

CHAPTER 7

Serial Correlation in the Residuals and the Durbin–Watson Test

In what follows it is assumed that the time order (or some other type of order) of the observations is known, and that observations are equally spaced in that ordering.

7.1. SERIAL CORRELATION IN RESIDUALS

In regression work, we typically assume that the observational errors are pairwise uncorrelated. If this assumption were substantially untrue, we would expect that the plot of residuals in time order, or some other sensible order defined by the practical circumstances, would help us to detect it. There are, of course, many ways in which the errors may be correlated. A common way is that they may be *serially correlated*, that is, the correlations between errors s steps apart are always the same. We shall use the notation ρ_s for this correlation, $s = 1, 2, \dots$.

More specifically, if residuals exhibit local positive serial correlation, successive residuals in a time sequence tend to be more alike than otherwise, and a time plot of them will have the general characteristics of Figure 7.1*a*, rising and falling but with close points more alike than otherwise. The correlation between residuals one (or two, or three, ...) step(s) apart is called the lag-1 (or 2, or 3, ...) serial correlation. The empirical lag-1 serial correlation can be examined by plotting each residual except the first against the one preceding it. The positive lag-1 serial correlation present in the data of Figure 7.1*a* reveals itself in the “lower-left to upper-right” tendency of such a plot, shown in Figure 7.1*b*. To view correlations for higher lags we can make similar plots for residuals two steps apart, three steps apart, and so on.

Negative serial correlation between successive residuals can also arise. One cause is a phenomenon known as carryover, which occurs in batch processes. This can happen as follows. Suppose that, for a particular batch of product in a process, incomplete recovery occurs because some of the product is left in the pipelines and pumps of the reactor system. The recorded yield for this batch will be unusually low. In the next batch, however, there would be a tendency for the material left behind to be recovered, thus giving an unusually high batch yield. A pattern of residuals may result like those plotted in Figure 7.2*a*, in which a positive value tends to be followed by a negative one, and vice versa. The existence of negative lag-1 serial correlation for these data is shown by the “lower-right to upper-left” pattern of Figure 7.2*b*.

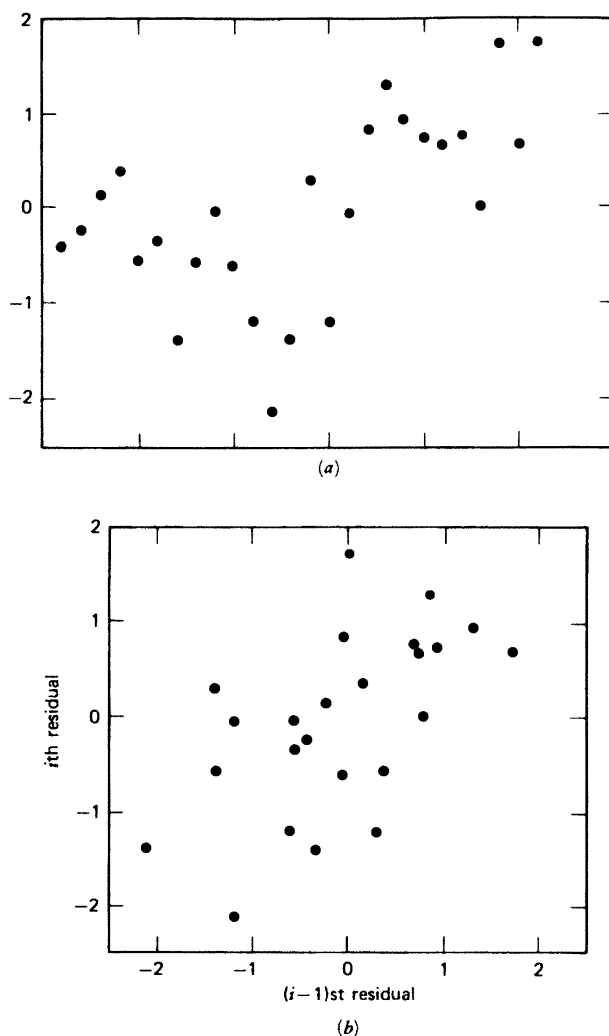


Figure 7.1. (a) A series of residuals exhibiting local positive serial correlation. (b) Lag-1 serial plot for this series.

Figure 7.3a shows a random series of residuals and Figure 7.3b shows the corresponding lag-1 serial plot that exhibits no tendency of trend at all.

We can characterize the behavior shown in the three figures as *attraction* (successive residuals are “like” one another but the plot wanders around), *repulsion* (successive residuals repel or are “unlike” one another), and *unrelated* (successive residuals are “almost independent”; they are of course related by the normal equations.)

The study of serial correlation patterns is one of the techniques used in time-series analysis. Such special analysis of correlated data can often be rewarding. The interested reader should look at texts such as Box, Jenkins and Reinsel (1994). Here we are concerned only with the detection of serial correlation in regression residuals.

A well-known way of checking for serial correlation patterns in an equally spaced sequence of residuals is via the Durbin–Watson test, which we now describe. A simpler, less sophisticated runs test is described in Section 7.3.

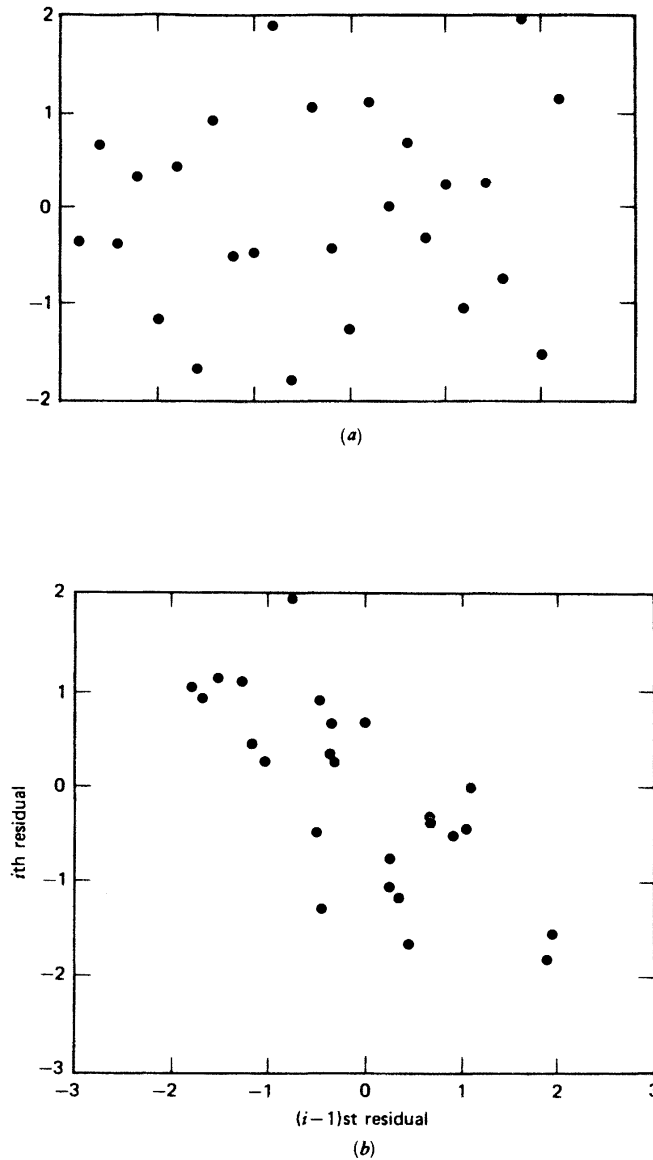


Figure 7.2. (a) A series of residuals exhibiting local negative serial correlation perhaps due to carryover. (b) Lag-1 serial plot for this series.

7.2. THE DURBIN-WATSON TEST FOR A CERTAIN TYPE OF SERIAL CORRELATION

A popular test for detecting a certain type of serial correlation is the *Durbin-Watson test*. (This is named after the two authors who discussed its use for testing regression residuals and provided suitable testing tables in 1951. It was originally put forward by Von Neumann for nonregression problems in 1941. Selected sources are listed at the end of the chapter.)

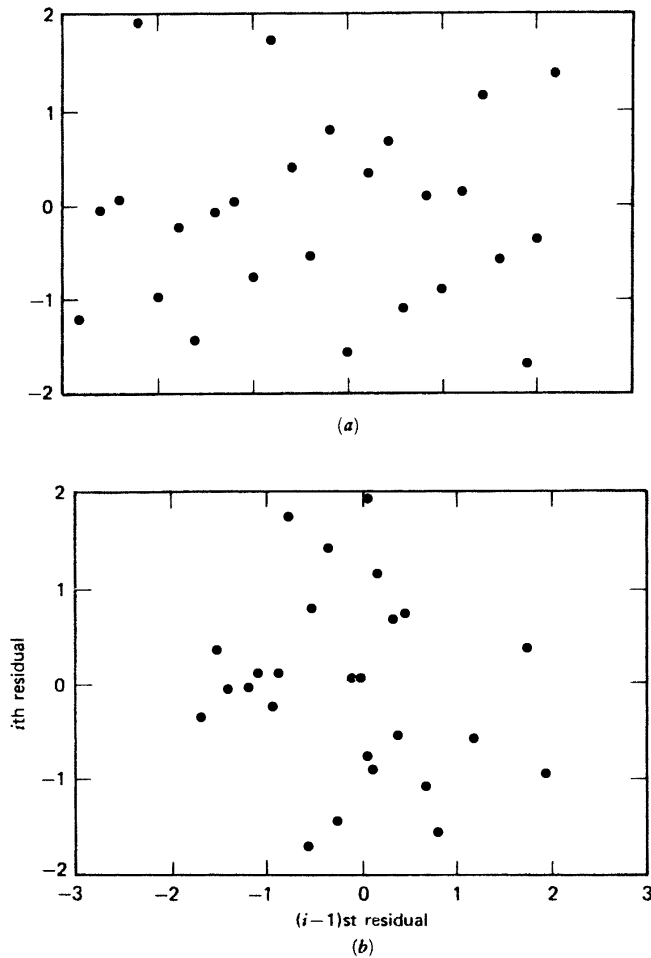


Figure 7.3. (a) An uncorrelated series of residuals. (b) Lag-1 serial plot for this series.

