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 $\underset{\sim}{x}_t$ are assumed to be generated according to the nonlinear regression model

$$y_t = f(\underset{\sim}{x}_t, \underset{\sim}{\theta}) + e_t \qquad (t = 1, 2, \ldots, n) .$$

The input variables are k-dimensional

$$\underset{\sim}{x}_t = (x_{1t}, x_{2t}, \ldots, x_{kt})$$

and the unknown parameter $\underset{\sim}{\theta}$ is p-dimensional

$$\underset{\sim}{\theta} = (\theta_1, \theta_2, \ldots, \theta_p) .$$

The least squares estimater is that value

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

which minimizes

$$SSE(\underset{\sim}{\theta}) = \Sigma_{t=1}^{n} \{y_t - f(\underset{\sim}{x}_t, \underset{\sim}{\theta})\}^2 .$$

The following vector and matrix notation will be useful in describing the modified Gauss-Newton method.

$$\underset{\sim}{y} = (y_1, y_2, \ldots, y_n)' \qquad (n \times 1)$$

$$\underset{\sim}{f}(\underset{\sim}{\theta}) = (f(\underset{\sim}{x}_1,\underset{\sim}{\theta}), f(\underset{\sim}{x}_2,\underset{\sim}{\theta}), \ldots, f(\underset{\sim}{x}_n,\underset{\sim}{\theta}))' \qquad (n \times 1)$$

$$\underset{\sim}{\nabla} f(\underset{\sim}{x},\underset{\sim}{\theta}) = \text{the } p \times 1 \text{ vector whose } j^{th} \text{ element is } \frac{\partial}{\partial\theta_j} f(\underset{\sim}{x},\underset{\sim}{\theta})$$

$$\underset{\sim}{F}(\underset{\sim}{\theta}) = \text{the } n \times p \text{ matrix whose } t^{th} \text{ row is } \underset{\sim}{\nabla}' f(\underset{\sim}{x}_t,\underset{\sim}{\theta})$$

To illustrate the notation, consider the 50 responses $(=y_t)$ and inputs $(=\underset{\sim}{x}_t)$ shown in Exhibit I. We assume that

$$y_t = \theta_1 e^{x_t \theta_2} + e_t \,.$$

Thus,

$$f(\underset{\sim}{x},\underset{\sim}{\theta}) = \theta_1 e^{x\theta_2}$$

$$\underset{\sim}{\theta} = (\theta_1, \theta_2)$$

$$\underset{\sim}{x}_t = x_t$$

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

$$\underset{\sim}{y} = \begin{bmatrix} .824 \\ .515 \\ \vdots \\ .576 \end{bmatrix} \qquad (50 \times 1)$$

$$\begin{bmatrix} \theta_1 e^{.883\ \theta_2} \\ \theta_1 e^{.249\ \theta_2} \\ \vdots \\ \theta_1 e^{.515\ \theta_2} \end{bmatrix} \quad (50 \times 1)$$

$$\begin{bmatrix} e^{.883\ \theta_2} & .833\ \theta_1 e^{.833\ \theta_2} \\ e^{.249\ \theta_2} & .249\ \theta_1 e^{.249\ \theta_2} \\ \vdots & \vdots \\ e^{.515\ \theta_2} & .515\ \theta_1 e^{.515\ \theta_2} \end{bmatrix} \quad (50 \times 2) \ .$$

The modified Gauss-Newton (Hartley, 1961) algorithm proceeds as follows. Find a start value $\underset{\sim}{\theta}_o$; methods of finding start values are discussed later.

0) From the starting estimate $\underset{\sim}{\theta}_o$ compute

$$\underset{\sim}{D}_o' = [\underset{\sim}{F}'(\underset{\sim}{\theta}_o)\ \underset{\sim}{F}(\underset{\sim}{\theta}_o)]^{-1}\underset{\sim}{F}'(\underset{\sim}{\theta}_o)[\underset{\sim}{y} - \underset{\sim}{f}(\underset{\sim}{\theta}_o)]$$

Find a λ_o between 0 and 1 such that

$$SSE(\underset{\sim}{\theta}_o + \lambda_o \underset{\sim}{D}_o) \le SSE(\underset{\sim}{\theta}_o) \ .$$

