

ESTIMATING AND TESTING AUTOCORRELATION WITH SMALL SAMPLES: A COMPARISON OF THE *C*-STATISTIC TO A MODIFIED ESTIMATOR

LAWRENCE T. DECARLO* and WARREN W. TRYON

Department of Psychology, Fordham University, Bronx, NY 10458, U.S.A.

(Received 5 February 1993)

Summary—Huitema and McKean (*Psychological Bulletin*, 110, 291–304, 1991) recently showed, in a Monte-Carlo study, that five conventional estimators of first-order autocorrelation perform poorly for small (< 50) sample sizes. They suggested a modified estimator and a test for autocorrelation. We examine an estimator not considered by Huitema and McKean: the *C*-statistic (Young, *Annals of Mathematical Statistics*, 12, 293–300, 1941). A Monte-Carlo study of the small sample properties of the *C*-statistic shows that it performs as well or better than the modified estimator suggested by Huitema and McKean (1991). The *C*-statistic is also shown to be closely related to the *d*-statistic of the widely used Durbin–Watson test.

INTRODUCTION

Research in the behavioral and social sciences often yields data that are ordered over time. Time series methods offer a means of recognizing the temporal order of the data in the analysis. A basic part of this analysis is the computation of the *autocorrelation* of the observations, that is, the correlation of the observation obtained at time or trial t with that obtained at time $t - 1$ (first-order autocorrelation), or at earlier times (higher-order autocorrelations). There are problems estimating autocorrelation, however, when the sample size is small (say less than 50 repeat observations). The autocorrelation estimator, for example, can be considerably biased. The performance of autocorrelation estimators for small sample sizes has implications for applied research, because the number of repeated observations is frequently small (for references relevant to behavioral analysis, see Matyas & Greenwood, 1991).

Huitema and McKean (1991) recently examined, in a Monte-Carlo study, five well-known estimators of first-order autocorrelation. They showed that the conventional estimators performed poorly for small sample sizes (less than 50 observations). Huitema and McKean (1991) proposed a modified estimator and showed, for positive autocorrelation, that it consistently out-performed the conventional estimators, although its performance for the smallest sample sizes was still poor. They also recommended a test for autocorrelation and showed that it offered an improvement over the conventional test.

A common aspect of all of the estimators considered by Huitema and McKean (1991) is that they are computed using sums of *cross-products*. In particular, if Y_t denotes the observation obtained at time t and Y_{t-1} is the previous observation, then all of the estimators compute the cross-product of the deviations of Y_t and Y_{t-1} about their means, where the means are computed using either all n observations or just $n - 1$ of them. There are several time series statistics, however, that use a related but nevertheless different approach: they are based on the sum of the squared *differences* between successive observations ($Y_t - Y_{t-1}$). An example is the *C*-statistic (Young, 1941), which is the focus of this article. Another example is the *d*-statistic of the Durbin–Watson test (Durbin & Watson, 1950, 1951) for autocorrelated residuals, which is discussed below. Statistics of this type were not considered by Huitema and McKean (1991). The purpose of the present article is to examine the small sample properties of the *C*-statistic. A Monte-Carlo study is used to compare the performance of the *C*-statistic for small sample sizes to that of the

*Author for correspondence.

conventional estimator and the modified estimator suggested by Huitema and McKean (1991). The next section introduces the estimators and briefly reviews Huitema and McKean's results.

AUTOCORRELATION AND ESTIMATION

A conventional estimator of first-order autocorrelation is:

$$r_1 = \frac{\sum_{t=1}^{n-1} (Y_t - \bar{Y})(Y_{t+1} - \bar{Y})}{\sum_{t=1}^n (Y_t - \bar{Y})^2}, \quad (1)$$

where Y_t and Y_{t+1} are the observations obtained at times (or on trials) t and $t + 1$, respectively, and \bar{Y} is the mean of all n observations. Note that r_1 is similar to Pearson's correlation r , with the difference being that the summation in the numerator of equation (1) includes $n - 1$ terms instead of n terms. This occurs because the cross-product of the last observation Y_n with Y_{n+1} cannot be computed (because Y_{n+1} is not observed). An important difference between r_1 and r is that, as shown by Fig. 1 of Huitema and McKean (1991), the bias of r_1 for small sample sizes can be large.