Suppose we wish to fit a postulated linear model

$$Y_u = \beta_0 + \sum_{i=1}^k \beta_i X_{iu} + \epsilon_u \quad (7.2.1)$$

by least squares to observations $(Y_u, X_{1u}, X_{2u}, \dots, X_{ku})$, $u = 1, 2, \dots, n$. We would usually assume that the errors ϵ_u are independent $N(0, \sigma^2)$ variables, so that all serial correlations $\rho_s = 0$. We want to see if this assumption is justified by checking the residuals. We shall test this null hypothesis H_0 : all $\rho_s = 0$ via the Durbin-Watson test against the alternative

$$H_1: \rho_s = \rho^s \quad (7.2.2)$$

($\rho \neq 0$ and $|\rho| < 1$), an alternative that arises from the assumption that the errors ϵ_u are such that

$$\epsilon_u = \rho \epsilon_{u-1} + z_u, \quad (7.2.3)$$

where $z_u \sim N(0, \sigma^2)$ and is independent of $\epsilon_{u-1}, \epsilon_{u-2}, \dots$, and of z_{u-1}, z_{u-2}, \dots . It is also assumed that both the mean and variance of ϵ_u are constant, independent of u ,

whereupon it follows, necessarily, that $\epsilon_u \sim N\{0, \sigma^2/(1 - \rho^2)\}$. Note that, under the null hypothesis $H_0: \rho = 0$, this reduces to $\epsilon_u \sim N(0, \sigma^2)$, our usual assumptions for all $u = 1, 2, \dots, n$.

Although the test is fully appropriate only against the specific alternative (7.2.2), it is typically applied in a general way without much thought of alternatives. This means that it will often lose power compared to when it is properly employed.

To test H_0 against H_1 we fit the model from Eq. (7.2.1) and find the residuals e_1, e_2, \dots, e_n . We then form the Durbin-Watson statistic

$$d = \sum_{u=2}^n (e_u - e_{u-1})^2 / \sum_{u=1}^n e_u^2 \quad (7.2.4)$$

and determine whether or not to reject the null hypothesis H_0 on the basis of the value of d . The distribution of d depends on the X -data and is not independent of them (as, for example, is the t -distribution, apart from degrees of freedom, in other regression contexts). The distribution of d lies between 0 and 4 and is symmetric about 2. Percentage points also depend on the X -data and would have to be calculated for each application to perform the test properly. Because of the difficulty of doing this routinely, the test is usually performed using tabled bounds (d_L, d_U), where L = lower and U = upper, on the percentage points. Thus instead of looking up a single critical value, we have to look up two critical values. Moreover, d is used only for a lower-tailed test against alternatives $\rho > 0$. To test against the alternative $\rho < 0$, we theoretically need an upper-tailed test; fortunately, this can simply be handled as a lower-tailed test using the statistic $(4 - d)$.

Note that the extremes 0 and 4 are attainable only for very large samples. The minimum attainable values depend on the sample size n in the following way:

n	15	30	50	100	200	300	500
Minimum d	0.0437	0.0110	0.0039	0.0010	0.0002	0.0001	0.0000

The corresponding maximum d values are $(4 - \text{minimum } d)$. This display and several portions of tables below are quoted and adapted from Savin and White (1977), fully referenced as (2) below.

Primary Test, Tables of d_L and d_U

Tables 7.1, 7.2, and 7.3 show pairs of 1%, 2.5%, and 5% lower-tail significance points, that is, critical values for probability levels $\alpha = 0.01, 0.025$, and 0.05 , respectively, called (d_L, d_U) . These are given for various numbers of observations n , and for $k = 1, 2, \dots, 20$ predictor variables X_i [see k in Eq. (7.2.1)]. The sources of these tables are two papers:

1. J. Durbin and G. S. Watson (1951). Testing for serial correlation in least squares regression, II. *Biometrika*, **38**, 159–178 (DW51).
2. N. E. Savin and K. J. White (1977). The Durbin-Watson test for serial correlation with extreme sample sizes or many regressors. *Econometrica*, **45**, 1989–1996 (SW77).

In our amended versions of tables reproduced from SW77, figures are rounded to two decimal places and some combinations of (n, k) values that result in relatively few residual degrees of freedom have been omitted. Also, we have omitted all d_U values that exceed 1.99. If the observed d statistic equaled or exceeded the distribution mean 2, there would be little point doing a lower-tailed test.

T A B L E 7.1. Significance Points of d_L and d_U : 1%

n	$k = 1$		$k = 2$		$k = 3$		$k = 4$		$k = 5$	
	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
15	0.81	1.07	0.70	1.25	0.59	1.46	0.49	1.70	0.39	1.96
16	0.84	1.09	0.74	1.25	0.63	1.44	0.53	1.66	0.44	1.90
17	0.87	1.10	0.77	1.25	0.67	1.43	0.57	1.63	0.48	1.85
18	0.90	1.12	0.80	1.26	0.71	1.42	0.61	1.60	0.52	1.80
19	0.93	1.13	0.83	1.26	0.74	1.41	0.65	1.58	0.56	1.77
20	0.95	1.15	0.86	1.27	0.77	1.41	0.68	1.57	0.60	1.74
21	0.97	1.16	0.89	1.27	0.80	1.41	0.72	1.55	0.63	1.71
22	1.00	1.17	0.91	1.28	0.83	1.40	0.75	1.54	0.66	1.69
23	1.02	1.19	0.94	1.29	0.86	1.40	0.77	1.53	0.70	1.67
24	1.04	1.20	0.96	1.30	0.88	1.41	0.80	1.53	0.72	1.66
25	1.05	1.21	0.98	1.30	0.90	1.41	0.83	1.52	0.75	1.65
26	1.07	1.22	1.00	1.31	0.93	1.41	0.85	1.52	0.78	1.64
27	1.09	1.23	1.02	1.32	0.95	1.41	0.88	1.51	0.81	1.63
28	1.10	1.24	1.04	1.32	0.97	1.41	0.90	1.51	0.83	1.62
29	1.12	1.25	1.05	1.33	0.99	1.42	0.92	1.51	0.85	1.61
30	1.13	1.26	1.07	1.34	1.01	1.42	0.94	1.51	0.88	1.61
31	1.15	1.27	1.08	1.34	1.02	1.42	0.96	1.51	0.90	1.60
32	1.16	1.28	1.10	1.35	1.04	1.43	0.98	1.51	0.92	1.60
33	1.17	1.29	1.11	1.36	1.05	1.43	1.00	1.51	0.94	1.59
34	1.18	1.30	1.13	1.36	1.07	1.43	1.01	1.51	0.95	1.59
35	1.19	1.31	1.14	1.37	1.08	1.44	1.03	1.51	0.97	1.59
36	1.21	1.32	1.15	1.38	1.10	1.44	1.04	1.51	0.99	1.59
37	1.22	1.32	1.16	1.38	1.11	1.45	1.06	1.51	1.00	1.59
38	1.23	1.33	1.18	1.39	1.12	1.45	1.07	1.52	1.02	1.58
39	1.24	1.34	1.19	1.39	1.14	1.45	1.09	1.52	1.03	1.58
40	1.25	1.34	1.20	1.40	1.15	1.46	1.10	1.52	1.05	1.58
45	1.29	1.38	1.24	1.42	1.20	1.48	1.16	1.53	1.11	1.58
50	1.32	1.40	1.28	1.45	1.24	1.49	1.20	1.54	1.16	1.59
55	1.36	1.43	1.32	1.47	1.28	1.51	1.25	1.55	1.21	1.59
60	1.38	1.45	1.35	1.48	1.32	1.52	1.28	1.56	1.25	1.60
65	1.41	1.47	1.38	1.50	1.35	1.53	1.31	1.57	1.28	1.61
70	1.43	1.49	1.40	1.52	1.37	1.55	1.34	1.58	1.31	1.61
75	1.45	1.50	1.42	1.53	1.39	1.56	1.37	1.59	1.34	1.62
80	1.47	1.52	1.44	1.54	1.42	1.57	1.39	1.60	1.36	1.62
85	1.48	1.53	1.46	1.55	1.43	1.58	1.41	1.60	1.39	1.63
90	1.50	1.54	1.47	1.56	1.45	1.59	1.43	1.61	1.41	1.64
95	1.51	1.55	1.49	1.57	1.47	1.60	1.45	1.62	1.42	1.64
100	1.52	1.56	1.50	1.58	1.48	1.60	1.46	1.63	1.44	1.65
150	1.61	1.64	1.60	1.65	1.58	1.67	1.57	1.68	1.56	1.69
200	1.66	1.68	1.65	1.69	1.64	1.70	1.63	1.72	1.62	1.72

Source: DW51 for $n \leq 100$ and SW77 for $n = 150, 200$ (see text).

The testing procedures are as follows:

1. *One-sided test against alternatives $\rho > 0$.*

If $d < d_L$, conclude d is significant, reject H_0 , at level α .

