Testing for Relationships Between Time Series

ALICE ORCUTT NAKAMURA, MASAO NAKAMURA and GUY H. ORCUTT*

The usual procedures for testing the significance of sample correlations between pairs of independently normally distributed series are not appropriate for testing sample correlations between pairs of autocorrelated series. We present sampling evidence supporting our hypothesis that the distributions of sample correlations between pairs of unrelated first-order Markov series conditional on the first lag sample autocorrelations of the series correlated are independent of the population first lag autocorrelations of these series. Based on this evidence, a new test of significance for correlations between autocorrelated series is proposed, which, although treating them as first-order Markov series, does not depend on the generally unknown generating properties of the series.

1. INTRODUCTION

Two related null hypotheses have been widely used by researchers in testing for significant relationships between pairs of time series:

I. H_0: $\beta = 0$ in the model $Y = \alpha + \beta X + \epsilon$, where the X series is assumed to be fixed, the disturbance term ϵ is independent of X, and, in particular, ϵ and, therefore, Y are assumed to be independently and normally distributed.

II. H_0: $\rho_{XY} = 0$ where ρ denotes the zero-order cross-correlation between two independently, normally distributed series X and Y.

The statistic used to test Hypothesis I is exactly the same t quantity that is commonly used to test Hypothesis II. Also it has been shown in [18, pp. 25–33] that even if both X and Y are assumed to be randomly, normally distributed in Hypothesis I, the usual t-test for this hypothesis is still appropriate. There is evidence, too, that the usual tests will frequently yield reasonable results even when the series correlated are drawings from nonnormal populations (see [12]). There is, however, one point at which these tests of significance for Hypotheses I and II clearly break down. Economic time series are usually autocorrelated, and ever since the article by Yule [34], it has been clear that higher correlations are to be expected by chance between unrelated autocorrelated series than between unrelated series which are independently distributed.

* Alice Orcutt Nakamura is assistant professor and Masao Nakamura is associate professor, both with Faculty of Business Administration and Commerce, The University of Alberta, Edmonton, Alberta, Canada. Guy H. Orcutt is Whitney Griswold Professor, Department of Economics and Institution for Social and Policy Studies, Yale University, New Haven, Conn. 06520. This work was supported by research grants from the Faculty of Graduate Studies and Research and the Faculty of Business Administration and Commerce, The University of Alberta. The authors would like to thank Arthur S. Goldberger, Phil Howrey, Jan Kmenta, Barry H. Margolin, Lester D. Taylor and Arnold Zellner, and two referees for their many helpful suggestions.

2. EVALUATING THE SIGNIFICANCE OF RELATIONSHIPS BETWEEN AUTOCORRELATED TIME SERIES

The problem of testing the significance of observed relationships between autocorrelated time series has been approached in various ways. For purposes of discussion we will categorize these approaches as

1. Spacing or selection of observations used.
2. Trend removal.
3. Autoregressive transformation, or differencing.
4. Direct evaluation.

The first three approaches are aimed at removing whatever autocorrelation is present from at least one of the series to be correlated. Approach 1 includes not only the practice of simply dropping out observations from the series of interest, but also the more sophisticated exact tests of significance for correlations between time series which have been developed (see [14, 15 and 26]). The main difficulty with these exact tests, as well as with simple spacing, is that not all the information in the original data is used, with a consequent reduction of degrees of freedom.[1] Among the difficulties associated with Approach 2 is the fact that trend-corrected data are not generally independently distributed (see [1, 11]). Approach 3 involves choosing an appropriate autoregressive transformation. The practical difficulties involved in making this choice are well-known (see [5, 19, 20, 30, 32]). Another difficulty, however, is that when we test for the significance of a linear relationship between, say, the first differences of two series, we are not testing exactly the same hypothesis as when we test for a linear relation between the two series themselves.[2]

[1] In [15], Hannan obtained an exact test of correlation between two series which can be applied whenever one of the two series is Markovian. However, for a Markov process of order h, only $n(h+1)^{-1}$ observations will be used in the correlation where n is the number of observations available. The remainder of the observations are used to reduce the series to independence in time. Also, as Hannan notes [15, p. 320], "When $\rho_1 \neq \rho_2$ the exact test of significance of the correlation between the two series there given is still an exact test, but the conclusions as to the power of this test and the others there considered, when x_t also comes from a Markov process, do not follow. . . ."

[2] A simple example may help to illustrate this point. Arbitrarily draw any two smooth curves. Divide each of these into N equal segments. On the vertical division lines between these segments, mark off for Curve 1 points alternately one unit above and one unit below the curve. Repeat with Curve 2. The two new curves may obviously be made to have highly positively correlated first differences. However, as is evident from the manner of construction, the original values of these new series may have zero or even a negative correlation. Another difficulty is that when an autoregressive transformation is performed on actual economic time series, random measurement error variances will likely be amplified relative to the phenomena under study.

March 1976, Volume 71, Number 353
Theory and Methods Section

The last approach is the only one aimed at directly testing the significance of observed relationships between autocorrelated time series, and most of this work has been focused on developing a test of significance for the more general null hypothesis (see [2, 3, 4, 17, 29, 31]):

III. H_0: $\rho_{XY} = 0$, where both X and Y are assumed to be autocorrelated.

Spectral analysis in [13, 24] represents one procedure which can be used for investigating and testing Hypothesis III. Basic questions have been raised, however, about the suitability of spectral methods to economic data and problems in [24, p. 340, and 33, p. 33].[3] Our work is aimed at developing a procedure for testing the significance of correlations between autocorrelated time series which is better suited to the data and problems of economists.

3. THE NULL DISTRIBUTION OF CORRELATIONS BETWEEN PAIRS OF FIRST ORDER MARKOV SERIES

McGregor [21] and McGregor and Bielenstein [22] have worked out approximate distributions for sample correlations between pairs of stationary first-order Markov series with both known and fitted means. Like Bartlett's large-sample variance approximations [2], these approximate distributions depend only on the length of the series correlated and the product of their first-lag population autocorrelations. An approximate density function of the type McGregor has derived could be numerically integrated to determine significance levels for sample correlations between stationary first-order Markov series. As one step in our exploration, we chose alternatively to approximate these significance levels using the Monte Carlo approach.

