NONLINEAR REGRESSION METHODS

by

A. RONALD GALLANT

Institute of Statistics
Mimeograph Series No. 890
Raleigh - September 1973
NONLINEAR REGRESSION METHODS

by

A. Ronald Gallant

ABSTRACT

The Modified Gauss-Newton method for finding the least squares estimates of the parameters appearing in a nonlinear regression model is described. A description of software implementing the method which is available to TUCC users is included.
NONLINEAR REGRESSION METHODS

by

A. Ronald Gallant

A sequence of responses $y_t$ to inputs $x_t$ are assumed to be generated according to the nonlinear regression model

$$y_t = f(x_t, \theta) + e_t \quad (t = 1, 2, \ldots, n).$$

The input variables are $k$-dimensional

$$x_t = (x_{1t}, x_{2t}, \ldots, x_{kt})$$

and the unknown parameter $\theta$ is $p$-dimensional

$$\theta = (\theta_1, \theta_2, \ldots, \theta_p).$$

The least squares estimator is that value

$$\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_p)$$

which minimizes

$$\text{SSE}(\theta) = \sum_{t=1}^n [y_t - f(x_t, \theta)]^2.$$

The following vector and matrix notation will be useful in describing the modified Gauss-Newton method.
\[ \chi = (y_1', y_2', \ldots, y_n')' \quad (n \times 1) \]

\[ \chi(\theta) = (f(x_1, \theta), f(x_2, \theta), \ldots, f(x_n, \theta))' \quad (n \times 1) \]

\[ \chi f(x, \theta) = \text{the } p \times 1 \text{ vector whose } j^{\text{th}} \text{ element is } \frac{\partial}{\partial \theta_j} f(x, \theta) \]

\[ \chi(\theta) = \text{the } n \times p \text{ matrix whose } t^{\text{th}} \text{ row is } \chi' f(x_{t'}, \theta) \]

To illustrate the notation, consider the 50 responses \((=y_t)\) and inputs \((=x_t)\) shown in Exhibit I. We assume that:

\[ y_t = \theta_1 x_t + e_t \]

Thus,

\[ f(x, \theta) = \theta_2 x \]

\[ \theta = (\theta_1, \theta_2) \]

\[ x_t = x_t \]

so that\( p=2 \) and \( k=1 \). The vectors and matrices introduced above are, for this example,

\[ \chi = \begin{bmatrix} .824 \\ .515 \\ \vdots \\ .576 \end{bmatrix} \quad (50 \times 1) \]
The modified Gauss-Newton (Hartley, 1961) algorithm proceeds as follows.
Find a start value \( \theta_0 \); methods of finding start values are discussed later.

0) From the starting estimate \( \theta_0 \) compute

\[
    D'_0 = [F'(\theta_0) F(\theta_0)]^{-1} F'(\theta_0) [X - f(\theta_0)]
\]

Find a \( \lambda_0 \) between 0 and 1 such that

\[
SSE(\theta_0 + \lambda_0 D_0) \leq SSE(\theta_0)
\]

1) Let \( \theta_1 = \theta_0 + \lambda_0 D_0 \). Compute

\[
    D'_1 = [F'(\theta_1) F(\theta_1)]^{-1} F'(\theta_1) [X - f(\theta_1)]
\]

Find a \( \lambda_1 \) between 0 and 1 such that

\[
SSE(\theta_1 + \lambda_1 D_1) \leq SSE(\theta_1)
\]
2) Let \( \theta_2 = \theta_1 + \lambda_1 d_1 \)

These iterations are continued until terminated according to some stopping rule; stopping rules are discussed later.

Hartley's (1961) paper sets forth assumptions such that these iterations will converge to \( \hat{\theta} \). Slightly weaker assumptions (Gallant, 1971) are listed in Appendix I.

As users of iterative methods are well aware, a mathematical proof of convergence is no guarantee that convergence will obtain on a computer. In this author's experience, convergence fails for two reasons: i) The model chosen does not fit the data, or ii) Poor start values are chosen.

When the inputs are scalars \((k=1)\) the first difficulty can be avoided. Plot the data. If your visual impression of the plotted data differs from the regression model chosen, expect difficulties. When the inputs are vector valued \((k > 1)\) the same considerations apply but are more difficult to verify.

The choice of start values is entirely an \textit{ad hoc} process. They may be obtained from prior knowledge of the situation, inspection of the data, grid search, or trial and error. For examples, see Gallant (1963) and Gallant and Fuller (1973). A more general approach to finding start values is given by Hartley and Booker (1965). A simpler variant of their idea is the following. Select \( p \) representative responses \( y_{t_1} \) and inputs \( x_{t_1} \) \((i = 1, \ldots, p)\). Solve the set of nonlinear equations

\[
y_{t_1} = f(x_{t_1}, \theta) \quad (i = 1, 2, \ldots, p)
\]

for \( \theta \).
Illustrating with our example, we will choose the observations with the largest and smallest inputs obtaining the equations

\[ 0.949 = \theta_1 e^{\theta_2} \]

\[ 0.481 = \theta_1 e^{\theta_2} \]

The solution of these equations is

\[ \theta_0 = (0.444, 0.823) \]

There are several ways of choosing \( \lambda_1 \) to satisfy \( \text{SSE}(\theta_1 + \lambda_1 D_1) < \text{SSE}(\theta_1) \) at each iterative step. Hartley (1961) suggests two methods in his paper. In this author's experience, it doesn't make very much difference how one chooses \( \lambda_1 \). What is important is that the computer program check that the condition

\[ \text{SSE}(\theta_1 + \lambda_1 D_1) < \text{SSE}(\theta_1) \]

is satisfied before taking the next iterative step.