If $d > d_U$, conclude d is not significant, do not reject H_0 .

If $d_L \leq d \leq d_U$, the test is said to be inconclusive.

TABLE 7.1. Significance Points of d_L and d_U :1% (Continued)

n	$k = 6$		$k = 7$		$k = 8$		$k = 9$		$k = 10$	
	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
20	0.52	1.92	0.44		0.36		0.29		0.23	
21	0.55	1.88	0.47		0.40		0.33		0.27	
22	0.59	1.85	0.51		0.44		0.37		0.30	
23	0.62	1.82	0.55	1.98	0.47		0.40		0.34	
24	0.65	1.80	0.58	1.94	0.51		0.44		0.38	
25	0.68	1.78	0.61	1.92	0.54		0.47		0.41	
26	0.71	1.76	0.64	1.89	0.57		0.51		0.44	
27	0.74	1.74	0.67	1.88	0.60		0.54		0.47	
28	0.76	1.73	0.70	1.85	0.63	1.97	0.57		0.50	
29	0.79	1.72	0.72	1.83	0.66	1.95	0.60		0.53	
30	0.81	1.71	0.75	1.81	0.68	1.93	0.62		0.56	
31	0.83	1.70	0.77	1.80	0.71	1.91	0.65		0.59	
32	0.86	1.69	0.79	1.79	0.73	1.89	0.67		0.61	
33	0.88	1.68	0.82	1.78	0.76	1.87	0.70	1.98	0.64	
34	0.90	1.68	0.84	1.77	0.78	1.86	0.72	1.96	0.67	
35	0.91	1.67	0.86	1.76	0.80	1.85	0.74	1.94	0.69	
36	0.93	1.67	0.88	1.75	0.82	1.84	0.77	1.93	0.71	
37	0.95	1.66	0.90	1.74	0.84	1.83	0.79	1.91	0.73	
38	0.97	1.66	0.91	1.74	0.86	1.82	0.81	1.90	0.75	1.99
39	0.98	1.66	0.93	1.73	0.88	1.81	0.83	1.89	0.77	1.97
40	1.00	1.65	0.95	1.72	0.90	1.80	0.84	1.88	0.79	1.96
45	1.07	1.64	1.02	1.70	0.97	1.77	0.97	1.83	0.88	1.90
50	1.12	1.64	1.08	1.69	1.04	1.75	1.00	1.81	0.96	1.86
55	1.17	1.64	1.13	1.69	1.10	1.73	1.06	1.79	1.02	1.84
60	1.21	1.64	1.18	1.68	1.14	1.73	1.11	1.77	1.07	1.82
65	1.25	1.64	1.22	1.68	1.19	1.72	1.15	1.76	1.12	1.80
70	1.28	1.65	1.25	1.68	1.22	1.72	1.19	1.75	1.16	1.79
75	1.31	1.65	1.28	1.68	1.26	1.72	1.23	1.75	1.20	1.79
80	1.34	1.65	1.31	1.68	1.29	1.71	1.26	1.75	1.23	1.78
85	1.36	1.66	1.34	1.69	1.31	1.71	1.29	1.74	1.26	1.77
90	1.38	1.66	1.36	1.69	1.34	1.71	1.31	1.74	1.29	1.77
95	1.40	1.67	1.38	1.69	1.36	1.72	1.34	1.74	1.31	1.77
100	1.42	1.67	1.40	1.69	1.38	1.72	1.36	1.74	1.34	1.77
150	1.54	1.71	1.53	1.72	1.52	1.74	1.50	1.75	1.49	1.77
200	1.61	1.74	1.60	1.75	1.59	1.76	1.58	1.77	1.57	1.80

Source: SW77 (see text). Copyright The Econometric Society.

2. One-sided test against alternative $\rho < 0$. Repeat (1) using $(4 - d)$ in place of d .
3. Two-sided equal-tailed test against alternatives $\rho \neq 0$.

If $d < d_L$ or $4 - d < d_L$, conclude d is significant, reject H_0 at level 2α .

If $d > d_U$ and $4 - d > d_U$, conclude d is not significant, do not reject H_0 at level 2α . Otherwise, the test is said to be inconclusive.

A Simplified Test

The inconclusive feature of the tests above is not attractive, but the problem is a difficult one. In later work, procedures for deciding inconclusive cases were formulated,

T A B L E 7.1. Significance Points of d_L and d_U :1% (Continued)

n	$k = 11$		$k = 12$		$k = 13$		$k = 14$		$k = 15$	
	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
25	0.35		0.29		0.24		0.19		0.15	
26	0.38		0.32		0.27		0.22		0.18	
27	0.41		0.36		0.30		0.25		0.21	
28	0.44		0.39		0.33		0.28		0.24	
29	0.47		0.42		0.36		0.31		0.27	
30	0.50		0.45		0.39		0.34		0.29	
31	0.53		0.48		0.42		0.37		0.32	
32	0.56		0.50		0.45		0.40		0.35	
33	0.59		0.53		0.48		0.43		0.38	
34	0.61		0.56		0.50		0.45		0.40	
35	0.63		0.58		0.53		0.48		0.43	
36	0.66		0.61		0.55		0.50		0.46	
37	0.68		0.63		0.58		0.53		0.48	
38	0.70		0.65		0.60		0.55		0.50	
39	0.72		0.67		0.62		0.58		0.53	
40	0.74		0.69		0.65		0.60		0.55	
45	0.84	1.97	0.79		0.74		0.70		0.66	
50	0.91	1.93	0.87	1.99	0.83		0.79		0.75	
55	0.98	1.89	0.94	1.95	0.90		0.86		0.83	
60	1.04	1.87	1.00	1.91	0.97	1.96	0.93		0.89	
65	1.09	1.85	1.05	1.89	1.02	1.93	0.99	1.98	0.95	
70	1.13	1.83	1.10	1.87	1.07	1.91	1.04	1.95	1.01	
75	1.17	1.82	1.14	1.86	1.11	1.89	1.08	1.93	1.05	1.97
80	1.21	1.81	1.18	1.84	1.15	1.88	1.12	1.91	1.09	1.95
85	1.24	1.80	1.21	1.83	1.18	1.87	1.16	1.90	1.13	1.93
90	1.26	1.80	1.24	1.83	1.22	1.86	1.19	1.89	1.17	1.92
95	1.29	1.79	1.27	1.82	1.24	1.85	1.22	1.88	1.20	1.91
100	1.31	1.79	1.29	1.82	1.27	1.84	1.25	1.87	1.23	1.90
150	1.47	1.78	1.46	1.80	1.44	1.81	1.43	1.83	1.41	1.85
200	1.56	1.79	1.55	1.80	1.54	1.81	1.53	1.82	1.52	1.84

Source: SW77 (see text). Copyright The Econometric Society

but they are more complicated, and we shall not discuss them here. It has been discovered, however, that, in many situations, treating the test as though d_L did not exist and d_U were the appropriate single critical value is a very good approximation¹ to the truth. Thus a simplified, approximate test procedure is the following

1S. *Simplified one-sided test against alternatives $\rho > 0$.* If $d < d_U$, reject H_0 at level α , otherwise do not reject.

2S. *Simplified one-sided test against alternatives $\rho < 0$.* If $4 - d < d_U$, reject H_0 at level α , otherwise do not reject.

3S. *Simplified two-sided test against alternatives $\rho \neq 0$.* If $d < d_U$ or $4 - d < d_U$, reject H_0 at level 2α .

There is no simple way to immediately determine if this simplified test is valid. For practical purposes at this level of complication, we suggest first applying the (d_L, d_U)

¹For a discussion of the accuracy of the d_U approximation, and alternatives, see Durbin and Watson (1971).

TABLE 7.1. Significance Points of d_L and d_U : 1% (Continued)

n	$k = 16$		$k = 17$		$k = 18$		$k = 19$		$k = 20$	
	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
30	0.25		0.21		0.17		0.14		0.11	
31	0.28		0.23		0.20		0.16		0.13	
32	0.30		0.26		0.22		0.18		0.15	
33	0.33		0.29		0.25		0.21		0.17	
34	0.36		0.31		0.27		0.23		0.20	
35	0.38		0.34		0.30		0.26		0.22	
36	0.41		0.36		0.32		0.28		0.24	
37	0.43		0.39		0.35		0.31		0.27	
38	0.46		0.41		0.37		0.33		0.29	
39	0.48		0.44		0.40		0.36		0.32	
40	0.51		0.46		0.42		0.38		0.34	
45	0.61		0.57		0.53		0.49		0.45	
50	0.71		0.67		0.63		0.59		0.55	
55	0.79		0.75		0.71		0.67		0.64	
60	0.86		0.82		0.79		0.75		0.72	
65	0.92		0.87		0.85		0.82		0.79	
70	0.97		0.94		0.91		0.88		0.85	
75	1.02		0.99		0.96		0.93		0.91	
80	1.07	1.98	1.04		1.01		0.98		0.96	
85	1.11	1.97	1.08		1.05		1.03		1.00	
90	1.14	1.95	1.12	1.98	1.09		1.07		1.04	
95	1.17	1.93	1.50	1.96	1.13	1.99	1.10		1.08	
100	1.20	1.92	1.18	1.95	1.16	1.98	1.14		1.11	
150	1.40	1.86	1.39	1.88	1.37	1.90	1.36	1.91	1.34	1.93
200	1.51	1.85	1.50	1.86	1.48	1.87	1.47	1.88	1.46	1.90

Source: SW77 (see text). Copyright The Econometric Society.

test to see if a clear decision is reached in that manner. Inconclusive results from this test will, of course, be judged significant by the simplified test, but this second-level decision can be regarded as having either a tentative “warning flag” attached to it or perhaps a slightly higher α -risk than the one indicated by the simplified test. Example 2 below illustrates this sort of judgment.