Our generating relationships were of the form

$$X_t = \rho_1 X_{t-1} + u_t \quad \text{and} \quad Y_t = \rho_2 Y_{t-1} + v_t \ , \qquad (3.1)$$

where u and v were generated by two Chen random normal number generators.[4] Both u and v have a mean of zero and a standard deviation of 25.[5] We set the initial values of X and Y in (3.1) equal to zero, and then generated pairs of long series of 600,000 items each for the following values of ρ_1 and ρ_2: (−.9, −.9), (−.7, −.7), (−.5, −.5), (−.3, −.3), (−.1, −.1), (0, 0), (.1, .1), (.3, .3), (.5, .5), (.7, .7) and (.9, .9). To minimize the effect of the initial values used in generating u and v, the first 30 items generated in each of these long series were discarded. Also, every other subsequent group of 30 items generated was discarded, leaving 10,000 paired subseries of length 30 for each pair of values of ρ_1 and ρ_2. For each of these pairs of subseries, the Pearson product moment correlation coefficient was calculated. The resulting 10,000 sample correlation coefficients for each pair of values of ρ_1 and ρ_2 were then ordered according to their absolute values, and two-tail significance levels of .01, .02, .05 and .10 were calculated for each group. Our results for $\rho_1 = \rho_2 = \pm.9$, $\pm.5$ and 0 are shown in Table 1, along with Fisher's significance levels [10] for the case where $\rho_1 = \rho_2 = 0$, as a check on our computational procedures. (Significance levels for $\rho_1 = \rho_2 = \pm.7$, $\pm.3$ and $\pm.1$ were calculated but are not shown.)[6]

In the absence of further information concerning factors which affect the probability of obtaining chance correlations between unrelated autocorrelated series, we might accept our Monte Carlo results shown in Table 1, or McGregor's analytic expressions, as a basis for a revised null test for correlations between autocorrelated economic time series. However, in 1948 Orcutt and James [29] found evidence that the conditional variances of sample correlations between pairs of unrelated autocorrelated series, given their first-lag sample autocorrelations, seem to vary systematically with the product of the values of these sample autocorrelations.

---

[3] Although spectral techniques can be advantageously used in analyzing certain time series, the following questions have been raised with respect to economic applications: (1) the amount of data (the number of items in a series) required before it becomes sensible to attempt to estimate a spectrum would seem to be greater than 100, which would form an important barrier to analyzing annual economic time series, particularly when the mean of the generating process must be estimated from the available data (see [24, 25]); (2) spectral and time domain (regression) methods are mathematically equivalent only if an observed time series is assumed to be identically zero outside the observation interval, which means that in spectral analysis emphasis is distributed equally over the entire observation interval, while in economic time series analysis a tangible point of the argument is that a good forecast is not made by assuming that the time series is zero outside the observation interval (see [33]); and (3) spectral methods are not attractive in the coordination between statistical and subject matter knowledge, since it is well-known that the spectral components defined by specific bands of the spectrum often do not lend themselves to a subject-matter interpretation (see [33]).

[4] See Chen [7, 8]. The initial values used for the starting integers were 748511649 and 147303541 for the u series and 180810529 and 536841077 for the v series. Satisfactory statistical properties are reported for random numbers generated using these initial numbers in Chen [7, 8]. The computer used was the IBM System/360 model 67 at the University of Alberta Computing Center.

[5] The results of our paper would have been identical regardless of the variances chosen for the disturbance terms.

Let $X_t = \rho X_{t-1} + u_t (t = 2, 3, 4, \cdots)$ and $X_1 = u_1$ (hence, $X_0 = 0$), where u_1 $u_2, u_3, \cdots$ are independently and normally distributed with mean 0 and standard deviation $k (k > 0)$. Then $X_1 = u_1$ and

$$X_t = \sum_{i=1}^{t} \rho^{t-i} u_i \quad \text{for} \quad t \geq 2 \ .$$

Suppose we generate $Z_1, Z_2, Z_3, \cdots$ which are independently and normally distributed with mean zero and standard deviation one. We can now generate the X_t-process by using $u_i = kZ_i (i = 1, 2, \cdots)$. Thus, we have

$$X_t = k \sum_{i=1}^{t} \rho^{t-i} Z_i (t \geq 2) \quad \text{and} \quad X_1 = kZ_1 \ .$$

From this it follows that the estimated first-order autocorrelations (denoted by r_1 and r_2 in our paper) and the estimated zero-order cross-correlations (denoted by r in our paper) should be identical for all values of the variances of the disturbance terms for our generating process, since

$$X_t = \sum_{i=1}^{t} \rho_1{}^{t-i} Z_i \ , \quad Y_t = \sum_{i=1}^{t} \rho^{t-i} Z_i \ , \quad X_t' = kX_t \ , \quad \text{and} \quad Y_t' = \ell Y_t \ ,$$

and, therefore,

$$r_{1X'} = \frac{\Sigma\,(kX_t - k\bar{X}_t)(kX_{t-1} - k\bar{X}_{t-1})}{\Sigma\,(kX_{t-1} - k\bar{X}_{t-1})^2} = \frac{k^2\,\Sigma\,(X_t - \bar{X}_t)(X_{t-1} - \bar{X}_{t-1})}{k^2\,\Sigma\,(X_{t-1} - \bar{X}_{t-1})^2} = r_{1X}$$

and

$$r_{X'Y'} = \frac{\Sigma\,(kX_t - k\bar{X}_t)(\ell Y_t - \ell\bar{Y}_t)}{k\ell(ns_X s_Y)} = \frac{k\ell\,\Sigma\,(X_t - \bar{X}_t)(Y_t - \bar{Y}_t)}{k\ell(ns_X s_Y)} = r_{XY} \ .$$

[6] The nonsymmetry about zero which is shown in Table 1 for $\rho_1 = \rho_2 = .9$ and $\rho_1 = \rho_2 = -.9$ was also observed for $\rho_1 = \rho_2 = .7$ and $\rho_1 = \rho_2 = -.7$. This nonsymmetry is due to the dependence, established in Sections 4, 5 and 6, of the distribution of the sample correlation coefficient on the first-lag sample autocorrelations of the series correlated.

1. Critical Points for Correlations Between Series of Length 30, Given the Population First-Lag Autocorrelations of These Series

$\rho_1 = \rho_2$	Level of significance for two-tailed test			
	.10	.05	.02	.01
−.9	.70	.77	.84	.86
−.5	.38	.44	.51	.55
0	.31	.36	.42	.47
.5	.38	.44	.51	.56
.9	.63	.71	.77	.81
	Fisher's approximation for series of length 27			
0	.3233	.3809	.4451	.4869
	Fisher's approximation for series of length 32			
0	.2960	.3494	.4093	.4487

4. EVIDENCE OF THE IMPORTANCE OF THE SAMPLE AUTOCORRELATIONS

Our generated series were used to carry out a test of the Orcutt-James hypothesis that the conditional variances of correlations between pairs of unrelated, autocorrelated series, given the first-lag sample autocorrelations of these series, are independent of the population first-lag autocorrelations of the series correlated.[7]

For each of our subseries of length 30, the autoregressive parameter was estimated using least squares regression,[8] and the Pearson product moment correlation coefficient was calculated for each pair of subseries.

Each of our 11 sets of 10,000 sample correlation coefficients corresponding to specified values for ρ_1 and ρ_2 was then cross-classified according to the respective values of the sample autocorrelation coefficients, r_1 and r_2, of the subseries used to compute each of the sample correlations. (In carrying out this classification we did not distinguish between (r_1, r_2) and (r_2, r_1).) The conditional variance of the sample correlations was calculated for each cell.[9]

Intervals of 0.1 were used in grouping the subseries according to their sample autocorrelation coefficients, except for sample autocorrelation coefficients ranging between −0.10 and +0.10. Intervals of 0.025 were used to group the series with sample autocorrelation coefficients falling in this range.

