

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 \tilde{x}_t are assumed to be generated according to the nonlinear regression model

$$y_t = f(\tilde{x}_t, \theta) + e_t \quad (t = 1, 2, \dots, n) .$$

The input variables are k-dimensional

$$\tilde{x}_t = (x_{1t}, x_{2t}, \dots, x_{kt})$$

and the unknown parameter θ is p-dimensional

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

The least squares estimator is that value

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

which minimizes

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

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

$$\underline{y} = (y_1, y_2, \dots, y_n)' \quad (n \times 1)$$

$$\underline{f}(\theta) = (f(x_1, \theta), f(x_2, \theta), \dots, f(x_n, \theta))' \quad (n \times 1)$$

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

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

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

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

Thus,

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

$$\theta = (\theta_1, \theta_2)$$

$$\underline{x}_t = x_t$$

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

$$\underline{y} = \begin{bmatrix} .824 \\ .515 \\ \vdots \\ .576 \end{bmatrix} \quad (50 \times 1)$$

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

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

The modified Gauss-Newton (Hartley, 1961) algorithm proceeds as follows.

Find a start value θ_0 ; methods of finding start values are discussed later.

0) From the starting estimate θ_0 compute

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

Find a λ_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

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

Find a λ_1 between 0 and 1 such that

$$SSE(\theta_1 + \lambda_1 D'_1) \leq SSE(\theta_1)$$

$$2) \text{ Let } \theta_2 = \theta_1 + \lambda_1 D_1$$

⋮

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 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 x_{t_i} ($i = 1, \dots, p$). Solve the set of nonlinear equations

$$y_{t_i} = f(x_{t_i}, \theta) \quad (i = 1, 2, \dots, p)$$

for θ .

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

$$\hat{\theta}_0 = (.444, .823)$$

There are several ways of choosing λ_i to satisfy $SSE(\hat{\theta}_i + \lambda_i D_i) < SSE(\hat{\theta}_i)$ 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

$$SSE(\hat{\theta}_i + \lambda_i D_i) < SSE(\hat{\theta}_i)$$

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

$$SSE(\hat{\theta}_i + \lambda D_i) < SSE(\hat{\theta}_i)$$

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

$$\beta_j = (\alpha)^j \quad j = 0, 1, \dots$$

and chose λ_i to be the largest β_j such that

$$SSE(\hat{\theta}_i + \beta_j D_i) < SSE(\hat{\theta}_i)$$

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

$$\|\varrho_i - \varrho_{i+1}\| \leq \epsilon \|\varrho_i\|$$

and simultaneously

$$|\text{SSE}(\varrho_i) - \text{SSE}(\varrho_{i+1})| \leq \epsilon |\text{SSE}(\varrho_i)|.$$

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

$$\text{SSE}(\varrho_i + \lambda \underline{D}_i) < \text{SSE}(\varrho_i).$$

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 ϱ_i and $\text{SSE}(\varrho_i)$ are printed out for each iteration until either this limit is reached or no λ can be found to improve ϱ_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

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

and

$$\hat{\Sigma} = [F'(\hat{\theta}) F(\hat{\theta})]^{-1} \hat{\sigma}^2$$

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

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 $\hat{\Sigma}$ rather than $\hat{\Sigma}^{-1}$. 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.

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 $Y_T = F(X_T, \theta) + \epsilon_T$.

USAGE

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

SUBROUTINES CALLED

FUNCT, DGMPRD, DSWEEEP

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 $F(X_T, \theta)$, 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 $F(X_T, \theta)$
INTEGER
- IP - NUMBER OF PARAMETERS IN THE MODEL $F(X_T, \theta)$
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 $L=0,1,2,\dots,40$.
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)
M0=1
M1=N           +M0
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(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,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	0.734936840 00	1	0.444000000 00
		2	0.823000000 00
1	0.454472620 00	1	0.447490530 00
		2	0.673350780 00
2	0.453567110 00	1	0.449330470 00
		2	0.659282920 00
3	0.453567080 00	1	0.449361500 00
		2	0.659161440 00
4	0.453567080 00	1	0.449361660 00
		2	0.659160910 00
5	0.453567080 00	1	0.449361660 00
		2	0.659160900 00
6	0.453567080 00	1	0.449361660 00
		2	0.659160900 00
7	0.453567080 00	1	0.449361660 00
		2	0.659160900 00
8	0.453567080 00	1	0.449361660 00
		2	0.659160900 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	0.2545280-01	17.6549
2	0.659161	0.6026480-01	8.2123