In an intermediate step in Hartley's (1961) proof one sees that there is an \( \epsilon > 0 \) such that for every \( \lambda \) between 0 and \( \epsilon \)

\[ \text{SSE}(\theta_1 + \lambda D_1) < \text{SSE}(\theta_1) \]

For this reason, the author prefers the following method of choosing \( \lambda_1 \).

For some \( \alpha \) between 0 and 1, say \( \alpha = 0.6 \), successively check the values

\[ \beta_j = (\alpha)^j \quad j = 0, 1, \ldots \]

and choose \( \lambda_1 \) to be the largest \( \beta_j \) such that

\[ \text{SSE}(\theta_1 + \beta_j D_1) < \text{SSE}(\theta_1) \]
In some cases, no \( \lambda_1 \) satisfying the requirements can be found within the computational limits of the machine. This situation is discussed in the next paragraph.

There are a variety of stopping rules or tests for convergence employed to decide when to terminate the iterations. For example, one might set some tolerance \( \varepsilon > 0 \) and terminate when

\[
\| \hat{\theta}_1 - \hat{\theta}_{1+1} \| \leq \varepsilon \| \hat{\theta}_1 \|
\]

and simultaneously

\[
|\text{SSE}(\hat{\theta}_1) - \text{SSE}(\hat{\theta}_{1+1})| \leq \varepsilon |\text{SSE}(\hat{\theta}_1)|.
\]

(The symbol \( \| \hat{Z} \| \) denotes the Euclidean norm; \( \| \hat{Z} \| = (\sum_{i=1}^{D} \hat{z}_i^2)^{\frac{1}{2}} \).) There is one situation where one must stop. This is when no \( \lambda \) can be found such that

\[
\text{SSE}(\hat{\theta}_1 + \lambda \hat{D}_1) < \text{SSE}(\hat{\theta}_1).
\]

The author's preference is not to use a stopping rule other than some pre-chosen limit on the number of iteratives. The values of \( \hat{\theta}_1 \) and \( \text{SSE}(\hat{\theta}_1) \) are printed out for each iteration until either this limit is reached or no \( \lambda \) can be found to improve \( \hat{\theta}_1 \). The observations, predicted values, and residuals from this last iteration are printed out as well. If the last few iterations are identical to seven significant digits and the predicted values and residuals indicate that these values are acceptable they are used. A further check is to try another start value and see if the same answers are obtained.

The value of this last iteration will be taken as \( \hat{\theta} \). From this last iteration are computed
\[ \hat{\sigma}^2 = \frac{1}{n} \text{SSE}(\hat{\theta}) \]

and

\[ \hat{\Sigma} = [\hat{F}'(\hat{\theta}) \hat{F}(\hat{\theta})]^{-1} \] .

Under the assumptions listed in Appendix II, the least squares estimator is approximately normally distributed with mean \( \hat{\theta} \) and variance-covariance matrix \( \hat{\sigma}^2 \hat{\Sigma} \).

A FORTRAN subroutine, DMGN, is available to TUCC users which will perform one modified Gauss-Newton iterative step. The user is free to handle his own input, output, and stopping rules. The documentation is displayed as Exhibit II. The requisite JCL is as follows:

```
//NONLIN JOB xxx.yyy.zzz,programmer-name
//STEPI EXEC FTGCG
//C.SYSIN DD *
   source code
//G.SYSLIB DD DSN=NCS.ES.24139.GALLANT.GALLANT,DISP=SHR
  DD DSN=SYSL.FORTLIB,DISP=SHR
  DD DSN=SYSL.SUBLIB,DISP=SHR
//G.SYSIN DD *
   data
```

Source coding which will handle input and output is shown in Exhibit III and the documentation is shown in Exhibit IV. Users at NCSU, Duke, and UNC may obtain copies of the source deck by calling the author.

The exponential example we have been considering will be used to illustrate the use of this program. Assume that the data of Exhibit I have been punched one observation per card according to
Subroutine INPUT is coded in Exhibit V and Subroutine FUNCT is coded in Exhibit VI.

