NAG Toolbox for MATLAB

g02dd

1 Purpose

g02dd calculates the regression parameters for a general linear regression model. It is intended to be called after g02dc, g02de or g02df.

2 Syntax

```
[rss, idf, b, se, cov, svd, irank, p, ifail] = g02dd(n, ip, q, rss, p,
'tol', tol)
```

3 Description

A general linear regression model fitted by g02da may be adjusted by adding or deleting an observation using g02dc, adding a new independent variable using g02de or deleting an existing independent variable using g02df. Alternatively a model may be constructed by a forward selection procedure using g02ee. These functions compute the vector c and the upper triangular matrix R. g02dd takes these basic results and computes the regression coefficients, $\hat{\beta}$, their standard errors and their variance-covariance matrix.

If R is of full rank, then $\hat{\beta}$ is the solution to

$$R\hat{\beta} = c_1,$$

where c_1 is the first p elements of c.

If R is not of full rank a solution is obtained by means of a singular value decomposition (**svd**) of R,

$$R = Q_* \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix} P^{\mathrm{T}},$$

where D is a k by k diagonal matrix with nonzero diagonal elements, k being the rank of R, and Q_* and P are p by p orthogonal matrices. This gives the solution

$$\hat{\beta} = P_1 D^{-1} Q_{*_1}^{\mathrm{T}} c_1.$$

P_1 being the first k columns of P, i.e., $P = (P_1 P_0)$, and Q_{*_1} being the first k columns of Q_*.

Details of the **svd** are made available in the form of the matrix P^*:

$$P^* = \begin{pmatrix} D^{-1} P_1^{\mathrm{T}} \\ P_0^{\mathrm{T}} \end{pmatrix}.$$

This will be only one of the possible solutions. Other estimates may be obtained by applying constraints to the parameters. These solutions can be obtained by calling g02dk after calling g02dd. Only certain linear combinations of the parameters will have unique estimates; these are known as estimable functions. These can be estimated using g02dn.

The residual sum of squares required to calculate the standard errors and the variance-covariance matrix can either be input or can be calculated if additional information on c for the whole sample is provided.

4 References

Golub G H and Van Loan C F 1996 *Matrix Computations* (3rd Edition) Johns Hopkins University Press, Baltimore

Hammarling S 1985 The singular value decomposition in multivariate statistics *SIGNUM Newsl.* **20 (3)** 2–25

Searle S R 1971 *Linear Models* Wiley

5 Parameters

5.1 Compulsory Input Parameters

1: **n – int32 scalar**

The number of observations.

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

2: **ip – int32 scalar**

p, the number of terms in the regression model.

Constraint: $\mathbf{ip} \geq 1$.

3: **q(ldq,ip + 1) – double array**

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

Must be the array **q** as output by g02dc, g02de, g02df or g02ee. If on entry $\mathbf{rss} \leq 0.0$ then all **n** elements of c are needed. This is provided by functions g02de, g02df or g02ee.

4: **rss – double scalar**

Either the residual sum of squares or a value less than or equal to 0.0 to indicate that the residual sum of squares is to be calculated by the function.

5: **p(ip × ip + 2 × ip) – double array**

The elements of **p** that are not referenced as described below remain unchanged.

5.2 Optional Input Parameters

1: **tol – double scalar**

The value of **tol** is used to decide if the independent variables are of full rank and, if not, what is the rank of the independent variables. The smaller the value of **tol** the stricter the criterion for selecting the singular value decomposition. If $\mathbf{tol} = 0.0$, the singular value decomposition will never be used, this may cause run time errors or inaccuracies if the independent variables are not of full rank.

Suggested value: $\mathbf{tol} = 0.000001$.

Default: 0.000001

Constraint: $\mathbf{tol} \geq 0.0$.

5.3 Input Parameters Omitted from the MATLAB Interface

ldq, wk

5.4 Output Parameters

1: **rss – double scalar**

If $\mathbf{rss} \leq 0.0$ on entry, then on exit **rss** will contain the residual sum of squares as calculated by g02dd.

