Chapter 3 Other Issues in Multiple regression (Part 2: Autocorrelation)

1 Overview: violation of independent random errors

In the least squares estimation (LSE), the optimal situation is that the errors are IID. But in many cases, esp, data observed over time (usually called time series data) and data observed over space, the observations nearby are not independent. In that cases, the LSE has the following problems

- The estimated coefficients are still unbiased, but no longer have minimum variance
- MSE may seriously underestimated the variance of the errors
- $s(b_k)$ may seriously underestimate the true standard deviation/error of the estimated regression coefficient
- The t-statistic and F-Statistic may no longer applicable
- However, only when the data can be ordered according to time or space, we can find remedies.

2 Detecting Autocorrelation

One (common) violation of the assumption is that

$$\varepsilon_1, \cdots, \varepsilon_n$$

have *first-order autocorrelation*, i.e.

$$Corr(\varepsilon_t, \varepsilon_{t+1}) = \rho(\neq 0), \quad t = 1, 2, \cdots, n.$$

or $\rho = Corr(\varepsilon_t, \varepsilon_{t+1})$ is defined as

$$\rho = \frac{Cov(\varepsilon_t, \varepsilon_{t+1})}{\sqrt{Var(\varepsilon_t)Var(\varepsilon_{t+1})}}$$

The estimators are

$$\widehat{Cov(\varepsilon_t, \varepsilon_{t+1})} = \frac{1}{n-1} \sum_{t=1}^{n-1} e_t e_{t+1}$$

$$\widehat{Var(\varepsilon_t)} = \frac{1}{n} \sum_{t=1}^{n} e_t^2$$

$$\widehat{Var(\varepsilon_{t+1})} = \frac{1}{n} \sum_{t=0}^{n-1} e_{t+1}^2$$

Thus, the estimator of ρ is then (approximately)

$$\hat{\rho} = \frac{\sum_{t=1}^{n-1} e_t e_{t+1}}{\sum_{t=1}^{n} e_t^2}$$

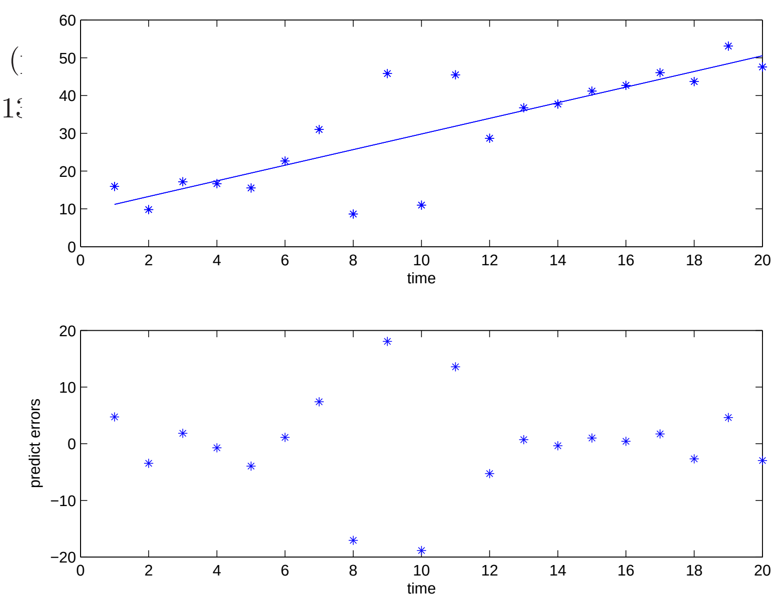
Example 2.1 *:* a time series { 15.91, 9.80, 17.16, 16.68, 15.53, 22.66, 31.01, 8.62, 45.82, 10.97, 45.46, 28.69, 36.75, 37.75, 41.18, 42.67, 46.05, 43.70, 53.08, 47.56 } We fit a linear trend model $y = \beta_0 + \beta_1 t + \varepsilon$ to the data and t is time $1, 2, ..., 20$.