Huitema and McKean (1991) suggested the following modified estimator of first-order autocorrelation:

$$r_1+ = r_1 + \frac{1}{n}. \quad (2)$$

With respect to positive autocorrelation, which is typically found for time series data, adding the term $1/n$ to r_1 corrects to some extent for negative bias, that is, underestimation of the positive autocorrelation. Huitema and McKean's (1991) Monte-Carlo study showed that the bias of r_1+ for positive autocorrelation is smaller than the bias of r_1 , particularly for small sample sizes ($n < 50$). In addition, the average (over autocorrelations from 0 to 0.9) mean-squared error (MSE) was shown to be smaller for r_1+ than for r_1 . The MSE is a function of an estimator's variance and bias:

$$MSE = VARIANCE + (BIAS)^2.$$

The MSE should be considered when choosing among estimators because there is a trade-off between bias and variance.

Equations (1) and (2) show that r_1 and r_1+ are both computed using sums of cross-products. The C -statistic (Young, 1941), on the other hand, is based on the sum of the squared differences between successive observations:

$$C = 1 - \frac{\sum_{t=1}^{n-1} (Y_t - Y_{t+1})^2}{2 \sum_{t=1}^n (Y_t - \bar{Y})^2}. \quad (3)$$

Although it may not be apparent at first glance, the C -statistic is an estimator of first-order autocorrelation. In particular, simple algebra (see the Appendix) shows that the C -statistic is related to r_1 as follows:

$$C = r_1 + \frac{(Y_n - \bar{Y})^2 + (Y_1 - \bar{Y})^2}{2 \sum_{t=1}^n (Y_t - \bar{Y})^2}, \quad (4)$$

where Y_1 and Y_n are the first and last observations, respectively. Equation (4) is important because it shows that the C -statistic is similar in several ways to the modified estimator r_1+ [Equation (2)]. For example, as shown by equation (2), r_1+ corrects for the underestimation of positive autocorrelation by adding $1/n$ to r_1 . Equation (4) shows that the C -statistic is similar in that it also adds a fraction to r_1 . In addition, it is apparent that r_1+ converges to r_1 as the sample size gets

Table 1. Empirical bias in r_1 , r_1+ and C

ρ	N														
	6			10			20			30			50		
	r_1	r_1+	C	r_1	r_1+	C	r_1	r_1+	C	r_1	r_1+	C	r_1	r_1+	C
0.9	-0.73	-0.57	-0.53	-0.50	-0.40	-0.36	-0.27	-0.22	-0.19	-0.18	-0.15	-0.13	-0.11	-0.09	-0.08
0.8	-0.66	-0.49	-0.46	-0.44	-0.34	-0.31	-0.23	-0.18	-0.16	-0.15	-0.12	-0.11	-0.09	-0.07	-0.07
0.7	-0.59	-0.42	-0.39	-0.39	-0.29	-0.26	-0.20	-0.15	-0.14	-0.13	-0.10	-0.09	-0.08	-0.06	-0.06
0.6	-0.52	-0.36	-0.34	-0.34	-0.24	-0.22	-0.17	-0.12	-0.12	-0.12	-0.08	-0.08	-0.07	-0.05	-0.05
0.5	-0.46	-0.29	-0.27	-0.29	-0.19	-0.18	-0.15	-0.10	-0.10	-0.10	-0.07	-0.07	-0.06	-0.04	-0.04
0.4	-0.39	-0.23	-0.22	-0.25	-0.15	-0.14	-0.13	-0.08	-0.08	-0.09	-0.05	-0.05	-0.05	-0.03	-0.03
0.3	-0.33	-0.16	-0.16	-0.22	-0.12	-0.11	-0.11	-0.06	-0.06	-0.07	-0.04	-0.04	-0.04	-0.02	-0.02
0.2	-0.27	-0.10	-0.10	-0.17	-0.07	-0.07	-0.09	-0.04	-0.04	-0.04	-0.03	-0.03	-0.04	-0.02	-0.01
0.1	-0.22	-0.05	-0.05	-0.13	-0.03	-0.03	-0.07	-0.02	-0.02	-0.05	-0.01	-0.01	-0.03	-0.01	-0.00
0	-0.16	0.01	0.00	-0.10	0.00	0.00	-0.05	0.00	0.00	-0.03	0.00	0.00	-0.02	0.00	0.00
-0.1	-0.10	0.06	0.06	-0.06	0.04	0.04	-0.03	0.02	0.02	-0.02	0.01	0.01	-0.01	0.01	0.01
-0.2	-0.05	0.11	0.10	-0.02	0.08	0.07	-0.01	0.04	0.04	0.00	0.03	0.03	0.00	0.02	0.02
-0.3	-0.01	0.16	0.15	0.00	0.10	0.10	0.00	0.05	0.05	0.00	0.04	0.04	0.00	0.02	0.02
-0.4	0.05	0.22	0.21	0.04	0.14	0.14	0.02	0.07	0.07	0.02	0.05	0.05	0.01	0.03	0.03
-0.5	0.10	0.27	0.25	0.07	0.17	0.17	0.04	0.09	0.09	0.03	0.06	0.06	0.02	0.04	0.04
-0.6	0.16	0.32	0.31	0.10	0.20	0.20	0.06	0.11	0.11	0.04	0.07	0.07	0.02	0.04	0.04
-0.7	0.20	0.37	0.35	0.14	0.24	0.23	0.08	0.13	0.13	0.06	0.09	0.09	0.03	0.05	0.05
-0.8	0.24	0.41	0.39	0.17	0.27	0.27	0.09	0.14	0.15	0.07	0.10	0.10	0.04	0.06	0.06
-0.9	0.27	0.44	0.43	0.19	0.29	0.29	0.11	0.16	0.16	0.08	0.11	0.11	0.05	0.07	0.07