Example 1. The residuals from a straight line fit to $n = 50$ pairs of values of (X, Y) gave rise to a d -statistic of value $d = 0.625$. For a two-sided test of $H_1: \rho = 0$ against the two-sided alternative $\rho \neq 0$ we first compare d and $4 - d = 3.375$ against suitable d_L and d_U values from Tables 7.1–7.3. From the $\alpha = 0.01$ table, with $k = 1$ and $n = 50$, we find

$$d = 0.625 < d_L = 1.32.$$

It follows, from applying procedure (3), that we reject H_0 at the $2\alpha = 0.02$ level and conclude that there *does* appear to be serial correlation, of the type tested against, present in the data. Doubt is thus cast on the fitted model and the data should be reconsidered in the light of this new information.

Example 2. For a set of 70 residuals from a linear model involving $k = 4$ predictor variables, we find a d -statistic of value $d = 1.51$. Test $H_0: \rho_s = 0$ against the one-sided alternative $H_1: \rho_s = \rho^s$, where $\rho > 0$.

T A B L E 7.2. Significance Points of d_L and d_U : 2.5%

n	$k = 1$		$k = 2$		$k = 3$		$k = 4$		$k = 5$	
	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
15	0.95	1.23	0.83	1.40	0.71	1.61	0.59	1.84	0.48	2.09
16	0.98	1.24	0.86	1.40	0.75	1.59	0.64	1.80	0.53	2.03
17	1.01	1.25	0.90	1.40	0.79	1.58	0.68	1.77	0.57	1.98
18	1.03	1.26	0.93	1.40	0.82	1.56	0.72	1.74	0.62	1.93
19	1.06	1.28	0.96	1.41	0.86	1.55	0.76	1.72	0.66	1.90
20	1.08	1.28	0.99	1.41	0.89	1.55	0.79	1.70	0.70	1.87
21	1.10	1.30	1.01	1.41	0.92	1.54	0.83	1.69	0.73	1.84
22	1.12	1.31	1.04	1.42	0.95	1.54	0.86	1.68	0.77	1.82
23	1.14	1.32	1.06	1.42	0.97	1.54	0.89	1.67	0.80	1.80
24	1.16	1.33	1.08	1.43	1.00	1.54	0.91	1.66	0.83	1.79
25	1.18	1.34	1.10	1.43	1.02	1.54	0.94	1.65	0.86	1.77
26	1.19	1.35	1.12	1.44	1.04	1.54	0.96	1.65	0.88	1.76
27	1.21	1.36	1.13	1.44	1.06	1.54	0.99	1.64	0.91	1.75
28	1.22	1.37	1.15	1.45	1.08	1.54	1.01	1.64	0.93	1.74
29	1.24	1.38	1.17	1.45	1.10	1.54	1.03	1.63	0.96	1.73
30	1.25	1.38	1.18	1.46	1.12	1.54	1.05	1.63	0.98	1.73
31	1.26	1.39	1.20	1.47	1.13	1.55	1.07	1.63	1.00	1.72
32	1.27	1.40	1.21	1.47	1.15	1.55	1.08	1.63	1.02	1.71
33	1.28	1.41	1.22	1.48	1.16	1.55	1.10	1.63	1.04	1.71
34	1.29	1.41	1.24	1.48	1.17	1.55	1.12	1.63	1.06	1.70
35	1.30	1.42	1.25	1.48	1.19	1.55	1.13	1.63	1.07	1.70
36	1.31	1.43	1.26	1.49	1.20	1.56	1.15	1.63	1.09	1.70
37	1.32	1.43	1.27	1.49	1.21	1.56	1.16	1.62	1.10	1.70
38	1.33	1.44	1.28	1.50	1.23	1.56	1.17	1.62	1.12	1.70
39	1.34	1.44	1.29	1.50	1.24	1.56	1.19	1.63	1.13	1.69
40	1.35	1.45	1.30	1.51	1.25	1.57	1.20	1.63	1.15	1.69
45	1.39	1.48	1.34	1.53	1.30	1.58	1.25	1.63	1.21	1.69
50	1.42	1.50	1.38	1.54	1.34	1.59	1.30	1.64	1.26	1.69
55	1.45	1.52	1.41	1.56	1.37	1.60	1.33	1.64	1.30	1.69
60	1.47	1.54	1.44	1.57	1.40	1.61	1.37	1.65	1.33	1.69
65	1.49	1.55	1.46	1.59	1.43	1.62	1.40	1.66	1.36	1.69
70	1.51	1.57	1.48	1.60	1.45	1.63	1.42	1.66	1.39	1.70
75	1.53	1.58	1.50	1.61	1.47	1.64	1.45	1.67	1.42	1.70
80	1.54	1.59	1.52	1.62	1.49	1.65	1.47	1.67	1.44	1.70
85	1.56	1.60	1.53	1.63	1.51	1.65	1.49	1.68	1.46	1.71
90	1.57	1.61	1.55	1.64	1.53	1.66	1.50	1.69	1.48	1.71
95	1.58	1.62	1.56	1.65	1.54	1.67	1.52	1.69	1.50	1.71
100	1.59	1.63	1.57	1.65	1.55	1.67	1.53	1.70	1.51	1.72

Source: Durbin and Watson (1951).

We obtain the following significance points from Tables 7.1–7.3:

	d_L	d_U
$\alpha = 0.05$	1.49	1.74
$\alpha = 0.025$	1.42	1.66
$\alpha = 0.01$	1.34	1.58

We see that, using procedure (1), the primary test is inconclusive at all levels. Applying

TABLE 7.3. Significance Points of d_L and d_U : 5%

n	$k = 1$		$k = 2$		$k = 3$		$k = 4$		$k = 5$	
	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
15	1.08	1.36	0.95	1.54	0.82	1.75	0.69	1.97	0.56	2.21
16	1.10	1.37	0.98	1.54	0.86	1.73	0.74	1.93	0.62	2.15
17	1.13	1.38	1.02	1.54	0.90	1.71	0.78	1.90	0.67	2.10
18	1.16	1.39	1.05	1.53	0.93	1.69	0.82	1.87	0.71	2.06
19	1.18	1.40	1.08	1.53	0.97	1.68	0.86	1.85	0.75	2.02
20	1.20	1.41	1.10	1.54	1.00	1.68	0.90	1.83	0.79	1.99
21	1.22	1.42	1.13	1.54	1.03	1.67	0.93	1.81	0.83	1.96
22	1.24	1.43	1.15	1.54	1.05	1.66	0.96	1.80	0.86	1.94
23	1.26	1.44	1.17	1.54	1.08	1.66	0.99	1.79	0.90	1.92
24	1.27	1.45	1.19	1.55	1.10	1.66	1.01	1.78	0.93	1.90
25	1.29	1.45	1.21	1.55	1.12	1.66	1.04	1.77	0.95	1.89
26	1.30	1.46	1.22	1.55	1.14	1.65	1.06	1.76	0.98	1.88
27	1.32	1.47	1.24	1.56	1.16	1.65	1.08	1.76	1.01	1.86
28	1.33	1.48	1.26	1.56	1.18	1.65	1.10	1.75	1.03	1.85
29	1.34	1.48	1.27	1.56	1.20	1.65	1.12	1.74	1.05	1.84
30	1.35	1.49	1.28	1.57	1.21	1.65	1.14	1.74	1.07	1.83
31	1.36	1.50	1.30	1.57	1.23	1.65	1.16	1.74	1.09	1.83
32	1.37	1.50	1.31	1.57	1.24	1.65	1.18	1.73	1.11	1.82
33	1.38	1.51	1.32	1.58	1.26	1.65	1.19	1.73	1.13	1.81
34	1.39	1.51	1.33	1.58	1.27	1.65	1.21	1.73	1.15	1.81
35	1.40	1.52	1.34	1.58	1.28	1.65	1.22	1.73	1.16	1.80
36	1.41	1.52	1.35	1.59	1.29	1.65	1.24	1.73	1.18	1.80
37	1.42	1.53	1.36	1.59	1.31	1.66	1.25	1.72	1.19	1.80
38	1.43	1.54	1.37	1.59	1.32	1.66	1.26	1.72	1.21	1.79
39	1.43	1.54	1.38	1.60	1.33	1.66	1.27	1.72	1.22	1.79
40	1.44	1.54	1.39	1.60	1.34	1.66	1.29	1.72	1.23	1.79
45	1.48	1.57	1.43	1.62	1.38	1.67	1.34	1.72	1.29	1.78
50	1.50	1.59	1.46	1.63	1.42	1.67	1.38	1.72	1.34	1.77
55	1.53	1.60	1.49	1.64	1.45	1.68	1.41	1.72	1.38	1.77
60	1.55	1.62	1.51	1.65	1.48	1.69	1.44	1.73	1.41	1.77
65	1.57	1.63	1.54	1.66	1.50	1.70	1.47	1.73	1.44	1.77
70	1.58	1.64	1.55	1.67	1.52	1.70	1.49	1.74	1.46	1.77
75	1.60	1.65	1.57	1.68	1.54	1.71	1.51	1.74	1.49	1.77
80	1.61	1.66	1.59	1.69	1.56	1.72	1.53	1.74	1.51	1.77
85	1.62	1.67	1.60	1.70	1.57	1.72	1.55	1.75	1.52	1.77
90	1.63	1.68	1.61	1.70	1.59	1.73	1.57	1.75	1.54	1.78
95	1.64	1.69	1.62	1.71	1.60	1.73	1.58	1.75	1.56	1.78
100	1.65	1.69	1.63	1.72	1.61	1.74	1.59	1.76	1.57	1.78
150	1.72	1.75	1.71	1.76	1.69	1.77	1.68	1.79	1.67	1.80
200	1.76	1.78	1.75	1.79	1.74	1.80	1.73	1.81	1.72	1.82

