1 Purpose

NAG\_\text{prob\_durbin\_watson} (g01epc) calculates upper and lower bounds for the significance of a Durbin–Watson statistic.

2 Specification

\[
\text{void nag\_prob\_durbin\_watson (Integer n, Integer ip, double d, double *pdl, double *pdu, NagError *fail)}
\]

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

\[
d = \frac{r^T Ar}{r^T r},
\]

where the \( n \) by \( n \) matrix \( A \) is given by

\[
A = \begin{bmatrix}
1 & -1 & 0 & \cdots \\
-1 & 2 & -1 & \cdots \\
0 & -1 & 2 & \cdots \\
\vdots & 0 & -1 & \cdots \\
\vdots & \vdots & \vdots & \ddots \\
0 & 0 & 0 & \cdots & 1
\end{bmatrix}
\]

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
409–428
Durbin J and Watson G S (1951) Testing for serial correlation in least-squares regression. II Biometrika 38
159–178
1–19
statistic Appl. Statist. 29 224–227
419–426
Pan Jie–Jian (1964) Distributions of the noncircular serial correlation coefficients Shuxue Jinzhan 7
328–337

5 Parameters
1: \( n \) – Integer
   \( \text{Input} \)
   On entry: the number of observations used in calculating the Durbin–Watson statistic, \( n \).
   Constraint: \( n > ip \).
2: \( ip \) – Integer
   \( \text{Input} \)
   On entry: the number, \( p \), of independent variables in the regression model, including the mean.
   Constraint: \( ip \geq 1 \).
3: \( d \) – double
   \( \text{Input} \)
   On entry: the Durbin–Watson statistic, \( d \).
   Constraint: \( d \geq 0.0 \).
4: \( pdl \) – double *
   \( \text{Output} \)
   On exit: lower bound for the significance of the Durbin–Watson statistic, \( p_l \).
5: \( pdu \) – double *
   \( \text{Output} \)
   On exit: upper bound for the significance of the Durbin–Watson statistic, \( p_u \).
6: \( fail \) – NagError *
   \( \text{Input/Output} \)
   The NAG error parameter (see the Essential Introduction).

6 Error Indicators and Warnings

NE_INT
   On entry, \( ip = \langle \text{value} \rangle \).
   Constraint: \( ip \geq 1 \).
NE_INT_2
On entry, \( n \) = \( \text{(value)} \), \( ip \) = \( \text{(value)} \).
Constraint: \( n > ip \).

NE_REAL
On entry, \( d \) = \( \text{(value)} \).
Constraint: \( d \geq 0.0 \).

NE_ALLOC_FAIL
Memory allocation failed.

NE_BAD_PARAM
On entry, parameter \( \text{(value)} \) had an illegal value.

NE_INTERNAL_ERROR
An internal error has occurred in this function. Check the function call and any array sizes. If the call is correct then please consult NAG for assistance.

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 nag_prob_lin_chi_sq (g01jdc) 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
/* nag_prob_durbin_watson (g01epc) Example Program. *
 * Copyright 2001 Numerical Algorithms Group.
 * * Mark 7, 2001.
 * */
#include <stdio.h>
#include <nag.h>
#include <nag_stdlib.h>
#include <nagg01.h>

int main(void)
{
    /* Scalars */
    double d, pdl, pdu;
    Integer exit_status, ip, n;
    NagError fail;

    INIT_FAIL(fail);
    exit_status = 0;
    Vprintf("g01epc Example Program Results\n");

    /* Skip heading in data file */
Vscanf("%*[\n] ");
Vscanf("%ld%ld%lf%*[\n] ", &n, &ip, &d);
g0lepc(n, ip, d, &pdl, &pdu, &fail);
if (fail.code != NE_NOERROR)
{
    Vprintf("Error from g0lepc.\n%\n", fail.message);
    exit_status = 1;
    goto END;
}
Vprintf("\n");
Vprintf(" Durbin-Watson statistic %10.4f\n", d);
Vprintf(" Probability for the lower bound = %10.4f\n", pdl);
Vprintf(" Probability for the upper bound = %10.4f\n", pdu);
END:
  return exit_status;
}

9.2 Program Data

9.3 Program Results