NAG Library Routine Document

G02LCF

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

G02LCF calculates parameter estimates for a given number of factors given the output from an orthogonal scores PLS regression (G02LAF or G02LBF).

2   Specification

```
SUBROUTINE G02LCF (IP, MY, MAXFAC, NFACT, P, LDP, C, LDC, W, LDW, RCOND,    &
                   B, LDB, ORIG, XBAR, YBAR, ISCALE, XSTD, YSTD, OB,        &
                   LDOB, VIPOPT, YCV, LDYCV, VIP, LDVIP, IFAIL)

INTEGER           IP, MY, MAXFAC, NFACT, LDP, LDC, LDW, LDB, ORIG,          &
                  ISCALE, LDOB, VIPOPT, LDYCV, LDVIP, IFAIL
REAL (KIND=nag_wp) P(LDP,MAXFAC), C(LDC,MAXFAC), W(LDW,MAXFAC), RCOND,      &
                  B(LDB,MY), XBAR(IP), YBAR(MY), XSTD(IP), YSTD(MY),        &
                  OB(LDOB,MY), YCV(LDYCV,MY), VIP(LDVIP,VIPOPT)
```

3   Description

The parameter estimates B for a l-factor orthogonal scores PLS model with m predictor variables and r response variables are given by,

$$B = W\left(P^{\mathrm{T}}W\right)^{-1}C^{\mathrm{T}}, \quad B \in \mathbb{R}^{m \times r},$$

where W is the m by k ($\geq l$) matrix of x-weights; P is the m by k matrix of x-loadings; and C is the r by k matrix of y-loadings for a fitted PLS model.

The parameter estimates B are for centred, and possibly scaled, predictor data X_1 and response data Y_1. Parameter estimates may also be given for the predictor data X and response data Y.

Optionally, G02LCF will calculate variable influence on projection (VIP) statistics, see Wold (1994).

4   References

Wold S (1994) PLS for multivariate linear modelling QSAR: chemometric methods in molecular design *Methods and Principles in Medicinal Chemistry* (ed van de Waterbeemd H) Verlag-Chemie

5   Arguments

1:      IP – INTEGER   *Input*

     On entry: m, the number of predictor variables in the fitted model.

     Constraint: IP > 1.

2:      MY – INTEGER   *Input*

     On entry: r, the number of response variables.

     Constraint: MY $\geq$ 1.

3:   MAXFAC – INTEGER   *Input*

  On entry: k, the number of factors available in the PLS model.

  Constraint: $1 \leq \text{MAXFAC} \leq \text{IP}$.

4:   NFACT – INTEGER   *Input*

  On entry: l, the number of factors to include in the calculation of parameter estimates.

  Constraint: $1 \leq \text{NFACT} \leq \text{MAXFAC}$.

5:   P(LDP, MAXFAC) – REAL (KIND=nag_wp) array   *Input*

  On entry: x-loadings as returned from G02LAF and G02LBF.

6:   LDP – INTEGER   *Input*

  On entry: the first dimension of the array P as declared in the (sub)program from which G02LCF is called.

  Constraint: $\text{LDP} \geq \text{IP}$.

7:   C(LDC, MAXFAC) – REAL (KIND=nag_wp) array   *Input*

  On entry: y-loadings as returned from G02LAF and G02LBF.

8:   LDC – INTEGER   *Input*

  On entry: the first dimension of the array C as declared in the (sub)program from which G02LCF is called.

  Constraint: $\text{LDC} \geq \text{MY}$.

9:   W(LDW, MAXFAC) – REAL (KIND=nag_wp) array   *Input*

  On entry: x-weights as returned from G02LAF and G02LBF.

10: LDW – INTEGER   *Input*

  On entry: the first dimension of the array W as declared in the (sub)program from which G02LCF is called.

  Constraint: $\text{LDW} \geq \text{IP}$.

11: RCOND – REAL (KIND=nag_wp)   *Input*

  On entry: singular values of $P^{\mathrm{T}}W$ less than RCOND times the maximum singular value are treated as zero when calculating parameter estimates. If RCOND is negative, a value of 0.005 is used.

