nag_robust_m_corr_user_fn (g02hlc)

1 Purpose

nag_robust_m_corr_user_fn (g02hlc) calculates a robust estimate of the covariance matrix for user-supplied weight functions and their derivatives.

2 Specification

```c
#include <nag.h>
#include <nagg02.h>

void nag_robust_m_corr_user_fn (Nag_OrderType order,
   void (*ucv)(double t, double *u, double *ud, double *w, double *wd,
   Nag_Comm *comm),
   Integer indm, Integer n, Integer m, const double x[],
   Integer pdx, double cov[], double a[], double wt[],
   double theta[], double bl, double bd, Integer maxit,
   Integer nitmon, const char *outfile,
   double tol, Integer *nit, Nag_Comm *comm, NagError *fail)
```

3 Description

For a set of \( n \) observations on \( m \) variables in a matrix \( X \), a robust estimate of the covariance matrix, \( C \), and a robust estimate of location, \( \theta \), are given by:

\[
C = \tau^2 (A^T A)^{-1},
\]

where \( \tau^2 \) is a correction factor and \( A \) is a lower triangular matrix found as the solution to the following equations.

\[
\frac{1}{n} \sum_{i=1}^{n} w(\|z_i\|_2) z_i = 0
\]

and

\[
\frac{1}{n} \sum_{i=1}^{n} u(\|z_i\|_2) z_i z_i^T - v(\|z_i\|_2) I = 0,
\]

where \( x_i \) is a vector of length \( m \) containing the elements of the \( i \)th row of \( X \),

\( z_i \) is a vector of length \( m \),

\( I \) is the identity matrix and \( 0 \) is the zero matrix,

and \( w \) and \( u \) are suitable functions.

nag_robust_m_corr_user_fn (g02hlc) covers two situations:

(i) \( v(t) = 1 \) for all \( t \),

(ii) \( v(t) = u(t) \).

The robust covariance matrix may be calculated from a weighted sum of squares and cross-products matrix about \( \theta \) using weights \( wt_i = u(||z_i||) \). In case (i) a divisor of \( n \) is used and in case (ii) a divisor of
\[ \sum_{i=1}^{n} w t_i \text{ is used. If } w(\cdot) = \sqrt{u(\cdot)}, \text{ then the robust covariance matrix can be calculated by scaling each row of } X \text{ by } \sqrt{w t_i} \text{ and calculating an unweighted covariance matrix about } \theta. \]

In order to make the estimate asymptotically unbiased under a Normal model a correction factor, \( \tau^2 \), is needed. The value of the correction factor will depend on the functions employed (see Huber (1981) and Marazzi (1987)).

\[ \text{nag}_\text{robust}_m_\text{corr}_\text{user}_\text{fn} (g02hlc) \text{ finds } A \text{ using the iterative procedure as given by Huber.} \]

\[ A_k = (S_k + I)A_{k-1} \]

and

\[ \theta_{ji} = \frac{b_j}{D_k}, \quad \theta_{ji-1}, \]

where \( S_k = (s_{jl}) \), for \( j, i = 1, 2, \ldots, m \) and \( l = 1, 2, \ldots, m \), is a lower triangular matrix such that:

\[ s_{jl} = \begin{cases} -\min\left[ \max\left( \frac{h_{jl}}{D_3}, -BL \right), BL \right], & j > l \\ -\min\left[ \max\left( \frac{h_{jj}}{(2D_3 - D_4)} \right), -BD \right], & j = l \end{cases} \]

where

\[ D_1 = \sum_{i=1}^{n} \left\{ w(\|z_i\|_2) + \frac{1}{m} u'(\|z_i\|_2)\|z_i\|_2 \right\} \]

\[ D_2 = \sum_{i=1}^{n} \left\{ \frac{1}{2}(u'(\|z_i\|_2)\|z_i\|_2 + 2u(\|z_i\|_2)\|z_i\|_2 - u'(\|z_i\|_2)) \right\} \|z_i\|_2 \]

\[ D_3 = \frac{1}{m+2} \sum_{i=1}^{n} \left\{ \frac{1}{m}(u'(\|z_i\|_2)\|z_i\|_2 + 2u(\|z_i\|_2)) + u(\|z_i\|_2) \right\} \|z_i\|_2^2 \]