We next tested the hypothesis that the observed variances of the sample correlations corresponding to any given cell in our two-way classification can be regarded as drawings from a single population. For instance, for the cell corresponding to $0.1 \leq r_1 \leq 0.2$ and $-0.2 \leq r_2 \leq -0.1$ or $0.1 \leq r_2 \leq 0.2$ and $-0.2 \leq r_1 \leq -0.1$, we tested the hypothesis that the sample correlations falling in this cell and corresponding to values of the generating autoregressive parameters of (−.7, −.7), (−.5, −.5), (−.3, −.3), (−.1, −.1), (0, 0), (.1, .1), (.3, .3), (.5, .5) and (.7, .7) can all be regarded as drawings from the same population.[10]

For each cell in our two-way classification containing sample correlations corresponding to two or more pairs of values of generating parameters, and more sample correlations than the number of pairs of corresponding generating parameters, we performed a simple one-way analysis of variance. For each cell the null hypothesis is

$$H_0\colon \rho_{1,k}^2 = \cdots = \rho_{11,k}^2 , \tag{4.1}$$

where $\rho_{i,k}$ denotes the population correlation coefficient for the sample correlations in the kth cell in our two-way classification, and corresponding to the ith pair of generating parameters ρ_1 and ρ_2 in (3.1).

Since intuition as well as a t-test confirm that

$$E(r \mid r_1, r_2, \rho_1, \rho_2) = E(r \mid r_1, r_2) = E(r) = 0 ,$$

where r denotes the sample correlation coefficient,

$$\overline{r^2} = \sum r^2/n = \sum [r - E(r)]^2/n$$

is an unbiased estimate of the variance of r. Thus, testing our null hypothesis H_0 is equivalent to testing the hypothesis that the sample correlations in any given cell of our two-way classification all have the same corresponding population conditional variances.

Out of 258 F-ratios, 13 or approximately 5.03 percent are significant at a 5-percent critical level, and five or approximately 1.93 percent are significant at a one-percent critical level. Thus, we feel that the hypothesis that the observed variances of sample correlations corresponding to any given cell in our two-way classification

[7] Orcutt and James worked with series generated by the model

$$Y_{t+1} = Y_t + 0.3(Y_t - Y_{t-1}) + \epsilon_{t+1} ,$$

where the random elements used, the ϵ_{t+1}, were two digit numbers drawn from a population having a rectangular distribution and a range of −49 through zero to 49.

[8] Since, in practice, one would have no way of knowing the true value of the constant term, we estimated a constant term along with the autoregressive parameter.

As to our choice of an estimator for the autoregressive parameter, Copas [9] has compared by Monte Carlo methods the performance for estimation in a stable Markov time series with unknown mean of a "mean likelihood" estimator, the first sample serial correlation coefficient, the first sample serial correlation coefficient bias corrected, and the least squares estimator. The "mean likelihood" estimator and the least squares estimator were found to perform almost equally well in terms of the mean squared error when the results were averaged over $\beta = -0.9\ (0.1)0.9$, where β denotes the true autoregressive parameter, and to provide better results than the other two estimators. Looking at the results for the individual values of β, the "mean likelihood" estimator was found to give the least mean squared error for the approximate range [−0.3, +0.6], the correlation coefficient for true values of $\beta < -0.3$, and the least squares estimator for true values of $\beta > 0.6$. These results are claimed to be valid for series of length 10 and 20, and accurate to approximately two decimal places. Because we are primarily interested in true values of the autoregressive parameter greater than 0.6, and because the least squares estimator is the most familiar to economists, we have chosen to base our study on this estimator. It should be noted, however, that we are not chiefly concerned with accurately estimating the true autoregressive parameter. We are rather trying to uncover a stable dependence of the distribution of the sample correlation coefficient on observable sample autoregressive properties of the series correlated. How well other estimators of the autoregressive parameter would perform in terms of this objective is not known.

[9] More precisely, the conditional variance referred to here is

$$\overline{r^2} = (\sum_R r^2)/\text{ number of elements in } R ,$$

where R is an index set over which the sum is taken. In this case, R is defined in terms of a cell in our two-way classification.

[10] No observations from the sets of r^2's with the generating autoregressive parameters of (−.9, −.9) and (.9, .9) fell in this particular cell. In general, for any given cell k, some of the r_{ik}^2, $i = 1, \cdots, 11$, may not exist.

can be regarded as drawings from a single population is supported.[11]

5. AUTOREGRESSIVE INFORMATION CONTAINED IN THE PRODUCT AND SIGNS OF r_1 AND r_2

The existence of a dependence of the conditional variances of sample correlations between unrelated autocorrelated series on the product of the associated first-lag sample autocorrelations is clearly demonstrated in Figure A. The observations in this figure are again based on our sets of pairs of subseries of length 30 constructed according to (3.1).

Each of our 11 sets of 10,000 r^2's was ordered by arranging the corresponding values of the product of the sample autoregressive coefficients from largest to smallest. Using this ordering, each set was divided into ten groups of 1,000 correlations each—the first group corresponding to the 1,000 largest values of the product r_1r_2 for that set, the second group to the next 1,000 values of r_1r_2, etc. For each group we calculated

$$\overline{r_1r_2} = \sum r_1r_2/1000 \ , \quad \overline{r^2} = \sum r^2/1000 \quad \text{and}$$
$$\sigma_{r^2}{}^2 = (1/999)\{\sum r^4 - (1/1000)(\sum r^2)^2\} \ .$$

Each of these groups of 1,000 r^2's is a subset of one of our 11 original samples of r^2's. Thus, $\overline{r^2}$ now stands for the variance of the correlation coefficients in each group, conditional on the mean value of the products of the sample autocorrelation coefficients, as well as on the population autocorrelations. In Figure A we have plotted and connected the resulting ten pairs of values of $\overline{r^2}$ and $\overline{r_1r_2}$ for each of the following pairs of values of ρ_1 and ρ_2: $(-.9, -.9)$, $(-.7, -.7)$, $(-.5, -.5)$, $(-.3, -.3)$, $(.3, .3)$, $(.5, .5)$, $(.7, .7)$, $(.9, .9)$.

