NAG Fortran Library Routine Document

G01EPF

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

G01EPF calculates upper and lower bounds for the significance of a Durbin–Watson statistic.

2 Specification

```
SUBROUTINE G01EPF(N, IP, D, PDL, PDU, WORK, IFAIL)
INTEGER           N, IP, IFAIL
real              D, PDL, PDU, WORK(N)
```

3 Description

Let $r = (r_1, r_2, \ldots, r_n)^T$ be the residuals from a linear regression of y on p independent variables, including the mean, where the y values $y_1, y_2, \ldots, y_n$ can be considered as a time series. The Durbin–Watson test (see Durbin and Watson (1950), Durbin and Watson (1951) and Durbin and Watson (1971)) can be used to test for serial correlation in the error term in the regression.

The Durbin–Watson test statistic is:

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

which can be written as

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

where the n by n matrix A is given by

$$A = \begin{bmatrix} 1 & -1 & 0 & \ldots & : \\ -1 & 2 & -1 & \ldots & : \\ 0 & -1 & 2 & \ldots & : \\ : & 0 & -1 & \ldots & : \\ : & : & : & \ldots & : \\ : & : & : & \ldots & -1 \\ 0 & 0 & 0 & \ldots & 1 \end{bmatrix}$$

with the non-zero eigenvalues of the matrix A being $\lambda_j = (1 - \cos(\pi j/n))$, for $j = 1, 2, \ldots, n-1$.

Durbin and Watson show that the exact distribution of d depends on the eigenvalues of a matrix HA, where H is the hat matrix of independent variables, i.e., the matrix such that the vector of fitted values, $\hat{y}$, can be written as $\hat{y} = Hy$. However, bounds on the distribution can be obtained, the lower bound being

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

and the upper bound being

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

where u_i are independent standard Normal variables.

Two algorithms are used to compute the lower tail (significance level) probabilities, p_l and p_u, associated with d_l and d_u. If $n \leq 60$ the procedure due to Pan (1964) is used, see Farebrother (1980), otherwise Imhof's method (Imhof (1961)) is used .

The bounds are for the usual test of positive correlation; if a test of negative correlation is required the value of d should be replaced by $4 - d$.

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

Farebrother R W (1980) Algorithm AS 153. Pan's procedure for the tail probabilities of the Durbin–Watson statistic *Appl. Statist.* **29** 224–227

Imhof J P (1961) Computing the distribution of quadratic forms in Normal variables *Biometrika* **48** 419–426

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

Pan Jie-Jian (1964) Distributions of the noncircular serial correlation coefficients *Shuxue Jinzhan* **7** 328–337

5 Parameters

1: N – INTEGER *Input*

On entry: the number of observations used in calculating the Durbin–Watson statistic, n.

Constraint: N > IP.

2: IP – INTEGER *Input*

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

Constraint: IP $\geq$ 1.

3: D – ***real*** *Input*

On entry: the Durbin–Watson statistic, d.

Constraint: D $\geq$ 0.0.

4: PDL – ***real*** *Output*

On exit: lower bound for the significance of the Durbin–Watson statistic, p_l.

5: PDU – ***real*** *Output*

On exit: upper bound for the significance of the Durbin–Watson statistic, p_u.

6: WORK(N) – ***real*** array *Workspace*

7: IFAIL – INTEGER *Input/Output*

On entry: IFAIL must be set to 0, -1 or 1. Users who are unfamiliar with this parameter should refer to Chapter P01 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, for users 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, D < 0.0.

7 Accuracy

On successful exit at least 4 decimal places of accuracy are achieved.

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 C set to 0.0 and D to the Durbin–Watson statistic.

9 Example

The values of n, p and the Durbin–Watson statistic d are input and the bounds for the significance level calculated and printed.

9.1 Program Text

Note: the listing of the example program presented below uses ***bold italicised*** terms to denote precision-dependent details. Please read the Users' Note for your implementation to check the interpretation of these terms. As explained in the Essential Introduction to this manual, the results produced may not be identical for all implementations.

```
*     G01EPF Example Program Text
*     Mark 15 Release. NAG Copyright 1991.
*     .. Parameters ..
      INTEGER          NIN, NOUT
      PARAMETER        (NIN=5,NOUT=6)
      INTEGER          NMAX
      PARAMETER        (NMAX=10)
*     .. Local Scalars ..
      real             D, PDL, PDU
      INTEGER          IFAIL, IP, N
*     .. Local Arrays ..
      real             WORK(NMAX)
*     .. External Subroutines ..
      EXTERNAL         G01EPF
*     .. Executable Statements ..
      WRITE (NOUT,*) 'G01EPF Example Program Results '
*     Skip heading in data file
      READ (NIN,*)
      READ (NIN,*) N, IP, D
*
      IF (N.LE.NMAX) THEN
         IFAIL = 0
*
```

```
         CALL G01EPF(N,IP,D,PDL,PDU,WORK,IFAIL)
*
         WRITE (NOUT,*)
         WRITE (NOUT,99999) ' Durbin-Watson statistic ', D
         WRITE (NOUT,*)
         WRITE (NOUT,99998) ' Probability for the lower bound = ', PDL
         WRITE (NOUT,99998) ' Probability for the upper bound = ', PDU
      ELSE
         WRITE (NOUT,*) ' N is larger than NMAX'
      END IF
      STOP
*
99999 FORMAT (1X,A,F10.4)
99998 FORMAT (1X,A,F10.4)
      END
```

9.2 Program Data

```
G01EPF Example Program Data
 10 2 0.9238
```

9.3 Program Results

```
 G01EPF Example Program Results

  Durbin-Watson statistic     0.9238

  Probability for the lower bound =     0.0610
  Probability for the upper bound =     0.0060
```