12: B(LDB, MY) – REAL (KIND=nag_wp) array   *Output*

  On exit: $\text{B}(i,j)$ contains the parameter estimate for the ith predictor variable in the model for the jth response variable, for $i = 1, 2, \ldots, \text{IP}$ and $j = 1, 2, \ldots, \text{MY}$.

13: LDB – INTEGER   *Input*

  On entry: the first dimension of the array B as declared in the (sub)program from which G02LCF is called.

  Constraint: $\text{LDB} \geq \text{IP}$.

14:     ORIG – INTEGER   *Input*

On entry: indicates how parameter estimates are calculated.

ORIG = −1
       Parameter estimates for the centred, and possibly, scaled data.

ORIG = 1
       Parameter estimates for the original data.

Constraint: ORIG = −1 or 1.

15:     XBAR(IP) – REAL (KIND=nag_wp) array   *Input*

On entry: if ORIG = 1, mean values of predictor variables in the model; otherwise XBAR is not referenced.

16:     YBAR(MY) – REAL (KIND=nag_wp) array   *Input*

On entry: if ORIG = 1, mean value of each response variable in the model; otherwise YBAR is not referenced.

17:     ISCALE – INTEGER   *Input*

On entry: if ORIG = 1, ISCALE must take the value supplied to either G02LAF or G02LBF; otherwise ISCALE is not referenced.

Constraint: if ORIG = 1, ISCALE = −1, 1 or 2.

18:     XSTD(IP) – REAL (KIND=nag_wp) array   *Input*

On entry: if ORIG = 1 and ISCALE ≠ −1, the scalings of predictor variables in the model as returned from either G02LAF or G02LBF; otherwise XSTD is not referenced.

19:     YSTD(MY) – REAL (KIND=nag_wp) array   *Input*

On entry: if ORIG = 1 and ISCALE ≠ −1, the scalings of response variables as returned from either G02LAF or G02LBF; otherwise YSTD is not referenced.

20:     OB(LDOB, MY) – REAL (KIND=nag_wp) array   *Output*

On exit: if ORIG = 1, $OB(1, j)$ contains the intercept value for the jth response variable, and $OB(i + 1, j)$ contains the parameter estimate on the original scale for the ith predictor variable in the model, for $i = 1, 2, \ldots, IP$ and $j = 1, 2, \ldots, MY$. Otherwise OB is not referenced.

21:     LDOB – INTEGER   *Input*

On entry: the first dimension of the array OB as declared in the (sub)program from which G02LCF is called.

Constraints:

       if ORIG = 1, LDOB ≥ IP + 1;
       otherwise LDOB ≥ 1.

22:     VIPOPT – INTEGER   *Input*

On entry: a flag that determines variable influence on projections (VIP) options.

VIPOPT = 0
       VIP are not calculated.

VIPOPT = 1
       VIP are calculated for predictor variables using the mean explained variance in responses.

VIPOPT = MY
       VIP are calculated for predictor variables for each response variable in the model.

Note that setting $\text{VIPOPT} = \text{MY}$ when $\text{MY} = 1$ gives the same result as setting $\text{VIPOPT} = 1$ directly.

Constraint: $\text{VIPOPT} = 0$, 1 or MY.

23:    $\text{YCV}(\text{LDYCV}, \text{MY})$ – REAL (KIND=nag_wp) array   *Input*

On entry: if $\text{VIPOPT} \neq 0$, $\text{YCV}(i, j)$ is the cumulative percentage of variance of the jth response variable explained by the first i factors, for $i = 1, 2, \ldots, \text{NFACT}$ and $j = 1, 2, \ldots, \text{MY}$; otherwise YCV is not referenced.

24:    LDYCV – INTEGER   *Input*

On entry: the first dimension of the array YCV as declared in the (sub)program from which G02LCF is called.

Constraint: if $\text{VIPOPT} \neq 0$, $\text{LDYCV} \geq \text{NFACT}$.

