

NAG Library Routine Document

G02MCF

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

G02MCF calculates additional parameter estimates following Least Angle Regression (LARS), forward stagewise linear regression or Least Absolute Shrinkage and Selection Operator (LASSO) as performed by G02MAF and G02MBF.

2 Specification

```
SUBROUTINE G02MCF (NSTEP, IP, B, LDB, FITSUM, KTYPE, NK, LNK, NB, LDNB,      &
                  IFAIL)
```

```
INTEGER          NSTEP, IP, LDB, KTYPE, LNK, LDNB, IFAIL
REAL (KIND=nag_wp) B(LDB,*), FITSUM(6,NSTEP+1), NK(LNK), NB(LDNB,*)
```

3 Description

G02MAF and G02MBF fit either a LARS, forward stagewise linear regression, LASSO or positive LASSO model to a vector of n observed values, $y = \{y_i : i = 1, 2, \dots, n\}$ and an $n \times p$ design matrix X , where the j th column of X is given by the j th independent variable x_j . The models are fit using the LARS algorithm of Efron *et al.* (2004).

The full solution path for all four of these models follow a similar pattern where the parameter estimate for a given variable is piecewise linear. One such path, for a LARS model with six variables ($p = 6$) can be seen in Figure 1. Both G02MAF and G02MBF return the vector of p parameter estimates, β_k , at K points along this path (so $k = 1, 2, \dots, K$). Each point corresponds to a step of the LARS algorithm. The number of steps taken depends on the model being fitted. In the case of a LARS model, $K = p$ and each step corresponds to a new variable being included in the model. In the case of the LASSO models, each step corresponds to either a new variable being included in the model or an existing variable being removed from the model; the value of K is therefore no longer bound by the number of parameters. For forward stagewise linear regression, each step no longer corresponds to the addition or removal of a variable; therefore the number of possible steps is often markedly greater than for a corresponding LASSO model.

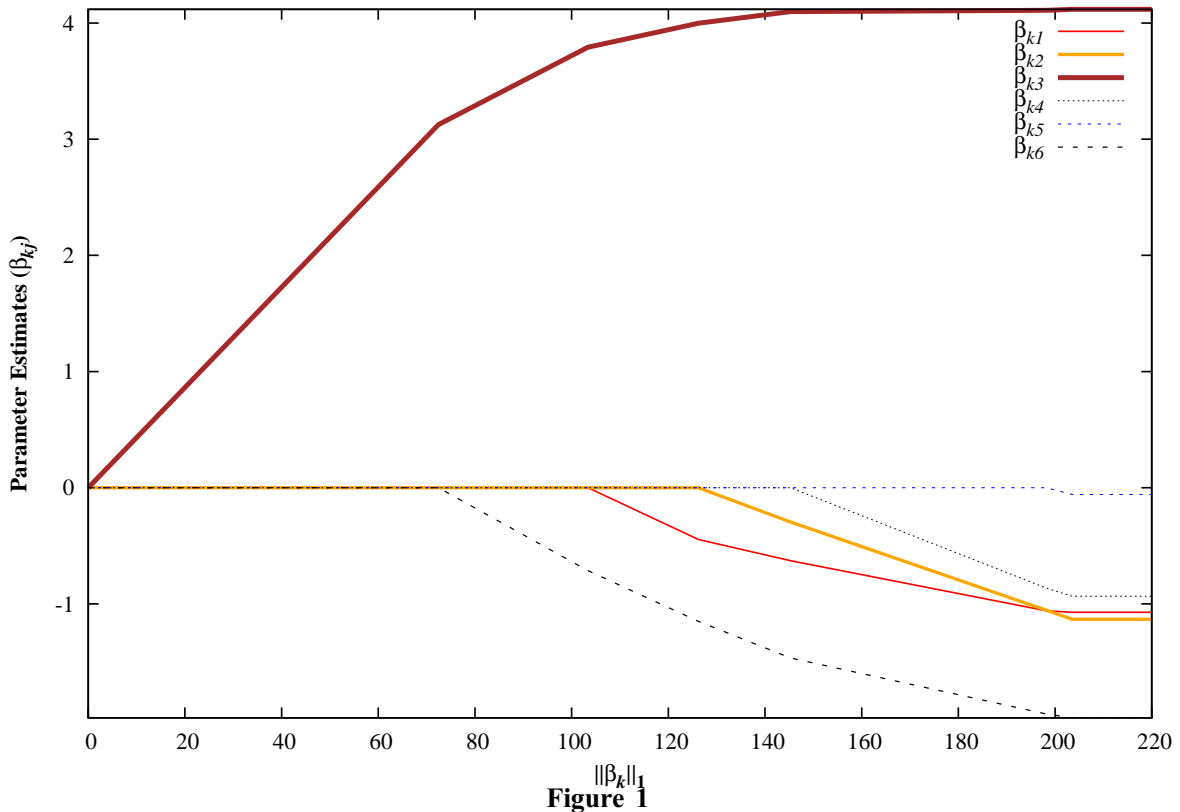


Figure 1

G02MCF uses the piecewise linear nature of the solution path to predict the parameter estimates, $\tilde{\beta}$, at a different point on this path. The location of the solution can either be defined in terms of a (fractional) step number or a function of the L_1 norm of the parameter estimates.