The estimators for the parameters are

$$b_0 = 9.12, \qquad b_1 = 2.07$$

The residuals (.2, 7.40, -17.05,

18.07, -18.84, 1



e_t	e_{t+1}	e_t	e_{t+1}
4.72	-3.45	13.57	-5.26
-3.45	1.83	-5.26	0.72
1.83	-0.71	0.72	-0.34
-0.71	-3.94	-0.34	1.01
-3.94	1.12	1.01	0.43
1.12	7.4	0.43	1.73
7.4	-17.05	1.73	-2.68
-17.05	18.07	-2.68	4.63
18.07	-18.84	4.63	-2.95
-18.84	13.57		

we have

$$\hat{\rho} = \frac{\sum_{t=1}^{n-1} e_t e_{t+1}}{\sum_{t=1}^{n} e_t^2} = \frac{-1152.9}{1335.7} = -0.86.$$

3 First order autoregressive error model

Suppose $\xi_1, \xi_2, ..., \xi_n$ are (random) data taken over time, denoted by $\xi_t : t = 1, 2, ...$ If the time series satisfies

$$\xi_t = \alpha_0 + \alpha_1 \xi_{t-1} + ... + \alpha_q \xi_{t-q} + u_t$$

for all t, where u_t are IID $N(0, \sigma^2)$ and independent of $\xi_{t-1}, \xi_{t-2}, ...$ Then we say ξ_t follows autoregressive model of order q, denoted by AR(q). [u_t is call white noise]

For AR(1) with $\alpha_0 = 0$,

$$\xi_t = \rho \xi_{t-1} + u_t$$

where $|\rho| < 1$, where u_t are IID $N(0, \sigma^2)$ and independent of ξ_{t-1}. we have

1. $E\xi_t = 0$, $\mathbf{Var}(\xi_t) = \sigma^2/(1-\rho^2)$

2. $\mathbf{Cov}(\xi_t, \xi_{t-1}) = \rho\sigma^2/(1-\rho^2)$, and $\mathbf{Cov}(\xi_t, \xi_{t-s}) = \rho^s \sigma^2/(1-\rho^2)$ for $s > 0$.

3. $\mathcal{E} = (\xi_1, ..., \xi_n)'$, then we have

$$\mathbf{Var}(\mathcal{E}) = \sigma^2 \begin{pmatrix} 1 & \rho & \rho^2 & ... & \rho^{n-1} \\ \rho & 1 & \rho & ... & \rho^{n-2} \\ \cdots & & & & \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & ... & 1 \end{pmatrix}$$

For a multiple regression model

$$Y_1 = \beta_0 + \beta_1 X_{11} + ... + \beta_1 X_{1p} + \xi_1$$

$$Y_2 = \beta_0 + \beta_1 X_{21} + ... + \beta_1 X_{2p} + \xi_2$$

$$...$$

$$Y_n = \beta_0 + \beta_1 X_{n1} + ... + \beta_1 X_{np} + \xi_n$$

The covariance matrix of $\mathcal{E} = (\xi_1, ..., \xi_n)'$ is no longer $\sigma^2\mathbf{I}$.

4 Durbin-Watson Test for autocorrelation

The estimator of ρ

$$\hat{\rho} = \frac{\sum_{t=2}^{n} e_t e_{t-1}}{\sqrt{\sum_{t=2}^{n} e_t^2 \sum_{t=2}^{n} e_{t-1}^2}} \approx \frac{\sum_{t=2}^{n} e_t e_{t-1}}{\sum_{t=2}^{n} e_t^2}$$

where $e_i = Y_i - \hat{Y}_i$. If $\rho = 0$, then $\hat{\rho} \approx 0$.

