Cumulative distribution function of the Durbin-Watson statistic

Stanisław Galus

9 July 2012

Let us consider the linear regression model

$$y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik} + \xi_i,$$

$i = 1, \ldots, n$, $k \geq 1$, $n > k + 1$, where ξ_i are independent identically distributed normal variables with zero mean, or $y = X\beta + \xi$. The least squares estimate of β is $b = (X^T X)^{-1} X^T y$ and the vector of residuals from the regression is $e = y - Xb$. The Durbin-Watson d statistic is defined by

$$d = \frac{\sum_{i=2}^{n} (e_i - e_{i-1})^2}{\sum_{i=1}^{n} e_i^2}$$

It is shown in [1, p. 426] that $d_L \leq d \leq d_U$, where

$$d_L = \frac{\sum_{i=1}^{n-k-1} \lambda_i \zeta_i^2}{\sum_{i=1}^{n-k-1} \zeta_i^2},$$

$$d_U = \frac{\sum_{i=1}^{n-k-1} \lambda_{i+k} \zeta_i^2}{\sum_{i=1}^{n-k-1} \zeta_i^2},$$

$$\lambda_i = 2\left(1 - \cos\frac{\pi i}{n}\right), \quad i = 1, \ldots, n-1,$$

and $\zeta_1, \ldots, \zeta_{n-1}$ are independent standard normal variables. The cumulative distribution functions of d_L and d_U are

$$P(d_L < x) = P\left(\sum_{i=1}^{n-k-1} (x - \lambda_i)\zeta_i^2 > 0\right), \tag{1}$$

$$P(d_U < x) = P\left(\sum_{i=1}^{n-k-1} (x - \lambda_{i+k})\zeta_i^2 > 0\right). \tag{2}$$

In both cases, the probabilities (1) and (2) are of the form

$$P\left(\sum_{i=1}^{m} w_i \zeta_i^2 > 0\right),$$

where w_i, $i = 1, \ldots, m$, $m > 0$, are those $x - \lambda_i$ or $x - \lambda_{i+k}$ which are not equal to 0. They can be calculated by the formula (3.2) of [3, p. 422], which states that if $w_i \neq 0$, $i = 1, \ldots, m$, then

$$P\left(\sum_{i=1}^{m} w_i \zeta_i^2 > 0\right) = \frac{1}{2} + \frac{1}{\pi}\int_0^{+\infty} f(u)du,$$

where

$$f(u) = \frac{\sin\theta(u)}{u\rho(u)},$$
$$\theta(u) = \frac{1}{2}\sum_{i=1}^{m} \arctan(w_i u),$$
$$\rho(u) = \prod_{i=1}^{m}(1 + w_i^2 u^2)^{1/4}.$$

It is shown in (3.3) of [3, p. 423] that

$$\lim_{u\to 0} f(u) = \frac{1}{2}\sum_{i=1}^{m} w_i,$$
$$\lim_{u\to\infty} \theta(u) = \frac{\pi}{4}\sum_{i=1}^{m} \operatorname{sgn} w_i,$$
$$|\frac{1}{\pi}\int_u^{+\infty} f(u)du| \leq \frac{1}{\pi}\frac{2}{m}u^{-m/2}\prod_{i=1}^{m}|w_i|^{-1/2}. \tag{3}$$

To achieve the result *prob* and the accuracy *eps* such that

$$prob - eps \leq P\left(\sum_{i=1}^{m} w_i \zeta_i^2 > 0\right) < prob + eps \tag{4}$$

we put

$$u \geq (\frac{1}{\pi}\frac{2}{m}(\prod_{i=1}^{m}|w_i|^{-1/2})/(eps/2))^{2/m} \tag{5}$$

so that by (3) we have

$$|\frac{1}{\pi}\int_u^{+\infty} f(u)du| \leq eps/2$$

and then calculate

$$\frac{1}{\pi}\int_0^{u} f(u)du \tag{6}$$

by the Simpson method until two consecutive approximations differ no more than $eps/2$. To have four significant digits after the comma, we put $eps = 0.000049$.

The function should not be used for low values of $n - k - 1$, say 1, 2, 3, as integration (6) fails. For $n - k - 1 = 1$ we have

$$\begin{aligned}
P(d_L < x) &= 0 \quad \text{for} \quad x \leq \lambda_1, \\
P(d_L < x) &= 1 \quad \text{for} \quad x > \lambda_1, \\
P(d_U < x) &= 0 \quad \text{for} \quad x \leq \lambda_{1+k}, \\
P(d_U < x) &= 1 \quad \text{for} \quad x > \lambda_{1+k}.
\end{aligned}$$

For possible solution to this problem, see [2]. See also [4].

References

[1] J. Durbin and G. S. Watson. Testing for serial correlation in least squares regression. I. *Biometrika*, 37:409–428, 1950.

[2] R. W. Farebrother. Remark AS R52: The distribution of a linear combination of central χ^2 random variables: A remark on AS 153: Pan's procedure for the tail probabilities of the Durbin-Watson statistic. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 33(3):363–366, 1984.

[3] J. P. Imhof. Computing the distribution of quadratic forms in normal variables. *Biometrika*, 48:419–426, 1961.

[4] N. E. Savin and K. J. White. The Durbin-Watson test for serial correlation with extreme sample sizes or many regressors. *Econometrica*, 45:1989–1996, 1977.