The deck arrangement is as follows:

//NONLIN JOB xxx.yyy.zzzz,programmer-name
//STEPL EXEC FTGCG
//C.SYSIN DD *
    source deck
    subroutine INPUT deck
    subroutine FUNCT deck
//G.SYSLIB DD DSN=NCSE.B4139.GALLANT.GALLANT,DISP=SHR
//    DD DSN=SYS1.FORTLIB,DISP=SHR
//    DD DSN=SYS1.SUBLIB,DISP=SHR
//G.SYSIN DD *
    data cards

The output for the example is shown in Exhibit VII. Note that the program prints a correlation matrix computed from $\hat{\Sigma}$ rather than $\hat{\Sigma}'$. The variance-covariance matrix $\hat{\sigma}^2 \hat{\Sigma}$ can be recovered by using the standard errors printed on the previous page.

If any ambiguities in documentation or difficulties with the program are encountered please call the author.
<table>
<thead>
<tr>
<th>t</th>
<th>( y_t )</th>
<th>( x_t )</th>
<th>t</th>
<th>( y_t )</th>
<th>( x_t )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.824</td>
<td>0.883</td>
<td>26</td>
<td>0.821</td>
<td>0.625</td>
</tr>
<tr>
<td>2</td>
<td>0.515</td>
<td>0.249</td>
<td>27</td>
<td>0.764</td>
<td>0.629</td>
</tr>
<tr>
<td>3</td>
<td>0.560</td>
<td>0.539</td>
<td>28</td>
<td>0.541</td>
<td>0.150</td>
</tr>
<tr>
<td>4</td>
<td>0.949</td>
<td>0.995</td>
<td>29</td>
<td>0.478</td>
<td>0.236</td>
</tr>
<tr>
<td>5</td>
<td>0.495</td>
<td>0.113</td>
<td>30</td>
<td>0.622</td>
<td>0.065</td>
</tr>
<tr>
<td>6</td>
<td>0.874</td>
<td>0.722</td>
<td>31</td>
<td>0.510</td>
<td>0.260</td>
</tr>
<tr>
<td>7</td>
<td>0.452</td>
<td>0.315</td>
<td>32</td>
<td>0.709</td>
<td>0.973</td>
</tr>
<tr>
<td>8</td>
<td>0.482</td>
<td>0.593</td>
<td>33</td>
<td>0.629</td>
<td>0.495</td>
</tr>
<tr>
<td>9</td>
<td>0.600</td>
<td>0.521</td>
<td>34</td>
<td>0.407</td>
<td>0.212</td>
</tr>
<tr>
<td>10</td>
<td>0.660</td>
<td>0.585</td>
<td>35</td>
<td>0.654</td>
<td>0.815</td>
</tr>
<tr>
<td>11</td>
<td>0.874</td>
<td>0.815</td>
<td>36</td>
<td>0.865</td>
<td>0.981</td>
</tr>
<tr>
<td>12</td>
<td>0.554</td>
<td>0.629</td>
<td>37</td>
<td>0.562</td>
<td>0.553</td>
</tr>
<tr>
<td>13</td>
<td>0.534</td>
<td>0.432</td>
<td>38</td>
<td>0.755</td>
<td>0.486</td>
</tr>
<tr>
<td>14</td>
<td>0.775</td>
<td>0.931</td>
<td>39</td>
<td>0.844</td>
<td>0.931</td>
</tr>
<tr>
<td>15</td>
<td>0.850</td>
<td>0.695</td>
<td>40</td>
<td>0.568</td>
<td>0.214</td>
</tr>
<tr>
<td>16</td>
<td>0.698</td>
<td>0.792</td>
<td>41</td>
<td>0.982</td>
<td>0.902</td>
</tr>
<tr>
<td>17</td>
<td>0.622</td>
<td>0.493</td>
<td>42</td>
<td>0.600</td>
<td>0.480</td>
</tr>
<tr>
<td>18</td>
<td>0.626</td>
<td>0.830</td>
<td>43</td>
<td>0.610</td>
<td>0.764</td>
</tr>
<tr>
<td>19</td>
<td>0.771</td>
<td>0.538</td>
<td>44</td>
<td>0.424</td>
<td>0.260</td>
</tr>
<tr>
<td>20</td>
<td>0.616</td>
<td>0.758</td>
<td>45</td>
<td>0.639</td>
<td>0.682</td>
</tr>
<tr>
<td>21</td>
<td>0.819</td>
<td>0.706</td>
<td>46</td>
<td>0.695</td>
<td>0.748</td>
</tr>
<tr>
<td>22</td>
<td>0.741</td>
<td>0.410</td>
<td>47</td>
<td>0.507</td>
<td>0.345</td>
</tr>
<tr>
<td>23</td>
<td>0.481</td>
<td>0.106</td>
<td>48</td>
<td>0.657</td>
<td>0.339</td>
</tr>
<tr>
<td>24</td>
<td>0.736</td>
<td>0.939</td>
<td>49</td>
<td>0.993</td>
<td>0.929</td>
</tr>
<tr>
<td>25</td>
<td>0.758</td>
<td>0.680</td>
<td>50</td>
<td>0.576</td>
<td>0.515</td>
</tr>
</tbody>
</table>
PURPOSE
COMPUTE A MODIFIED GAUSS-NEWTON ITERATIVE STEP FOR THE REGRESSION
MODEL YT=F(XT, THETA)+ET.