1) Let $\underset{\sim}{\theta}_1 = \underset{\sim}{\theta}_o + \lambda_o \underset{\sim}{D}_o$. Compute

$$\underset{\sim}{D}_1' = [\underset{\sim}{F}'(\underset{\sim}{\theta}_1)\ \underset{\sim}{F}(\underset{\sim}{\theta}_1)]^{-1}\underset{\sim}{F}'(\underset{\sim}{\theta}_1)[\underset{\sim}{y} - f(\underset{\sim}{\theta}_1)]$$

Find a λ_1 between 0 and 1 such that

$$SSE(\underset{\sim}{\theta}_1 + \lambda_1 \underset{\sim}{D}_1) \le SSE(\underset{\sim}{\theta}_1)$$

2) Let $\underset{\sim}{\theta}_2 = \underset{\sim}{\theta}_1 + \lambda_1 \underset{\sim}{D}_1$

$$\vdots$$

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 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 (1968) 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_i} and inputs $\underset{\sim}{x}_{t_i}$ $(i = 1, \ldots, p)$. Solve the set of nonlinear equations

$$y_{t_i} = f(\underset{\sim}{x}_{t_i}, \underset{\sim}{\theta}) \qquad (i = 1, 2, \ldots, p)$$

for $\underset{\sim}{\theta}$.

Illustrating with our example, we will choose the observations with the largest and smallest inputs obtaining the equations

$$.949 = \theta_1 e^{.995\,\theta_2}$$

$$.481 = \theta_1 e^{.106\,\theta_2} \; .$$

The solution of these equations is

$$\underset{\sim}{\theta}_o = (.444, .823) \; .$$

There are several ways of chosing λ_i to satisfy $SSE(\underset{\sim}{\theta}_i + \lambda_i \underset{\sim}{D}_i) < SSE(\underset{\sim}{\theta}_i)$ 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 λ_i. What is important is that the computer program check that the condition

$$SSE(\underset{\sim}{\theta}_i + \lambda_i \underset{\sim}{D}_i) < SSE(\underset{\sim}{\theta}_i)$$

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 λ between 0 and ϵ

$$SSE(\underset{\sim}{\theta}_i + \lambda \underset{\sim}{D}_i) < SSE(\underset{\sim}{\theta}_i) \; .$$

For this reason, the author prefers the following method of choosing λ_i. For some α between 0 and 1, say $\alpha = .6$, successively check the values

$$\beta_j = (\alpha)^j \qquad j = 0, 1, \ldots$$

and chose λ_i to be the largest β_j such that

$$SSE(\underset{\sim}{\theta}_i + \beta_j \underset{\sim}{D}_i) < SSE(\underset{\sim}{\theta}_i) \; .$$

In some cases, no λ_i 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 $\epsilon > 0$ and terminate when

$$\| \underset{\sim}{\theta}_i - \underset{\sim}{\theta}_{i+1} \| \leq \epsilon \, \| \underset{\sim}{\theta}_i \|$$

and simultaneously

$$|SSE(\underset{\sim}{\theta}_i) - SSE(\underset{\sim}{\theta}_{i+1})| \leq \epsilon \, |SSE(\underset{\sim}{\theta}_i)| \, .$$

(The symbol $\| \underset{\sim}{Z} \|$ denotes the Euclidean norm; $\| \underset{\sim}{Z} \| = (\Sigma_{i=1}^{p} \, z_i^2)^{\frac{1}{2}}$.) There is one situation where one must stop. This is when no λ can be found such that

$$SSE(\underset{\sim}{\theta}_i + \lambda \underset{\sim}{D}_i) < SSE(\underset{\sim}{\theta}_i) \, .$$

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 $\underset{\sim}{\theta}_i$ and $SSE(\underset{\sim}{\theta}_i)$ are printed out for each iteration until either this limit is reached or no λ can be found to improve $\underset{\sim}{\theta}_i$. 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

$$\hat{\sigma}^2 = \frac{1}{n}\,\mathrm{SSE}(\hat{\theta})$$

and

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

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

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.zzzz,programmer-name
//STEP1   EXEC   FTGCG
//C.SYSIN   DD *
     source code
//G.SYSLIB  DD  DSN=NCS.ES.B4139.GALLANT.GALLANT,DISP=SHR
//          DD  DSN=SYS1.FORTLIB,DISP=SHR
//          DD  DSN=SYS1.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

```
100 FORMAT(2F10.3) .
```

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.zzzzz,programmer-name
//STEP1  EXEC   FTGCG
//C.SYSIN  DD *
    source deck
    subroutine INPUT deck
    subroutine FUNCT deck
//G.SYSLIB DD DSN=NCS.ES.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 $\hat{\underset{\sim}{\Sigma}}$ rather than $\hat{\underset{\sim}{\Sigma}}$. The variance-covariance matrix $\hat{\sigma}^2 \hat{\underset{\sim}{\Sigma}}$ 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.