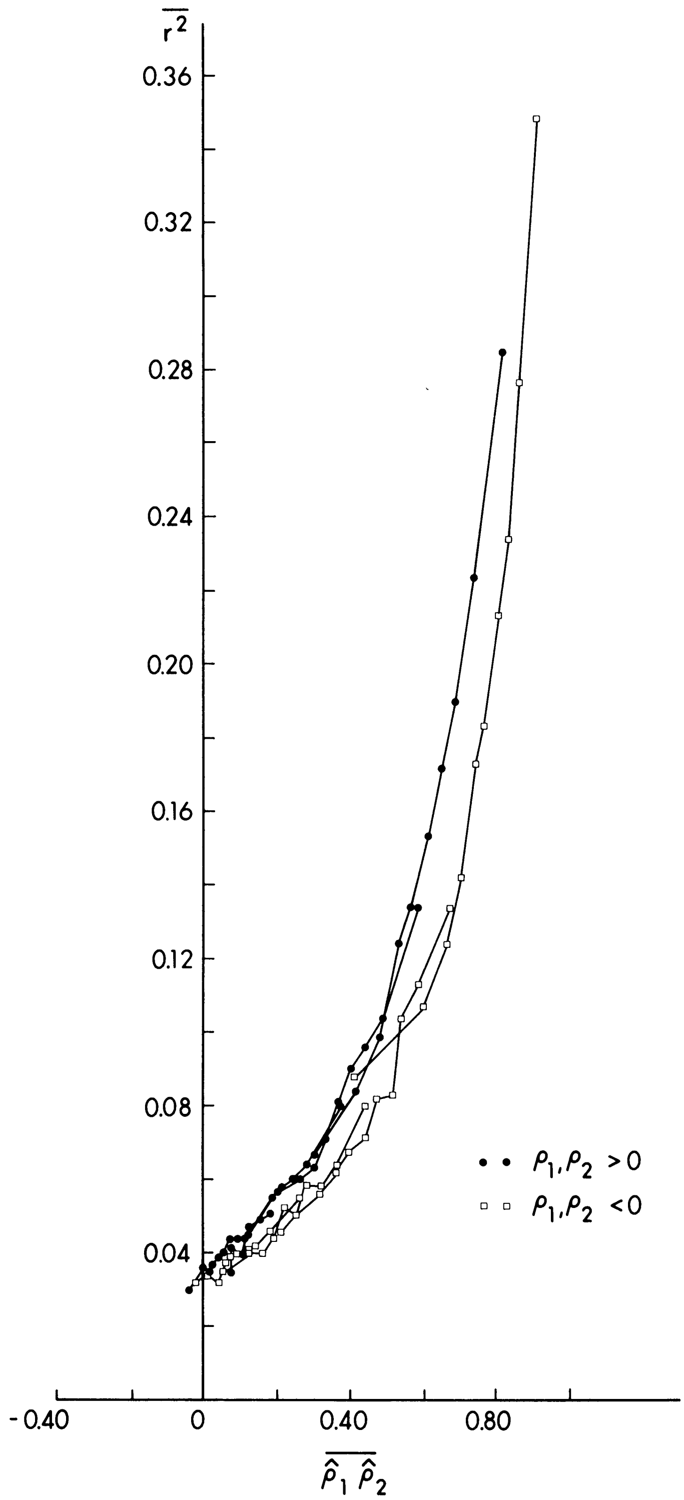
This figure shows clearly that the conditional variances of the sample correlations generated in this study are, in fact, dependent on the corresponding values of the product of the sample autoregressive coefficients. There are, however, small but systematic differences between those line segments corresponding to positive and negative pairs of values of the generating autoregressive parameters.

6. SIMPLE AND MODIFIED PRODUCT CLASSIFICATIONS

These systematic differences noted in Figure A were verified by an analysis of variance test. We grouped all our sample correlation coefficients according only to the values of the products of the sample autoregressive coefficients for the subseries yielding these correlations, using the same intervals as in Section 4. We then tested Hypothesis (4.1), where k now denotes the cell in our simple product classification. Out of 19 F-ratios, ten or approximately 52.6 percent are significant at a 5-percent critical level, and eight or approximately 42.1 percent are significant at a one-percent critical level. Thus, the hypothesis must be rejected that the sample correlations corresponding to each cell in our simple product classification are drawings from a single population.

We next grouped our sample correlation coefficients by the values of the products of the sample autoregressive coefficients for the subseries yielding these correlations, with a distinction being made between $r_1, r_2 > 0$ and $r_1, r_2 < 0$. We again tested Hypothesis (4.1), where k now

A. A Demonstration of the Dependence of r^2 on r_1r_2



[11] Distributions of the sample correlation coefficient conditional on various other sample properties of the series correlated may also display this property.

denotes a cell in our modified product classification. These F-ratios are shown in Table 2. Out of 29 F-ratios, two, or approximately 6.9 percent are significant at a 5-percent critical level, and zero are significant at a one-percent level. Thus, using this modified product classification, we feel that the hypothesis is supported that the distributions of sample correlations between unrelated first-order Markov series conditional on the values of the product and signs of the sample first-lag autocorrelations of the series correlated are independent of the population first-lag autocorrelations of these series.

2. F-Tests for Modified Product Classification

Cell	F-Ratio	Degrees of freedom between	Degrees of freedom within
−.2 to −.1	.813	7	346
−.1 to −.075	1.318	8	466
−.075 to .050	.466	8	1169
−.050 to −.025	.450	8	3053
−.025 to 0	.368	8	11165
	$r_1, r_2 > 0$		
0 to .025	.470	7	6511
.025 to .050	.654	7	3364
.050 to .075	.836	6	2520
.075 to .1	.804	6	2173
.1 to .2	.330	6	6895
.2 to .3	.881	4	5323
.3 to .4	1.215	3	4678
.4 to .5	.279	3	4003
.5 to .6	3.327 [a]	2	3512
.6 to .7	1.118	2	2734
.7 to .8	.054	1	1681
.8 to .9	—	—	—
.9 to 1.0	.052	1	40
	$r_1, r_2 < 0$		
0 to .025	.466	7	7273
.025 to .050	2.072	6	3683
.050 to .075	1.887	5	2742
.075 to .1	1.293	6	2338
.1 to .2	.884	6	6786
.2 to .3	2.364	4	5183
.3 to .4	.515	3	4521
.4 to .5	.952	3	3837
.5 to .6	1.048	2	3517
.6 to .7	.588	2	3087
.7 to .8	4.989 [a]	1	3325
.8 to .9	2.426	1	2782

[a] Significant at the five-percent level.

Having established our hypothesis for the 45-degree line $-1 \leq \rho_1 = \rho_2 \leq 1$, we next explored the parameter rectangle $0 \leq \rho_1, \rho_2 \leq 1$. Twenty-one sets of 1,000 pairs of subseries of length 30 were generated by the procedure described in Section 3, using the parameter combinations (.9, .9), (.9, .7), (.9, .5), (.9, .3), (.9, .1), (.9, 0), (.7, .7), (.7, .5), (.7, .3), (.7, .1,), (.7, 0), (.5, .5), (.5, .3), (.5, .1), (.5, 0), (.3, .3), (.3, .1), (.3, 0) (.1, .1), (.1, 0), and (0, 0). As before, analysis of variance was used to test the hypothesis that the sample correlations in any given cell in our modified product classification all have the same population conditional variance, regardless of the true values of ρ_1 and ρ_2. Out of 25 F-ratios, one, or approximately 4 percent, are significant at a 5-percent critical level and zero are significant at a one-percent critical level. Thus, we feel that our hypothesis is supported for the entire parameter region $-1 \leq \rho_1, \rho_2 \leq 1$ that the distributions of sample correlations between unrelated first-order Markov series, conditional on the values of the product and signs of the sample first-lag autocorrelations of the series correlated, are independent of the population first-lag autocorrelations of these series.

