NAG Toolbox for MATLAB

g02fc

1 Purpose

g02fc calculates the Durbin–Watson statistic, for a set of residuals, and the upper and lower bounds for its significance.

2 Syntax

```
[d, pdl, pdu, ifail] = g02fc(ip, res, 'n', 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

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

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

$$d_1 = \frac{\sum_{i=1}^{n-p}\lambda_i u_i^2}{\sum_{i=1}^{n-p} u_i^2}$$

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 g01ep. 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 g01ep 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

5.1 Compulsory Input Parameters

1: **ip – int32 scalar**

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

Constraint: **ip** ≥ 1.

2: **res(n) – double array**

The residuals, $r_1, r_2, \ldots, r_n$.

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

5.2 Optional Input Parameters

1: **n – int32 scalar**

Default: The dimension of the array **res**.

n, the number of residuals.

Constraint: **n** $>$ **ip**.

5.3 Input Parameters Omitted from the MATLAB Interface

work

5.4 Output Parameters

1: **d – double scalar**

The Durbin–Watson statistic, d.

2: **pdl – double scalar**

Lower bound for the significance of the Durbin–Watson statistic, p_l.

3: **pdu – double scalar**

Upper bound for the significance of the Durbin–Watson statistic, p_u.

4: **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, $\mathbf{n} \leq \mathbf{ip}$,
or $\mathbf{ip} < 1$.

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 g01jd 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

```
ip = int32(2);
res = [3.735719;
     0.912755;
     0.683626;
     0.416693;
     1.9902;
     -0.444816;
     -1.283088;
     -3.666035;
     -0.426357;
     -1.918697];
[d, pdl, pdu, ifail] = g02fc(ip, res)
```

```
d =
```

```
    0.9238
pdl =
    0.0610
pdu =
    0.0060
ifail =
           0
```