Durbin-Watson statistic

$$\begin{aligned} DW &= \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2} \\ &= \frac{\sum_{t=1}^{n-1} e_t^2}{\sum_{t=1}^{n} e_t^2} + \frac{\sum_{t=1}^{n-1} e_{t+1}^2}{\sum_{t=1}^{n} e_t^2} - 2\hat{\rho} \\ &\approx 2 - 2\hat{\rho} \end{aligned}$$

It is easy to see that when $\rho = 0$, then DW $= 2$; when $\rho < 0$, then DW > 2; when $\rho > 0$ then $DW < 2$

The critical value d_L, d_U depends on n, p' (the number of coefficient parameters) and significant level α. See the statistical table.

1. The effect of positive autocorrelation $\rho > 0$ is more serious than that of negative autocorrelation. Suppose we need to test

$$H_0 : \rho = 0, \qquad H_a : \rho > 0$$

   - If $DW > d_U$, conclude H_0.
   - If $DW < d_L$, conclude H_a.
   - if $d_L \le DW \le d_U$, no conclusion

2. Suppose we need to test

$$H_0 : \rho = 0, \qquad H_a : \rho < 0$$

   - If $4 - DW > d_U$, conclude H_0.
   - If $4 - DW < d_L$, conclude H_a.
   - if $d_L \le 4 - DW \le d_U$, no conclusion

For the above example 2.1, we have

$$\sum_{t=1}^{n} e_t^2 = 1335.71, \qquad \sum_{t=1}^{n-1}(e_t - e_{t+1})^2 = 4946.2$$

Thus

$$DW = \sum_{t=1}^{n-1}(e_t - e_{t+1})^2 / \sum_{t=1}^{n} e_t^2 = 3.70.$$

Note that $n = 20$, $n_p = 2$ and $\alpha = 0.05$, we have

$$d_{L,\alpha} = 1.2$$

Because $4 - DW = 0.3 < d_L$, we reject H_0, and there is negative autocorrelation in the error terms.

Example 4.1 The Blaisdell Company wish to predict its sales (Y) by using industry sales as a predictor variable (X), (in data, **data**) . The model is

$$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t$$

The fitted model is

$$\hat{Y} = -1.45 + 0.176X$$

with

$$DW = 0.735$$

with $\alpha = 0.01$, $p = 1$ and $n = 20$, we have

$$d_{L,\alpha} = 0.95, d_{U,\alpha} = 1.15$$

Since $DW < d_{L,\alpha}$, there is significant positive autocorrelation in the residuals. see **(code)**

5 Remedial measure for autocorrelation

Suppose the true model is

$$\begin{aligned} Y_t &= \beta_0 + \beta_1 X_t + \varepsilon_t, \qquad \varepsilon_t = \rho\varepsilon_{t-1} + u_t \\ &\text{or} \quad Y_t = \beta_0 + \beta_1 X_t + \rho\varepsilon_{t-1} + u_t \end{aligned}$$

Define

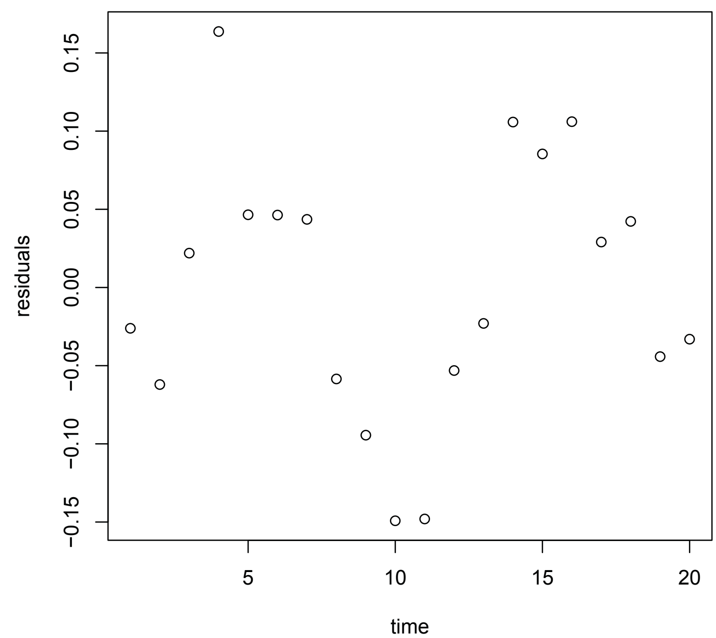
$$\tilde{Y}_t = Y_t - \rho Y_{t-1}$$

and

$$\tilde{X}_t = X_t - \rho X_{t-1}$$

or $\tilde{X}_t = X_t - \hat{\rho} X_{t-1}$ in calculation. Then the model becomes transformed model