Source: DW 51 for $n \leq 100$ and SW77 for $n = 150, 200$ (see text).

procedure (1S) we then come to the secondary conclusion that H_0 should be rejected because it falls below d_U at the $\alpha = 0.01$ level. Our true rejection level is perhaps not as low as $\alpha = 0.01$ because we are using the simplified test. However, we also note that we would *almost* reject at the $\alpha = 0.05$ level using an ordinary test because 1.51 is close to $d_L = 1.49$. Thus we can think, with reasonable safety, of the rejection of

T A B L E 7.3. Significance Points of d_L and d_U :5% (Continued)

n	$k = 6$		$k = 7$		$k = 8$		$k = 9$		$k = 10$	
	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
20	0.69		0.60		0.50		0.42		0.34	
21	0.73		0.64		0.55		0.46		0.38	
22	0.77		0.68		0.59		0.50		0.42	
23	0.80		0.72		0.63		0.55		0.47	
24	0.84		0.75		0.67		0.58		0.51	
25	0.87		0.78		0.70		0.62		0.54	
26	0.90	1.99	0.82		0.74		0.66		0.58	
27	0.93	1.97	0.85		0.77		0.69		0.62	
28	0.95	1.96	0.87		0.80		0.72		0.65	
29	0.98	1.94	0.90		0.83		0.75		0.68	
30	1.00	1.93	0.93		0.85		0.78		0.71	
31	1.02	1.92	0.95		0.88		0.81		0.74	
32	1.04	1.91	0.97		0.90		0.84		0.77	
33	1.06	1.90	0.99	1.99	0.93		0.86		0.80	
34	1.08	1.89	1.02	1.98	0.95		0.89		0.82	
35	1.10	1.88	1.03	1.97	0.97		0.91		0.85	
36	1.11	1.88	1.05	1.96	0.99		0.93		0.87	
37	1.13	1.87	1.07	1.95	1.01		0.95		0.89	
38	1.15	1.86	1.09	1.94	1.03		0.97		0.91	
39	1.16	1.86	1.10	1.93	1.05		0.99		0.93	
40	1.18	1.85	1.12	1.92	1.06		1.01		0.95	
45	1.24	1.84	1.19	1.90	1.14	1.96	1.09		1.04	
50	1.29	1.82	1.25	1.88	1.20	1.93	1.16	1.99	1.11	
55	1.33	1.81	1.29	1.86	1.25	1.91	1.21	1.96	1.17	
60	1.37	1.81	1.34	1.85	1.30	1.89	1.26	1.94	1.22	1.98
65	1.40	1.81	1.37	1.84	1.34	1.88	1.30	1.92	1.27	1.96
70	1.43	1.80	1.40	1.84	1.37	1.87	1.34	1.91	1.31	1.95
75	1.46	1.80	1.43	1.83	1.40	1.87	1.37	1.90	1.34	1.94
80	1.48	1.80	1.45	1.83	1.43	1.86	1.40	1.89	1.37	1.93
85	1.50	1.80	1.47	1.83	1.45	1.86	1.42	1.89	1.40	1.92
90	1.52	1.80	1.49	1.83	1.47	1.85	1.44	1.88	1.42	1.91
95	1.54	1.80	1.51	1.83	1.49	1.85	1.47	1.88	1.44	1.90
100	1.55	1.80	1.53	1.83	1.51	1.85	1.48	1.87	1.46	1.90
150	1.65	1.82	1.64	1.83	1.62	1.85	1.61	1.86	1.59	1.88
200	1.71	1.83	1.70	1.84	1.69	1.85	1.68	1.86	1.67	1.87

Source: SW77 (see text). Copyright The Econometric Society.

H_0 being at a level of somewhat between $\alpha = 0.05$ and $\alpha = 0.01$.² Doubt is cast on the fitted model and a reanalysis of the data, taking account of the indicated serial correlation, is appropriate.

Width of the Primary Test Inconclusive Region

Figure 7.4 shows a plot of the 5% values of d_L and d_U versus the number of observations n , for $15 \leq n \leq 100$ and $1 \leq k \leq 5$, joined up by smooth lines to clarify the plot. (See Durbin and Watson, 1951.) This plot could be extended using Table 7.3, of course.

²An alternative interpretation is that the primary test would be significant at about the 0.06 level.

TABLE 7.3. Significance Points of d_L and d_U : 5% (Continued)

n	$k = 11$		$k = 12$		$k = 13$		$k = 14$		$k = 15$	
	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
25	0.47		0.40		0.34		0.28		0.22	
26	0.51		0.44		0.37		0.31		0.26	
27	0.54		0.48		0.41		0.35		0.29	
28	0.58		0.51		0.45		0.38		0.33	
29	0.61		0.54		0.48		0.42		0.36	
30	0.64		0.58		0.51		0.45		0.39	
31	0.67		0.61		0.55		0.48		0.43	
32	0.70		0.64		0.58		0.52		0.46	
33	0.73		0.67		0.61		0.55		0.49	
34	0.76		0.70		0.63		0.58		0.52	
35	0.78		0.72		0.66		0.60		0.55	
36	0.81		0.75		0.69		0.63		0.58	
37	0.83		0.77		0.71		0.66		0.60	
38	0.85		0.80		0.74		0.68		0.63	
39	0.88		0.82		0.76		0.71		0.65	
40	0.90		0.84		0.79		0.73		0.68	
45	0.99		0.94		0.89		0.84		0.79	
50	1.06		1.02		0.97		0.93		0.88	
55	1.13		1.09		1.05		1.00		0.96	
60	1.18		1.15		1.11		1.07		1.03	
65	1.23		1.20		1.16		1.12		1.09	
70	1.27	1.99	1.24		1.21		1.17		1.14	
75	1.31	1.97	1.28		1.25		1.22		1.18	
80	1.34	1.96	1.31	1.99	1.28		1.25		1.22	
85	1.37	1.95	1.34	1.98	1.32		1.29		1.26	
90	1.40	1.94	1.37	1.97	1.34		1.32		1.29	
95	1.42	1.93	1.39	1.96	1.37	1.98	1.35		1.32	
100	1.44	1.92	1.42	1.95	1.39	1.97	1.37		1.35	
150	1.58	1.89	1.56	1.91	1.55	1.92	1.54	1.94	1.52	1.96
200	1.65	1.89	1.64	1.90	1.63	1.91	1.62	1.92	1.61	1.93

Source: SW77 (see text). Copyright The Econometric Society.

Note that the vertical distance between pairs of correspondingly numbered curves is the region of indecision involved in the standard test, and that the width of this region becomes smaller as n increases. The moral is obvious: the more observations we have, the more likely it is that we shall be able to make a definite decision via the Durbin-Watson test. Workers in time series have a rule of thumb that $n \geq 50$ observations are needed in order for their analyses to produce worthwhile conclusions. As Figure 7.4 shows, such a rule of thumb would not be out of place for application of the Durbin-Watson test.

Mean Square Successive Difference

Readers viewing Figure 7.4 may be curious about the dashed line lying between the various upper and lower pairs of curves. This is a join of the 5% points for testing for serial correlation in a model $Y = \beta_0 + \epsilon$. In other words, it is the Durbin-Watson test for $k = 0$ X 's in the model. Table 7.4 shows selected lower-tail percentage points

T A B L E 7.3. Significance Points of d_L and d_U :5% (Continued)

n	$k = 16$		$k = 17$		$k = 18$		$k = 19$		$k = 20$	
	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
30	0.34		0.29		0.24		0.20		0.16	
31	0.37		0.32		0.27		0.22		0.18	
32	0.40		0.35		0.30		0.25		0.21	
33	0.43		0.38		0.33		0.28		0.24	
34	0.46		0.41		0.36		0.31		0.27	
35	0.49		0.44		0.39		0.34		0.30	
36	0.52		0.47		0.42		0.37		0.32	
37	0.55		0.50		0.45		0.40		0.35	
38	0.58		0.52		0.47		0.42		0.38	
39	0.60		0.55		0.50		0.45		0.40	
40	0.63		0.58		0.53		0.48		0.43	
45	0.74		0.69		0.64		0.60		0.55	
50	0.84		0.79		0.75		0.70		0.66	
55	0.92		0.88		0.84		0.80		0.75	
60	0.99		0.95		0.91		0.87		0.84	
65	1.05		1.02		0.98		0.94		0.91	
70	1.11		1.07		1.04		1.01		0.97	
75	1.15		1.12		1.09		1.06		1.03	
80	1.20		1.17		1.14		1.11		1.08	
85	1.23		1.21		1.18		1.15		1.12	
90	1.27		1.24		1.21		1.19		1.16	
95	1.30		1.27		1.25		1.22		1.20	
100	1.32		1.30		1.28		1.25		1.23	
150	1.50	1.97	1.49	1.99	1.47		1.46		1.44	
200	1.60	1.94	1.59	1.96	1.58	1.97	1.57	1.98	1.55	1.99

Source: SW77 (see text). Copyright The Econometric Society.

given by Nelson (1980), here rounded to two decimal places. Nelson's figures allow the dashed line to be extended for higher n . We have omitted Nelson's one-tailed 10% values, which do not match with any of the other tables shown here. The blank 2.5% column in Table 7.4 indicates the absence of these values, not given by Nelson.