USAGE
CALL DMIN(FUNCT, Y, X, T1, N, K, IP, T2, E, D, C, VAR, IER)

SUBROUTINES CALLED
FUNCT, DMPRD, DSWEEP

ARGUMENTS

FUNCT - USER WRITTEN SUBROUTINE CONCERNING THE FUNCTION TO BE
FITTED. FOR AN INPUT VECTOR XT AND PARAMETER THETA FUNCT
SUPPLIES THE VALUE OF THE FUNCTION F(XT, THETA), STORED IN
THE ARGUMENT VAL, AND THE PARTIAL DERIVATIVES WITH RESPECT
TO THETA, STORED IN THE ARGUMENT DEL.
SUBROUTINE FUNCT IS OF THE FORM:
SUBROUTINE FUNCT(XT, THETA, VAL, DEL, ISW)
REAL*8 XT(K), THETA(IP), VAL, DEL(IP)
(STATEMENTS TO SUPPLY VAL)
IF(ISW.EQ.1) RETURN
(STATEMENTS TO SUPPLY DEL)
RETURN
END
THIS STATEMENT MUST BE INCLUDED IN THE CALLING PROGRAM:
EXTERNAL FUNCT

Y - AN N BY 1 VECTOR OF OBSERVATIONS.
ELEMENTS OF Y ARE REAL*8.

X - A K BY N MATRIX CONTAINING THE INPUT K BY 1 VECTORS STORED
AS COLUMNS OF X. COLUMN I OF X CONTAINS THE INPUT VECTOR
CORRESPONDING TO OBSERVATION Y(I). STORED COLUMNWISE
(STORAGE MODE 0).
ELEMENTS OF X ARE REAL*8.

T1 - INPUT IP BY 1 VECTOR CONTAINING THE START VALUE OF THETA
OR THE VALUE COMPUTED IN THE PREVIOUS ITERATIVE STEP.
ELEMENTS OF T1 ARE REAL*8.

N - NUMBER OF OBSERVATIONS
INTEGER

K - DIMENSION OF AN INPUT VECTOR IN THE MODEL F(XT, THETA)
INTEGER

IP - NUMBER OF PARAMETERS IN THE MODEL F(XT, THETA)
INTEGER

T2 - AN IP BY 1 VECTOR CONTAINING THE NEW ESTIMATE OF THETA.
THE ELEMENTS OF T2 ARE REAL*8.

E - AN N BY 1 VECTOR OF RESIDUALS FROM THE MODEL WITH THETA=T1.
ELEMENTS OF E ARE REAL*8.

D - AN IP BY 1 VECTOR CONTAINING AN UNMODIFIED GAUSS-NEWTON
CORRECTION VECTOR. SET THETA=T1+D TO OBTAIN AN UNMODIFIED
NEW ESTIMATE.
ELEMENTS OF D ARE REAL*8.
C - AN IP BY IP MATRIX CONTAINING THE ESTIMATED VARIANCE-COVARIANCE MATRIX OF T1 PROVIDED T1 IS THE LEAST SQUARES ESTIMATE OF THETA. C=VAR*INVERSE(SUM(DEL1*DELI')). STORED COLUMNWISE (STORAGE MODE O). ELEMENTS OF C ARE REAL*8.

VAR - ESTIMATED VARIANCE OF OBSERVATIONS PROVIDED T1 IS THE LEAST SQUARES ESTIMATE.
REAL*8.

IER - INTEGER ERROR PARAMETER CODED AS FOLLOWS:
IER=0 NO ERROR
IER=1 T2=T1+V*D WHERE V=.6**L FAILED TO REDUCE THE RESIDUAL SUM OF SQUARES FOR I=0,1,2,...,40.
IER.GT.9 AN INVERSION ERROR OCCURRED, UNITS POSITION OF IER HAS THE SAME MEANING AS ABOVE.

REFERENCE
HARTLEY, H. O. THE MODIFIED GAUSS-NEWTON METHOD FOR THE FITTING OF NON-LINEAR REGRESSION FUNCTIONS BY LEAST SQUARES. TECHNOMETRICS, 3.