7. A SUMMARY OF OUR FINDINGS

Thus, to summarize our findings,[12] let

$$p(r|r_1, r_2, \rho_1, \rho_2) = p(r, r_1, r_2|\rho_1, \rho_2)/p(r_1, r_2|\rho_1, \rho_2) \quad (7.1)$$

denote the distribution of the sample correlation coefficient r conditional on the sample first-lag autocorrelations r_1, r_2 and the population autocorrelations ρ_1, ρ_2 (and implicitly on $\rho = 0$, i.e., $\sigma_{uv} = 0$). The conventional tests of significance for r rely on

$$p(r|\rho_1, \rho_2) = \int\int p(r, r_1, r_2|\rho_1, \rho_2)dr_1, dr_2 \ , \quad (7.2)$$

and it appears from Bartlett's and McGregor's work that

$$p(r|\rho_1, \rho_2) \cong p(r|\rho_1\rho_2) \ . \quad (7.3)$$

We note that (7.1) is more informative than (7.2). In particular, Monte Carlo evidence is presented supporting the hypothesis that

$$V(r|r_1, r_2, \rho_1, \rho_2) = V(r|r_1r_2, \text{sign } (r_1), \text{sign } (r_2)) \ , \quad (7.4)$$

and on the basis of this evidence we conjecture that

$$p(r|r_1, r_2, \rho_1, \rho_2) = p(r|r_1, r_2) = p(r|r_1r_2, \text{sign } (r_1), \text{sign } (r_2)) \ . \quad (7.5)$$

Monte Carlo tabulation of (7.5) is presented in Section 8.

8. A NEW TEST OF SIGNIFICANCE FOR CORRELATIONS BETWEEN TIME SERIES OF THIRTY OBSERVATIONS

On the basis of our findings in Section 6 we have proceeded to use Monte Carlo methods to approximate significance levels for correlations between autocorrelated time series of length 30, with given first-lag sample autocorrelations, and which can be viewed as drawings from first-order Markov processes. Our observed sample correlations for all 11 of our sets of 10,000 correlations were grouped according to the product of the first-lag sample autocorrelations of the subseries correlated, with a distinction being made between correlations corresponding to $r_1, r_2 > 0$ and those corresponding to $r_1, r_2 < 0$. This is the modified product classification used in Section 6.

The observed sample correlations in each group containing 100 or more observations were then ordered from

[12] This formalization of our findings was suggested to us by Arthur S. Goldberger.

smallest to largest according to their absolute values, and two-tail significance levels of .10, .05, .02 and .01 were calculated for each group. Our results are shown in Table 3.

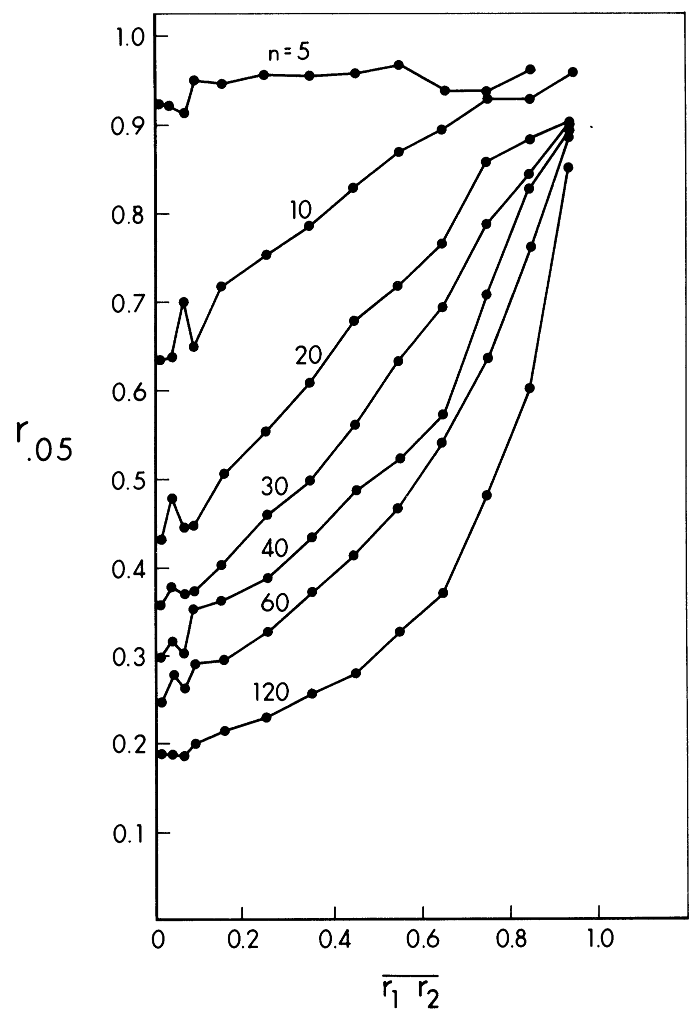
3. Critical Points for Correlations Between Series of Length 30, Given the First-Lag Sample Autocorrelations of These Series

Class intervals for $r_1 r_2$	Level of significance for two-tailed test .10	.05	.02	.01	Number of observations
−.2 to −.1	.30	.35	.40	.42	354
−.1 to −.075	.27	.32	.39	.40	475
−.075 to −.050	.28	.34	.41	.43	1178
−.050 to −.025	.30	.36	.41	.44	3062
−.025 to 0	.30	.36	.43	.47	11174
		$r_1, r_2 > 0$			
0 to .025	.31	.36	.43	.47	6519
.025 to .050	.32	.38	.46	.49	3372
.050 to .075	.34	.39	.46	.49	2527
.075 to .1	.34	.39	.46	.50	2180
.1 to .2	.36	.42	.48	.54	6902
.2 to .3	.40	.46	.53	.58	5328
.3 to .4	.45	.51	.57	.61	4682
.4 to .5	.50	.57	.63	.67	4007
.5 to .6	.57	.63	.70	.74	3515
.6 to .7	.65	.71	.76	.78	2737
.7 to .8	.72	.77	.82	.84	1683
.8 to .9	.78	.84	.88	.90	508
		$r_1, r_2 < 0$			
0 to 0.25	.30	.36	.42	.45	7281
.025 to .050	.31	.36	.42	.47	3690
.050 to .075	.32	.38	.44	.48	2748
.075 to .1	.32	.38	.44	.49	2345
.1 to .2	.34	.40	.46	.50	6793
.2 to .3	.38	.44	.50	.54	5188
.3 to .4	.41	.47	.54	.58	4525
.4 to .5	.46	.52	.58	.62	3841
.5 to .6	.52	.58	.64	.67	3520
.6 to .7	.57	.63	.69	.73	3090
.7 to .8	.66	.72	.78	.81	3327
.8 to .9	.78	.81	.85	.87	2784
.9 to 1.0	.86	.88	.90	.92	600

Of course, a test which is only applicable for series of length 30 would be of limited usefulness. Thus, 30 more samples of 1,000 pairs of subseries of length 130 were now generated by the procedure described in Section 3, using the parameter combinations of (1.0, 1.0), (1.0, .9), (1.0, .7), (1.0, .5), (1.0, .3), (.9, .9), (.9, .7), (.9, .5), (.7, .3), (.7, .1), (−1.0, −1.0), (−1.0, −.9), (−1.0, −.7), (−1.0, −.5), (−1.0, −.3), (−.9, −.9), (−.9, −.7), (−.9, −.5), (−.7, −.3), (−.7, −.1), (1.0, −1.0), (1.0, −.9), (1.0, −.7), (1.0, −.5), (1.0, −.3), (.9, −.9), (.9, −.7), (.9, −.5), (.7, −.3), and (.7, −.1). For the first 5, 10, 15, 20, 25, 30, 40, 60 and 120 items in each pair of subseries, the Pearson product moment correlation coefficient was calculated, and the autoregressive parameter was estimated as before for each of the subseries correlated. We next classified our 30,000 values of the correlation coefficient, r, for each value of n according to the product of the estimated first-lag autoregressive parameters $r_1 r_2$, with a distinction being made between $r_1, r_2 > 0$ and $r_1, r_2 < 0$. As before, intervals of 0.1 were used, except between −0.10 and +0.10, where intervals of 0.025 were used.

