NAG Fortran Library Routine Document

G02HFF

Note: before using this routine, please read the Users' Note for your implementation to check the interpretation of ***bold italicised*** terms and other implementation-dependent details.

1 Purpose

G02HFF calculates an estimate of the asymptotic variance-covariance matrix for the bounded influence regression estimates (M-estimates). It is intended for use with G02HDF.

2 Specification

```
 SUBROUTINE G02HFF(PSI, PSP, INDW, INDC, SIGMA, N, M, X, IX, RS, WGT, C,
1                  IC, WK, IFAIL)
 INTEGER           INDW, INDC, N, M, IX, IC, IFAIL
 real              PSI, PSP, SIGMA, X(IX,M), RS(N), WGT(N), C(IC,M),
1                  WK(M*(N+M+1)+2*N)
 EXTERNAL          PSI, PSP
```

3 Description

For a description of bounded influence regression see G02HDF. Let θ be the regression parameters and let C be the asymptotic variance-covariance matrix of $\hat{\theta}$. Then for Huber type regression

$$C = f_H (X^T X)^{-1} \hat{\sigma}^2,$$

where

$$f_H = \frac{1}{n-m} \frac{\sum_{i=1}^n \psi^2 (r_i/\hat{\sigma})}{\left(\frac{1}{n} \sum \psi' \left(\frac{r_i}{\hat{\sigma}}\right)\right)^2} \kappa^2$$

$$\kappa^2 = 1 + \frac{m}{n} \frac{\frac{1}{n} \sum_{i=1}^n \left(\psi'(r_i/\hat{\sigma}) - \frac{1}{n} \sum_{i=1}^n \psi'(r_i/\hat{\sigma})\right)^2}{\left(\frac{1}{n} \sum_{i=1}^n \psi' \left(\frac{r_i}{\hat{\sigma}}\right)\right)^2},$$

see Huber (1981) and Marazzi (1987b).

For Mallows and Schweppe type regressions, C is of the form

$$\frac{\hat{\sigma}^2}{n} S_1^{-1} S_2 S_1^{-1},$$

where $S_1 = \frac{1}{n} X^T D X$ and $S_2 = \frac{1}{n} X^T P X$.

D is a diagonal matrix such that the ith element approximates $E(\psi'(r_i/(\sigma w_i)))$ in the Schweppe case and $E(\psi'(r_i/\sigma) w_i)$ in the Mallows case.

P is a diagonal matrix such that the ith element approximates $E(\psi^2(r_i/(\sigma w_i)) w_i^2)$ in the Schweppe case and $E(\psi^2(r_i/\sigma) w_i^2)$ in the Mallows case.

Two approximations are available in G02HFF:

1. Average over the r_i

Schweppe	Mallows
$D_i = \left(\frac{1}{n} \sum_{j=1}^n \psi' \left(\frac{r_j}{\hat{\sigma} w_i}\right)\right) w_i$	$D_i = \left(\frac{1}{n} \sum_{j=1}^n \psi' \left(\frac{r_j}{\hat{\sigma}}\right)\right) w_i$
$P_i = \left(\frac{1}{n} \sum_{j=1}^n \psi^2 \left(r \frac{j}{\hat{\sigma} w_i}\right)\right) w_i^2$	$P_i = \left(\frac{1}{n} \sum_{j=1}^n \psi^2 \left(\frac{r_j}{\hat{\sigma}}\right)\right) w_i^2$

2. Replace expected value by observed

$$\begin{array}{cc} \text{Schweppe} & \text{Mallows} \\ D_i = \psi'\left(\frac{r_i}{\hat{\sigma} w_i}\right) w_i & D_i = \psi'\left(\frac{r_i}{\hat{\sigma}}\right) w_i \\ P_i = \psi^2\left(\frac{r_i}{\hat{\sigma} w_i}\right) w_i^2 & P_i = \psi^2\left(\frac{r_i}{\hat{\sigma}}\right) w_i^2 \end{array}$$

See Hampel *et al.* (1986) and Marazzi (1987b).

In all cases $\hat{\sigma}$ is a robust estimate of σ.

G02HFF is based on routines in ROBETH; see Marazzi (1987b).

4 References

Hampel F R, Ronchetti E M, Rousseeuw P J and Stahel W A (1986) *Robust Statistics. The Approach Based on Influence Functions* Wiley

Huber P J (1981) *Robust Statistics* Wiley

Marazzi A (1987b) Subroutines for robust and bounded influence regression in ROBETH *Cah. Rech. Doc. IUMSP, No. 3 ROB 2* Institut Universitaire de Médecine Sociale et Préventive, Lausanne

5 Parameters

1: PSI – ***real*** FUNCTION, supplied by the user. *External Procedure*

PSI must return the value of the ψ function for a given value of its argument.