large, because $1/n$ approaches zero. It follows from equation (4) that the C -statistic also converges to r_1 as the sample size gets large, because the term to the right of r_1 approaches zero as n increases (only the denominator increases as n increases). In short, the relation between C and r_1+ implies that a comparison of their properties for small sample sizes is in order. We conducted a Monte-Carlo study with this goal in mind.

HYPOTHESIS TESTING

Although the focus of the present article is on estimation, a comparison of the statistics with respect to tests of hypotheses is also of interest. The null hypothesis of zero autocorrelation is commonly tested by using r_1 divided by $1/\sqrt{n}$ as a test statistic, where n is the number of observations. The test statistic is compared to critical values from the normal distribution. Huitema and McKean (1991, Fig. 4) showed that the Type I error rates for this test, however, deviate considerably from their normal values (for the right tail). For example, for a sample size of 10, the empirical Type I error rate for the right tail (positive autocorrelation) was 0.008 instead of the chosen 0.05 value. The deviation from the nominal value persisted for sample sizes of 20, 50 and 100.

Huitema and McKean (1991) suggested using an estimator of the standard error (SE) of r_1 proposed by Moran (1948), which is:

$$SE(r_1) = \frac{n - 2}{n \cdot \sqrt{n - 1}}$$

The recommended test procedure was to use r_1+ divided by the above quantity as a test for autocorrelation. Huitema and McKean (1991) showed that the empirical probability of a Type I error for this test was close to the nominal value (0.05) across all sample sizes. The test also had greater power than the conventional test across all sample sizes, although the power was low for the smallest sample sizes: (e.g. $n = 6$ and $n = 10$).

Table 2. Average MSE for r_1 , r_1+ and C

Sample size	r_1	r_1+	C
6	0.3133	0.1969	0.2029
10	0.1684	0.1217	0.1187
20	0.0653	0.0531	0.0509
30	0.0377	0.0322	0.0311
50	0.0194	0.0175	0.0171