\[ D_4 = \sum_{i=1}^{n} \left\{ \frac{1}{m}u(\|z_i\|_2)\|z_i\|_2^2 - u'(\|z_i\|_2) \right\} \]

\[ h_{jl} = \sum_{i=1}^{n} u(\|z_i\|_2)z_{ij}z_{il}, \text{ for } j > l \]

\[ h_{jj} = \sum_{i=1}^{n} u(\|z_i\|_2)\left( z_{ij}^2 - \|z_{ij}\|_2^2 / m \right) \]

\[ b_j = \sum_{i=1}^{n} w(\|z_i\|_2)(x_{ij} - b_i) \]

and \( BD \) and \( BL \) are suitable bounds.

\text{nag}_\text{robust}_m_\text{corr}_\text{user}_\text{fn} (g02hlc) \text{ is based on routines in ROBETH; see Marazzi (1987).}

4 References


Arguments

1: order – Nag_OrderType

On entry: the order argument specifies the two-dimensional storage scheme being used, i.e., row-major ordering or column-major ordering. C language defined storage is specified by order = Nag_RowMajor. See Section 3.2.1.3 in the Essential Introduction for a more detailed explanation of the use of this argument.

Constraint: order = Nag_RowMajor or Nag_ColMajor.

2: ucv – function, supplied by the user

ucv must return the values of the functions $u$ and $w$ and their derivatives for a given value of its argument.

The specification of ucv is:

```c
void ucv (double t, double *u, double *ud, double *w, double *wd,
          Nag_Comm *comm)
```

1: t – double

On entry: the argument for which the functions $u$ and $w$ must be evaluated.

2: u – double *

On exit: the value of the $u$ function at the point $t$.

Constraint: $u \geq 0.0$.

3: ud – double *

On exit: the value of the derivative of the $u$ function at the point $t$.

4: w – double *

On exit: the value of the $w$ function at the point $t$.

Constraint: $w \geq 0.0$.

5: wd – double *

On exit: the value of the derivative of the $w$ function at the point $t$.

6: comm – Nag_Comm *

Pointer to structure of type Nag_Comm; the following members are relevant to ucv.

user – double *
iuser – Integer *
p – Pointer

The type Pointer will be void *. Before calling nag_robust_m_corr_user_fn (g02hlc) you may allocate memory and initialize these pointers with various quantities for use by ucv when called from nag_robust_m_corr_user_fn (g02hlc) (see Section 3.2.1.1 in the Essential Introduction).

3: indm – Integer

On entry: indicates which form of the function $v$ will be used.

indm = 1

$v = 1$. 
\begin{verbatim}

\textbf{g02hlc} \hspace{1cm} \textit{NAG Library Manual}

\begin{align*}
\text{indm} \neq 1 \\
v = u.
\end{align*}

4: \hspace{1cm} \textbf{n} \quad \text{Integer} \quad \textit{Input}

\textit{On entry:} n, the number of observations.

\textit{Constraint:} n > 1.

5: \hspace{1cm} \textbf{m} \quad \text{Integer} \quad \textit{Input}

\textit{On entry:} m, the number of columns of the matrix X, i.e., number of independent variables.

\textit{Constraint:} 1 \leq m \leq n.

6: \hspace{1cm} x[\text{dim}] \quad \text{const double} \quad \textit{Input}

\textit{Note:} the dimension, \text{dim}, of the array \text{x} must be at least
\begin{align*}
\max(1, \text{pdx} \times m) & \text{ when order} = \text{Nag\_ColMajor}; \\
\max(1, n \times \text{pdx}) & \text{ when order} = \text{Nag\_RowMajor}.
\end{align*}

Where \( X(i, j) \) appears in this document, it refers to the array element
\begin{align*}
x[(j-1) \times \text{pdx} + i - 1] & \text{ when order} = \text{Nag\_ColMajor}; \\
x[(i-1) \times \text{pdx} + j - 1] & \text{ when order} = \text{Nag\_RowMajor}.
\end{align*}

\textit{On entry:} \( X(i, j) \) must contain the \( i \)th observation on the \( j \)th variable, for \( i = 1, 2, \ldots, n \) and \( j = 1, 2, \ldots, m \).

