2\pi U_{j+1}$$

$$e_{j+1} = (-2 \ln U_j)^{1/2} \sin 2\pi U_{j+1},$$

Box and Muller (1958) show that $e_1, e_2, \ldots$ is a sequence of independent, normally distributed random numbers satisfying conditions (8) with density function $g(e) = (2\pi)^{-1/2} \exp(-e^2/2)$ on $(-\infty, \infty)$.

Pearson Type III

If $U_1, U_2, \ldots$ is defined as above and $e_k = -(10)^{-1/2} \sum_{i=1}^{10} \ln(1-U_{10(k-1)+i}) - (10)^{1/2}$, then $e_1, e_2, \ldots$ is a sequence of independent random numbers satisfying conditions (8) and having

a Pearson Type III density function (gamma) defined by

$$g(e) = \frac{[10^{1/2}\ e+10]^9\ \exp[-(10^{1/2}\ e+10)]}{10^{-1/2}\ 9!}$$

on the interval $(-10^{1/2}, \infty)$.

Notice that the rectangular distribution might correspond to a situation where extreme outliers have been brought in closer to the mean, i.e. the Wisorization procedure, while the Pearson Type III corresponds to a skewed distribution of errors.

Measures of Efficiency

In order to compare the methods of estimation, we shall need a quantitative measure of their behavior. An obvious approach is to compare their **bias** and **error mean squares.**

Suppose $\hat{t}$ is an estimator of some parameter t. Then if $E(\hat{t}) = t + d_t$, d_t is called the **bias** in $\hat{t}$. If $d_t = 0$, then $\hat{t}$ is said to be an unbiased estimator of t. In any case, $E(\hat{t})$ provides a mean value for the estimator.

The **error mean square** (EMS) for $\hat{t}$ is defined by

$$EMS(\hat{t}) = E[(\hat{t} - t)^2] = var(\hat{t}) + d_t^2 \quad .$$

If the bias $d_t = 0$, then the $EMS(\hat{t})$ and variance of $\hat{t}$ are identical. In any case, EMS provides a measure of dispersion of the estimator.

If we want to use mean value and EMS as a basis for comparing the estimation procedures then we need estimates of these two expressions for our various estimators of A and B. This is where the Monte Carlo approach enters in. Suppose from known values of A and B we generate m independent sets of data which give rise to m independent estimates $\hat{A}_1, \hat{A}_2, \ldots, \hat{A}_m$ of A and m independent estimates $\hat{B}_1, \hat{B}_2, \ldots \hat{B}_m$ of B. It may be shown that the following are consistent estimators:

$$\begin{aligned}
&\sum_{i=1}^{m} \hat{A}_i/m \quad \text{for} \quad A + d_A \\
&\sum_{i=1}^{m} \hat{B}_i/m \quad \text{for} \quad B + d_B \\
&\sum_{i=1}^{m} (\hat{A}_i - A)^2/m \quad \text{for} \quad EMS(\hat{A}) \\
&\sum_{i=1}^{m} (\hat{B}_i - B)^2/m \quad \text{for} \quad EMS(\hat{B})
\end{aligned} \tag{9}$$

These four functions, calculated for each of the estimation procedures, form the basis of the numerical comparisons which follow.

Numerical Results

Two forms of the model $y_i = t_i^A \exp(-Bt_i) + e_i$ were generated by varying A and B. Expected values of y for various settings of t are given in Table 1 for each form of the model. Model 1 corresponds to

near completion of a full growth-decay cycle. Model 2 is essentially pure growth with only slight decay in the last measurements.

Tables 2 and 3 give estimates of the mean values and EMS's calculated from equations (9) for each of the estimation procedures. In these results, m = 1000 corresponding to the number of data sets which were generated from each of the error distributions previously described. With respect to the Monte Carlo procedure, from an inspection of the sequence of these estimates, as $m \to 1000$, it appears that the figures given in tables are correct except possibly for some instability in the last digit shown.

Examination of tables 2 and 3 yields some interesting conclusions. All of the procedures yield essentially unbiased estimates for A and B regardless of the error distribution used or the model specified. We feel that the discrepancies which appear between $E(\hat{A})$ and A or $E(\hat{B})$ and B in some cases can be largely accounted for by the fact that we stopped with m = 1000 data sets rather than, say m = 100,000 or more. There was no indication as $m \to 1000$ that any of the sequences of estimates was converging to a parameter other than that specified by the model.

Striking differences do appear, however, in comparing the EMS among methods. In every case both the linearized-weighted and Gauss-Newton procedures yield estimates of both A and B with considerably smaller EMS than the corresponding estimates by the linearized-unweighted procedure — in some cases by more than an order of magnitude. Equally remarkable is the fact that the linearized-weighted procedure and the Gauss-Newton yielded estimates with essentially identical EMS's, i.e. the procedure of using a linearized model and empirically calculated weights apparently is quite a good competitor with the Gauss-Newton iteration, particularly when we consider the extra effort required by the latter procedure. Certainly our results lend evidence to the argument that one should not simply ignore the disturbance in variance structure which occurs when subjecting data to nonlinear transformations.