4 References

Efron B, Hastie T, Johnstone I and Tibshirani R (2004) Least Angle Regression *The Annals of Statistics (Volume 32)* **2** 407–499

Hastie T, Tibshirani R and Friedman J (2001) *The Elements of Statistical Learning: Data Mining, Inference and Prediction* Springer (New York)

Tibshirani R (1996) Regression Shrinkage and Selection via the Lasso *Journal of the Royal Statistics Society, Series B (Methodological)* (Volume 58) **1** 267–288

Weisberg S (1985) *Applied Linear Regression* Wiley

5 Parameters

1: NSTEP – INTEGER Input

On entry: K , the number of steps carried out in the model fitting process, as returned by G02MAF and G02MBF.

Constraint: $\text{NSTEP} \geq 0$.

2: IP – INTEGER Input

On entry: p , number of parameter estimates, as returned by G02MAF and G02MBF.

Constraint: $\text{IP} \geq 1$.

- 3: B(LDB,*) – REAL (KIND=nag_wp) array Input
Note: the second dimension of the array B must be at least NSTEP + 1.
On entry: β the parameter estimates, as returned by G02MAF and G02MBF, with $B(j, k) = \beta_{kj}$, the parameter estimate for the j th variable, for $j = 1, 2, \dots, p$, at the k th step of the model fitting process.
Constraint: B should be unchanged since the last call to G02MAF or G02MBF.
- 4: LDB – INTEGER Input
On entry: the first dimension of the array B as declared in the (sub)program from which G02MCF is called.
Constraint: $LDB \geq IP$.
- 5: FITSUM(6, NSTEP + 1) – REAL (KIND=nag_wp) array Input
On entry: summaries of the model fitting process, as returned by G02MAF and G02MBF.
Constraint: FITSUM should be unchanged since the last call to G02MAF or G02MBF..
- 6: KTYPE – INTEGER Input
On entry: indicates what target values are held in NK.
 KTYPE = 1
 NK holds (fractional) LARS step numbers.
 KTYPE = 2
 NK holds values for L_1 norm of the (scaled) parameters.
 KTYPE = 3
 NK holds ratios with respect to the largest (scaled) L_1 norm.
 KTYPE = 4
 NK holds values for the L_1 norm of the (unscaled) parameters.
 KTYPE = 5
 NK holds ratios with respect to the largest (unscaled) L_1 norm.
 If G02MAF was called with PRED = 0 or 1 or G02MBF was called with PRED = 0 then the model fitting routine did not rescale the independent variables, X , prior to fitting the model and therefore there is no difference between KTYPE = 2 or 3 and KTYPE = 4 or 5.
Constraint: KTYPE = 1, 2, 3, 4 or 5.
- 7: NK(LNK) – REAL (KIND=nag_wp) array Input
On entry: target values used for predicting the new set of parameter estimates.
Constraints:
 if KTYPE = 1, $0 \leq NK(i) \leq NSTEP$, for $i = 1, 2, \dots, LNK$;
 if KTYPE = 2, $0 \leq NK(i) \leq FITSUM(1, NSTEP)$, for $i = 1, 2, \dots, LNK$;
 if KTYPE = 3 or 5, $0 \leq NK(i) \leq 1$, for $i = 1, 2, \dots, LNK$;
 if KTYPE = 4, $0 \leq NK(i) \leq \|\beta_K\|_1$, for $i = 1, 2, \dots, LNK$.
- 8: LNK – INTEGER Input
On entry: number of values supplied in NK.
Constraint: $LNK \geq 1$.

9: NB(LDNB,*) – REAL (KIND=nag_wp) array Output

Note: the second dimension of the array NB must be at least LNK.

On exit: $\tilde{\beta}$ the predicted parameter estimates, with $B(j, i) = \tilde{\beta}_{ij}$, the parameter estimate for variable j , $j = 1, 2, \dots, p$ at the point in the fitting process associated with $NK(i)$, $i = 1, 2, \dots, LNK$.

10: LDNB – INTEGER Input

On entry: the first dimension of the array NB as declared in the (sub)program from which G02MCF is called.

Constraint: $LDNB \geq IP$.