It has been argued that, when the model contains a lagged response variable as well as predictors, the Durbin–Watson test is inappropriate. An example of such a model is

$$Y_{i+1} = \beta_0 + \beta Y_i + \beta_1 X_{1i} + \beta_2 X_{2i} + \epsilon_i.$$

Rayner (1994) concluded that, in spite of the fact that the Durbin–Watson test has a bias toward nonrejection for such models, it may well be better than competitors. Details and related references are given in the quoted paper.

7.3. EXAMINING RUNS IN THE TIME SEQUENCE PLOT OF RESIDUALS: RUNS TEST

This test provides a quick but approximate alternative to the Durbin–Watson test. Since it ignores the actual sizes of the residuals and uses only their signs in time sequence, it throws away a lot of information, but it is easy to apply.

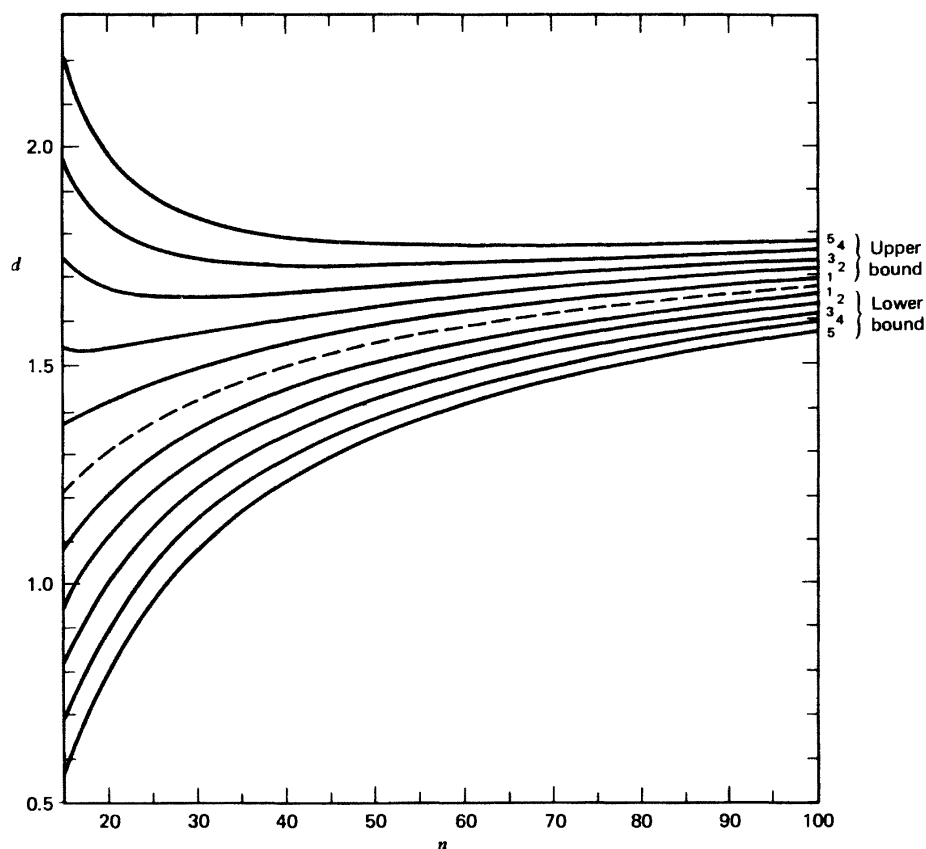


Figure 7.4. Graphs of 5% values of d_u and d_l against n for $k = 1, 2, 3, 4, 5$.

When the time sequence of a set of residuals is known, it is sometimes noticeable that groups of positive or negative residuals occur in what might be an unusual pattern. To take an extreme case, if 30 residuals in time sequence consisted of 16 negative followed by 14 positive residuals, we might first suspect that an unconsidered variable had changed levels between the 16th and 17th runs. Such behavior can also arise from a positive serial correlation between successive residuals, however. Similarly, a very large number of sign switches in a sequence might arise from negative serial correlation. When there is a sequence of such runs it is useful to have a method that will enable a decision to be made on whether the run pattern is “unusual” or not. We now explore how this might be done.

Runs

Suppose we have a sequence of signs such as

$$+ + - + - - - + + - + + + +.$$

These may be the signs of residuals in time sequence (which will be our application) but equally well the “plus and minus” signs might denote “male and female,” “head and tail,” “better and worse,” “treatment A and treatment B,” or the two levels of any other dichotomous classification. Suppose there are n signs in all, n_1 plus signs

T A B L E 7.4. Selected Lower Significance Points for the Mean Square Successive Difference Test (Durbin–Watson Test with $k = 0$ Predictor Variables): The Corresponding Upper Points Are 4 – (Those Shown)

n	1%	2.5%	5%
10	0.75		1.06
20	1.04		1.30
30	1.20		1.42
40	1.29		1.49
50	1.36		1.54
60	1.42		1.58
70	1.46		1.61
80	1.49		1.64
90	1.52		1.66
100	1.54		1.67
150	1.62		1.73
200	1.67		1.77
300	1.73		1.81
400	1.77		1.84
500	1.79		1.85
600	1.81		1.87
800	1.84		1.88
1000	1.85		1.90

and n_2 minus signs, and there are r runs. In the above example, $n_1 = 9$, $n_2 = 6$, and there are $r = 7$ runs indicated by the parentheses below

$$(+ \ +) (-) (+) (- \ - \ -) (+ \ +) (-) (+ \ + \ + \ +) .$$

We can ask if the particular arrangement of signs we observe is an “extreme” arrangement or not.

We first examine a case where there are only 15 possible sign sequences. Suppose there are six signs, two of which are plus. The following sign arrangements are possible:

Arrangement	Number of Runs, u
+ + - - - -	2
+ - + - - -	4
+ - - + - -	4
+ - - - + -	4
+ - - - - +	3
- + + - - -	3
- + - + - -	5
- + - - + -	5
- + - - - +	4
- - + + - -	3
- - + - + -	5
- - + - - +	4
- - - + + -	3
- - - + - +	4
- - - - + +	2

The distribution of runs is as follows:

	$r = 2$	3	4	5	
Frequency	= 2	4	6	3	(Total = 15)
Cumulative Probability	= 0.133	0.400	0.800	1.000	

Thus five runs would occur in $\frac{3}{15}$ or 20% of the possible cases, that is, with a probability of 0.2. Alternatively, two runs would occur in $\frac{2}{15}$ or 13.3% of the possible cases, that is, with a probability of 0.133. A low number of runs in the residuals sequence might indicate positive serial correlation, while a high number might arise from negative serial correlation. We discuss the situation in the “too few runs” context first. If we observed only $r = 2$ runs in a set of six residuals of which two were positive, we would have observed an event that occurs with a probability of 0.133. Obviously, nothing significant will occur in such a small example. For any given sequence of signs we can find the probability that the observed value of r (or a lesser value) will occur. (*Example:* When $n_1 = 2$, $n_2 = 4$, and the observed number of runs is 3, $\text{Prob}(r \leq 3) = (2 + 4)/15 = 0.4$, a not unusual event occurring in 40% of cases.) On the basis of such a probability level we can decide whether or not we believe that a random arrangement of signs has occurred. [We might, for example, compare the probability with a preassigned value, say, $\alpha = 0.05$, and reject the idea of a random arrangement if $\text{Prob}(u \leq \text{observed } u) \leq 0.05$.] For a “too many runs” test we would cumulate the frequencies downward from the high- r end.

Tables for Modest n_1 and n_2

Tables 7.5 and 7.6 show, respectively, lower-tail and upper-tail cumulative probabilities for selected n_1 , n_2 , with $3 \leq n_1 \leq n_2 \leq 10$. Only probability values less than or equal to 0.10 are shown. Typically, n_1 and n_2 would roughly be equal in a set of residuals. (When $n_1 > n_2$, interchange n_1 and n_2 .) These distributions were given originally by Swed and Eisenhart (1943). In that paper more decimal places are given and the arrangement of cases is different. Related tables also appear in Lindley and Scott (1984, see pp. 60–62).

Example. Twenty residuals—half positive, half negative—show five runs of signs in their time sequence. Is this an unusually small number? We see that $p(r \leq 5) = 0.004$ from the (10, 10) line, so the answer is yes. (At the $\alpha = 0.10$ level, seven or fewer runs would be a small number.) We can now reexamine the residuals and data and evaluate the lag-1 serial correlation. In the upper tail for this case, 15 or more runs would be a large number at the $\alpha = 0.10$ level.

Note that $(\frac{1}{2}n, \frac{1}{2}n)$ cases have a symmetric distribution, that is, $p(r \leq m) = p(r \geq n + 2 - m)$. Also, the distribution mean is at $(\frac{1}{2}n + 1)$; see below.