Exhibit I. Observations

t	y_t	x_t	t	y_t	x_t
1	0.824	0.883	26	0.821	0.625
2	0.515	0.249	27	0.764	0.629
3	0.560	0.539	28	0.541	0.150
4	0.949	0.995	29	0.478	0.236
5	0.495	0.113	30	0.622	0.065
6	0.874	0.722	31	0.510	0.260
7	0.452	0.315	32	0.709	0.973
8	0.482	0.393	33	0.629	0.495
9	0.600	0.521	34	0.407	0.212
10	0.660	0.585	35	0.654	0.815
11	0.874	0.815	36	0.865	0.981
12	0.554	0.629	37	0.562	0.553
13	0.534	0.432	38	0.755	0.486
14	0.775	0.931	39	0.844	0.931
15	0.850	0.695	40	0.568	0.214
16	0.698	0.792	41	0.982	0.902
17	0.622	0.493	42	0.600	0.480
18	0.626	0.830	43	0.610	0.764
19	0.771	0.538	44	0.424	0.260
20	0.616	0.758	45	0.639	0.682
21	0.819	0.706	46	0.695	0.748
22	0.741	0.410	47	0.507	0.345
23	0.481	0.106	48	0.657	0.339
24	0.736	0.939	49	0.993	0.929
25	0.758	0.680	50	0.576	0.515

Exhibit II

DMGN 4/20/72

PURPOSE
COMPUTE A MODIFIED GAUSS-NEWTON ITERATIVE STEP FOR THE REGRESSION MODEL YT=F(XT,THETA)+ET.

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

SUBROUTINES CALLED
FUNCT, DGMPRD, 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(DEL*DEL')).  STORED
         COLUMNWISE (STORAGE MODE 0).
         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 L=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

Exhibit III

```
      REAL*8 R(3000)
      ISW=1
      CALL INPUT(ISW,N,K,IP,ITER,R,R,R)
      MO=1
      M1=N                 +MO
      M2=K*N               +M1
      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(MO),R(M1),R(M2),R(M3),R(M4),R(M5),R(M6),R(M7),
     *           N,K,IP,ITER,IOUT)
C     SUBROUTINE NONLIN(Y,X,T1,T2,E,D,C,VAR,N,K,IP,ITER,IOUT)
C     REAL*8 Y(N),X(K,N),T1(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 M8 =',I9,   '.                        *'/
     *36X,'*                                                          *'/
     *36X,'************************************************************')
2     FORMAT(' '////////
     *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
```

Exhibit IV

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 DMGN 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)=.444D0
      TO(2)=.823D0
      READ(1,100) (Y(I),X(1,I),I=1,50)
  100 FORMAT(2F10.3)
      RETURN
      END
```

Exhibit VI

```
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
```

Exhibit VII

```
***********************************************************
*                                                         *
*      THE VECTOR R MUST BE DIMENSIONED AT LEAST          *
*      AS LARGE AS MB =       162.                        *
*                                                         *
***********************************************************


