underlying regression model is linear, and $\boldsymbol{X}$ contains only fixed regressors. This distribution necessarily depends on $\boldsymbol{X}$. The calculation uses the fact that the d statistic can be written as

$$\frac{\boldsymbol{u}^{\top}\boldsymbol{M}_X\boldsymbol{A}\boldsymbol{M}_X\boldsymbol{u}}{\boldsymbol{u}^{\top}\boldsymbol{M}_X\boldsymbol{u}}, \tag{10.82}$$

where $\boldsymbol{A}$ is the $n \times n$ matrix

$$\begin{bmatrix} 1 & -1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & -1 & 2 & -1 \\ 0 & 0 & 0 & 0 & \cdots & 0 & -1 & 1 \end{bmatrix}.$$

From (10.82), the d statistic is seen to be a ratio of quadratic forms in normally distributed random variables, and the distributions of such ratios can be evaluated using several numerical techniques; see Durbin and Watson (1971) and Savin and White (1977) for references.

Most applied workers never attempt to calculate the exact distribution of the d statistic corresponding to their particular $\boldsymbol{X}$ matrix. Instead, they use the fact that the critical values for its distribution are known to fall between two bounding values, d_L and d_U, which depend on the sample size, n, the number of regressors, k, and whether or not there is a constant term. Tables of d_L and d_U may be found in some econometrics textbooks and in papers such as Durbin and Watson (1951) and Savin and White (1977). As an example, when $n = 50$ and $k = 6$ (counting the constant term as one of the regressors), for a test against $\rho > 0$ at the .05 level, $d_L = 1.335$ and $d_U = 1.771$. Thus, if one calculated a d statistic for this sample size and number of regressors and it was less than 1.335, one could confidently decide to reject the null hypothesis of no serial correlation at the .05 level. If the statistic was greater than 1.771, one could confidently decide not to reject. However, if the statistic was in the "inconclusive region" between 1.335 and 1.771, one would be unsure of whether to reject or not. When the sample size is small, and especially when it is small relative to the number of regressors, the inconclusive region can be very large. This means that the d statistic may not be very informative when used in conjunction with the tables of d_L and d_U.[9] In such cases, one may have no choice but to calculate the exact distribution of the statistic, if one wants to make inferences from the d statistic in a small sample. A few software packages, such as SHAZAM, allow one to do this. Of course,

[9] There is reason to believe that when the regressors are slowly changing, a situation which may often be the case with time-series data, d_U provides a better approximation than d_L. See Hannan and Terrell (1966).