PROGRAMMER
DR. A. RONALD GALLANT
DEPARTMENT OF STATISTICS
NORTH CAROLINA STATE UNIVERSITY
RALEIGH, NORTH CAROLINA 27607
REAL*8 R(3000)
ISW=1
CALL INPUT(ISW,N,K,IP,ITER,R,R)
M0=1
ML=N +M0
M2=K*N +ML
M3=IP +M2
M4=IP +M3
M5=N +M4
M6=IP +M5
M7=IP*IP +M6
M8=1 +M7
WRITE(3,1)M8
WRITE(3,2)
DO 10 I=1,M8
10 R(I)=0. DO
IOUT=1
IOUT=2
CALL NONLIN(R(M0),R(M1),R(M2),R(M3),R(M4),R(M5),R(M6),R(M7),
* N,K,IP,ITER,IOUT)
C SUBROUTINE NONLIN(Y,X,TL,T2,E,D,C,VAR,N,K,IP,ITER,IOUT)
C REAL*8 Y(N),X(K,N),TL(IP),T2(IP),E(N),D(IP),C(IP,IP),VAR
STOP
1 FORMAT('1'/////////
*36X,'***********************************************************************/
*36X,'*/
*36X,'* THE VECTOR R MUST BE DIMENSIONED AT LEAST */
*36X,'* AS LARGE AS M3 =',I9, '*' /
*36X,'*'/
*36X,'***********************************************************************/
2 FORMAT('1'/////////
*36X,'***********************************************************************/
*36X,'*'/
*36X,'* PLEASE REPORT ANY PROBLEMS WITH THIS PROGRAM TO; */
*36X,'* DR. A. RONALD GALLANT */
*36X,'* DEPARTMENT OF STATISTICS */
*36X,'* NORTH CAROLINA STATE UNIVERSITY */
*36X,'* RALEIGH, NORTH CAROLINA 27607 */
*36X,'* (919) 737-2531 */
*36X,'*'/
*36X,'***********************************************************************/
END
SOURCE DECK FOR MODIFIED GAUSS-NEWTON NONLINEAR ESTIMATION
USER SUPPLIED SUBROUTINES INPUT AND FUNCT REQUIRED.

SUBROUTINE INPUT IS OF THE FORM:
SUBROUTINE INPUT(ISW,N,K,IP,ITER,Y,X,TO)
REAL*8 Y(N),X(K,N),TO(IP)
IF(ISW.EQ.1) GO TO 1
IF(ISW.EQ.2) GO TO 2
1 CONTINUE
(CODING TO SUPPLY N, K, IP, ITER)
RETURN
2 CONTINUE
(CODING TO SUPPLY Y, X, TO)
END

SUBROUTINE FUNCT IS DESCRIBED IN THE DOCUMENTATION OF SUBROUTINE
DGN AND IS OF THE FORM:
SUBROUTINE FUNCT(XT,THETA,VAL,DEL,ISW)
REAL*8 XT(K),THETA(IP),VAL,DEL(IP)
(CODING TO SUPPLY VAL)
IF(ISW.EQ.1) RETURN
(CODING TO SUPPLY DEL)
RETURN
END

ARGUMENTS OF INPUT AND FUNCT:
ISW - INTEGER SWITCH.
      SUPPLIED BY CALLING PROGRAM.
      INTEGER*4
N - NUMBER OF OBSERVATIONS.
    SUPPLIED BY USER, AVAILABLE TO CALLING PROGRAM ON RETURN.
    INTEGER*4
K - DIMENSION OF AN INPUT VECTOR IN THE MODEL F(XT,THETA).
    SUPPLIED BY USER, AVAILABLE TO CALLING PROGRAM ON RETURN.
    INTEGER*4
IP - NUMBER OF PARAMETERS IN THE MODEL F(XT,THETA).
    SUPPLIED BY USER, AVAILABLE TO CALLING PROGRAM ON RETURN.
    INTEGER*4
ITER - NUMBER OF ITERATIONS DESIRED.
       SUPPLIED BY USER, AVAILABLE TO CALLING PROGRAM ON RETURN.
       INTEGER*4
Y - AN N BY 1 VECTOR OF OBSERVATIONS.
      SUPPLIED BY USER, AVAILABLE TO CALLING PROGRAM ON RETURN.
      REAL*8
X - A K BY N MATRIX CONTAINING THE K BY 1 INPUT VECTORS STORED
    AS COLUMNS OF X. COLUMN I OF X CONTAINS THE INPUT VECTOR
    CORRESPONDING TO OBSERVATIONS Y(I).
      SUPPLIED BY USER, AVAILABLE TO CALLING PROGRAM ON RETURN.
      REAL*8
TO - INPUT IP BY 1 VECTOR CONTAINING THE START VALUE OF THETA.
SUPPLIED BY USER, AVAILABLE TO CALLING PROGRAM ON RETURN.
REAL*8

XT - A K BY 1 VECTOR CONTAINING AN INPUT VECTOR.
SUPPLIED BY CALLING PROGRAM.
REAL*8

THETA - AN IP BY 1 VECTOR CONTAINING PARAMETER VALUES.
SUPPLIED BY CALLING PROGRAM.
REAL*8

VAL - VAL = F(XT, THETA).
SUPPLIED BY USER, AVAILABLE TO CALLING PROGRAM ON RETURN.
REAL*8

DEL - AN IP BY 1 VECTOR CONTAINING THE PARTIAL DERIVATIVES OF
F(XT, THETA) WITH RESPECT TO THETA.
SUPPLIED BY USER, AVAILABLE TO CALLING PROGRAM ON RETURN.
Exhibit V