7: \hspace{1cm} \textbf{pdx} \quad \text{Integer} \quad \textit{Input}

\textit{On entry:} the stride separating row or column elements (depending on the value of \text{order}) in the array \text{x}.

\textit{Constraints:}
\begin{align*}
\text{if order} = \text{Nag\_ColMajor}, \text{pdx} \geq n; \\
\text{if order} = \text{Nag\_RowMajor}, \text{pdx} \geq m.
\end{align*}

8: \hspace{1cm} \textbf{cov}[m \times (m + 1)/2] \quad \text{double} \quad \textit{Output}

\textit{On exit:} contains a robust estimate of the covariance matrix, \( C \). The upper triangular part of the matrix \( C \) is stored packed by columns (lower triangular stored by rows), \( C_{ij} \) is returned in \( \text{cov}[(j \times (j-1)/2 + i) - 1], i \leq j \).

9: \hspace{1cm} \textbf{a}[m \times (m + 1)/2] \quad \text{double} \quad \textit{Input/Output}

\textit{On entry:} an initial estimate of the lower triangular real matrix \( A \). Only the lower triangular elements must be given and these should be stored row-wise in the array.

The diagonal elements must be \( \neq 0 \), and in practice will usually be \( > 0 \). If the magnitudes of the columns of \( X \) are of the same order, the identity matrix will often provide a suitable initial value for \( A \). If the columns of \( X \) are of different magnitudes, the diagonal elements of the initial value of \( A \) should be approximately inversely proportional to the magnitude of the columns of \( X \).

\textit{Constraint:} \( a[j \times (j-1)/2 + j] \neq 0.0 \), for \( j = 0, 1, \ldots, m - 1 \).

\textit{On exit:} the lower triangular elements of the inverse of the matrix \( A \), stored row-wise.

10: \hspace{1cm} \textbf{wt}[n] \quad \text{double} \quad \textit{Output}

\textit{On exit:} \( \text{wt}[i-1] \) contains the weights, \( wt_i = u(\|z_i\|_2) \), for \( i = 1, 2, \ldots, n \).

11: \hspace{1cm} \textbf{theta}[m] \quad \text{double} \quad \text{Input/Output}

\textit{On entry:} an initial estimate of the location argument, \( \theta_j \), for \( j = 1, 2, \ldots, m \).
\end{verbatim}
In many cases an initial estimate of \( \theta_j = 0 \), for \( j = 1, 2, \ldots, m \), will be adequate. Alternatively medians may be used as given by \texttt{nag_median_1var} (g07dac).

\textit{On exit}: contains the robust estimate of the location argument, \( \theta_j \), for \( j = 1, 2, \ldots, m \).

12: \textbf{bl} – double \hspace{10cm} \textit{Input}

\textit{On entry}: the magnitude of the bound for the off-diagonal elements of \( S_k \), \( BL \).

\textit{Suggested value}: \texttt{bl} = 0.9.

\textit{Constraint}: \texttt{bl} > 0.0.

13: \textbf{bd} – double \hspace{10cm} \textit{Input}

\textit{On entry}: the magnitude of the bound for the diagonal elements of \( S_k \), \( BD \).

\textit{Suggested value}: \texttt{bd} = 0.9.

\textit{Constraint}: \texttt{bd} > 0.0.

14: \textbf{maxit} – Integer \hspace{10cm} \textit{Input}

\textit{On entry}: the maximum number of iterations that will be used during the calculation of \( A \).

\textit{Suggested value}: \texttt{maxit} = 150.

\textit{Constraint}: \texttt{maxit} > 0.

15: \textbf{nitmon} – Integer \hspace{10cm} \textit{Input}

\textit{On entry}: indicates the amount of information on the iteration that is printed.

\texttt{nitmon} > 0

The value of \( A \), \( \theta \) and \( \delta \) (see Section 7) will be printed at the first and every \texttt{nitmon} iterations.

\texttt{nitmon} \leq 0

No iteration monitoring is printed.

16: \textbf{outfile} – const char * \hspace{10cm} \textit{Input}

\textit{On entry}: a null terminated character string giving the name of the file to which results should be printed. If \texttt{outfile} = \texttt{NULL} or an empty string then the \texttt{stdout} stream is used. Note that the file will be opened in the append mode.