$$\tilde{Y}_t = \beta_0(1-\rho) + \beta_1 \tilde{X}_t + u_t$$



or

$$\tilde{Y}_t = \tilde{\beta}_0 + \tilde{\beta}_1 \tilde{X}_t + u_t \qquad \text{(called transformed model)}$$

NOTE THAT, the transformed model has independent random errors Estimate the transformed model, denote the estimator by

$$\tilde{b}_0 \quad \tilde{b}_1$$

Then the estimator for the original models is

$$b_0 = \frac{\tilde{b}_0}{1-\rho} \qquad b_1 = \tilde{b}_1$$

or when ρ is unknown

$$b_0 = \frac{\tilde{b}_0}{1-\hat{\rho}} \qquad b_1 = \tilde{b}_1$$

where $\hat{\rho}$ is the estimator ρ. Simply speaking, we also have

$$s(b_0) = \frac{s(\tilde{b}_0)}{1-\hat{\rho}} \qquad s(b_1) = s(\tilde{b}_1)$$

Example 5.1 (4.1 continued) For model

$$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t$$

we have the first order autocorrelation of errors is estimated as

$$\hat{\rho} = r = 0.63$$

Because the positive autocorrelation exists, we need to consider transformation. Define

$$\tilde{Y}_t = Y_t - \hat{\rho} * Y_{t-1}, \quad \tilde{X}_t = X_t - \hat{\rho} * X_{t-1}$$

consider the new model

$$\tilde{Y}_t = \tilde{\beta}_0 + \tilde{\beta}_1 \tilde{X}_t + \varepsilon_t \qquad \text{(transformed model)}$$

The estimated coefficient

$$\tilde{b}_0 = -0.397, \tilde{b}_1 = 0.17$$

with

$$s(\tilde{b}_0) = 0.1673, \quad s(b_1) = 0.00294$$

for the transformed model, $D\tilde{W} = 1.73 > d_U = 1.13$, there is no autocorrelation in the transformed model.

For the original model, we have $\beta_0 = -0.397/(1-r) = -1.07$

$$\begin{array}{cccccccc} \hat{Y}_t & = & -1.07 & + & 0.17X_t & + & 0.63e_{t-1} \\ (SE) & & (0.45) & & (0.0029) & & \end{array}$$

6 Forecasting with autocorrelation Error terms

Actually, for regression model with first order autorrelated error terms, the model is

$$Y_t = \beta^\top X_t + \varepsilon_t$$

where $X_t = (1, X_{t1}, ..., X_{tp})'$ and $\beta = (\beta_0, \beta_1, ..., \beta_p)^\top$ with

$$\varepsilon_t = \rho \varepsilon_{t-1} + u_t$$

After transformation $\tilde{Y} = Y_t - \rho Y_{t-1}, \tilde{X}_t = X_t - \rho X_{t-1}$ The transformed model is

$$\tilde{Y}_t = \tilde{\beta} \tilde{X}_t + u_t$$

It is easy to see that

$$\beta_0 = \tilde{\beta}_0/(1-\rho), \beta_1 = \tilde{\beta}_1, ..., \beta_p = \tilde{\beta}_p$$