SUBROUTINE INPUT(ISW,N,K,IP,ITER,Y,X,TO)
  REAL*8 Y(50),X(1,50),TO(2)
  IF(ISW.EQ.1) GO TO 1
  IF(ISW.EQ.2) GO TO 2
1 CONTINUE
  N=50
  K=1
  IP=2
  ITER=25
  RETURN
2 CONTINUE
  TO(1)=.4444D0
  TO(2)=.823D0
  READ(1,100) (Y(I),X(1,I),I=1,50)
100 FORMAT(2F10.3)
  RETURN
END
SUBROUTINE FUNCT(XT, THETA, VAL, DEL, ISW)
REAL*8 XT(1), THETA(2), VAL, DEL(2)
VAL=THETA(1)*DEXP(THETA(2)*XT(1))
IF(ISW.EQ.1) RETURN
DEL(1)=DEXP(THETA(2)*XT(1))
DEL(2)=THETA(1)*XT(1)*DEXP(THETA(2)*XT(1))
RETURN
END
THE VECTOR A MUST BE DIMENSIONED AT LEAST AS LARGE AS MB = 162.

PLEASE REPORT ANY PROBLEMS WITH THIS PROGRAM TO:

DR. A. RONALD GALLAGHER
DEPARTMENT OF STATISTICS
DEPARTMENT OF COMPUTER SCIENCE
UNIVERSITY OF PENNSYLVANIA
PHILADELPHIA, PENNSYLVANIA
(215) 726-2531
## MODIFIED GAUSS-NEWTON ITERATIONS

<table>
<thead>
<tr>
<th>STEP</th>
<th>ERROR SUM OF SQUARES</th>
<th>PARAMETER NUMBER</th>
<th>INTERMEDIATE ESTIMATE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.734936840 00</td>
<td>1</td>
<td>0.444000000 00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>0.823000000 00</td>
</tr>
<tr>
<td>1</td>
<td>0.454472620 00</td>
<td>1</td>
<td>0.447405300 00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>0.673350780 00</td>
</tr>
<tr>
<td>2</td>
<td>0.453567110 00</td>
<td>1</td>
<td>0.449304470 00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>0.659292920 00</td>
</tr>
<tr>
<td>3</td>
<td>0.453567080 00</td>
<td>1</td>
<td>0.449361500 00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>0.659164400 00</td>
</tr>
<tr>
<td>4</td>
<td>0.453567080 00</td>
<td>1</td>
<td>0.449361660 00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>0.659160910 00</td>
</tr>
<tr>
<td>5</td>
<td>0.453567080 00</td>
<td>1</td>
<td>0.449361660 00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>0.659160900 00</td>
</tr>
<tr>
<td>6</td>
<td>0.453567080 00</td>
<td>1</td>
<td>0.449361660 00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>0.659160900 00</td>
</tr>
<tr>
<td>7</td>
<td>0.453567080 00</td>
<td>1</td>
<td>0.449361660 00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>0.659160900 00</td>
</tr>
<tr>
<td>8</td>
<td>0.453567080 00</td>
<td>1</td>
<td>0.449361660 00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>0.659160900 00</td>
</tr>
</tbody>
</table>