Its specification is:

```
real FUNCTION PSI(T)
real              T
```

1: T – ***real*** *Input*

On entry: the argument for which PSI must be evaluated.

PSI must be declared as EXTERNAL in the (sub)program from which G02HFF is called. Parameters denoted as *Input* must **not** be changed by this procedure.

2: PSP – ***real*** FUNCTION, supplied by the user. *External Procedure*

PSP must return the value of $\psi'(t) = \frac{d}{dt}\psi(t)$ for a given value of its argument.

Its specification is:

```
real FUNCTION PSP(T)
real              T
```

1: T – ***real*** *Input*

On entry: the argument for which PSP must be evaluated.

PSP must be declared as EXTERNAL in the (sub)program from which G02HFF is called. Parameters denoted as *Input* must **not** be changed by this procedure.

3: INDW – INTEGER *Input*

On entry: the type of regression for which the asymptotic variance-covariance matrix is to be calculated.

If INDW $= 0$, Huber type regression.

If INDW < 0, Mallows type regression.

If INDW > 0, Schweppe type regression.

4: INDC – INTEGER *Input*

On entry: if INDW $\neq 0$, INDC must specify the approximation to be used.

If INDC $= 1$, averaging over residuals.

If INDC $\neq 1$, replacing expected by observed.

If INDW $= 0$, INDC is not referenced.

5: SIGMA – ***real*** *Input*

On entry: the value of $\hat{\sigma}$, as given by G02HDF.

Constraint: SIGMA > 0.

6: N – INTEGER *Input*

On entry: the number, n, of observations.

Constraint: N > 1.

7: M – INTEGER *Input*

On entry: the number, m, of independent variables.

Constraint: $1 \leq$ M $<$ N.

8: X(IX,M) – ***real*** array *Input*

On entry: the values of the X matrix, i.e., the independent variables. X(i, j) must contain the ijth element of X, for $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, m$.

9: IX – INTEGER *Input*

On entry: the first dimension of the array X as declared in the (sub)program from which G02HFF is called.

Constraint: IX $\geq$ N.

10: RS(N) – ***real*** array *Input*

On entry: the residuals from the bounded influence regression. These are given by G02HDF.

11: WGT(N) – ***real*** array *Input*

On entry: if INDW $\neq 0$, WGT must contain the vector of weights used by the bounded influence regression. These should be used with G02HDF.

If INDW $= 0$, WGT is not referenced.

Constraint: if INDW $\neq 0$, WGT$(i) \geq 0.0$, for $i = 1, 2, \ldots n$.

12: C(IC,M) – ***real*** array *Output*

On exit: the estimate of the variance-covariance matrix.

13: IC – INTEGER *Input*

On entry: the first dimension of the array C as declared in the (sub)program from which G02HFF is called.

Constraint: IC $\geq$ M.

14: WK(M*(N+M+1)+2*N) – ***real*** array *Output*

On exit: if INDW $\neq 0$, WK(i), for $i = 1, 2, \ldots, n$, will contain the diagonal elements of the matrix D and WK(i), for $i = n+1, n+2, \ldots, 2n$, will contain the diagonal elements of matrix P.

The rest of the array is used as workspace.

15: IFAIL – INTEGER *Input/Output*

On entry: IFAIL must be set to 0, -1 or 1. Users who are unfamiliar with this parameter should refer to Chapter P01 for details.

On exit: IFAIL $= 0$ unless the routine detects an error (see Section 6).

For environments where it might be inappropriate to halt program execution when an error is detected, the value -1 or 1 is recommended. If the output of error messages is undesirable, then the value 1 is recommended. Otherwise, for users not familiar with this parameter the recommended value is 0. **When the value -1 or 1 is used it is essential to test the value of IFAIL on exit.**

6 Error Indicators and Warnings

If on entry IFAIL $= 0$ or -1, explanatory error messages are output on the current error message unit (as defined by X04AAF).

Errors or warnings detected by the routine:

IFAIL $= 1$

On entry, N ≤ 1,
or M < 1,
or N $\leq$ M,
or IC $<$ M,
or IX $<$ N.

IFAIL $= 2$

On entry, SIGMA ≤ 0.0,
or INDW $\neq 0$ and WGT(i) < 0.0 for some $i = 1, 2, \ldots n$.

IFAIL $= 3$

If INDW $= 0$ then the matrix $X^T X$ is either not positive-definite, possibly due to rounding errors, or is ill-conditioned.

If INDW $\neq 0$ then the matrix S_1 is singular or almost singular. This may be due to many elements of D being zero.

IFAIL $= 4$

Either the value of $\frac{1}{n}\sum_{i=1}^{n} \psi'\left(\frac{r_i}{\hat{\sigma}}\right) = 0$,

or $\kappa = 0$,

or $\sum_{i=1}^{n} \psi^2\left(\frac{r_i}{\hat{\sigma}}\right) = 0$.