We estimate the transformed model and denote it by $\tilde{b}$, and thus

$$b_0 = \tilde{b}_0/(1-\hat{\rho}), b_1 = \tilde{b}_1, ..., b_p = \tilde{b}_p$$

and

$$\hat{\tilde{Y}}_t = \tilde{b}'\tilde{X}_t$$

or the estimated model for the original model is

$$\hat{Y}_t = b'X_t,$$

where $b = (b_0, b_1, ..., b_p)^\top$. Let

$$e_t = Y_t - b'X_t, t = 1, ..., n$$

We can then predict the $(n+1)$th subject

$$\hat{Y}_{n+1} = b'X_{n+1} + \hat{\rho} * e_n$$

and the $(n+2)$th subject

$$\hat{Y}_{n+2} = b'X_{n+2} + \hat{\rho}^2 * e_n$$

For predicting the confidence interval of Y_{n+1} or preciction interval with confidence $1-\alpha$, we have

$$\hat{Y}_{n+1} \pm t(1-\alpha/2, n-p-2)s(pred)$$

where $s(pred)$ is the standard error of transformed model, it can be approximated by

$$s(pred) = \sqrt{(1 + \tilde{X}_{new}'(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}\tilde{X}_{new})MSE}$$

where

$$MSE = \sum_{t=1}^{n}(Y_t - \hat{Y}_t)^2/(n-p-1-1)$$

and

$$\tilde{\mathbf{X}} = \begin{pmatrix} \tilde{X}_2' \\ \tilde{X}_3' \\ \vdots \\ \tilde{X}_n' \end{pmatrix}$$

Similarly, if we consider the confidence interval for EY_{n+1}, then it is

$$\hat{Y}_{n+1} \pm t(1-\alpha/2, n-p-2)s(\hat{Y}_{n+1})$$

where $s(\hat{Y}_{n+1})$ is the standard error of transformed model, it can be approximated by

$$s(\hat{Y}_{n+1}) = \sqrt{(\tilde{X}_{new}'(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}\tilde{X}_{new})MSE}$$

Example 6.1 (cont.) Suppose $X_{21} = 175.3$ the estimated model is

$$\hat{Y} = -0.17 + 0.17X$$

the error for e_{20} is

$$e_{20} = Y_{20} - \hat{Y}_{20} = 0.0139$$

The forecasted value for period 21 is

$$\hat{Y}_{21} = -0.17 + 0.17 * 175.3 + 0.63 * 0.0139 = 29.4$$

For the transformed model with $\tilde{Y}_t = Y_t - \hat{\rho}Y_{t-1}, \tilde{X}_t = X_t - \hat{\rho}X_{t-1}$ as response and predictor respectively , we need to consider the prediction standard error for $\tilde{X}_{21} = X_{21} - 0.63 * X_{20} = 66.93$. Its standard error is

$$s(pred) = \{MSE[1 + \frac{1}{19} + \frac{(\bar{\tilde{X}} - \tilde{X}_{21})^2}{\sum_{i=2}^{20}(\tilde{X}_i - \bar{\tilde{X}})^2}]\}^{1/2} = 0.0757$$

Thus, with confidence 95%, $t(1 - 0.05/2, 19 - 2) = 2.11$, the prediction interval is

$$\hat{Y}_{21} \pm 2.11 * 0.0757 = [29.24, 29.56]$$

(code)

7 Model test with autocorrelation Error terms

Actually, for regression model with first order autorrelated error terms, the model is

$$Y_t = \beta^\top X_t + \varepsilon_t, \qquad t = 1, ..., n$$

where $X_t = (1, X_{t1}, ..., X_{tp})'$ and $\beta = (\beta_0, \beta_1, ..., \beta_p)^\top$ with

$$\varepsilon_t = \rho\varepsilon_{t-1} + u_t$$

After transformation

$$\tilde{Y}_t = Y_t - \rho Y_{t-1}, \qquad \tilde{X}_t = X_t - \rho X_{t-1}, \qquad t = 2, ..., n$$

The transformed model is

$$\tilde{Y}_t = \tilde{\beta}\tilde{X}_t + u_t, \qquad t = 2, ..., n$$

It is easy to see that

$$\beta_0 = \tilde{\beta}_0/(1 - \rho), \beta_1 = \tilde{\beta}_1, ..., \beta_p = \tilde{\beta}_p$$