--- UNABLE TO IMPROVE ON THIS ESTIMATE

---

CHECK THESE ITERATIONS TO SEE IF CONVERGENCE HAS BEEN ACHIEVED BEFORE USING THE FOLLOWING RESULTS.
<table>
<thead>
<tr>
<th>PARAMETER NUMBER</th>
<th>ESTIMATE</th>
<th>STANDARD ERROR</th>
<th>ESTIMATED VARIANCE</th>
<th>RESIDUAL SUM OF SQUARES</th>
<th>NUMBER OF OBSERVATIONS</th>
<th>NUMBER OF PARAMETERS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.449462</td>
<td>0.669661</td>
<td>0.669661</td>
<td>0.907174</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>0.634250-01</td>
<td>0.597240-01</td>
<td>0.597240-01</td>
<td>0.4596706</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>17.059</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.523</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y</td>
<td>THAT</td>
<td>RESIDUAL</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>--------------</td>
<td>--------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.8240000</td>
<td>0.8042150</td>
<td>0.1978460</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.8150000</td>
<td>0.5295140</td>
<td>-0.4513370</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5600000</td>
<td>0.4410590</td>
<td>-0.1199530</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.9490000</td>
<td>0.8658360</td>
<td>-0.0316660</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.4950000</td>
<td>0.4841210</td>
<td>-0.0108950</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.8740000</td>
<td>0.7237410</td>
<td>-0.1507590</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.4520000</td>
<td>0.5530840</td>
<td>-0.1010560</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.4820000</td>
<td>0.5822370</td>
<td>-0.1002370</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.4600000</td>
<td>0.6326910</td>
<td>-0.3349430</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.6600000</td>
<td>0.6609110</td>
<td>-0.0700710</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.8740000</td>
<td>0.7689460</td>
<td>-0.1050300</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.8560000</td>
<td>0.6007360</td>
<td>-0.1262390</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5340000</td>
<td>0.5973990</td>
<td>-0.0632990</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.7750000</td>
<td>0.6300670</td>
<td>-0.5506740</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.7050000</td>
<td>0.7104330</td>
<td>-0.1395170</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.6500000</td>
<td>0.7573940</td>
<td>-0.9839390</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.6220000</td>
<td>0.6214990</td>
<td>-0.9069320</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.6760000</td>
<td>0.7766550</td>
<td>-0.1506050</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.6600000</td>
<td>0.6406130</td>
<td>-0.1203370</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.7710000</td>
<td>0.7406040</td>
<td>-0.1266080</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.6160000</td>
<td>0.7156510</td>
<td>-0.1033670</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.8190000</td>
<td>0.5887840</td>
<td>-0.1520200</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.7610000</td>
<td>0.4818220</td>
<td>-0.0819160</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.7360000</td>
<td>0.8344480</td>
<td>-0.9245610</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.7360000</td>
<td>0.7034110</td>
<td>-0.5540730</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.6870000</td>
<td>0.7084450</td>
<td>-0.1425550</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.7640000</td>
<td>0.6602830</td>
<td>-0.8376370</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.6410000</td>
<td>0.6967610</td>
<td>-0.4497940</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.4760000</td>
<td>0.5295610</td>
<td>-0.6409570</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.6220000</td>
<td>0.6640310</td>
<td>-0.1529600</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5370000</td>
<td>0.5333367</td>
<td>-0.2366700</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.7090000</td>
<td>0.6533690</td>
<td>-0.1144390</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.6290000</td>
<td>0.6227320</td>
<td>-0.6270220</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.4450000</td>
<td>0.5157864</td>
<td>-0.1097580</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.6540000</td>
<td>0.7689640</td>
<td>-0.1149640</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.8650000</td>
<td>0.6578810</td>
<td>-0.7119440</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.9620000</td>
<td>0.6669900</td>
<td>-0.0949340</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.7950000</td>
<td>0.6190640</td>
<td>-0.1359460</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.8440000</td>
<td>0.8300670</td>
<td>-0.1392260</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.9680000</td>
<td>0.5714370</td>
<td>-0.5098260</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.9820000</td>
<td>0.6143510</td>
<td>-0.1074880</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.6000000</td>
<td>0.6166030</td>
<td>-0.1660290</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.6100000</td>
<td>0.7435430</td>
<td>-0.1339530</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.6240000</td>
<td>0.5333670</td>
<td>-0.1053370</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.4390000</td>
<td>0.7044210</td>
<td>-0.6542070</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.6950000</td>
<td>0.7357450</td>
<td>-0.1407420</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5070000</td>
<td>0.6523330</td>
<td>-0.0610400</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.6520000</td>
<td>0.5619770</td>
<td>-0.0952270</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.9930000</td>
<td>0.6289740</td>
<td>-0.1648260</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5760000</td>
<td>0.6309940</td>
<td>-0.0949390</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
REFERENCES


Gallant, A. R. and Fuller, W. A. (1973) "Fitting segmented polynomial models whose join points have to be estimated," Journal of the American Statistical Association, 68. p. 144-47.


APPENDIX I

Theorem. We are given a regression model

\[ y_t = f(x_t; \theta) + e_t \]

and the data pairs \((y_t, x_t)\) \((t = 1, 2, \ldots, n)\). Let \(Q(\theta) = \text{SSE}(\theta)\).

Conditions. There is a convex, bounded subset \(S\) of \(\mathbb{R}^p\) and a \(\theta_0\) interior to \(S\) such that:

1) \(\nabla f(x_t, \theta)\) exists and is continuous over \(\bar{S}\) for \(t = 1, 2, \ldots, n\).
2) \(\theta \in S\) implies the rank of \(F(\theta)\) is \(p\).
3) \(Q(\theta_0) < \tilde{Q} = \inf\{Q(\theta) : \theta\text{ a boundary point of } S\}\).
4) There does not exist \(\theta', \theta''\) in \(S\) such that

\[ \nabla Q(\theta') = \nabla Q(\theta'') = 0 \quad \text{and} \quad Q(\theta') = Q(\theta'') \]

Construction. Construct the sequence \(\{\theta_\alpha\}_{\alpha=1}^\infty\) as follows:

0) Compute \(D'_0 = [F'(\theta_0) F(\theta_0)]^{-1} F'(\theta_0) [y - f(\theta_0)]\).

Find \(\lambda_0\) which minimizes \(Q(\theta_0 + \lambda D_0)\) over \(\Lambda_0 = \{\lambda : 0 \leq \lambda \leq 1, \theta_0 + \lambda D_0 \in \bar{S}\}\).

1) Set \(\theta_1 = \theta_0 + \lambda_0 D_0\).

Compute \(D'_1 = [F'(\theta_1) F(\theta_1)]^{-1} F'(\theta_1) [y - f(\theta_1)]\).