***********************************************************
*                                                         *
*   PLEASE REPORT ANY PROBLEMS WITH THIS PROGRAM TO:      *
*   DR. A. RONALD GALLANT                                 *
*   DEPARTMENT OF STATISTICS                              *
*   NORTH CAROLINA STATE UNIVERSITY                       *
*   RALEIGH, NORTH CAROLINA 27607                         *
*   (919) 737-2531                                        *
*                                                         *
***********************************************************
```

MODIFIED GAUSS-NEWTON ITERATIONS

STEP	ERROR SUM OF SQUARES	PARAMETER NUMBER	INTERMEDIATE ESTIMATE
0	0.73493684D 00	1	0.44400000D 00
		2	0.82300000D 00
1	0.45447262D 00	1	0.44749053D 00
		2	0.67335078D 00
2	0.45356711D 00	1	0.44933047D 00
		2	0.65928292D 00
3	0.45356708D 00	1	0.44936150D 00
		2	0.65916144D 00
4	0.45356708D 00	1	0.44936166D 00
		2	0.65916091D 00
5	0.45356708D 00	1	0.44936166D 00
		2	0.65916090D 00
6	0.45356708D 00	1	0.44936166D 00
		2	0.65916090D 00
7	0.45356708D 00	1	0.44936166D 00
		2	0.65916090D 00
8	0.45356708D 00	1	0.44936166D 00
		2	0.65916090D 00
----	UNABLE TO IMPROVE ON THIS ESTIMATE		

CHECK THESE ITERATIONS TO SEE IF CONVERGENCE HAS
BEEN ACHIEVED BEFORE USING THE FOLLOWING RESULTS

MODIFIED GAUSS-NEWTON PARAMETER ESTIMATES

PARAMETER NUMBER	ESTIMATE	STANDARD ERROR	Z-STATISTIC
1	0.449362	0.254525D-01	17.6549
2	0.659161	0.802648D-01	8.2123

ESTIMATED VARIANCE	0.907134D-02
RESIDUAL SUM OF SQUARES	0.45356708
NUMBER OF OBSERVATIONS	50
NUMBER OF PARAMETERS	2

CORRELATION MATRIX OF THE PARAMETER ESTIMATES

$$\begin{vmatrix} 1.000000 & -0.935921 \\ -0.935921 & 1.000000 \end{vmatrix}$$

OBSERVED RESPONSES, PREDICTED RESPONSES, RESIDUALS

Y	YHAT	RESIDUAL
0.8240000D 00	0.8042150D 00	0.197846D-01
0.5150000D 00	0.5295140D 00	-0.145137D-01
0.5600000D 00	0.6410550D 00	-0.810553D-01
0.9490000D 00	0.8658340D 00	0.831661D-01
0.4950000D 00	0.4841110D 00	0.108895D-01
0.8740000D 00	0.7232410D 00	0.150759D 00
0.4520000D 00	0.5530580D 00	-0.101058D 00
0.4820000D 00	0.5822370D 00	-0.100237D 00
0.6000000D 00	0.6334940D 00	-0.334947D-01
0.6600000D 00	0.6607910D 00	-0.790714D-03
0.8740000D 00	0.7689640D 00	0.105036D 00
0.5540000D 00	0.6802360D 00	-0.126236D 00
0.5340000D 00	0.5973990D 00	-0.633992D-01
0.7750000D 00	0.8300670D 00	-0.550674D-01
0.8500000D 00	0.7104830D 00	0.139517D 00
0.6980000D 00	0.7573940D 00	-0.593939D-01
0.6220000D 00	0.6219090D 00	0.906352D-04
0.6260000D 00	0.7766050D 00	-0.150605D 00
0.7710000D 00	0.6406330D 00	0.130367D 00
0.6160000D 00	0.7406080D 00	-0.124608D 00
0.8190000D 00	0.7156530D 00	0.103347D 00
0.7410000D 00	0.5887980D 00	0.152202D 00
0.4810000D 00	0.4818820D 00	-0.881916D-03
0.7360000D 00	0.8344560D 00	-0.984561D-01
0.7580000D 00	0.7034930D 00	0.545073D-01
0.8210000D 00	0.6784450D 00	0.142555D 00
0.7640000D 00	0.6802360D 00	0.837637D-01
0.5410000D 00	0.4960630D 00	0.449374D-01
0.4780000D 00	0.5249960D 00	-0.469957D-01
0.6220000D 00	0.4690330D 00	0.152967D 00
0.5100000D 00	0.5333670D 00	-0.233671D-01
0.7090000D 00	0.8533690D 00	-0.144369D 00
0.6290000D 00	0.6227300D 00	0.627022D-02
0.4070000D 00	0.5167560D 00	-0.109756D 00
0.6540000D 00	0.7689640D 00	-0.114964D 00
0.8650000D 00	0.8578810D 00	0.711944D-02
0.5620000D 00	0.6469990D 00	-0.849985D-01
0.7550000D 00	0.6190460D 00	0.135954D 00
0.8440000D 00	0.8300670D 00	0.139326D-01
0.5680000D 00	0.5174370D 00	0.505626D-01
0.9820000D 00	0.8143510D 00	0.167649D 00
0.6000000D 00	0.6166030D 00	-0.166029D-01
0.6100000D 00	0.7435430D 00	-0.133543D 00
0.4240000D 00	0.5333670D 00	-0.109367D 00
0.6390000D 00	0.7044210D 00	-0.654207D-01
0.6950000D 00	0.7357430D 00	-0.407426D-01
0.5070000D 00	0.5641040D 00	-0.571040D-01
0.6570000D 00	0.5618770D 00	0.951227D-01
0.9930000D 00	0.8289740D 00	0.164026D 00