25:    $\text{VIP}(\text{LDVIP}, \text{VIPOPT})$ – REAL (KIND=nag_wp) array   *Output*

On exit: if $\text{VIPOPT} = 1$, $\text{VIP}(i, 1)$ contains the VIP statistic for the ith predictor variable in the model for all response variables, for $i = 1, 2, \ldots, \text{IP}$.

If $\text{VIPOPT} = \text{MY}$, $\text{VIP}(i, j)$ contains the VIP statistic for the ith predictor variable in the model for the jth response variable, for $i = 1, 2, \ldots, \text{IP}$ and $j = 1, 2, \ldots, \text{MY}$.

Otherwise VIP is not referenced.

26:    LDVIP – INTEGER   *Input*

On entry: the first dimension of the array VIP as declared in the (sub)program from which G02LCF is called.

Constraint: if $\text{VIPOPT} \neq 0$, $\text{LDVIP} \geq \text{IP}$.

27:    IFAIL – INTEGER   *Input/Output*

On entry: IFAIL must be set to 0, -1 or 1. If you are unfamiliar with this argument you should refer to Section 3.4 in How to Use the NAG Library and its Documentation for details.

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, if you are not familiar with this argument, the recommended value is 0. **When the value -1 or 1 is used it is essential to test the value of IFAIL on exit.**

On exit: $\text{IFAIL} = 0$ unless the routine detects an error or a warning has been flagged (see Section 6).

6   Error Indicators and Warnings

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

Errors or warnings detected by the routine:

$\text{IFAIL} = 1$

On entry, $\text{IP} = \langle value \rangle$.
Constraint: $\text{IP} > 1$.

On entry, $\text{ISCALE} = \langle value \rangle$.
Constraint: if $\text{ORIG} = 1$, $\text{ISCALE} = -1$ or 1.

On entry, $\text{MY} = \langle value \rangle$.
Constraint: $\text{MY} \geq 1$.

On entry, ORIG $= \langle value \rangle$.
Constraint: ORIG $= -1$ or 1.

On entry, VIPOPT $= \langle value \rangle$ and MY $= \langle value \rangle$.
Constraint: VIPOPT $= 0$, 1 or MY.

IFAIL $= 2$

On entry, LDB $= \langle value \rangle$ and IP $= \langle value \rangle$.
Constraint: LDB $\geq$ IP.

On entry, LDC $= \langle value \rangle$ and MY $= \langle value \rangle$.
Constraint: LDC $\geq$ MY.

On entry, LDOB $= \langle value \rangle$ and IP $= \langle value \rangle$.
Constraint: if ORIG $= 1$, LDOB $\geq$ IP $+ 1$.

On entry, LDP $= \langle value \rangle$ and IP $= \langle value \rangle$.
Constraint: LDP $\geq$ IP.

On entry, LDVIP $= \langle value \rangle$ and IP $= \langle value \rangle$.
Constraint: if VIPOPT $\neq 0$, LDVIP $\geq$ IP.

On entry, LDW $= \langle value \rangle$ and IP $= \langle value \rangle$.
Constraint: LDW $\geq$ IP.

On entry, LDYCV $= \langle value \rangle$ and NFACT $= \langle value \rangle$.
Constraint: if VIPOPT $\neq 0$, LDYCV $\geq$ NFACT.

On entry, MAXFAC $= \langle value \rangle$ and IP $= \langle value \rangle$.
Constraint: $1 \leq$ MAXFAC $\leq$ IP.

On entry, NFACT $= \langle value \rangle$ and MAXFAC $= \langle value \rangle$.
Constraint: $1 \leq$ NFACT $\leq$ MAXFAC.

IFAIL $= -99$

An unexpected error has been triggered by this routine. Please contact NAG.

See Section 3.9 in How to Use the NAG Library and its Documentation for further information.

IFAIL $= -399$

Your licence key may have expired or may not have been installed correctly.

See Section 3.8 in How to Use the NAG Library and its Documentation for further information.

IFAIL $= -999$

Dynamic memory allocation failed.

See Section 3.7 in How to Use the NAG Library and its Documentation for further information.