11: IFAIL – INTEGER Input/Output

On entry: IFAIL must be set to 0, -1 or 1. If you are unfamiliar with this parameter you should refer to Section 3.3 in the Essential Introduction 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, because for this routine the values of the output parameters may be useful even if $IFAIL \neq 0$ on exit, the recommended value is -1. **When the value -1 or 1 is used it is essential to test the value of IFAIL on exit.**

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

6 Error Indicators and Warnings

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

Note: G02MCF may return useful information for one or more of the following detected errors or warnings.

Errors or warnings detected by the routine:

IFAIL = 11

On entry, NSTEP = $\langle value \rangle$.
Constraint: $NSTEP \geq 0$.

IFAIL = 21

On entry, IP = $\langle value \rangle$.
Constraint: $IP \geq 1$.

IFAIL = 31

B has been corrupted since the last call to G02MAF or G02MBF.

IFAIL = 41

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

IFAIL = 51

FITSUM has been corrupted since the last call to G02MAF or G02MBF.

IFAIL = 61

On entry, KTYPE = $\langle value \rangle$.
Constraint: KTYPE = 1, 2, 3, 4 or 5.

IFAIL = 71

On entry, KTYPE = 1, $NK(\langle value \rangle) = \langle value \rangle$ and $NSTEP = \langle value \rangle$
 Constraint: $0 \leq NK(i) \leq NSTEP$ for all i .

IFAIL = 72

On entry, KTYPE = 2, $NK(\langle value \rangle) = \langle value \rangle$, $NSTEP = \langle value \rangle$ and
 $FITSUM(1, NSTEP) = \langle value \rangle$.
 Constraint: $0 \leq NK(i) \leq FITSUM(1, NSTEP)$ for all i .

IFAIL = 73

On entry, KTYPE = 3 or 5, $NK(\langle value \rangle) = \langle value \rangle$.
 Constraint: $0 \leq NK(i) \leq 1$ for all i .

IFAIL = 74

On entry, KTYPE = 4, $NK(\langle value \rangle) = \langle value \rangle$ and $\|\beta_K\|_1 = \langle value \rangle$
 Constraint: $0 \leq NK(i) \leq \|\beta_K\|_1$ for all i .

IFAIL = 81

On entry, $LNK = \langle value \rangle$.
 Constraint: $LNK \geq 1$.

IFAIL = 101

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

IFAIL = -99

An unexpected error has been triggered by this routine. Please contact NAG.
 See Section 3.8 in the Essential Introduction for further information.

IFAIL = -399

Your licence key may have expired or may not have been installed correctly.
 See Section 3.7 in the Essential Introduction for further information.

IFAIL = -999

Dynamic memory allocation failed.
 See Section 3.6 in the Essential Introduction for further information.

7 Accuracy

Not applicable.

8 Parallelism and Performance

G02MCF is threaded by NAG for parallel execution in multithreaded implementations of the NAG Library.

G02MCF 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

None.

10 Example

This example performs a LARS on a set a simulated dataset with 20 observations and 6 independent variables.

Additional parameter estimates are obtained corresponding to a LARS step number of 0.2, 1.2, 3.2, 4.5 and 5.2. Where, for example, 4.5 corresponds to the solution halfway between that obtained at step 4 and that obtained at step 5.

10.1 Program Text

```

Program g02mcfe

!      G02MCF Example Program Text

!      Mark 25 Release. NAG Copyright 2014.

!      .. Use Statements ..
Use nag_library, Only: g02maf, g02mcf, nag_wp
!      .. Implicit None Statement ..
Implicit None
!      .. Parameters ..
Integer, Parameter          :: nin = 5, nout = 6
!      .. Local Scalars ..
Integer                     :: i, ifail, ip, k, ktype, ldb, ldd,      &
                             ldnb, lisx, lnk, lropt, m, mnstep,      &
                             mtype, n, nstep, pred, prey
!      .. Local Arrays ..
Real (Kind=nag_wp), Allocatable :: b(:,,:), d(:,,:), fitsum(:,,:),      &
                             nb(:,,:), nk(:,), ropt(:,), y(:,)
Integer, Allocatable          :: isx(:)
!      .. Intrinsic Procedures ..
Intrinsic                    :: max, repeat
!      .. Executable Statements ..
Write (nout,*) 'G02MCF Example Program Results'
Write (nout,*)

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

!      Read in the problem size
Read (nin,*) n, m

!      Read in the model specification
Read (nin,*) mtype, pred, prey, mnstep

!      Use all of the variables
lisx = 0
Allocate (isx(lisx))

!      Optional arguments (using defaults)
lropt = 0
Allocate (ropt(lropt))

!      Read in the data
ldd = n
Allocate (y(n),d(ldd,m))
Read (nin,*)(d(i,1:m),y(i),i=1,n)

!      Allocate output arrays
ldb = m
Allocate (b(ldb,mnstep+2),fitsum(6,mnstep+1))

!      Call the model fitting routine

