

Acta Cryst. (1980). A36, 495–496

Testing for serial correlation in intensity data. By H. D. FLACK and M. G. VINCENT,* *Laboratoire de Cristallographie aux Rayons X, Université de Genève, 24 Quai Ernest Ansermet, CH-1211 Genève 4, Switzerland* and J. A. VINCENT, *4 South View Drive, South Woodford, London E18, England*

(Received 8 November 1979; accepted 26 November 1979)

Abstract

An example of the use of the Durbin–Watson d statistic to test for positive and negative serial correlation is presented. It is found that the diffractometer data tested do not suffer from serial correlation.

One of the basic assumptions of the ordinary least-squares model is that the error or disturbance terms contained within certain successive experimentally determined values are independent, *i.e.* they are uncorrelated with each other. Statistically, one says there is serial independence of disturbance or error terms.

It can be shown (this is proved in the Appendix) that even when the assumption of serial independence is not true, the results of applying the ordinary least-squares method to data still give unbiased estimates of the parameters, but that the variances of these estimates are liable to serious error. Because of this, it is of some importance to establish the presence or absence of serial correlation between disturbance terms and this is most easily done with the Durbin–Watson (D–W) d statistic (Durbin & Watson, 1950, 1951). This statistic is widely used in econometrics and it is common practice when publishing econometric work to quote the relevant D–W figure.

We have recently been treating a set of intensity data of very high quality and in the course of the analysis it seemed desirable either to establish the absence of serial correlation or to correct for it in some suitable manner. In this note we present the necessary details of the calculation.

The intensity data were measured on a sphere of Mg_2Si [radius 63(1) μm , $\mu(\text{Ag } K\alpha) = 0.479 \text{ mm}^{-1}$] out to $\sin \theta/\lambda = 1.64 \text{ \AA}^{-1}$ with $\text{Ag } K\alpha$ radiation reflected from a graphite monochromator. A Philips PW1100 four-circle diffractometer was used in the background–peak–background mode with an ω – 2θ scan. Three reference reflections were measured at regular intervals (~ 100 min) and these observations were used to correct the other measurements for any long-term drift in the output of the X-ray tube (a maximum drift of 1.5% was observed). The 5320 intensity measurements were corrected for Lorentz, polarization and absorption factors and converted to $|F|$ values. These were used in a least-squares refinement of nine parameters: one scale factor, two isotropic temperature factors and six anisotropic extinction parameters of type 1 with a Lorentzian distribution (Becker & Coppens, 1974, 1975). A weighting scheme $1/\sigma_{F_o}^2$ was used. The following values of certain statistical parameters were obtained; goodness of fit = 1.38,

$$\Delta F = |F_o| - |F_c|, w = 1/\sigma_{F_o}^2,$$

$$R(F) = \sum \Delta F / \sum |F_o| = 2.4\%,$$

$$R_w(F) = (\sum w \Delta F^2 / \sum w F_o^2)^{1/2} = 1.5\%,$$

$$R(J) = \sum |I_o - I_c| / \sum I_o = 1.8\%.$$

These calculations were carried out with the XRAY system (1976).

The D–W d statistic is calculated from

$$d = \sum_{i=2}^N (\Delta F_i - \Delta F_{i-1})^2 / \sum_{i=1}^N \Delta F_i^2,$$

where N is the number of observations. This value is used in their statistical test based on the null hypothesis that there is no serial correlation between terms, *viz.* that the disturbance terms are independent of each other. Durbin & Watson's (1951) tables of values only extend to 100 observations and five least-squares parameters or 'explanatory variables'. Theil & Nagar (1961) show, however, that when $N > 100$, d has an approximately normal distribution. In these circumstances the value of d may be tested against significance points (Theil & Nagar, 1961) calculated with the formula

$$Q = 2 \left(\frac{N-1}{N-M} - \frac{p}{\sqrt{N+2}} \right),$$

where M is the number of least-squares parameters estimated and p are the ordinates for the one-tailed cumulative normal distribution at the significance level required: for 0.1%, $p = 3.0902$; for 1%, $p = 2.32635$ and for 5%, $p = 1.64485$. The D–W test is based on a linear least-squares model. In the present case, the model is non-linear but has been linearized by use of a first-order Taylor expansion. For this reason, it is as well to be very conservative in applying the test by choosing a significance level of 0.1%. In the test for positive serial correlation, if $d > Q$ the null hypothesis cannot be rejected at the chosen significance level (*i.e.* $d > Q$, there is no positive serial correlation), whereas if $d \leq Q$ the null hypothesis is rejected in favour of an alternative hypothesis ($d \leq Q$, there is evidence at the chosen level of significance that there is positive serial correlation). Durbin & Watson (1950) show that to test for negative serial correlation d is replaced by $4 - d$ in the above test (*i.e.* $4 - d > Q$, there is no negative serial correlation).