7 Accuracy

The calculations are based on the singular value decomposition of $P^{\mathrm{T}}W$.

8 Parallelism and Performance

G02LCF is threaded by NAG for parallel execution in multithreaded implementations of the NAG Library.

G02LCF 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 routine. Please also consult the Users' Note for your implementation for any additional implementation-specific information.

9   Further Comments

G02LCF allocates internally $l(l + r + 4) + \max(2l, r)$ elements of real storage.

10   Example

This example reads in details of a PLS model, and a set of parameter estimates are calculated along with their VIP statistics.

10.1 Program Text

```
    Program g02lcfe

!      G02LCF Example Program Text

!      Mark 26 Release. NAG Copyright 2016.

!      .. Use Statements ..
       Use nag_library, Only: g02lcf, nag_wp, x04caf
!      .. Implicit None Statement ..
       Implicit None
!      .. Parameters ..
       Integer, Parameter                :: nin = 5, nout = 6
!      .. Local Scalars ..
       Real (Kind=nag_wp)                :: rcond
       Integer                           :: i, ifail, ip, iscale, ldb, ldc,   &
                                            ldob, ldp, ldvip, ldw, ldycv,      &
                                            maxfac, my, nfact, orig, vipopt
!      .. Local Arrays ..
       Real (Kind=nag_wp), Allocatable   :: b(:,:), c(:,:), ob(:,:), p(:,:),   &
                                            vip(:,:), w(:,:), xbar(:), xstd(:), &
                                            ybar(:), ycv(:,:), ystd(:)
!      .. Executable Statements ..
       Write (nout,*) 'G02LCF Example Program Results'
       Write (nout,*)
       Flush (nout)

!      Skip heading in data file
       Read (nin,*)

!      Read problem size
       Read (nin,*) ip, my, maxfac, nfact, orig, iscale, vipopt

       ldp = ip
       ldc = my
       ldw = ip
       ldb = ip
       If (orig==1) Then
         ldob = ip + 1
       Else
         ldob = 1
       End If
       If (vipopt/=0) Then
         ldycv = nfact
         ldvip = ip
       Else
         ldycv = 0
         ldvip = 0
       End If
       Allocate (p(ldp,maxfac),c(ldc,maxfac),w(ldw,maxfac),b(ldb,my),xbar(ip),  &
         ybar(my),xstd(ip),ystd(my),ob(ldob,my),ycv(ldycv,my),                  &
         vip(ldvip,vipopt))
```

```
!       Read in data
        Read (nin,*)(p(i,1:maxfac),i=1,ip)
        Read (nin,*)(c(i,1:maxfac),i=1,my)
        Read (nin,*)(w(i,1:maxfac),i=1,ip)
        If (vipopt/=0) Then
          Read (nin,*)(ycv(i,1:my),i=1,nfact)
        End If

!       Read means and scalings
        If (orig==1) Then
          Read (nin,*) xbar(1:ip)
          Read (nin,*) ybar(1:my)
          If (iscale/=-1) Then
            Read (nin,*) xstd(1:ip)
            Read (nin,*) ystd(1:my)
          End If
        End If

!       Calculate predictions
        rcond = -1.0E0_nag_wp
        ifail = 0
        Call g02lcf(ip,my,maxfac,nfact,p,ldp,c,ldc,w,ldw,rcond,b,ldb,orig,xbar,  &
          ybar,iscale,xstd,ystd,ob,ldob,vipopt,ycv,ldycv,vip,ldvip,ifail)

!       Display results
        ifail = 0
        Call x04caf('General',' ',ip,my,b,ldb,'B',ifail)
        If (orig==1) Then
          Write (nout,*)
          Flush (nout)
          ifail = 0
          Call x04caf('General',' ',ip+1,my,ob,ldob,'OB',ifail)
        End If
        If (vipopt/=0) Then
          Write (nout,*)
          Flush (nout)
          ifail = 0
          Call x04caf('General',' ',ip,vipopt,vip,ldvip,'VIP',ifail)
        End If

    End Program g02lcfe
```