17: \textbf{tol} – double \hspace{10cm} \textit{Input}

\textit{On entry}: the relative precision for the final estimates of the covariance matrix. Iteration will stop when maximum \( \delta \) (see Section 7) is less than \texttt{tol}.

\textit{Constraint}: \texttt{tol} > 0.0.

18: \textbf{nit} – Integer * \hspace{10cm} \textit{Output}

\textit{On exit}: the number of iterations performed.

19: \textbf{comm} – Nag_Comm * \hspace{10cm} \textit{Input/Output}

The NAG communication argument (see Section 3.2.1.1 in the Essential Introduction).

20: \textbf{fail} – NagError * \hspace{10cm} \textit{Input/Output}

The NAG error argument (see Section 3.6 in the Essential Introduction).
6 Error Indicators and Warnings

NE_ALLOC_FAIL
Dynamic memory allocation failed.
See Section 3.2.1.2 in the Essential Introduction for further information.

NE_BAD_PARAM
On entry, argument \langle value \rangle had an illegal value.

NE_CONST_COL
Column \langle value \rangle of \mathbf{x} has constant value.

NE_CONVERGENCE
Iterations to calculate weights failed to converge.

NE_FUN_RET_VAL
\( u \) value returned by \texttt{ucv} < 0.0: \( u(\langle value \rangle) = \langle value \rangle \).
\( w \) value returned by \texttt{ucv} < 0.0: \( w(\langle value \rangle) = \langle value \rangle \).

NE_INT
On entry, \( m = \langle value \rangle \).
Constraint: \( m \geq 1 \).

On entry, \( \texttt{maxit} = \langle value \rangle \).
Constraint: \( \texttt{maxit} > 0 \).

On entry, \( n = \langle value \rangle \).
Constraint: \( n > 1 \).

On entry, \( \texttt{pdx} = \langle value \rangle \).
Constraint: \( \texttt{pdx} > 0 \).

NE_INT_2
On entry, \( m = \langle value \rangle \) and \( n = \langle value \rangle \).
Constraint: \( 1 \leq m \leq n \).

On entry, \( n = \langle value \rangle \) and \( m = \langle value \rangle \).
Constraint: \( n \geq m \).

On entry, \( \texttt{pdx} = \langle value \rangle \) and \( m = \langle value \rangle \).
Constraint: \( \texttt{pdx} \geq m \).

On entry, \( \texttt{pdx} = \langle value \rangle \) and \( n = \langle value \rangle \).
Constraint: \( \texttt{pdx} \geq n \).

NE_INTERNAL_ERROR
An internal error has occurred in this function. Check the function call and any array sizes. If the call is correct then please contact NAG for assistance.

An unexpected error has been triggered by this function. Please contact NAG.
See Section 3.6.6 in the Essential Introduction for further information.

NE_NO_LICENCE
Your licence key may have expired or may not have been installed correctly.
See Section 3.6.5 in the Essential Introduction for further information.
NE_NOT_CLOSE_FILE
Cannot close file ⟨value⟩.

NE_NOT_WRITE_FILE
Cannot open file ⟨value⟩ for writing.

NE_REAL
On entry, bd = ⟨value⟩.
Constraint: bd > 0.0.
On entry, bl = ⟨value⟩.
Constraint: bl > 0.0.
On entry, tol = ⟨value⟩.
Constraint: tol > 0.0.

NE_ZERO_DIAGONAL
On entry, diagonal element ⟨value⟩ of a is 0.0.

NE_ZERO_SUM
The sum D1 is zero.
The sum D2 is zero.
The sum D3 is zero.

7 Accuracy
On successful exit the accuracy of the results is related to the value of tol; see Section 5. At an iteration let
(i) \( d_1 \) = the maximum value of \( |s_{ij}| \)
(ii) \( d_2 \) = the maximum absolute change in \( wt(i) \)
(iii) \( d_3 \) = the maximum absolute relative change in \( \theta_j \)
and let \( \delta = \max(d_1, d_2, d_3) \). Then the iterative procedure is assumed to have converged when \( \delta < \text{tol} \).

8 Parallelism and Performance
nag_robust_m_corr_user_fn (g02hlc) is not threaded by NAG in any implementation.

nag_robust_m_corr_user_fn (g02hlc) makes calls to BLAS and/or LAPACK routines, which may be threaded within the vendor library used by this implementation. Consult the documentation for the vendor library for further information.