0.5760000D 00	0.6309940D 00	-0.549937D-01

REFERENCES

Gallant, A. R. (1968) "A note on the measurement of cost-quantity relationships in the aircraft industry," Journal of the American Statistical Association, 63. p. 1247-52.

Gallant, A. R. (1971) "Statistical inference for nonlinear regression models," Ph.D. dissertation, Iowa State University.

Gallant, A. R. (1973) "Inference for nonlinear models," Institute of Statistics Mimeo Series, No. 875, Raleigh.

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.

Hartley, H. O. (1961) "The modified Gauss-Newton method for the fitting of non-linear regression functions by least squares," Technometrics, 3. p. 269-80.

Hartley, H. O. and Booker, A. (1965) "Non-linear least squares estimation," The Annals of Mathematical Statistics, 36. p. 638-50.

Jennrich, R. L. (1969) "Asymptotic properties of non-linear least squares estimators," The Annals of Mathematical Statistics, 40. p. 633-43.

Malinvaud, E. (1970) "The consistency of nonlinear regressions," The Annals of Mathematical Statistics, 41. p. 956-69.

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 $Q(\theta) = SSE(\theta)$.

Conditions. There is a convex, bounded subset S of R^p and a θ_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) $Q(\theta_o) < \tilde{Q} = \inf\{Q(\theta): \theta$ a boundary point of $S\}$.
4) There does not exist θ', θ'' 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 $D_o' = [F'(\theta_o)F(\theta_o)]^{-1}F'(\theta_o)[y - f(\theta_o)]$.

Find λ_o which minimizes $Q(\theta_o + \lambda D_o)$ over $\Lambda_o = \{\lambda: 0 \le \lambda \le 1, \theta_o + \lambda D_o \in \bar{S}\}$.

1) Set $\theta_1 = \theta_o + \lambda_o D_o$.

Compute $D_1' = [F'(\theta_1)F(\theta_1)]^{-1}F'(\theta_1)[y - f(\theta_1)]$.

Find λ_1 which minimizes $Q(\theta_1 + \lambda D_1)$ over $\Lambda_1 = \{\lambda: 0 \le \lambda \le 1, \theta_1 + \lambda D_1 \in \bar{S}\}$.

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) θ_α is an interior point of S for $\alpha = 1, 2, \ldots$

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

3) $\nabla Q(\theta^*) = 0$.

<u>Proof</u>. Gallant (1971).

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 R^k from which the inputs x_t $(t = 1, 2, \ldots)$ are to be chosen.

<u>Definition</u>. Let $\mathcal{G}$ 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 μ_n on $(X, \mathcal{G})$ is defined by

$$\mu_n(A) = n^{-1} \Sigma_{t=1}^{n} I_A(x_t)$$

for each $A \in \mathcal{G}$.

<u>Definition</u>. A sequence of measures $\{\mu_n\}$ on $(X, \mathcal{G})$ is said to converge weakly to a measure μ on $(X, \mathcal{G})$ 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.

<u>Assumptions</u>. The parameter space Ω 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 μ

defined over $(X, \mathcal{G})$. The true value of θ, denoted by θ^o, is contained in an open set which, in turn, is contained in Ω. If $f(x,\theta) = f(x,\theta^o)$ except on a set of μ measure zero, it is assumed that $\theta = \theta^o$. The $p \times p$ matrix

$$\Sigma = \left[\int \frac{\partial}{\partial\theta_i} f(x,\theta^o) \frac{\partial}{\partial\theta_j} f(x,\theta^o) \, d\mu(x) \right]$$

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

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

1) The estimator $\hat{\theta}$ is consistent for θ^o.
2) The estimator $\hat{\sigma}^2$ is consistent for σ^2.
3) $n^{-1} F'(\hat{\theta}) F(\hat{\theta})$ is consistent for Σ.
4) $\sqrt{n}\,(\hat{\theta} - \theta^o)'$ is asymptotically normal with mean zero and variance-covariance matrix $\sigma^2 \Sigma^{-1}$.

<u>Proof</u>. Gallant (1971 or 1973).

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

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

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