10.2 Program Data

```
G02LCF Example Program Data
15 1 4 2 1 1 1 : IP, MY, MAXFAC, NFACT, ORIG, SCALE, VIPOPT
-0.6708     -1.0047      0.6505      0.6169
 0.4943      0.1355     -0.9010     -0.2388
-0.4167     -1.9983     -0.5538      0.8474
 0.3930      1.2441     -0.6967     -0.4336
 0.3267      0.5838     -1.4088     -0.6323
 0.0145      0.9607      1.6594      0.5361
-2.4471      0.3532     -1.1321     -1.3554
 3.5198      0.6005      0.2191      0.0380
 1.0973      2.0635     -0.4074     -0.3522
-2.4466      2.5640     -0.4806      0.3819
 2.2732     -1.3110     -0.7686     -1.8959
-1.7987      2.4088     -0.9475     -0.4727
 0.3629      0.2241     -2.6332      2.3739
 0.3629      0.2241     -2.6332      2.3739
-0.3629     -0.2241      2.6332     -2.3739 : P
 3.5425      1.0475      0.2548      0.1866 : C
-1.5764E-01  -1.5935E-01   1.7774E-01   5.4029E-02
 8.5680E-02  -1.5240E-04  -1.2179E-01   1.0989E-01
-1.6931E-01  -3.7431E-01   9.4348E-02   3.1878E-01
 1.2153E-01   2.0589E-01  -1.8144E-01  -4.4610E-02
 7.1133E-02   5.5884E-02  -2.6916E-01   5.4912E-02
 6.5188E-02   2.4170E-01   2.3365E-01  -1.8849E-01
-4.2481E-01  -1.8798E-03  -3.2413E-01  -1.1600E-01
 6.5370E-01   1.6725E-01   2.1908E-01   2.5461E-01
```

```
 2.8504E-01    3.6549E-01  -1.9244E-01  -1.5430E-01
-2.9341E-01    5.0464E-01  -1.0952E-02   1.3881E-01
 2.9829E-01   -3.6979E-01  -4.9942E-01  -4.9355E-01
-2.0313E-01    4.1952E-01  -2.5684E-01  -7.5647E-02
 5.6905E-02   -2.3197E-02  -3.0503E-01   3.9673E-01
 5.6905E-02   -2.3197E-02  -3.0503E-01   3.9673E-01
-5.6905E-02    2.3197E-02   3.0503E-01  -3.9673E-01 : W
89.638060 97.476270 97.939839 98.188474 : YCV
-2.6137 -2.3614 -1.0449 2.8614  0.3156
-0.2641 -0.3146 -1.1221 0.2401  0.4694
-1.9619  0.1691  2.5664 1.3741 -2.7821  : XBAR
 0.4520                                  : YBAR
 1.4956  1.3233  0.5829 0.7735  0.6247
 0.7966  2.4113  2.0421 0.4678  0.8197
 0.9420  0.1735  1.0475 0.1359  1.3853  : XSTD
 0.9062                                  : YSTD
```

10.3 Program Results

```
G02LCF Example Program Results

 B
          1
  1  -0.1383
  2   0.0572
  3  -0.1906
  4   0.1238
  5   0.0591
  6   0.0936
  7  -0.2842
  8   0.4713
  9   0.2661
 10  -0.0914
 11   0.1226
 12  -0.0488
 13   0.0332
 14   0.0332
 15  -0.0332

 OB
          1
  1  -0.4374
  2  -0.0838
  3   0.0392
  4  -0.2964
  5   0.1451
  6   0.0857
  7   0.1065
  8  -0.1068
  9   0.2091
 10   0.5155
 11  -0.1011
 12   0.1180
 13  -0.2548
 14   0.0287
 15   0.2214
 16  -0.0217

 VIP
          1
  1   0.6111
  2   0.3182
  3   0.7513
  4   0.5048
  5   0.2712
  6   0.3593
  7   1.5777
  8   2.4348
  9   1.1322
 10   1.2226
```

```
11      1.1799
12      0.8840
13      0.2129
14      0.2129
15      0.2129
```