Larger n_1 and n_2 Values

Outside the range of small n_1 , n_2 values in the tables, it is convenient to use a normal approximation to the actual distribution. Let

$$\mu = \frac{2n_1n_2}{n_1 + n_2} + 1, \quad (7.3.1)$$

$$\sigma^2 = \frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)}. \quad (7.3.2)$$

T A B L E 7.5. Cumulative Lower-Tail Areas in the Distribution of the Total Number of Runs r in Samples of Size (n_1, n_2) for $3 \leq n_1 \leq n_2 \leq 10$, $n_1 + n_2 \geq 10$, and Tail Areas ≤ 0.10 only (for $n_1 \geq n_2$, Simply Interchange n_1 and n_2)

$(n_1, n_2) \ r =$	2	3	4	5	6	7
(3, 7)	0.017	0.083				
(3, 8)	0.012	0.067				
(3, 9)	0.009	0.055				
(3, 10)	0.007	0.045				
(4, 6)	0.010	0.048				
(4, 7)	0.006	0.033				
(4, 8)	0.004	0.024				
(4, 9)	0.003	0.018	0.085			
(4, 10)	0.002	0.014	0.068			
(5, 5)	0.008	0.040				
(5, 6)	0.004	0.024				
(5, 7)	0.003	0.015	0.076			
(5, 8)	0.002	0.010	0.054			
(5, 9)	0.001	0.007	0.039			
(5, 10)	0.001	0.005	0.029	0.095		
(6, 6)	0.002	0.013	0.067			
(6, 7)	0.001	0.008	0.043			
(6, 8)	0.001	0.005	0.028	0.086		
(6, 9)	0.000	0.003	0.019	0.063		
(6, 10)	0.000	0.002	0.013	0.047		
(7, 7)	0.001	0.004	0.025	0.078		
(7, 8)	0.000	0.002	0.015	0.051		
(7, 9)	0.000	0.001	0.010	0.035		
(7, 10)	0.000	0.001	0.006	0.024	0.080	
(8, 8)	0.000	0.001	0.009	0.032	0.100	
(8, 9)	0.000	0.001	0.005	0.020	0.069	
(8, 10)	0.000	0.000	0.003	0.013	0.048	
(9, 9)	0.000	0.000	0.003	0.012	0.044	
(9, 10)	0.000	0.000	0.002	0.008	0.029	0.077
(10, 10)	0.000	0.000	0.001	0.004	0.019	0.051

Source: Adapted from Swed and Eisenhart (1943).

It can be shown that these are the actual mean and variance of the discrete distribution of r . Then approximately, for a lower-tail test,

$$z = \frac{(r - \mu + \frac{1}{2})}{\sigma} \quad (7.3.3)$$

is a unit normal deviate where the $\frac{1}{2}$ is the usual *continuity correction*, which helps compensate for the fact that a continuous distribution is being used to approximate

T A B L E 7.6. Cumulative Upper-Tail Areas in the Distribution of the Total Number of Runs r in Samples of Size (n_1, n_2) for $3 \leq n_1 \leq n_2 \leq 10$, $n_1 + n_2 \geq 10$, and Tail Areas ≤ 0.10 only (for $n_1 \geq n_2$, simply interchange n_1 and n_2)

$(n_1, n_2) \ r =$	9	10	11	12	13	14	15	16	17	18	19	20
(4, 6)	0.024											
(4, 7)	0.046											
(4, 8)	0.071											
(4, 9)	0.098											
(4, 10)												
(5, 5)	0.040	0.008										
(5, 6)	0.089	0.024	0.002									
(5, 7)		0.045	0.008									
(5, 8)		0.071	0.016									
(5, 9)		0.098	0.028									
(5, 10)			0.042									
(6, 6)		0.067	0.013	0.002								
(6, 7)			0.034	0.008	0.001							
(6, 8)			0.063	0.016	0.002							
(6, 9)			0.098	0.028	0.006							
(6, 10)				0.042	0.010							
(7, 7)			0.078	0.025	0.004	0.001						
(7, 8)				0.051	0.012	0.002	0.000					
(7, 9)				0.084	0.025	0.006	0.001					
(7, 10)					0.043	0.010	0.002					
(8, 8)				0.100	0.032	0.009	0.001	0.000				
(8, 9)					0.061	0.020	0.004	0.001	0.000			
(8, 10)					0.097	0.036	0.010	0.002	0.000			
(9, 9)						0.044	0.012	0.003	0.000	0.000		
(9, 10)						0.077	0.026	0.008	0.001	0.000	0.000	
(10, 10)							0.051	0.019	0.004	0.001	0.000	0.000

Source: Adapted from Swed and Eisenhart (1943).

to a discrete distribution in the lower tail. For a “too many runs” upper-tail test, the continuity correction is $-\frac{1}{2}$ so that we use instead

$$z = \frac{r - \mu - \frac{1}{2}}{\sigma} \quad (7.3.4)$$

and look up the upper tail of the $N(0, 1)$ distribution. How do we know which tail we want? We first evaluate μ . If $r > \mu$ we use the upper-tail test, and if $r < \mu$ the lower-tail test.

What if there are zeros in the residuals? Do they receive a minus sign or a plus sign? Exact zeros are unlikely in a regression fit but, if one occurs, the easiest way out is to assume first a plus and then a minus and see if the results are the same. This situation is unlikely to arise except in constructed class examples.

Example. Examination of a set of 27 residuals, 15 of which were of one sign and 12 of which were of the opposite sign, arranged in time sequence, revealed $r = 7$ runs. Does the arrangement of signs appear to have “too few runs”?

Here $n_1 = 15$, $n_2 = 12$, $r = 7$. From Eqs. (7.3.1) and (7.3.2), $\mu = \frac{43}{9}$, $\sigma^2 = \frac{740}{117}$. Thus the observed value of z from Eq. (7.3.3) is

$$z = \frac{(7 - \frac{44}{9} + \frac{1}{2})}{(\frac{740}{117})^{1/2}} = -2.713. \quad (7.3.5)$$

The probability of obtaining a unit normal deviate of value -2.713 or smaller is 0.0033 (or 0.33%) so that an unusually low number of runs appears to have occurred. We should reject the idea that the arrangement of signs is random. The model would be suspect and we would now search for an assignable cause for the pattern of residuals.

Comments

Strictly speaking, the test for runs is applicable only when the occurrences that produce the pattern of runs are independent. In a time sequence of residuals this is not true due to the correlations that exist among the residuals, and the probability level obtained from the procedure will be affected in a way that depends on the particular structure of the data. In most practical regression situations, unless the ratio $(n - p)/n$, that is, (number of degrees of freedom in residuals)/(number of residuals), is quite small, the effect can be ignored.

Time plots of residuals can also be subjected to calculations suggested by Cleveland and Kleiner (1975). Three curves of moving statistics are drawn, involving (1) the *midmean* (the average of all observations between the quartiles of the data to that point in time), (2) the *lower semi-midmean* (the midmean of all observations *below* the median of the data to that point in time), and (3) the *upper midmean* (. . . *above* . . .). “These three statistics summarize the location, spread, and skewness of the data.” (See p. 449 of the reference cited.)

REFERENCES

Box, Jenkins, and Reinsel (1994); Diggle (1990); Durbin (1969, 1970); Durbin and Watson (1950, 1951, 1971); Savin and White (1977); Wei (1990).

EXERCISES FOR CHAPTER 7

- A. Fit, to the appropriate portion of the steam data in Appendix 1A, the model $Y = \beta_0 + \beta_5 X_5 + \beta_6 X_6 + \beta_8 X_8 + \epsilon$. This will give you the following fitted model and analysis of variance table, and also the fitted values and residuals in Table A:

$$\hat{Y} = -2.968 + 0.4020X_5 + 0.19892X_6 - 0.073924X_8$$

Source	df	SS	MS	F
Regression b_0	3	56.472	18.824	53.83
Residual	21	7.344	0.350	
Total, corrected	24	63.816		

TABLE A. Observations, Fitted Values, and Residuals

Row	Y	\hat{Y}	Residual
1	10.98	10.86	0.12
2	11.13	10.47	0.66
3	12.51	11.79	0.72
4	8.40	8.72	-0.32
5	9.27	9.13	0.14
6	8.73	8.20	0.53
7	6.36	6.18	0.18
8	8.50	8.40	0.10
9	7.82	8.04	-0.22
10	9.14	9.22	-0.08
11	8.24	9.64	-1.40
12	12.19	11.54	0.65
13	11.88	11.60	0.28
14	9.57	9.18	0.39
15	10.94	10.61	0.33
16	9.58	9.49	0.09
17	10.09	9.49	0.60
18	8.11	8.29	-0.18
19	6.83	6.51	0.32
20	8.88	8.56	0.32
21	7.68	7.74	-0.06
22	8.47	9.38	-0.91
23	8.86	9.77	-0.91
24	10.36	11.00	-0.64
25	11.08	11.76	-0.68