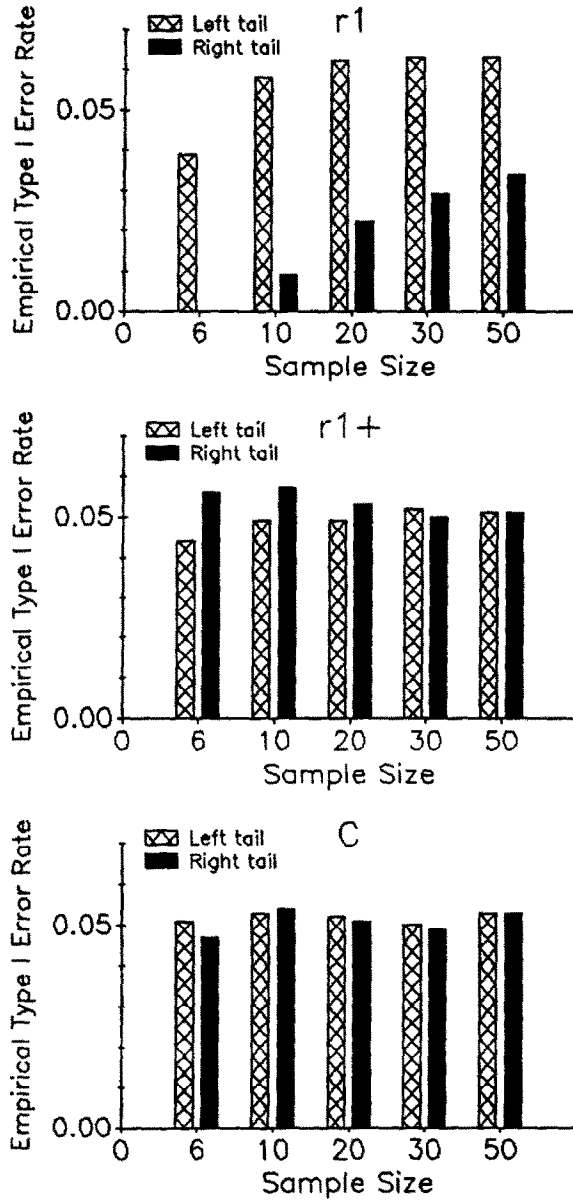


Fig. 1. Empirical Type I error rates (two-tailed, $\alpha = 0.10$) for r_1 divided by $1/\sqrt{n}$, r_1+ divided by $(n-2)/n(n-1)^{1/2}$ and C divided by $[(n-2)/(n^2-1)]^{1/2}$.

The estimator of the SE of the C -statistic is:

$$SE(C) = \sqrt{\frac{n-2}{n^2-1}}$$

(Young, 1941). The test statistic in this case is the ratio of C to its SE, which is compared to critical values from the normal distribution. We examine the Type I error rates and power of this test for several sample sizes.

A MONTE-CARLO STUDY

Method

The simulation was run on a DEC VAX using VMS Version 5.4. The normal probability

generator RANNOR of the SAS system (SAS Institute Inc., 1990) was used to generate observations from a first-order autoregressive process:

$$Y_t = \rho Y_{t-1} + e_t,$$

where ρ is the autocorrelation parameter and e_t is a random normal variate with a mean of zero and a variance of one. Each series was started with a normal variate, Y_0 , having a mean of zero and a variance of $1/(1 - \rho^2)$. Autocorrelation parameters of -0.9 to 0.9 in steps of 0.1 were examined. The sample sizes investigated were $6, 10, 20, 30$ and 50 . For each parameter value and sample size, $10,000$ samples were generated. The first 30 observations of each sample were dropped (to eliminate any dependency between the samples). The estimators r_1 , r_{1+} and C , plus their associated test statistics were computed for each sample. The empirical bias and MSE for each estimator were determined, along with the number of times the null hypothesis of zero autocorrelation was rejected. The latter was used to compute the empirical Type I error rates and power.