* Present address: Biozentrum der Universität Basel, Klingelbergstrasse 70, CH4056 Basel, Switzerland.

† In positive serial correlation, successive values of the residuals ΔF_i tend to have the same sign. In negative serial correlation, the tendency is for opposite signs to occur.

For our intensity data with $N = 5320$, $M = 9$ we find $d = 1.997$, $4 - d = 2.003$ and the percentage point 0.1%: $Q = 1.918$, 1%: $Q = 1.939$, 5%: $Q = 1.958$. Thus we cannot reject the null hypothesis of serial independence, *viz.* we find no evidence of serial correlation in our data set.

In conclusion it would seem desirable to incorporate the test for serial correlation into crystallographic least-squares programs as a means of detecting unsuspected errors in the data set. In the main, these could either be associated with the treatment of the reference reflections or be due to any long-term instability in the measuring apparatus or crystal.

APPENDIX

Proof of statement that even where there is serial correlation the ordinary least-squares method gives unbiased estimates of the parameters

Consider the simple linear relationship

$$Y_t = \alpha + \beta X_t + U_t, \quad (A1)$$

where α and β are parameters and U_t is a disturbance or error term. It is assumed for simplicity that U_t follows a first-order Markov auto-regressive scheme, *i.e.*

$$U_t = \rho U_{t-1} + e_t$$

where $|\rho| < 1$ and e_t is an individual error disturbance term with the expectations that

$$\left. \begin{aligned} E(e_t) &= 0 \\ E(e_t \cdot e_{t+s}) &= \sigma_e^2 \text{ when } s = 0 \\ &= 0 \text{ when } s \neq 0 \end{aligned} \right\} \text{for all } t.$$

Then

$$\begin{aligned} U_t &= \rho(U_{t-1}) + e_t \\ &= \rho(\rho U_{t-2} + e_{t-1}) + e_t \\ &= \rho^2(\rho U_{t-3} + e_{t-2}) + \rho e_{t-1} + e_t \\ &= \rho^3(\rho U_{t-4} + e_{t-3}) + \rho^2 e_{t-2} + \rho e_{t-1} + e_t \end{aligned}$$

$= \rho^4(\rho U_{t-5} + e_{t-4}) + \rho^3 e_{t-3} + \rho^2 e_{t-2} + \rho e_{t-1} + e_t$
and so on in this iterative manner.

$$U_t = e_t + \rho e_{t-1} + \rho^2 e_{t-2} + \rho^3 e_{t-3} + \dots + \rho^r e_{t-r} + \dots$$

or

$$U_t = \sum_{r=0}^{\infty} \rho^r e_{t-r}.$$

Since

$$E(e_t) = 0 \quad (\text{assumed above}),$$

it follows that

$$E(U_t) = E\left[\sum_{r=0}^{\infty} \rho^r e_{t-r}\right] = 0.$$

Taking expectations of (A1), then

$$E(Y_t) = E(\alpha + \beta X_t + U_t),$$

$$E(Y_t) = E(\alpha) + E(\beta X_t) + E(U_t)$$

or

$$E(Y_t) = \alpha + \beta \mu_x + 0,$$

i.e. we still get unbiased estimates of the parameters even when U_t follows an auto-regressive scheme.

References

- BECKER, P. J. & COPPENS, P. (1974). *Acta Cryst.* **A30**, 129–153.
 BECKER, P. J. & COPPENS, P. (1975). *Acta Cryst.* **A31**, 417–425.
 DURBIN, J. & WATSON, G. S. (1950). *Biometrika*, **37**, 409–428.
 DURBIN, J. & WATSON, G. S. (1951). *Biometrika*, **38**, 159–178.
 THEIL, H. & NAGAR, A. L. (1961). *J. Am. Stat. Assoc.* **56**, 793–806.
 XRAY system (1976). Edited by J. M. STEWART. Technical Report TR-446, Computer Science Center, University of Maryland.

Acta Cryst. (1980). **A36**, 496

A new least-squares refinement technique based on the fast Fourier transform algorithm: *erratum*. By

RAMESH C. AGARWAL,* IBM T. J. Watson Research Center, Yorktown Heights, NY 10598, USA

(Received 12 December 1979)

Abstract

In Agarwal [*Acta Cryst.* (1978), **A34**, 791–809], equation (61) should read

$$c_3 = 2C_m^1 C_m^2.$$

All information is given in the *Abstract*.

* Present address: Centre for Applied Research in Electronics, Indian Institute of Technology, New Delhi 110029, India.