NAG Library Routine Document

G02FCF

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

G02FCF calculates the Durbin–Watson statistic, for a set of residuals, and the upper and lower bounds for its significance.

2 Specification

```
SUBROUTINE G02FCF(N, IP, RES, D, PDL, PDU, WORK, IFAIL)
INTEGER             N, IP, IFAIL
double precision    RES(N), D, PDL, PDU, WORK(N)
```

3 Description

For the general linear regression model

$$y = X\beta + \epsilon,$$

where y is a vector of length n of the dependent variable,

X is a n by p matrix of the independent variables,

β is a vector of length p of unknown parameters,

and ϵ is a vector of length n of unknown random errors.

The residuals are given by

$$r = y - \hat{y} = y - X\hat{\beta}$$

and the fitted values, $\hat{y} = X\hat{\beta}$, can be written as Hy for a n by n matrix H. Note that when a mean term is included in the model the sum of the residuals is zero. If the observations have been taken serially, that is $y_1, y_2, \ldots, y_n$ can be considered as a time series, the Durbin–Watson test can be used to test for serial correlation in the ϵ_i, see Durbin and Watson (1950), Durbin and Watson (1951) and Durbin and Watson (1971).

The Durbin–Watson statistic is

$$d = \frac{\sum_{i=1}^{n-1} (r_{i+1} - r_i)^2}{\sum_{i=1}^{n} r_i^2}.$$

Positive serial correlation in the ϵ_i will lead to a small value of d while for independent errors d will be close to 2. Durbin and Watson show that the exact distribution of d depends on the eigenvalues of the matrix HA where the matrix A is such that d can be written as

$$d = \frac{r^{\mathrm{T}} A r}{r^{\mathrm{T}} r}$$

and the eigenvalues of the matrix A are $\lambda_j = (1 - \cos(\pi j/n))$, for $j = 1, 2, \ldots, n-1$.

However bounds on the distribution can be obtained, the lower bound being

$$d_{\mathrm{l}} = \frac{\sum_{i=1}^{n-p} \lambda_i u_i^2}{\sum_{i=1}^{n-p} u_i^2}$$

and the upper bound being

$$d_{\mathrm{u}} = \frac{\sum_{i=1}^{n-p} \lambda_{i-1+p} u_i^2}{\sum_{i=1}^{n-p} u_i^2},$$

where the u_i are independent standard Normal variables. The lower tail probabilities associated with these bounds, p_{l} and p_{u}, are computed by G01EPF. The interpretation of the bounds is that, for a test of size (significance) α, if $p_l \leq \alpha$ the test is significant, if $p_u > \alpha$ the test is not significant, while if $p_{\mathrm{l}} > \alpha$ and $p_{\mathrm{u}} \leq \alpha$ no conclusion can be reached.

The above probabilities are for the usual test of positive auto-correlation. If the alternative of negative auto-correlation is required, then a call to G01EPF should be made with the parameter D taking the value of $4 - d$; see Newbold (1988).

4 References

Durbin J and Watson G S (1950) Testing for serial correlation in least-squares regression. I *Biometrika* **37** 409–428

Durbin J and Watson G S (1951) Testing for serial correlation in least-squares regression. II *Biometrika* **38** 159–178

Durbin J and Watson G S (1971) Testing for serial correlation in least-squares regression. III *Biometrika* **58** 1–19

Granger C W J and Newbold P (1986) *Forecasting Economic Time Series* (2nd Edition) Academic Press

Newbold P (1988) *Statistics for Business and Economics* Prentice–Hall

5 Parameters

1: N – INTEGER *Input*

On entry: n, the number of residuals.

Constraint: N > IP.

2: IP – INTEGER *Input*

On entry: p, the number of independent variables in the regression model, including the mean.

Constraint: IP ≥ 1.

3: RES(N) – ***double precision*** array *Input*

On entry: the residuals, $r_1, r_2, \ldots, r_n$.

Constraint: the mean of the residuals $\leq \sqrt{\epsilon}$, where $\epsilon =$ ***machine precision***.

4: D – ***double precision*** *Output*

On exit: the Durbin–Watson statistic, d.

5: PDL – ***double precision*** *Output*

On exit: lower bound for the significance of the Durbin–Watson statistic, p_{l}.

6: PDU – ***double precision*** *Output*

On exit: upper bound for the significance of the Durbin–Watson statistic, p_{u}.

7: WORK(N) – ***double precision*** array *Workspace*

8: 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.

On exit: IFAIL $= 0$ unless the routine detects an error (see Section 6).

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 parameter, the recommended value is 0. **When the value -1 or 1 is used it is essential to test the value of IFAIL on exit.**

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).

Errors or warnings detected by the routine:

IFAIL $= 1$

On entry, N $\leq$ IP,
or IP < 1.

IFAIL $= 2$

On entry, the mean of the residuals was $> \sqrt{\epsilon}$, where $\epsilon =$ ***machine precision***.

IFAIL $= 3$

On entry, all residuals are identical.

7 Accuracy

The probabilities are computed to an accuracy of at least 4 decimal places.

8 Further Comments

If the exact probabilities are required, then the first $n - p$ eigenvalues of HA can be computed and G01JDF used to compute the required probabilities with the parameter C set to 0.0 and the parameter D set to the Durbin–Watson statistic d.

9 Example

A set of 10 residuals are read in and the Durbin–Watson statistic along with the probability bounds are computed and printed.

9.1 Program Text

```
*     G02FCF Example Program Text
*     Mark 15 Release. NAG Copyright 1991.
*     .. Parameters ..
      INTEGER          NIN, NOUT
      PARAMETER        (NIN=5,NOUT=6)
      INTEGER          N
      PARAMETER        (N=10)
*     .. Local Scalars ..
      DOUBLE PRECISION D, PDL, PDU
      INTEGER          I, IFAIL, IP
*     .. Local Arrays ..
      DOUBLE PRECISION RES(N), WORK(N)
*     .. External Subroutines ..
      EXTERNAL         G02FCF
*     .. Executable Statements ..
      WRITE (NOUT,*) 'G02FCF Example Program Results'
*     Skip heading in data file
      READ (NIN,*)
      READ (NIN,*) IP
      READ (NIN,*) (RES(I),I=1,N)
*
      IFAIL = 1
*
      CALL G02FCF(N,IP,RES,D,PDL,PDU,WORK,IFAIL)
*
      IF (IFAIL.EQ.0) THEN
         WRITE (NOUT,*)
         WRITE (NOUT,99999) ' Durbin-Watson statistic ', D
         WRITE (NOUT,*)
         WRITE (NOUT,99998) ' Lower and upper bound ', PDL, PDU
      ELSE
         WRITE (NOUT,*)
         WRITE (NOUT,99997) ' ** G02FCF returned with IFAIL = ', IFAIL
      END IF
*
99999 FORMAT (1X,A,F10.4)
99998 FORMAT (1X,A,2F10.4)
99997 FORMAT (1X,A,I5)
      END
```

9.2 Program Data

```
G02FCF Example Program Data
 2
3.735719 0.912755 0.683626 0.416693 1.9902
-0.444816 -1.283088 -3.666035 -0.426357 -1.918697
```

9.3 Program Results

```
 G02FCF Example Program Results

  Durbin-Watson statistic     0.9238

  Lower and upper bound     0.0610    0.0060
```