Extensions

Several extensions to these methods are immediately evident. Though we had some theoretical basis for using the weights defined by (7), Box (1957) has suggested using weights of the form $w_i = [E(y_i)]^r$ where r as well as the unknowns of the model are involved in the iteration. He had made no quantitative study of the behavior of the resulting estimates at the time of this writing.

Looking again at (7), we notice that $\sigma_i^2 = \text{var}(y_i)$ appears in the denominator of the w_i. In the foregoing work, we were not concerned with σ_i^2 since we constructed our error distributions such that $\sigma_i^2 = 1$ for all i. On the other hand with real data, one cannot always assume

that these variances are all constant and, in fact, it may be evident that they are not. What should be done? Though Box's (op. cit.) procedure may have merit here, we want to suggest an alternative. Following the steps of the linearized-weighted procedure, at each stage of the iteration, set $\sigma_i^2 = [y_i - t_i^{\hat{A}} \exp(-\hat{B}t_i)]^2$ in (7) where $\hat{A}$ and $\hat{B}$ are the current estimates of A and B. This should have the effect of gradually reducing the influence of outliers and/or aberrant data points and more heavily weighting those observations which are quite consistent with the model. Regrettably, we cannot give a quantitative evaluation of this suggestion at present.

REFERENCES

Box, G. E. P. 1967. Private Communication. University of Wisconsin. Madison, Wisconsin.

Box, G. E. P. and M. E. Muller. 1958. A note on the generation of random normal deviates. Annals of Math. Stat. 29:610-611.

Hartley, H. O. 1961. The modified Gauss-Newton method for fitting of nonlinear regression functions by least squares. Technometrics 3:269-280.

I.B.M. 1959. Reference Manual, Random Number Generation and Testing. I.B.M. Data Processing Division, White Plains, N. Y.

Rao, C. R. 1952. Advanced Statistical Methods in Biometric Research. John Wiley and Sons. New York. Page 209.

Table I. Expected values of observations for various t and under two parameter settings.

t	Model 1 A=4, B=O.5	Model 2 A=5, B=O.5
2	5.886	11.77
3	18.07	54.22
4	34.65	138.6
5	51.30	256.5
6	64.52	387.1
7	72.50	507.5
8	75.02	600.2
9	72.89	656.0
10	67.38	673.8
11	59.83	658.2
12	51.40	
13	42.94	
14	35.03	
15	28.00	
16	21.98	
17	17.00	

Table 2. Mean values and EMS for estimators of A=4 and B=0.5 (Model 1)

RECTANGULAR ERRORS

Method	$E(\hat{A})$	$EMS(\hat{A})$	$E(\hat{B})$	$EMS(\hat{B})$
Linearized-unweighted	3.998	7.16×10^{-4}	.4997	3.26×10^{-5}
Linearized-weighted	4.000	2.16×10^{-4}	.4999	1.13×10^{-5}
Modified Gauss-Newton	3.999	2.16×10^{-4}	.5000	1.13×10^{-5}

NORMAL ERRORS

Method	$E(\hat{A})$	$EMS(\hat{A})$	$E(\hat{B})$	$EMS(\hat{B})$
Linearized-unweighted	3.998	7.30×10^{-4}	.4995	3.34×10^{-5}
Linearized-weighted	4.000	2.08×10^{-4}	.4999	1.10×10^{-5}
Modified Gauss-Newton	3.999	2.07×10^{-4}	.4999	1.10×10^{-5}

PEARSON TYPE III ERRORS

Method	$E(\hat{A})$	$EMS(\hat{A})$	$E(\hat{B})$	$EMS(\hat{B})$
Linearized-unweighted	4.000	6.80×10^{-4}	.5000	3.09×10^{-5}
Linearized-weighted	4.000	2.05×10^{-4}	.5002	1.09×10^{-5}
Modified Gauss-Newton	4.000	2.07×10^{-4}	.5002	1.10×10^{-5}

Table 3. Mean values and EMS for estimators of A=5 and B=0.5 (Model 2)

RECTANGULAR ERRORS

Method	$E(\hat{A})$	$EMS(\hat{A})$	$E(\hat{B})$	$EMS(\hat{B})$
Linearized-unweighted	4.999	5.47×10^{-4}	.4998	6.13×10^{-5}
Linearized-weighted	5.000	1.14×10^{-5}	.5000	6.66×10^{-7}
Modified Gauss-Newton	5.000	1.14×10^{-5}	.5000	6.66×10^{-7}

NORMAL ERRORS

Method	$E(\hat{A})$	$EMS(\hat{A})$	$E(\hat{B})$	$EMS(\hat{B})$
Linearized-unweighted	4.999	5.67×10^{-4}	.4998	3.25×10^{-5}
Linearized-weighted	5.000	1.07×10^{-5}	.5000	6.22×10^{-7}
Modified Gauss-Newton	5.000	1.07×10^{-5}	.5000	6.21×10^{-7}

PEARSON TYPE III ERRORS

Method	$E(\hat{A})$	$EMS(\hat{A})$	$E(\hat{B})$	$EMS(\hat{B})$
Linearized-unweighted	4.999	5.25×10^{-4}	.4998	3.01×10^{-5}
Linearized-weighted	5.000	1.08×10^{-5}	.5000	6.63×10^{-7}
Modified Gauss-Newton	5.000	1.08×10^{-5}	.5000	6.26×10^{-7}