In this situation G02HFF returns C as $(X^T X)^{-1}$.

7 Accuracy

In general, the accuracy of the variance-covariance matrix will depend primarily on the accuracy of the results from G02HDF.

8 Further Comments

This routine is only for situations in which X has full column rank.

Care has to be taken in the choice of the ψ function since if $\psi'(t) = 0$ for too wide a range then either the value of f_H will not exist or too many values of D_i will be zero and it will not be possible to calculate C.

9 Example

The asymptotic variance-covariance matrix is calculated for a Schweppe type regression. The values of X, $\hat{\sigma}$ and the residuals and weights are read in. The averaging over residuals approximation is used.

9.1 Program Text

Note: the listing of the example program presented below uses ***bold italicised*** terms to denote precision-dependent details. Please read the Users' Note for your implementation to check the interpretation of these terms. As explained in the Essential Introduction to this manual, the results produced may not be identical for all implementations.

```
*     G02HFF Example Program Text
*     Mark 14 Revised.  NAG Copyright 1989.
*     .. Parameters ..
      INTEGER          NIN, NOUT
      PARAMETER        (NIN=5,NOUT=6)
      INTEGER          NMAX, MMAX
      PARAMETER        (NMAX=5,MMAX=3)
*     .. Local Scalars ..
      real             SIGMA
      INTEGER          I, IC, IFAIL, INDC, INDW, IX, J, K, M, N
*     .. Local Arrays ..
      real             C(MMAX,MMAX), RS(NMAX), WGT(NMAX),
     +                 WK(MMAX*(NMAX+MMAX+1)+2*NMAX), X(NMAX,MMAX)
*     .. External Functions ..
      real             PSI, PSP
      EXTERNAL         PSI, PSP
*     .. External Subroutines ..
      EXTERNAL         G02HFF
*     .. Executable Statements ..
      WRITE (NOUT,*) 'G02HFF Example Program Results'
*     Skip heading in data file
      READ (NIN,*)
*     Read in the dimensions of X
      READ (NIN,*) N, M
      WRITE (NOUT,*)
      IF (N.GT.0 .AND. N.LE.NMAX .AND. M.GT.0 .AND. M.LE.MMAX) THEN
*        Read in the X matrix
         DO 20 I = 1, N
            READ (NIN,*) (X(I,J),J=1,M)
   20    CONTINUE
*        Read in SIGMA
         READ (NIN,*) SIGMA
*        Read in weights and residuals
         DO 40 I = 1, N
            READ (NIN,*) WGT(I), RS(I)
   40    CONTINUE
*        Set other parameter values
         IX = NMAX
         IC = MMAX
*        Set parameters for Schweppe type regression
         INDW = 1
         INDC = 1
         IFAIL = 0
*
         CALL G02HFF(PSI,PSP,INDW,INDC,SIGMA,N,M,X,IX,RS,WGT,C,IC,WK,
     +               IFAIL)
*
         WRITE (NOUT,*) 'Covariance matrix'
         DO 60 J = 1, M
            WRITE (NOUT,99999) (C(J,K),K=1,M)
   60    CONTINUE
```

```
      END IF
      STOP
*
99999 FORMAT (1X,6F10.4)
      END
*
      real  FUNCTION PSI(T)
*     .. Parameters ..
      real              C
      PARAMETER         (C=1.5e0)
*     .. Scalar Arguments ..
      real              T
*     .. Intrinsic Functions ..
      INTRINSIC         ABS
*     .. Executable Statements ..
      IF (T.LE.-C) THEN
         PSI = -C
      ELSE IF (ABS(T).LT.C) THEN
         PSI = T
      ELSE
         PSI = C
      END IF
      RETURN
      END
*
      real  FUNCTION PSP(T)
*     .. Parameters ..
      real              C
      PARAMETER         (C=1.5e0)
*     .. Scalar Arguments ..
      real              T
*     .. Intrinsic Functions ..
      INTRINSIC         ABS
*     .. Executable Statements ..
      PSP = 0.0e0
      IF (ABS(T).LT.C) PSP = 1.0e0
      RETURN
      END
```

9.2 Program Data

```
G02HFF Example Program Data

    5    3              : N  M

  1.0 -1.0 -1.0         : X1  X2  X3
  1.0 -1.0  1.0
  1.0  1.0 -1.0
  1.0  1.0  1.0
  1.0  0.0  3.0         : End of X1 X2 and X3 values

    20.7783             : SIGMA

    0.4039    0.5643    : Weights and residuals, WGT and RS
    0.5012   -1.1286
    0.4039    0.5643
    0.5012   -1.1286
    0.3862    1.1286    : End of weights and residuals
```

9.3 Program Results

```
 G02HFF Example Program Results

 Covariance matrix
     0.2070    0.0000   -0.0478
     0.0000    0.2229   -0.0000
    -0.0478   -0.0000    0.0796
```