Please consult the X06 Chapter Introduction for information on how to control and interrogate the OpenMP environment used within this function. Please also consult the Users’ Note for your implementation for any additional implementation-specific information.

9 Further Comments
The existence of A will depend upon the function u (see Marazzi (1987)); also if X is not of full rank a value of A will not be found. If the columns of X are almost linearly related, then convergence will be slow.
10 Example

A sample of 10 observations on three variables is read in along with initial values for $A$ and $\theta$ and argument values for the $u$ and $w$ functions, $c_u$ and $c_w$. The covariance matrix computed by nag_robust_m_corr_user_fn (g02hlc) is printed along with the robust estimate of $\theta$. ucv computes the Huber’s weight functions:

\[
u(t) = \begin{cases} 
1 & \text{if } t \leq c_u^2 \\
\frac{c_u}{t^2} & \text{if } t > c_u^2
\end{cases}
\]

and

\[
w(t) = \begin{cases} 
1 & \text{if } t \leq c_w \\
\frac{c_w}{t} & \text{if } t > c_w
\end{cases}
\]

and their derivatives.

10.1 Program Text

/* nag_robust_m_corr_user_fn (g02hlc) Example Program. *
* Copyright 2014 Numerical Algorithms Group.
* Mark 7, 2002.
* Mark 7b revised, 2004. */

#include <stdio.h>
#include <nag.h>
#include <nagg02.h>
#include <nag_stdlib.h>

#ifdef __cplusplus
extern "C" {
#endif

static void NAG_CALL ucv(double t, double *u, double *ud, double *w, double *wd, Nag_Comm *comm);

#ifdef __cplusplus
}
#endif

int main(void)
{

/* Scalars */
double bd, bl, tol;
Integer exit_status, i__, indm, j, k, l1, l2, m, maxit, mm, n, nit, nitmon;
Integer pdx;
NagError fail;
Nag_OrderType order;
Nag_Comm comm;

/* Arrays */
double *a = 0, *cov = 0, *theta = 0, *userp = 0, *wt = 0, *x = 0;

#ifdef NAG_COLUMN_MAJOR
#define X(I, J) x[(J-1)*pdx + I - 1]
#else
#define X(I, J) x[(I-1)*pdx+J-1]
#endif

/* Mark 25 */
INIT_FAIL(fail);
exit_status = 0;
printf(
    "nag_robust_m_corr_user_fn (g02hlc) Example Program Results\n\n"); /* Skip heading in data file */
#ifdef _WIN32
    scanf_s("%*[\n"]
#else
    scanf("%*[\n"]
#endif
#endif
/* Read in the dimensions of X */
#ifdef _WIN32
    scanf_s("%"NAG_IFMT"%"NAG_IFMT"%*[\n"]", &n, &m);
#else
    scanf("%"NAG_IFMT"%"NAG_IFMT"%*[\n"]", &n, &m);
#endif
/* Allocate memory */
if (!(a = NAG_ALLOC(m*(m+1)/2, double)) ||
    !(cov = NAG_ALLOC(m*(m+1)/2, double)) ||
    !(theta = NAG_ALLOC(m, double)) ||
    !(userp = NAG_ALLOC(2, double)) ||
    !(wt = NAG_ALLOC(n, double)) ||
    !(x = NAG_ALLOC(n * m, double))
{
    printf("Allocation failure\n");
    exit_status = -1;
    goto END;
}
#ifdef NAG_COLUMN_MAJOR
    pdx = n;
#else
    pdx = m;
#endif
/* Read in the X matrix */
for (i__ = 1; i__ <= n; ++i__)
{  
    for (j = 1; j <= m; ++j)
#ifdef _WIN32
        scanf_s("%lf", &X(i__, j));
#else
        scanf("%lf", &X(i__, j));
#endif
#ifdef _WIN32
        scanf_s("%*[\n"]
#else
        scanf("%*[\n"]
#endif
#ifdef _WIN32
    scanf_s("%*[\n"]
#else
    scanf("%*[\n"]
#endif
/* Read in the initial value of A */
    mm = (m + 1) * m / 2;
    for (j = 1; j <= mm; ++j)
#ifdef _WIN32
        scanf_s("%lf", &a[j - 1]);
#else
        scanf("%lf", &a[j - 1]);
#endif
#ifdef _WIN32
        scanf_s("%*[\n"]
#else
        scanf("%*[\n"]
#endif
/* Read in the initial value of theta */
    for (j = 1; j <= m; ++j)
#ifdef _WIN32
        scanf_s("%lf", &theta[j - 1]);
#else
        scanf("%lf", &theta[j - 1]);
#endif
#else
scanf("%lf", &theta[j - 1]);
#endif