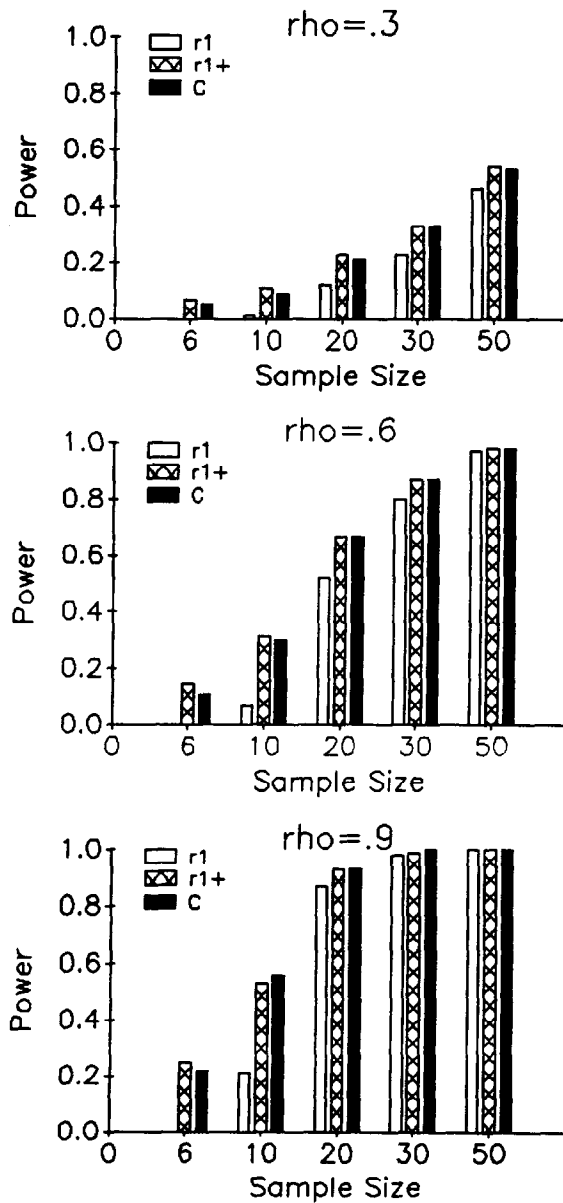


Fig. 2. Power for the tests using r_1 , r_{1+} and C for autocorrelation parameters of $0.3, 0.6$ and 0.9 .

Results

Estimation. Table 1 shows the empirical bias of r_1 , r_1+ and C for each autocorrelation parameter and each sample size. The table shows that the biases of r_1 and r_1+ are close to those obtained by Huitema and McKean (1991, Tables 2 and 5), thus replicating their results (using SAS in lieu of IMSL). For values of ρ greater than zero, the bias of r_1+ is consistently less than the bias of r_1 , which confirms Huitema and McKean's (1991) findings. In addition, the bias of C for positive values of ρ is less than or equal to the bias of r_1+ for all sample sizes. Thus, with respect to bias in estimating positive autocorrelation, C performs better than r_1+ .

Table 2 presents the average MSE (across autocorrelation values of 0 to 0.9 in steps of 0.1) for r_1 , r_1+ and C . The values obtained for r_1 and r_1+ are close to those obtained by Huitema and McKean (1991, Table 3). The table shows that, for all sample sizes, the average MSEs of both r_1+ and C are smaller than the average MSE of r_1 . In addition, the average MSE of C is smaller than the average MSE of r_1+ for all sample sizes except the smallest ($n = 6$). Inspection of the individual MSEs for each ρ suggested that the MSE of C tended to be smaller than the MSE of r_1+ for $\rho > 0.5$, whereas the MSE of r_1+ tended to be smaller for $\rho < 0.5$.

In sum, with respect to estimation of positive autocorrelation, Tables 1 and 2 show that C performs better than r_1+ .

Hypothesis testing. Figure 1 shows the empirical Type I error rates for a two-tailed test of the null hypothesis of no autocorrelation with $\alpha = 0.10$ (0.05 in each tail). The test statistics are r_1 divided by $1/\sqrt{n}$ (top panel), r_1+ divided by Moran's SE (middle panel), which is the test recommended by Huitema and McKean (1991) and C divided by its SE (bottom panel). The proportion of cases in the left and right tails are shown separately. Ideally, each bar should have a height of 0.05. The top panel shows, for the test using r_1 , that the proportion of cases in the right tail deviates considerably from the nominal 0.05 value across all sample sizes. For example, for a sample size of 10, the proportion of cases in the right tail was 0.009 instead of 0.05, which agrees with Huitema and McKean's (1991) finding of 0.008. In contrast, the middle panel shows that the empirical Type I error rates for the test using r_1+ are close to the nominal value (0.05 per tail) across all sample sizes. The results for r_1 and r_1+ replicate Huitema and McKean's (1991) findings (Figs 4 and 9). The bottom panel shows, for the test using C , that the empirical Type I error rates are close to the nominal value (0.05 per tail) across all sample sizes. Thus, with respect to Type I error rates, the test using C performs as well as the test proposed by Huitema and McKean (1991).