For each cell containing at least 50 observations in our modified product classification for each value of n, the values of r were ordered according to their absolute values from smallest to largest, and critical values of r were calculated for two-tailed significance levels of .10, .05, .02 and .01. Due to an improved choice of values for ρ_1 and ρ_2, we were able to calculate these critical values of r for all cells in our modified product classification, such that $-1.0 < r_1, r_2 < 1.0$, for n = 5, 10, 15, 20, 25, 30, 40, 60 and 120. The critical values of r for a two-tailed significance level of .05 are shown in Figure B for n = 5, 10, 20, 30, 40, 60 and 120.

B. Critical Points for Correlations Between Series of Lengths 5, 10, 20, 30, 40, 60, and 120, Given the First Lag Sample Autocorrelations of These Series for $r_1, r_2 > 0$ and $\alpha = .05$



Based on these calculations, a small FORTRAN program has been developed for calculating and testing the significance of correlations between autocorrelated time series. The program is applicable for $20 \leq n \leq 120$ and $-1.0 < r_1, r_2 < 1.0$.[13]

It is reassuring to note that these new critical values of r for the two-tailed significance levels of .10, .05, .02 and .01 for $n = 30$ are virtually identical to the values presented in Table 3 for all areas of overlap, except for $-0.10 \leq r_1r_2 \leq 0.10$, where the results are seen to be somewhat unstable. This is further evidence that relationship (7.5) is true, and that a reliable testing procedure has been developed.

Questions have been raised as to whether the tables which we have developed might have been more cheaply derived without resorting to the Monte Carlo method. In particular, it has been pointed out that Hannan [16, p. 368] suggests that a satisfactory procedure for testing the significance of correlations between autocorrelated time series "should be to use r as an ordinary correlation from $N(1 - \rho_1\rho_2)/(1 + \rho_1\rho_2)$ observations. (Of course, ρ_1 and ρ_2 would need to be estimated from the data and mean corrections would have to be made.)" This approach is suggested by the form of McGregor's [21] and McGregor and Bielenstein's [22] approximations of the distributions of sample correlations between pairs of stationary first-order Markov series with known and fitted means, and by the fact that the variance of these distributions is near to $(1 + \rho_1\rho_2)/n(1 - \rho_1\rho_2)$. As Hannan notes, this procedure was suggested by Bartlett in 1935. Tables facilitating the application of this procedure for $10 \leq n \leq 100$ and $.1 \leq r_1r_2 \leq .9$ were also presented by Orcutt [28] at a meeting of the American Statistical Association in 1949. The estimator used for ρ_1 in Orcutt's paper is

$$r_1' = 1 - (\tfrac{1}{2})(\sigma^2/S^2) \ ,$$

where

$$\sigma^2 = \sum_{i=1}^{n-1} (Y_{i+1} - Y_i)^2/(n-1) \text{ and } S^2 = \sum_{i=1}^{n} (Y_i - \bar{Y})^2/n \ ,$$

with no adjustment being made for bias.

Since the hypothesis that

$$V(r \mid r_1, r_2, \rho_1, \rho_2) \cong V(r \mid \rho_1, \rho_2) \cong V(r \mid \rho_1\rho_2) \cong V(r \mid r_1r_2)$$

is at the heart of this suggested approach, Nakamura and Nakamura [23] examined the percentage errors which would result from using the approximation

$$V(r \mid r_1, r_2, \rho_1, \rho_2) \cong ((1 + \overline{r_1r_2})/n(1 - \overline{r_1r_2})) - (2(\overline{r_1r_2})/n^2(1 - \overline{r_1r_2})^2) \ , \quad (8.1)$$

for $n = 30$ and values of the generating autoregressive parameters of (−.9, .9), (−.7, .7), (−.3, .3), (−.3, −.3), (.3, .3), (−.7, −.7), (.7, .7), (−.9, −.9), and (.9, .9). Expression (8.1) was found to provide a reasonably good, though far from perfect, approximation of $V(r \mid r_1, r_2, \rho_1, \rho_2)$. Moreover, the approximation provided by (8.1) was found to be substantially better on the whole than the approximation provided by

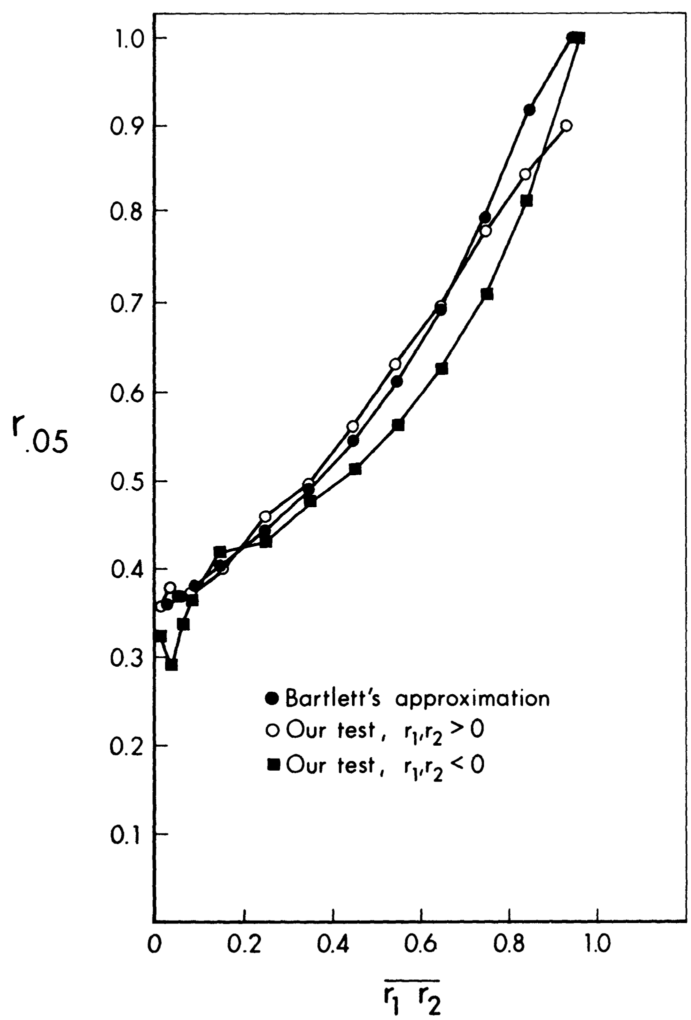
$$V(r \mid r_1, r_2, \rho_1, \rho_2) \cong \frac{1 + \rho_1\rho_2}{n(1 - \rho_1\rho_2)} - \frac{2\rho_1\rho_2}{n^2(1 - \rho_1\rho_2)^2} .$$

Close examination of the findings of this study also revealed marked differences, however, in the results obtained for $\rho_1 = \rho_2 = .3$ and $\rho_1 = \rho_2 = -.3$, for $\rho_1 = \rho_2 = .7$ and $\rho_1 = \rho_2 = -.7$, and for $\rho_1 = \rho_2 = .9$ and $\rho_1 = \rho_2 = -.9$. These differences are substantiated in Sections 5 and 6.