#ifdef _WIN32
scanf_s("%*[^\n] ");
#else
scanf("%*[^\n] ");
#endif

/* Read in the values of the parameters of the ucv functions */
#ifdef _WIN32
scanf_s("%lf%lf%*[^\n] ", &userp[0], &userp[1]);
#else
scanf("%lf%lf%*[^\n] ", &userp[0], &userp[1]);
#endif

/* Set the values of remaining parameters */
indm = 1;
b1 = 0.9;
b2 = 0.9;
maxit = 50;
tol = 5e-5;
/* Change nitmon to a positive value if monitoring information */
* is required */
nitmon = 0;
comm.p = (void *) userp;

/* nag_robust_m_corr_user_fn (g02hlc). */
* Calculates a robust estimation of a correlation matrix,
* user-supplied weight function plus derivatives */
nag_robust_m_corr_user_fn(order, ucv, indm, n, m, x, pdx, cov, a, wt,
theta, bl, bd, maxit, nitmon, 0, tol, &nit, &comm,
&fail);

if (fail.code != NE_NOERROR)
{
printf("Error from nag_robust_m_corr_user_fn (g02hlc).\n%s\n",
fail.message);
exit_status = 1;
goto END;
}

printf("\n");
printf("nag_robust_m_corr_user_fn (g02hlc) required %4"NAG_IFMT" 
iterations to converge\n\n", nit);

printf("Robust covariance matrix\n");
l2 = 0;
for (j = 1; j <= m; ++j)
{
   11 = 12 + 1;
   l2 += 3;
   for (k = 11; k <= l2; ++k)
      printf("%10.3f\n", cov[k - 1], k%6 == 0 || k == l2?"\n": "");
}
printf("\n");

printf("Robust estimates of theta\n");
for (j = 1; j <= m; ++j)
   printf(" %10.3f\n", theta[j - 1]);

END:
NAG_FREE(a);
NAG_FREE(cov);
NAG_FREE(theta);
NAG_FREE(userp);
NAG_FREE(wt);
NAG_FREE(x);
return exit_status;

#endif

void NAG_CALL ucv(double t, double *u, double *ud, double *w, double *wd,
Nag_Comm *comm)
{
    double t2, cu, cw;
    double *userp = (double *) comm->p;

    /* Function Body */
    cu = userp[0];
    *u = 1.0;
    *ud = 0.0;
    if (t != 0.0)
    {
        t2 = t * t;
        if (t2 > cu)
        {
            *u = cu / t2;
            *ud = *u * -2.0 / t;
        }
    }

    /* w function and derivative */
    cw = userp[1];
    if (t > cw)
    {
        *w = cw / t;
        *wd = -(*w) / t;
    }
    else
    {
        *w = 1.0;
        *wd = 0.0;
    }
    return;
}

10.2 Program Data

nag_robust_m_corr_user_fn (g02hlc) Example Program Data

10 3 : N M
3.4 6.9 12.2 : X1 X2 X3
6.4 2.5 15.1
4.9 5.5 14.2
7.3 1.9 18.2
8.8 3.6 11.7
8.4 1.3 17.9
5.3 3.1 15.0
2.7 8.1 7.7
6.1 3.0 21.9
5.3 2.2 13.9 : End of X1 X2 and X3 values
1.0 0.0 1.0 0.0 0.0 1.0 : A
0.0 0.0 0.0 : THETA
4.0 2.0 : CU CW

10.3 Program Results

nag_robust_m_corr_user_fn (g02hlc) Example Program Results

nag_robust_m_corr_user_fn (g02hlc) required 25 iterations to converge

Robust covariance matrix
3.278
-3.692 5.284
4.739 -6.409 11.837

Robust estimates of theta
5.700
3.864
14.704