ESTIMATED VARIANCE 0.9071340-02

RESIDUAL SUM OF SQUARES 0.45356708

NUMBER OF OBSERVATIONS 50

NUMBER OF PARAMETERS 2

*
*
*
*
*

CORRELATION MATRIX OF THE PARAMETER ESTIMATES

1.00000 -0.93592
-0.93592 1.00000

 * OBSERVED RESPONSES, PREDICTED RESPONSES, RESIDUALS *
 * *****

Y	YHAT	RESIDUAL
0.824000 00	0.804215 00	0.197846D-01
0.515000 00	0.529514 00	-0.145137D-01
0.560000 00	0.641055 00	-0.810553D-01
0.949000 00	0.865834 00	0.831661D-01
0.495000 00	0.484111 00	0.108895D-01
0.874000 00	0.723241 00	0.150759D 00
0.452000 00	0.553058 00	-0.101058D 00
0.482000 00	0.582237 00	-0.100237D 00
0.660000 00	0.633494 00	-0.333494D-01
0.660000 00	0.660791 00	-0.790714D-03
0.874000 00	0.768964 00	0.105036D 00
0.534000 00	0.680236 00	-0.126236D 00
0.534000 00	0.597399 00	-0.63392D-01
0.775000 00	0.830067 00	-0.550674D-01
0.850000 00	0.710483 00	0.139517D 00
0.680000 00	0.757344 00	-0.593939D-01
0.622000 00	0.621909 00	0.906352D-04
0.626000 00	0.776605 00	-0.150605D 00
0.771000 00	0.640633 00	0.130367D 00
0.616000 00	0.740608 00	-0.124608D 00
0.819000 00	0.715653 00	0.103347D 00
0.741000 00	0.580798 00	0.152202D 00
0.481000 00	0.481828 00	-0.881916D-03
0.736000 00	0.834456 00	-0.984561D-01
0.758000 00	0.703473 00	0.545073D-01
0.821000 00	0.678445 00	0.142555D 00
0.764000 00	0.680236 00	0.837637D-01
0.541000 00	0.496630 00	0.449374D-01
0.478000 00	0.524976 00	-0.469957D-01
0.622000 00	0.469031 00	0.152967D 00
0.510000 00	0.531367 00	-0.233671D-01
0.769000 00	0.853369 00	-0.144369D 00
0.629000 00	0.622730 00	0.627022D-02
0.407000 00	0.515756 00	-0.109756D 00
0.654000 00	0.766968 00	-0.114964D 00
0.865000 00	0.857891 00	-0.711944D-02
0.562000 00	0.646999 00	-0.849985D-01
0.755000 00	0.619046 00	0.135954D 00
0.844000 00	0.830067 00	0.133226D-01
0.568000 00	0.517437 00	0.505626D-01
0.982000 00	0.814351 00	0.167649D 00
0.660000 00	0.616603 00	-0.166029D-01
0.610000 00	0.743543 00	-0.133543D 00
0.424000 00	0.533367 00	-0.109367D 00
0.639000 00	0.704421 00	-0.654207D-01
0.695000 00	0.735743 00	-0.407426D-01
0.507000 00	0.564104 00	-0.571040D-01
0.637000 00	0.618977 00	-0.951227D-01
0.993000 00	0.828974 00	0.164026D 00
0.576000 00	0.630994 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) ($t = 1, 2, \dots, n$). Let $Q(\theta) = \text{SSE}(\theta)$.

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 $t = 1, 2, \dots, n$.
- 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 θ', θ'' in S such that

$$\nabla Q(\theta') = \nabla Q(\theta'') = 0 \text{ and } Q(\theta') = Q(\theta'') .$$

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 λ_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 λ_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 $\theta_2 = \theta_1 + \lambda_1 D'_1$.

⋮

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

- 1) θ_α is an interior point of S for $\alpha = 1, 2, \dots$
- 2) The sequence $\{\theta_\alpha\}$ converges to a limit θ^* which is interior to S .
- 3) $\nabla Q(\theta^*) = 0$.

Proof. 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 ($t = 1, 2, \dots$) are to be chosen.

Definition. 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

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

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

Definition. 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

$$\int g(x) d\mu_n(x) \rightarrow \int g(x) d\mu(x)$$

as $n \rightarrow \infty$.

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

Assumptions. 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, G) . The true value of θ , denoted by θ^0 , is contained in an open set which, in turn, is contained in Ω . If $f(x, \theta) = f(x, \theta^0)$ except on a set of μ measure zero, it is assumed that $\theta = \theta^0$. The $p \times p$ matrix

$$\Sigma = \left[\int \frac{\partial}{\partial \theta_i} f(x, \theta^0) \frac{\partial}{\partial \theta_j} f(x, \theta^0) d\mu(x) \right]$$

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).

Theorem. 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^0 .
- 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) $\sqrt{n}(\hat{\theta} - \theta^0)'$ is asymptotically normal with mean zero and variance-covariance matrix $\sigma^2 \Sigma^{-1}$.

Proof. Gallant (1971 or 1973).

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

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

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