In Table 4 we show the critical points for a two-tailed significance level of .05, which are obtained by treating r as an ordinary correlation from $30(1 - r_1r_2)/(1 + r_1r_2)$ observations. These results have been plotted in Figure C against our results for $r_1, r_2 > 0$ and $r_1, r_2 < 0$ for $n = 30$.

Bartlett's approximation does not capture the differences which we have established depending on whether $r_1, r_2 > 0$ or $r_1, r_2 < 0$, nor is the fit particularly good for values of $r_1r_2 > .7$. However, this approximation clearly is a tremendous improvement over ignoring autocorrela-

C. Values of $r_{.05}$ Found Using Bartlett's Approximation with r_1r_2 Substituted for $\rho_1\rho_2$, Compared with Our Values for $r_{.05}$ for $r_1, r_2 > 0$ and $r_1, r_2 < 0$



[13] This program, as well as the data tapes containing the values of r, r_1 and r_2, which were calculated for n = 5, 10, 15, 20, 25, 30, 40, 60 and 120, are available on request from M. Nakamura.

4. Bartlett's Approximation

r_1r_2	$\frac{30(1-r_1r_2)}{1+r_1r_2}$	$r_{.05}$ [a]
.03	28.25	.360
.06	26.60	.371
.09	25.04	.380
.15	22.17	.404
.25	18.00	.443
.35	14.44	.491
.45	11.37	.545
.55	8.70	.611
.65	6.36	.692
.75	4.28	.796
.85	2.43	.999

[a] These values for $r_{.05}$ were obtained by using linear interpolation between the values of the correlation coefficient shown in Table V-A of [10].

tion when testing the significance of sample correlations. As would be expected, the goodness of the approximation achieved appears to improve as n becomes larger. It is also interesting to note that Bartlett's approximation with r_1 and r_2 substituted for ρ_1 and ρ_2 seems to provide better estimates of the critical points for the distribution $p(r|r_1, r_2)$ than when this approximation is used in its original form to find estimates of the critical points of the distribution $p(r|\rho_1, \rho_2)$. For instance, according to Bartlett's approximation, when $\rho_1 = \rho_2 = .9$, then $n(1-\rho_1\rho_2)/(1+\rho_1\rho_2) = 3.14$ and $r_{.05} \cong .87$. However, looking at Table 1, we find that for this case $r_{.05} = .71$. This finding casts doubt on whether making a bias correction in our estimates of ρ_1 and ρ_2 (as suggested by Hannan and others) would improve the comparison illustrated in Figure C between Bartlett's approach to estimating critical points for the distribution $p(r|r_1, r_2)$ and ours.

9. POWER COMPARISONS

Finally, for series of length 30 we compared the power of our test with the power of a test based on the population autoregressive parameters. For $\rho_1 = \rho_2 = 0$, $\pm.5$, and $\pm.9$, 1,000 subseries of length 30 were generated using the relationship

$$X_t = \rho_1 X_{t-1} + u_t \ ,$$

and ten sets of 1,000 subseries of length 30 were generated using the relationship

$$Y_t = \rho_2 Y_{t-1} + a u_t + v_t \ .$$

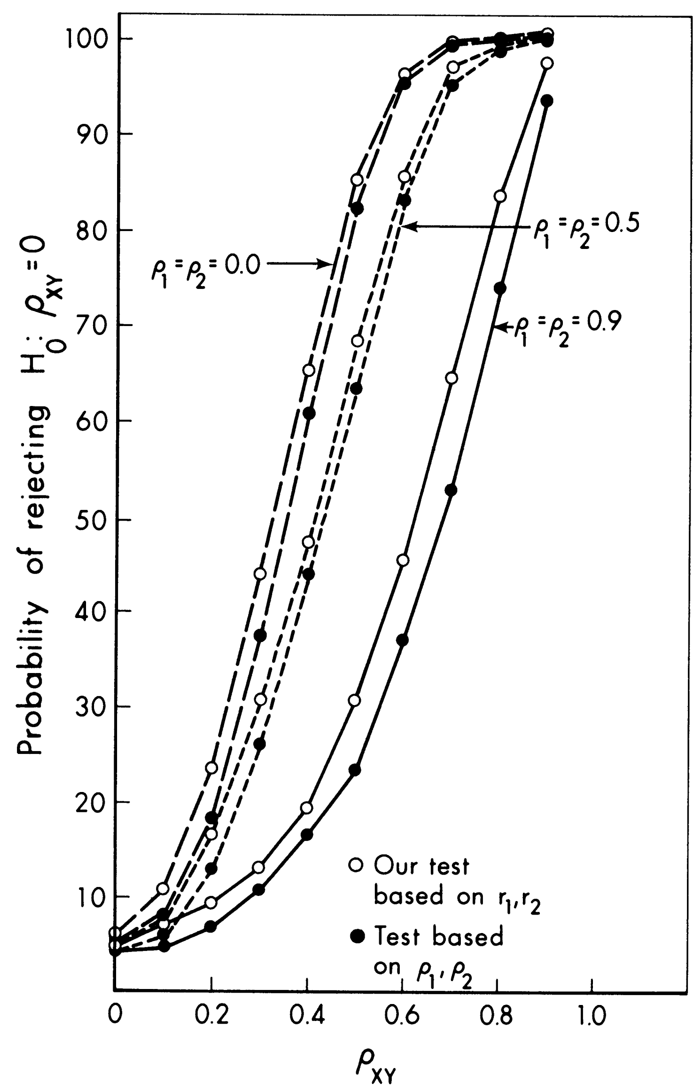
The values of a were chosen so as to generate ten sets of 1,000 subseries for each pair of values of ρ_1 and ρ_2 with population cross-correlations, ρ_{XY}, between the pairs of X and Y subseries in each set of 0, .1, .2, .3, .4, .5, .6, .7, .8 and .9, respectively. For each pair of subseries we calculated r, r_1 and r_2 just as before.

We now tested the significance of each of the resulting 50,000 sample correlations by first applying our test of significance for $\alpha = .10$, .05, .02 and .01, and then by applying the corresponding critical values based on the true values of ρ_1 and ρ_2 which are shown in Table 1. For each set of 1,000 sample correlations corresponding to each pair of values of ρ_1 and ρ_2 we counted the number of times the null hypothesis

$$H_0\colon \rho_{XY} = 0$$

was rejected using each test of significance. (The complete tables of our results for $\rho_1 = \rho_2 = 0, \pm.5$, and $\pm.9$, $\alpha = .10$, .05, .02 and .01, and $r_1, r_2 > 0$ are available on request from the authors.) Power functions comparing these two testing procedures for $\alpha = .05$; $\rho_1 = \rho_2 = 0$, .5, and .9; and $r_1, r_2 > 0$ are shown in Figure D.