Figure 2 shows, for the five sample sizes, the power of the tests using r_1 , r_1+ and C for ρ s of 0.3 (top panel), 0.6 (middle panel), and 0.9 (bottom panel). The test in this case was two-tailed with $\alpha = 0.05$ (0.025 in each tail). The figure shows that the tests using r_1+ and C have similar power. In addition, the two tests have greater power than the conventional test (using r_1).

DISCUSSION

The simulation shows that the C -statistic has properties comparable to those of r_1+ for small sample sizes. With respect to estimation of positive autocorrelation, the bias and MSE of C were smaller than those of r_1+ . With respect to testing for positive autocorrelation, the empirical Type I error rates and power of the test using C were similar to those of the test using r_1+ . The results show that the search for improved small-sample estimators should consider the approach offered by the C -statistic, in addition to modifications or r_1 . Overall, statistics based on differences and cross-products are both useful for the analysis of correlated data.

If interest centers on estimation, then the results suggest that the C -statistic is preferable to r_1+ , because the bias and MSE tend to be smaller, although the difference between the two estimators is small. It should be kept in mind that the bias of both r_1+ and C is still quite large for small sample sizes (< 30). With respect to hypothesis testing, the power of both tests tends to be low for $n < 50$ (for $\rho < 0.5$). In light of the results, Box and Jenkins' (1976) recommendation of sample sizes of at least 50–100 repeat observations is shown to be reasonable. In situations where only small sample sizes can be obtained, as is often the case in behavioral analysis, one option is to pool the time series across S s, which is referred to as *pooled time series* or *cross-sectional time series* in statistics and econometrics (e.g. see Sayrs, 1989).

Huitema and McKean's (1991) study and our study have focused on testing for autocorrelation of the observations Y_t . It is often the case, however, that a model has been fitted to the data. Regression models, for example, are frequently used to determine how other variables affect the response, and to study trends or intervention effects. The test for autocorrelation in this case is applied to the residuals, which by definition are the difference between the observed and fitted values. The situation differs from the above in that the distribution of the test statistic depends on the sample values of the regressors.

Here we note that the C -statistic is closely related to the d -statistic of the widely used Durbin–Watson (DW) test (Durbin & Watson, 1950, 1951), which tests for autocorrelated residuals. The d -statistic is computed as follows:

$$d = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2}, \quad (5)$$

where e_t and e_{t-1} are the regression residuals for times t and $t - 1$, respectively. The d -statistic has been extensively investigated in statistics and econometrics, and has been shown to have good small sample properties (for references, see Kmenta, 1986; Judge, Griffiths, Hill, Lutkepohl & Lee, 1985).

Equation (5) shows that the d -statistic, which is based on the difference between successive residuals, is similar to the C -statistic, which is based on the difference between successive observations. In fact, if the C -statistic is computed on the residuals obtained from a regression analysis, so that Y in equation (3) is replaced by e , then C has an exact relation to d . In this situation, the term \bar{e} (i.e. \bar{Y}) in the denominator of equation (3) equals zero (the mean of the residuals is zero), so the denominator of equation (3) is simply twice that of equation (5). Writing out the terms summed in the numerator of equations (3) and (5) shows that they are identical. It therefore follows that:

$$C = 1 - \frac{1}{2}d.$$

The above shows that, if the C -statistic is computed using residuals from a regression analysis, then it has an exact relation to the d -statistic of the Durbin–Watson test. It should be noted that the test of the null hypothesis of zero autocorrelation differs, however, in that the d -statistic is compared to upper and lower critical bounds, whereas the C -statistic is compared to critical values from the standard normal distribution. The bounds for the d -statistic arise because of difficulties in determining an exact critical value which is independent of the sample values of the regressors.

In sum, there are two basic approaches to estimating and testing autocorrelation. One approach is based on the cross-product of successive observations; r_1 and r_1+ are examples. The second approach is based on the squared differences between successive observations; the C -statistic is an example. The two approaches are closely related, and an understanding of this relationship is important for future studies of small sample autocorrelation estimators.