We estimate the transformed model and denote it by $\tilde{b}$, and thus

$$b_0 = \tilde{b}_0/(1-\hat{\rho}), b_1 = \tilde{b}_1, ..., b_p = \tilde{b}_p$$

and the estimated transformed model is

$$\hat{\tilde{Y}}_t = \tilde{b}'\tilde{X}_t$$

their fitted error is

$$\tilde{e}_t = \tilde{Y}_t - \hat{\tilde{Y}}_t$$

SSE, **SSR** and **SST** of the transformed model

$$SSE = \sum_{t=2}^{n} \tilde{e}_t^2, \text{with DF } (n-1)-p-1 = n-p-2$$

$$SSR = \sum_{t=2}^{n} (\hat{\tilde{Y}}_t - \bar{\tilde{Y}})^2, \text{with DF } p$$

$$SST = \sum_{t=2}^{n} (\tilde{Y}_t - \bar{\tilde{Y}})^2, \text{with DF } (n-1)-1 = n-2$$

where

$$\bar{\tilde{Y}} = \frac{1}{n-1}\sum_{t=2}^{n} \tilde{Y}_t$$

TEST of $H_0 : \beta_k = 0$

Note that the $H_0 : \beta_k = 0$ is equivalent to $H_0 : \tilde{\beta}_k = 0$. Therefore, we change to test the transformed model

1. Approach 1: t-statistic; see Chapter 2, part 2 (for the transformed model with n-1 observations)

2. Approach 2: F-statistic; see Chapter 2, part 2 (for the transformed model with n-1 observations)

TEST of $H_0 : \beta_0 = c$

Note that the $H_0 : \beta_0 = c$ is equivalent to $H_0 : \tilde{\beta}_0 = c(1-\rho)$. ; see Chapter 2, part 2 (for the transformed model with n-1 observations)

TEST of $H_0 : \beta_k = \beta_j = 0$ or more complicated hypothesis

We can adopt the full-model and reduced model approach as before, the only difference is the sample size is now $n-1$ instead of n; see Chapter 2, part 2

Example 7.1 For **(data)** with Y, X_1, X_2, X_3. We first fit a model

$$Y_t = \beta_0 + \beta_1 X_{t1} + \beta_2 X_{t2} + \beta_3 X_{t3} + \varepsilon_t$$

The $DW = 0.53 < L = 1.00$, thus there is positive autocorrelation

Now we consider transformed model

$$\tilde{Y}_t = \tilde{\beta}_0 + \tilde{\beta}_1 \tilde{X}_{t1} + \tilde{\beta}_2 \tilde{X}_{t2} + \tilde{\beta}_3 \tilde{X}_{t3} + \tilde{\varepsilon}_t$$

where $\tilde{Y}_t = Y_t - Y_{t-1}, \quad \tilde{X}_{t1} = X_{t1} - X_{t-1,1}, \quad \tilde{X}_{t2} = X_{t2} - X_{t-1,2}, \quad \tilde{X}_{t3} = X_{t3} - X_{t-1,3}$
$DW = 1.38 > L = 0.97$, thus there is no significant positive autocorrelation

It is easy to see from the calculation that $H_0 : \beta_3 = 0$ can be accepted with $\alpha = 0.05$ (based on the transformed model)

R-Code

Critical values for the Durbin–Watson d statistic ($\alpha = .05$)

	$n_p - 1 = 1$		$n_p - 1 = 2$		$n_p - 1 = 3$		$n_p - 1 = 4$		$n_p - 1 = 5$	
n	$d_{L,.05}$	$d_{U,.05}$	$d_{L,.05}$	$d_{U,.05}$	$d_{L,.05}$	$d_{U,.05}$	$d_{L,.05}$	$d_{U,.05}$	$d_{L,.05}$	$d_{U,.05}$
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

Source: From J. Durbin and G.S. Watson, "Testing for Serial Correlation in Least Squares Regression, II," *Biometrika* 30 (1951), 159–178. Reproduced by permission of the *Biometrika* Trustees.