If **rss** was positive on entry, it will be unchanged.

2: **idf – int32 scalar**

The degrees of freedom associated with the residual sum of squares.

3: **b(ip) – double array**

The estimates of the p parameters, $\hat{\beta}$.

4: **se(ip) – double array**

The standard errors of the p parameters given in **b**.

5: **cov(ip × (ip + 1)/2) – double array**

The upper triangular part of the variance-covariance matrix of the p parameter estimates given in **b**. They are stored packed by column, i.e., the covariance between the parameter estimate given in $\mathbf{b}(i)$ and the parameter estimate given in $\mathbf{b}(j)$, $j \geq i$, is stored in $\mathbf{cov}(j \times (j-1)/2 + i)$.

6: **svd – logical scalar**

If a singular value decomposition has been performed, **svd** = **true**, otherwise **svd** = **false**.

7: **irank – int32 scalar**

The rank of the independent variables.

If **svd** = **false**, **irank** = **ip**.

If **svd** = **true**, **irank** is an estimate of the rank of the independent variables.

irank is calculated as the number of singular values greater than **tol** × (largest singular value). It is possible for the **svd** to be carried out but **irank** to be returned as **ip**.

8: **p(ip × ip + 2 × ip) – double array**

Contains details of the singular value decomposition if used.

If **svd** = **false**, **p** is not referenced.

If **svd** = **true**, the first **ip** elements of **p** will not be referenced, the next **ip** values contain the singular values. The following **ip** × **ip** values contain the matrix P^* stored by columns.

9: **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:

ifail = 1

On entry, **n** < 1,
or **ip** < 1,
or **ldq** < **ip**,
or **ldq** < **n**,
or **tol** < 0.0.

ifail = 2

The degrees of freedom for error are less than or equal to 0. In this case the estimates of β are returned but not the standard errors or covariances.

ifail = 3

The singular value decomposition, if used, has failed to converge, see f02wu. This is an unlikely error exit.

7 Accuracy

The accuracy of the results will depend on the accuracy of the input R matrix, which may lose accuracy if a large number of observations or variables have been dropped.

8 Further Comments

None.

9 Example

```
n = int32(12);
ip = int32(4);
q = [-62.35960257517148, -1.732050807568877, 0, 0, 0;
     64.60549512231913, 0, 1.732050807568877, 0, 0;
     72.0590870975572, 0, -0.5773502691896258, 1.732050807568877, 0;
     -65.60431108801717, 0, -0, -0.5773502691896258, -1.732050807568877;
     -0.3146182873523283, 0, -0, -0, 0.5287121214812434;
     -0.7278845221914594, 0, -0.5773502691896258, -0, 0.305252085661014;
     -2.34461828735233, 0, -0, -0, 0.5287121214812434;
     1.44536695831506, 0.4597008433809831, -0, -0, 0;
     2.253195474436084, 0, -0, -0.5773502691896258, 0.1762373738270812;
     2.665366958315059, 0.4597008433809831, -0, -0, 0;
     -0.2368045255639181, 0, -0, -0.5773502691896258, 0.1762373738270812;
     1.332115477808543, 0, -0.5773502691896258, -0, 0.305252085661014];
rss = 0;
p = zeros(24, 1);
[rssOut, idf, b, se, cov, svd, irank, pOut, ifail] = g02dd(n, ip, q, rss,
p)
```

```
rssOut =
   22.2268
idf =
          8
b =
   36.0033
   37.3000
   41.6033
   37.8767
se =
    0.9623
    0.9623
    0.9623
    0.9623
cov =
    0.9261
         0
    0.9261
         0
         0
    0.9261
         0
         0
         0
    0.9261
svd =
     0
irank =
          4
pOut =
     0
     0
     0
     0
     0
```

```
      0
      0
      0
      0
      0
      0
      0
      0
      0
      0
      0
      0
      0
      0
      0
      0
      0
      0
      0
ifail =
           0
```