D. Power Functions Comparing Our Test with a Test Based on ρ_1 and ρ_2 for $n = 30$, $\alpha = .05$, $\rho_1 = \rho_2 = 0$, .5, .9, and $r_1, r_2 > 0$



Our test was found to be uniformly more powerful than the test based on ρ_1 and ρ_2 for all pairs of values of ρ_1 and ρ_2, and all values of α, for which power functions were computed. The differences between the two tests were found to be greater for larger absolute values of ρ_1 and ρ_2 and for smaller values of α. Although we did not compute power functions for any other values of n, presumably the advantage of our test over a test based on ρ_1 and ρ_2 would be smaller for larger values of n. In addition to being more powerful than a test of significance based on ρ_1 and ρ_2, of course, our proposed test also has the advantage that it is based on the estimated autoregressive properties of the series correlated rather than on the underlying generating properties which, in general, are unknown.

[*Received February 1974. Revised March 1975.*]

REFERENCES

[1] Ames, E. and Reiter, S., "Distribution of Correlation Coefficients in Economic Time Series," *Journal of the American Statistical Association*, 56 (September 1961), 637–56.

[2] Bartlett, M.S., "Some Aspects of the Time Correlation Problem in Regard to Tests of Significance," *Journal of the Royal Statistical Society*, 98 (1935), 536–43.

[3] ———, "On the Theoretical Specification and Sampling Properties of Autocorrelated Time-Series," *Journal of the Royal Statistical Society Supplement, Ser. B*, 8, No. 1 (1946), 27–41.

[4] ———, *Stochastic Processes*, Cambridge, England: Cambridge University Press, 1955.

[5] Basawa, I.V., "Estimation of the Autocorrelation Coefficient in Simple Markov Chains," *Biometrika*, 59, No. 1 (1972), 85–9.

[6] Box, G.E.P. and Jenkins, G.M., *Time Series Analysis; Forecasting and Control*, San Francisco: Holden-Day, Inc., 1970.

[7] Chen, E.H., "A Random Normal Number Generator for 32-Bit-Word Computers," *Journal of the American Statistical Association*, 66 (June 1971), 400–3.

[8] ———, "Supplement to 'A Random Normal Number Generator for 32-Bit-Word Computers,' " Health Sciences Computing Facility, University of California, Los Angeles, Calif., 1971.

[9] Copas, J.B., "Monte Carlo Results for Estimation in a Stable Markov Time Series," *Journal of the Royal Statistical Society, Ser. A*, 129, Part 1 (1966), 110–6.

[10] Fisher, R.A., *Statistical Methods for Research Workers*, 13th ed., London; Oliver & Boyd, Ltd., 1948.

[11] Gartaganis, A.J., "Autoregression in the United States Economy, 1870–1929," *Econometrica*, 22 (April 1954), 228–43.

[12] Gayen, A.K., "The Frequency Distribution of the Product-Moment Correlation Coefficient in Random Samples of Any Size Drawn from Nonnormal Universes," *Biometrika*, 38 (June 1951), 219–47.

[13] Granger, C.W.J. and Hatanaka, M., *Spectral Analysis of Economic Time Series*, Princeton, N.J.: Princeton University Press, 1964.

[14] Hannan, E.J., "Exact Tests for Serial Correlation," *Biometrika*, 42, Nos. 1 and 2 (1955), 133–42.

[15] ———, "An Exact Test for Correlation Between Time Series," *Biometrika*, 42, Nos. 3 and 4 (1955), 316–26.

[16] ———, *Multiple Time Series*, New York: John Wiley & Sons, Inc., 1970.

[17] Jenkins, G.M. and Watts, D.G., *Spectral Analysis and Its Applications*, San Francisco: Holden-Day, Inc., 1968.

[18] Johnston, J., *Econometric Methods*, New York: McGraw-Hill Book Co., 1960.

[19] Levenbach, H., "Estimation of Autoregressive Parameter from a Marginal Likelihood Function," *Biometrika*, 59, No. 1 (1972), 61–71.

[20] Lomnicki, Q.A. and Zaremba, B.K., "On the Estimation of Autocorrelation in Time Series," *Annals of Mathematical Statistics*, 28, No. 1 (1957), 140–58.

[21] McGregor, J.R., "The Approximate Distribution of the Correlation Between Two Stationary Linear Markov Series," *Biometrika*, 49, Nos. 3 and 4 (1962), 379–88.

[22] ——— and Bielenstein, U.M., "The Approximate Distribution of the Correlation Between Two Stationary Linear Markov Series, II," *Biometrika*, 52, Nos. 1 and 2 (1965), 301–2.

[23] Nakamura, A.O. and Nakamura, M., "Estimating the Variances of Sample Correlations," *Proceedings of the Social Statistics Section of the Annual Meeting of the American Statistical Association*, New York City (December 27–30, 1973), 365–9.

[24] Naylor, T.H., Wertz, K. and Wonnacott, T.H., "Spectral Analysis of Data Generated by Simulation Experiments with Econometric Models," *Econometrica*, 37 (April 1969), 333–52.

[25] Neave, H.R., "Observations on 'Spectral Analysis of Short Series—A Simulation Study' by Granger and Hughes," *Journal of the Royal Statistical Society, Ser. A*, 135, Part 3 (1972), 393–405.

[26] Ogawara, M., "A Note on the Test of Serial Correlations Coefficients," *Annals of Mathematical Statistics*, 22, No. 1 (1951), 115–8.

[27] Orcutt, G.H., "A Study of the Autoregressive Nature of the Time Series Used for Tinbergen's Model of the Economic System of the United States, 1919–1932," *Journal of the Royal Statistical Society, Ser. B*, 10, No. 1 (1948), 1–45.

[28] ———, "Demonstrating the Existance of Relationship," Paper presented at the 109th Annual Meeting of the American Statistical Association, New York City, December 27–30, 1949.

[29] ——— and James, S.F., "Testing the Significance of Correlation Between Time Series," *Biometrika* 35, Nos. 3 and 4 (1948), 397–413.

[30] ——— and Winokur, H.S., Jr., "First-Order Autoregression: Inference, Estimation, and Prediction," *Econometrica*. 37 (January 1969), 1–14.

[31] Walker, S.G., "Apparent Correlation Between Independent Series of Autocorrelated Observations," *Biometrika*, 37, Nos. 1 and 2 (1950), 184–5.

[32] Weinstein, A.S., "Alternative Definitions of the Serial Correlation Coefficient in Short Autoregressive Sequences," *Journal of the American Statistical Association*, 53 (December 1958), 881–92.

[33] Wold, H.O.A. (Editor), *Econometric Model Building*, Amsterdam: North-Holland Publishing Co., 1964.

[34] Yule, G.U., "Why Do We Sometimes Get Nonsense Correlations Between Time Series?" *Journal of the Royal Statistical Society*, 89 (1926), 1–64.