1. Plot the residuals in a histogram, in a normal probability (nscore) plot, versus order 1–25, and versus \hat{Y} .
 2. Comment on what you see in the four plots in part 1.
 3. Evaluate the Durbin–Watson statistic and use it to test $H_0: \rho = 0$ versus the two-sided alternative $H_1: \rho \neq 0$, assuming that $\rho_s = \rho^s$ in the usual notation.
 4. Carry out a test for “too few runs” using the runs tests, on the residuals in the sequence order displayed in Table A.
- B.** (Source: “Using an hyperbola as a transition model to fit two-regime straight-line data,” by D. G. Watts and D. W. Bacon, *Technometrics*, **16**, 1974, 369–373.) A set of sediment settling data was subjected to three different regression calculations, using three different models. The residuals from these three separate calculations, multiplied by 1000, are shown in Table B, in the time order in which the data occurred; the actual times of observation appear in the first column. Plot the residuals against time and analyze their behavior by applying a two-sided runs test to each set. What are your conclusions?
- C.** (Source: “Car accidents—environmental aspects,” by D. F. Andrews, *International Statistical Review*, **41**, 1973, 235–239.) The data in Table C consist of 50 observations of the response variable Y = “driving deaths” and six possible predictor variables X_1, X_2, \dots, X_6 for 49 states and the District of Columbia. Figure C1 shows the plot of Y versus X_1 = 1964 drivers $\times 10^{-4}$, while Figure C2 shows $y = \log Y$ versus $Z_1 = \log X_1$. To the latter data the model

$$\hat{y} = -0.101 + 0.938Z_1$$

T A B L E B. Three Sets of Residuals (Multiplied by 1000) Versus Time t

Time, t	Set 1	Set 2	Set 3
0.5	-19	^a	^a
1	-19	0	2
1.5	-18	0	2
2	-28	-10	-8
2.5	-27	0	2
3	-27	0	2
4.5	-45	-20	-15
6	-23	19	25
9	-19	2	10
12	-5	12	20
14	18	23	27
16	-9	-25	-23
18	4	13	16
20	-3	-5	-4
22	-9	-6	-5
24	-6	4	4
26	8	14	13
28	3	-4	-7
30	7	6	2
32	22	17	11
34	27	8	0
36	33	9	-1
40	26	1	-24
42	14	-9	-24
44	3	-9	-25
46	3	1	-16
48	4	2	-16
50	-4	-6	-26
52	-9	-5	-25
54	-11	-3	-23
56	-11	0	-21
58	-17	-6	-27
60	-17	-2	-23
62	-11	5	-17
64	-6	3	-20
66	-12	-6	-31
68	-7	5	-22
70	-9	-2	-32
72	-8	1	-31
74	4	14	-20
76	7	6	-29
78	8	5	-27
80	17	12	-17
82	31	19	-6
90	45	34	-16
106	-11	-17	11
120	-23	-4	47
150	-36	-9	33

^aThe particular analysis used did not provide a first residual.

T A B L E C. Fifty Observations of Motor Vehicle Death Data Together with Some Possible Explanatory Predictor Variables

Region	Y, 1964 Deaths	X_1 , 1964 Drivers $\times 10^{-4}$	X_2 , 1960 Persons /sq. mi.	X_3 , 1963 Road (Rural) Mileage $\times 10^{-3}$	X_4 , 1960 More Males than Females	X_5 , normal January Maximum Temperature	X_6 , 1964 Highway Fuel Consumption gallons $\times 10^7$
AL	968	158	64	66	No	62	119
AK	43	11	0.4	5.9	Yes	30	6.2
AZ	588	91	12	33	Y	64	65
AR	640	92	34	73	N	51	74
CA	4743	952	100	118	N	65	105
CO	566	109	17	73	N	42	78
CT	325	167	518	5.1	N	37	95
DE	118	30	226	3.4	N	41	20
DC	115	35	12524	—	N	44	23
FL	1545	298	91	57	N	67	216
GA	1302	203	68	83	N	54	162
ID	262	41	8.1	40	Y	36	29
IL	2207	544	180	102	N	33	350
IN	1410	254	129	89	N	37	196
IA	833	150	49	100	N	30	109
KS	669	136	27	124	N	42	94
KY	911	147	76	65	N	44	104
LA	1037	146	72	40	N	65	109
ME	196	46	31	19	N	30	37
MD	616	157	314	29	N	44	113
MA	766	255	655	17	N	37	166
MI	2120	403	137	95	N	33	306
MN	841	189	43	110	N	22	132
MS	648	85	46	59	N	57	77
MO	1289	234	63	100	N	40	180
MT	259	38	4.6	72	Y	29	31
NE	450	89	18.4	97	N	32	61
NV	215	23	2.6	44	Y	40	24
NH	158	37	67	13	N	32	23
NJ	1071	329	807	21	N	43	231
NM	387	54	7.8	62	Y	46	48
NY	2745	744	350	84	N	31	439
NC	1580	226	93	71	N	51	177
ND	185	38	9.1	102	Y	20	24
OH	2096	530	237	84	N	41	358
OK	785	137	34	94	N	46	107
OR	575	108	18	73	N	45	81
PA	1889	570	252	89	N	39	353
RI	100	46	812	1.3	N	38	27
SC	870	122	79	52	N	61	86
SD	270	40	9	87	Y	23	28
TN	1059	177	85	67	N	49	135
TX	3006	515	37	196	N	50	448
UT	295	57	10.8	32	N	37	38
VT	131	20	42	13	N	25	15
VA	1050	208	100	50	N	50	150
WA	730	160	43	59	Y	46	109
WV	467	88	77	32	N	43	54
WI	1059	207	72	87	N	26	141
WY	148	22	3.4	67	Y	37	20

is fitted, giving rise to the residuals in Figure C3. By examining the data corresponding to the residuals marked with the names of states, suggest what variables appear to have influence on the data. Which would be the most logical candidate for entry into regression next?

- D.** A set of 56 residuals equally spaced in time order contains 26 positive residuals and 30 negative residuals. There are 38 runs. Is this an “unusually large” number of runs, do you think?
- E.** A set of 25 residuals equally spaced in time order has 12 positive values and 13 negative values and exhibits five runs. Is that an unusually small number in your opinion?
- F.** A regression fit $\hat{Y} = b_0 + b_1X_1 + b_2X_2 + b_3X_3$ on 85 observations equally spaced in time produces a Durbin–Watson statistic of $d = 2.33$. Might this indicate serial correlation? Test at a two-tailed $\alpha = 0.05$ level.

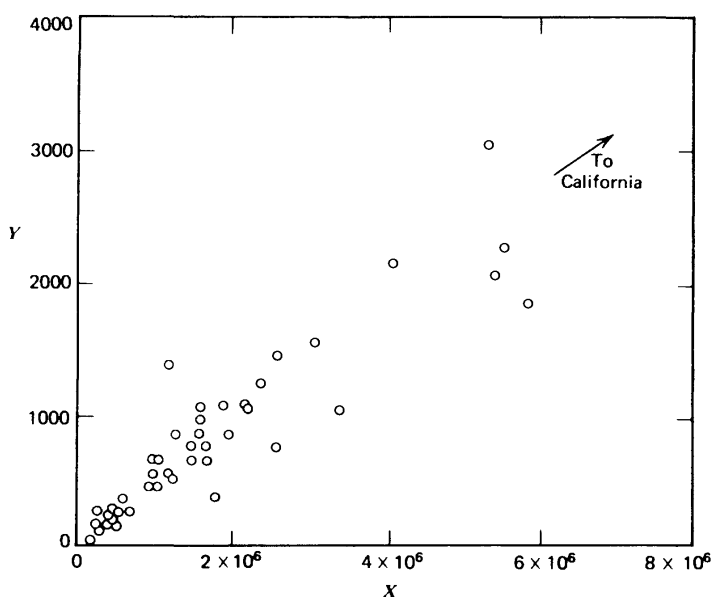


Figure C1. Motor vehicle deaths and drivers by state.

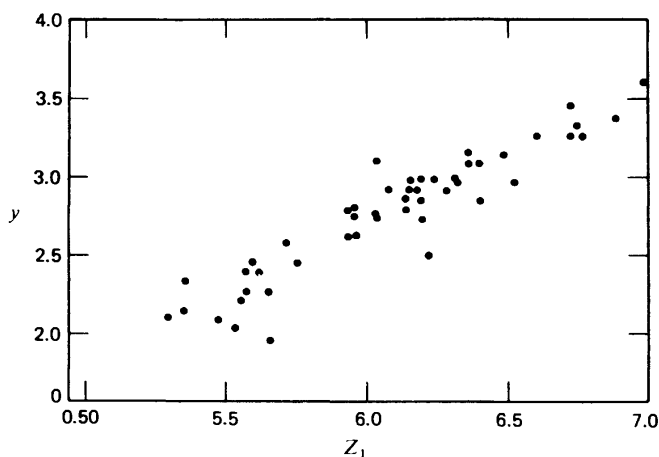


Figure C2. Log (motor vehicle deaths) and log (drivers by state).

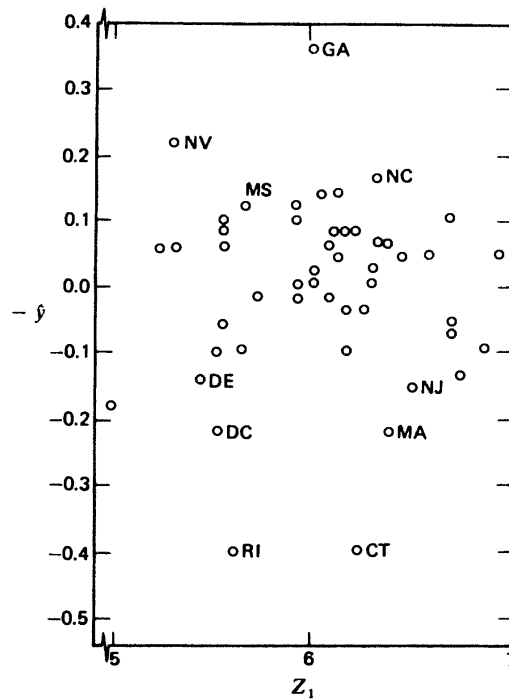


Figure C3. Residuals with some identifiers added.

- G.** An experimenter tells you: “I have looked at my 51 residuals in time order. Each residual differs from the previous one by an amount that is always within the range $(-1 \leq d_i \leq 1)$. Also, the sum of squares of the residuals is 50. My regression equation involves five predictors. What would you advise me to do?”
- H.** The following 24 residuals from a straight line fit are equally spaced in time and are given in time sequential order. Is there any evidence of lag-1 serial correlation, do you think? (Use a two-sided test at level $\alpha = 0.05$.)

8, -5, 7, 1, -3, -6, 1, -2, 10, 1, -1, 8, -6, 1, -6, -8, 10, -6, 9, -3, 3, -5, 1, -9