NAG Toolbox for MATLAB

g02hf

1 Purpose

g02hf calculates an estimate of the asymptotic variance-covariance matrix for the bounded influence regression estimates (M-estimates). It is intended for use with g02hd.

2 Syntax

```
[c, wk, ifail] = g02hf(psi, psp, indw, indc, sigma, x, rs, wgt, 'n', n,
'm', m)
```

3 Description

For a description of bounded influence regression see g02hd. Let θ be the regression parameters and let C be the asymptotic variance-covariance matrix of $\hat{\theta}$. Then for Huber type regression

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

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

$$\frac{\hat{\sigma}^2}{n}S_1^{-1}S_2S_1^{-1},$$

where $S_1 = \frac{1}{n}X^{\mathrm{T}}DX$ and $S_2 = \frac{1}{n}X^{\mathrm{T}}PX$.

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

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

Two approximations are available in g02hf:

1. Average over the r_i

Schweppe

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

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

Mallows

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

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

2. Replace expected value by observed

Schweppe

$$D_i = \psi'\left(\frac{r_i}{\hat{\sigma}w_i}\right)w_i$$

$$P_i = \psi^2\left(\frac{r_i}{\hat{\sigma}w_i}\right)w_i^2$$

Mallows

$$D_i = \psi'\left(\frac{r_i}{\hat{\sigma}}\right)w_i$$

$$P_i = \psi^2\left(\frac{r_i}{\hat{\sigma}}\right)w_i^2$$

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

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

g02hf 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

5.1 Compulsory Input Parameters

1: **psi – string containing name of m-file**

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

Its specification is:

```
[result] = psi(t)
```

Input Parameters

1: **t – double scalar**

The argument for which **psi** must be evaluated.

Output Parameters

1: **result – double scalar**

The result of the function.

2: **psp – string containing name of m-file**

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

Its specification is:

```
[result] = psp(t)
```

Input Parameters

1: **t – double scalar**

The argument for which **psp** must be evaluated.

Output Parameters

1: **result – double scalar**

The result of the function.

3: **indw – int32 scalar**

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

indw < 0

Mallows type regression.

indw $= 0$

Huber type regression.

indw > 0

Schweppe type regression.

4: **indc – int32 scalar**

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.

Constraint: **indc** $= 1$ or 0.

5: **sigma – double scalar**

The value of $\hat{\sigma}$, as given by g02hd.

Constraint: **sigma** > 0.

6: **x(ldx,m) – double array**

ldx, the first dimension of the array, must be at least **n**.

The values of the X matrix, i.e., the independent variables. $\mathbf{x}(i,j)$ must contain the ijth element of X, for $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, m$.

7: **rs(n) – double array**

The residuals from the bounded influence regression. These are given by g02hd.

8: **wgt(n) – double array**

If **indw** $\neq 0$, **wgt** must contain the vector of weights used by the bounded influence regression. These should be used with g02hd.

If **indw** $= 0$, **wgt** is not referenced.

Constraint: **wgt**$(i) \geq 0.0$ if **indw** $\neq 0$, for $i = 1, 2, \ldots,$ Ai.

5.2 Optional Input Parameters

1: **n – int32 scalar**

Default: The dimension of the arrays **rs**, **wgt**. (An error is raised if these dimensions are not equal.)

n, the number of observations.

Constraint: $\mathbf{n} > 1$.

2: **m – int32 scalar**

Default: The dimension of the arrays **x**, **c**. (An error is raised if these dimensions are not equal.)

m, the number of independent variables.

Constraint: $1 \leq \mathbf{m} < \mathbf{n}$.

5.3 Input Parameters Omitted from the MATLAB Interface

ldx, ldc

5.4 Output Parameters

1: **c(ldc,m) – double array**

The estimate of the variance-covariance matrix.

2: **wk($\mathbf{m} \times (\mathbf{n} + \mathbf{m} + \mathbf{1}) + \mathbf{2} \times \mathbf{n}$) – double array**

If $\mathbf{indw} \neq 0$, $\mathbf{wk}(i)$, for $i = 1, 2, \ldots, n$, will contain the diagonal elements of the matrix D and $\mathbf{wk}(i)$, 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.

3: **ifail – int32 scalar**

0 unless the function detects an error (see Section 6).

6 Error Indicators and Warnings

Errors or warnings detected by the function:

$\mathbf{ifail} = 1$

On entry, $\mathbf{n} \leq 1$,
or $\mathbf{m} < 1$,
or $\mathbf{n} \leq \mathbf{m}$,
or $\mathbf{ldc} < \mathbf{m}$,
or $\mathbf{ldx} < \mathbf{n}$.

$\mathbf{ifail} = 2$

On entry, $\mathbf{sigma} \leq 0.0$,
or $\mathbf{indw} \neq 0$ and $\mathbf{wgt}(i) < 0.0$ for some $i = 1, 2, \ldots n$.

$\mathbf{ifail} = 3$

If $\mathbf{indw} = 0$ then the matrix $X^{\mathrm{T}}X$ is either not positive-definite, possibly due to rounding errors, or is ill-conditioned.

If $\mathbf{indw} \neq 0$ 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 g02hf returns C as $\left(X^{\mathrm{T}}X\right)^{-1}$.

7 Accuracy

In general, the accuracy of the variance-covariance matrix will depend primarily on the accuracy of the results from g02hd.

8 Further Comments

g02hf 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

g02hf_psi.m

```
function [result] = psi(t)
  if t < -1.5
    result = -1.5;
  elseif abs(t) < 1.5
    result = t;
  else
    result = 1.5;
  end;
```

g02hf_psp.m

```
function [result] = psp(t)
  if (abs(t) < 1.5)
    result=1;
  else
    result=0;
  end
```

```
indw = int32(1);
indc = int32(1);
sigma = 20.7783;
x = [1, -1, -1;
     1, -1, 1;
     1, 1, -1;
     1, 1, 1;
     1, 0, 3];
rs = [0.5643;
     -1.1286;
     0.5643;
     -1.1286;
     1.1286];
wgt = [0.4039;
```

```
     0.5012;
     0.4039;
     0.5012;
     0.3862];
[c, wk, ifail] = g02hf('g02hf_psi', 'g02hf_psp', indw, indc, sigma, x,
rs, wgt)
```

```
c =
    0.2070         0   -0.0478
         0    0.2229         0
   -0.0478         0    0.0796
wk =
    1.0000
    1.0000
    1.0000
    1.0000
    1.0000
    0.0021
    0.0021
    0.0021
    0.0021
    0.0021
    1.1607
         0
   -0.2679
         0
    1.2500
         0
   -0.2679
         0
    0.4464
         0
   -1.0000
    1.0000
    0.0021
    0.0021
         0
    0.0000
         0
    0.0021
         0
    0.0000
         0
    0.0021
   -0.0021
    0.0021
   -0.0021
    0.0021
    0.0062
ifail =
           0
```