Find \(\lambda_1\) which minimizes \(Q(\theta_1 + \lambda D_1)\) over \(\Lambda_1 = \{\lambda : 0 \leq \lambda \leq 1, \theta_1 + \lambda D_1 \in \bar{S}\}\).

2) Set \(\theta_2 = \theta_1 + \lambda_1 D_1\).

\vdots
Conclusions. Then for the sequence \( \{\theta_{\alpha}\}_{\alpha=1}^{\infty} \) it follows that:

1) \( \theta_{\alpha} \) is an interior point of \( S \) for \( \alpha = 1, 2, \ldots \).

2) The sequence \( \{\theta_{\alpha}\} \) converges to a limit \( \theta^* \) which is interior to \( S \).

3) \( vQ(\theta^*) = 0 \).

APPENDIX II

In order to obtain asymptotic results, it is necessary to specify the behavior of the inputs \( x_t \) as \( n \) becomes large. A general way of specifying the limiting behavior of nonlinear regression inputs is due to Malinvaud (1970). His definitions are merely stated here, a complete discussion and examples are contained in his paper.

Let \( X \) be a subset of \( \mathbb{R}^k \) from which the inputs \( x_t \) \( (t = 1, 2, \ldots) \) are to be chosen.

Definition. Let \( G \) be the Borel subsets of \( X \) and \( \{x_t\}_{t=1}^{\infty} \) be the sequence of inputs chosen from \( X \). Let \( I_A(x_t) \) be the indicator function of a subset \( A \) of \( X \). The measure \( \mu_n \) on \( (X,G) \) is defined by

\[
\mu_n(A) = n^{-1} \sum_{t=1}^{n} I_A(x_t)
\]

for each \( A \in G \).

Definition. A sequence of measures \( \{\mu_n\} \) on \( (X,G) \) is said to converge weakly to a measure \( \mu \) on \( (X,G) \) if for every bounded, continuous function \( g \) with domain \( X \)

\[
\int g(x) \, d\mu_n(x) \to \int g(x) \, d\mu(x)
\]

as \( n \to \infty \).

Asymptotic results may be obtained under the following set of Assumptions.

Assumptions. The parameter space \( \Omega \) and the set \( X \) are compact subsets of the \( p \)-dimensional and \( k \)-dimensional reals, respectively. The response function \( f(x,\theta) \) and the partial derivatives \( \frac{\partial}{\partial \theta_i} f(x,\theta) \) and \( \frac{\partial^2}{\partial \theta_i \partial \theta_j} f(x,\theta) \) are continuous on \( X \times \Omega \). The sequence of inputs \( \{x_t\}_{t=1}^{\infty} \) is chosen such that the sequence of measures \( \{\mu_n\}_{n=1}^{\infty} \) converges weakly to a measure \( \mu \).
defined over \((X,\mathcal{G})\). The true value of \(\theta\), denoted by \(\theta^0\), is contained in an open set which, in turn, is contained in \(\Omega\). If \(f(x,\theta) = f(x,\theta^0)\) except on a set of \(\mu\) measure zero, it is assumed that \(\theta = \theta^0\). The \(p \times p\) matrix

\[
\mathbf{L} = \left[ \int \frac{\partial}{\partial \theta_i} f(x,\theta^0) \frac{\partial}{\partial \theta_j} f(x,\theta^0) \, d\mathbf{u}(x) \right]
\]

is non-singular. The errors \(\{e_t\}\) are independent and identically distributed with mean zero and finite, non-zero variance \(\sigma^2\).

These Assumptions are patterned after those used by Malinvaud (1970) to show that the least squares estimator is consistent. A similar set of Assumptions which does not require that \(\Omega\) be bounded or that the second partial derivatives of \(f(x,\theta)\) exist may be found in Gallant (1971 or 1973). An alternative set of Assumptions may be found in Jennrich (1969).

Theorem. Let \(\hat{\theta}\) denote the function of \(y\) which minimizes SSE(\(\theta\)). Let \(\hat{\sigma}^2 = n^{-1}\text{SSE}(\hat{\theta})\). Under the Assumptions listed above:

1) The estimator \(\hat{\theta}\) is consistent for \(\theta^0\).
2) The estimator \(\hat{\sigma}^2\) is consistent for \(\sigma^2\).
3) \(n^{-1}F'(\hat{\theta})F(\hat{\theta})\) is consistent for \(\mathbf{L}\).
4) \(\sqrt{n} (\theta - \theta^0)\)' is asymptotically normal with mean zero and variance-covariance matrix \(\sigma^2 \mathbf{L}^{-1}\).


Remark. In computations, it is customary to absorb the term \(\sqrt{n}\) of \(\sqrt{n} (\theta - \theta^0)'\) in the variance-covariance matrix. Thus, one enters the tables by taking

\[
(\theta - \theta^0)' \sim N_p(0, \hat{\sigma}^2 \hat{\Sigma})
\]

where \(\hat{\Sigma} = [F'(\hat{\theta})F(\hat{\theta})]^{-1}\).