REFERENCES

- Box, G. E. P. & Jenkins, G. M. (1976). *Time series analysis: Forecasting and control* (rev. Ed.). San Francisco: Holden-Day.
- Durbin, J. & Watson, G. S. (1950). Testing for serial correlation in least squares regression, I. *Biometrika*, 37, 409–427.
- Durbin, J. & Watson, G. S. (1951). Testing for serial correlation in least squares regression, II. *Biometrika*, 38, 159–178.
- Huitema, B. E. & McKean, J. W. (1991). Autocorrelation estimation and inference with small samples. *Psychological Bulletin*, 110, 291–304.
- Judge, G. G., Griffiths, W. E., Hill, R. C., Lutkepohl, H. & Lee, T. C. (1985). *The theory and practice of econometrics* (2nd Ed.). New York: Wiley.
- Kmenta, J. (1986). *Elements of econometrics* (2nd Ed.). New York: Macmillan.
- Matyas, T. A. & Greenwood, K. M. (1991). Problems in the estimation of autocorrelation in brief time series and some implications for behavioral data. *Behavioral Assessment*, 13, 137–157.
- Moran, P. A. P. (1948). Some theorems on time series II. The significance of the series correlation coefficient. *Biometrika*, 35, 255–260.
- SAS Institute Inc. (1990). *SAS Language: Reference* (Version 6, First Ed.). Cary, NC: Author.
- Says, L. W. (1989). *Pooled time series analysis*. Sage University Series on Quantitative Applications in the Social Sciences, 07–070. Beverly Hills: Sage Publications.
- Young, L. C. (1941). On randomness in ordered sequences. *Annals of Mathematical Statistics*, 12, 293–300.

APPENDIX

Relation Between C and r_1

The C -statistic is:

$$C = 1 - \frac{\sum_{t=1}^{n-1} (Y_t - Y_{t+1})^2}{2 \sum_{t=1}^n (Y_t - \bar{Y})^2}.$$

The numerator of the term on the far right can be expanded as:

$$(Y_t - Y_{t+1})^2 = [(Y_t - \bar{Y}) - (Y_{t+1} - \bar{Y})]^2.$$

Squaring and rearranging terms gives:

$$C = 1 - \frac{\sum_{t=1}^{n-1} (Y_t - \bar{Y})^2 + \sum_{t=1}^{n-1} (Y_{t+1} - \bar{Y})^2 - 2 \sum_{t=1}^{n-1} (Y_t - \bar{Y})(Y_{t+1} - \bar{Y})}{2 \sum_{t=1}^n (Y_t - \bar{Y})^2}.$$

Writing the last term in the numerator on the right as a separate fraction shows that it is simply r_1 , so that:

$$C = r_1 + 1 - \frac{\sum_{t=1}^{n-1} (Y_t - \bar{Y})^2 + \sum_{t=1}^{n-1} (Y_{t+1} - \bar{Y})^2}{2 \sum_{t=1}^n (Y_t - \bar{Y})^2}.$$

Next, the two terms in the numerator on the right can be replaced using the following identities:

$$\sum_{t=1}^{n-1} (Y_t - \bar{Y})^2 = \sum_{t=1}^n (Y_t - \bar{Y})^2 - (Y_n - \bar{Y})^2$$

and

$$\sum_{t=1}^{n-1} (Y_{t+1} - \bar{Y})^2 = \sum_{t=1}^n (Y_t - \bar{Y})^2 - (Y_1 - \bar{Y})^2,$$

which gives

$$C = r_1 + 1 - \frac{2 \sum_{t=1}^n (Y_t - \bar{Y})^2 - (Y_n - \bar{Y})^2 - (Y_1 - \bar{Y})^2}{2 \sum_{t=1}^n (Y_t - \bar{Y})^2}.$$

Substituting

$$\frac{2 \sum_{t=1}^n (Y_t - \bar{Y})^2}{2 \sum_{t=1}^n (Y_t - \bar{Y})^2},$$

for 1 and cancelling terms gives

$$C = r_1 + \frac{(Y_n - \bar{Y})^2 + (Y_1 - \bar{Y})^2}{2 \sum_{t=1}^n (Y_t - \bar{Y})^2}$$

which is equation (4).