```

```

ifail = -1
Call g02maf(mtype,pred,prey,n,m,d,ldd,lsx,lisx,y,mnstep,ip,nstep,b,ldb, &
  fitsum,ropt,lropt,ifail)
If (ifail/=0) Then
  If (ifail/=112 .And. ifail/=161 .And. ifail/=162 .And. ifail/=163) &
    Then
!       IFAIL = 112, 161, 162 and 163 are warnings, so no need to terminate
!       if they occur
      Go To 100
    End If
  End If
End If

!   Read in the number of additional parameter estimates required and the
!   way they will be specified
Read (nin,*) ktype, lnk
ldnb = ip
Allocate (nk(lnk),nb(ip,lnk))

!   Read in the target values
Read (nin,*) nk(1:lnk)

!   Calculate the additional parameter estimates
ifail = 0
Call g02mcf(nstep,ip,b,ldb,fitsum,ktype,nk,lnk,nb,ldnb,ifail)

Write (nout,*) 'Parameter Estimates from G02MAF'
Write (nout,*) ' Step ', repeat(' ',max((ip-2),0)*5), &
  ' Parameter Estimate'
Write (nout,*) repeat('-',5+ip*10)
Do k = 1, nstep
  Write (nout,99999) k, b(1:ip,k)
End Do
Write (nout,*)

Write (nout,*) 'Additional Parameter Estimates from G02MCF'
Write (nout,*) ' NK ', repeat(' ',max((ip-2),0)*5), &
  ' Parameter Estimate'
Write (nout,*) repeat('-',5+ip*10)
Do k = 1, lnk
  Write (nout,99998) nk(k), nb(1:ip,k)
End Do

100 Continue
99999 Format (2X,I3,10(1X,F9.3))
99998 Format (1X,F4.1,10(1X,F9.3))
End Program g02mcfe

```

10.2 Program Data

G02MCF Example Program Data

```

20 6                                :: N,M
1 3 1 6                             :: MTYPE,PRED,PREY,MNSTEP
10.28 1.77 9.69 15.58 8.23 10.44    -46.47
 9.08 8.99 11.53 6.57 15.89 12.58   -35.80
17.98 13.10 1.04 10.45 10.12 16.68  -129.22
14.82 13.79 12.23 7.00 8.14 7.79    -42.44
17.53 9.41 6.24 3.75 13.12 17.08    -73.51
 7.78 10.38 9.83 2.58 10.13 4.25    -26.61
11.95 21.71 8.83 11.00 12.59 10.52  -63.90
14.60 10.09 -2.70 9.89 14.67 6.49    -76.73
 3.63 9.07 12.59 14.09 9.06 8.19    -32.64
 6.35 9.79 9.40 12.79 8.38 16.79    -83.29
 4.66 3.55 16.82 13.83 21.39 13.88  -16.31
 8.32 14.04 17.17 7.93 7.39 -1.09    -5.82
10.86 13.68 5.75 10.44 10.36 10.06  -47.75
 4.76 4.92 17.83 2.90 7.58 11.97    18.38
 5.05 10.41 9.89 9.89 9.04 7.90 13.12 -54.71
 5.41 9.32 5.27 15.53 5.06 19.84    -55.62
 9.77 2.37 9.54 20.23 9.33 8.82     -45.28

```

```

14.28  4.34 14.23 14.95 18.16 11.03   -22.76
10.17  6.80  3.17  8.57 16.07 15.93  -104.32
 5.39  2.67  6.37 13.56 10.68  7.35   -55.94  :: End of D, Y
1 5                                     :: KTYPE,LNK
0.2 1.2 3.2 4.5 5.2                   :: End of NK

```

10.3 Program Results

G02MCF Example Program Results

Parameter Estimates from G02MAF

Step	Parameter Estimate					
1	0.000	0.000	3.125	0.000	0.000	0.000
2	0.000	0.000	3.792	0.000	0.000	-0.713
3	-0.446	0.000	3.998	0.000	0.000	-1.151
4	-0.628	-0.295	4.098	0.000	0.000	-1.466
5	-1.060	-1.056	4.110	-0.864	0.000	-1.948
6	-1.073	-1.132	4.118	-0.935	-0.059	-1.981

Additional Parameter Estimates from G02MCF

NK	Parameter Estimate					
0.2	0.000	0.000	0.625	0.000	0.000	0.000
1.2	0.000	0.000	3.258	0.000	0.000	-0.143
3.2	-0.483	-0.059	4.018	0.000	0.000	-1.214
4.5	-0.844	-0.676	4.104	-0.432	0.000	-1.707
5.2	-1.062	-1.071	4.112	-0.878	-0.012	-1.955