

UNCLASSIFIED

AD 259 869

*Reproduced
by the*

ARMED SERVICES TECHNICAL INFORMATION AGENCY
ARLINGTON HALL STATION
ARLINGTON 12, VIRGINIA



UNCLASSIFIED

NOTICE: When government or other drawings, specifications or other data are used for any purpose other than in connection with a definitely related government procurement operation, the U. S. Government thereby incurs no responsibility, nor any obligation whatsoever; and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data is not to be regarded by implication or otherwise as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use or sell any patented invention that may in any way be related thereto.

TESTS BASED ON THE MOVEMENTS IN AND THE
COMOVEMENTS BETWEEN m -DEPENDENT TIME SERIES ¹

by

Leo A. Goodman

Columbia University
Mathematical Statistics

June 1961

CATALOGED BY ASTIA
AS AD NO. _____

259869

205 800



This research was sponsored in part by the Office of Naval Research under contract Number Nonr-266(33), Project Number NR042-034. Reproduction in whole or in part is permitted for any purpose of the United States Government.

¹ This paper is dedicated to the memory of Yehuda Grunfeld. It was written while the author was a Visiting Professor of Mathematical Statistics and Sociology at Columbia University on leave of absence from the University of Chicago. For their helpful comments, the author is indebted to Jacob Mincer and W. Allen Wallis.

XEROX

1. INTRODUCTION AND SUMMARY

Tests of the existence of correlation between the movements in two time series, which eliminate at least the primary effects of trends in the series, will be presented in this article. These tests are modifications and generalizations of a test for correlation proposed by Moore and Wallis [14] and of tests proposed by Goodman and Grunfeld [9], [8]. A test of the existence of trend in a time series will also be presented here. This test is a generalization of a test for trend proposed by Cox and Stuart [2].

The tests presented here can be applied to certain kinds of time series that are m -dependent. A time series $X = \{X_1, X_2, \dots, X_{n+1}\}$ is defined as m -dependent if $(X_1, X_2, \dots, X_j)$ and $(X_j, X_{j+1}, \dots, X_{n+1})$ are independent whenever $j > 1+m$ (see [13]). For an m -dependent time series, any two observations X_i, X_j will be independent whenever $j > i+m$. Two examples of m -dependent time series are the following: (I) if the $X_1, X_2, \dots, X_{n+1}$ are mutually independent, then X is 0-dependent; (II) if X is a process of moving averages of the form

$$X_i = \sum_{j=0}^m \alpha_j \xi_{i-j},$$

where the ξ_i are (unobserved) mutually independent random variables ($i=0, \pm 1, \pm 2, \dots$) (see [1], [20]), then X is m -dependent. Other examples of m -dependent time series could be presented here, but this will not be necessary.

The earlier papers by Moore and Wallis [14], Goodman and Grunfeld [9], [8], Cox and Stuart [2], present tests that can be applied to certain kinds of time series that are 0-dependent. (The case where the series of first differences are 0-dependent is also discussed in [14].) For time series that are not 0-dependent but that are m -dependent for some $m > 0$, these tests require modification. The purpose of the present paper is to present simple modifications of these tests that can be applied to certain kinds of m -dependent ($m \geq 0$) time series of long duration.

It is often the case that observed time series are not 0-dependent. The earlier papers discussed tests for 0-dependent time series partly because mathematical results were more readily available for this special case; i.e., the case of mutually independent random variables. The earlier authors were of course aware of the fact that the observed time series, to which tests appropriate for 0-dependent series were applied, might actually not be 0-dependent. For example, Cox and Stuart [2], in discussing their test for trend, point out that "positive serial correlation among the observations would increase the chance of a significant answer even in the absence of a trend"; i.e., a significant answer obtained by an application of their test for trend to a (serially correlated) time series might be due to the presence of serial correlation rather than to the presence of a trend. The modification of their test presented in the present article does not have this disadvantage.

Moore and Wallis [14] point out that some of the techniques presented in their paper are "restricted in scope by the fact that in many time series problems neither the hypothesis of sequential randomness in observations nor that of sequential randomness in the differences is tenable". The modifications of their test presented herein are not so restricted in scope.

In order to apply the tests presented here, it is not necessary to know the specific value of m that describes the order of dependence of the observed time series. The value of m could be zero or any positive integer. The magnitude of m should, however, be small relative to the duration of the observed time series. The tests presented here under the assumption that the value of m is specified will actually be suitable even if the true value of m is less than that specified. Thus, even if there is doubt concerning the true value of m for a particular m -dependent time series, it will still be possible to apply these tests. For time series that are not m -dependent for any value of m (e.g., where X_i and X_j are not independent for any $i \neq j$),

the methods presented herein are not to be recommended.

If time series X is m_1 -dependent and time series Y is m_2 -dependent, where m_1 and m_2 may have different values, then the definition of m -dependence given above implies that both series are η -dependent, where $\eta = \max[m_1, m_2]$. The tests of the existence of correlation between the movements in X and Y presented herein under the assumption that both X and Y are m -dependent can be applied by taking $m = \eta$. Denoting $\min[m_1, m_2]$ by μ , the statistic appropriate for testing the existence of correlation between the movements in X and Y in the case where X and Y are η -dependent, will converge in probability (under the null hypothesis when X and Y are η -dependent) to the statistic appropriate in the case where both X and Y are μ -dependent. The test statistic appropriate where both X and Y are μ -dependent can be used (instead of the test statistic appropriate when X and Y are η -dependent) even if the order of dependence of X or Y is greater than μ . Thus the test presented in [9], which was justified there in the case where both X and Y are 0-dependent, can also be justified in the case where one of the m -dependent series is not 0-dependent but the other one is.

The tests presented in the earlier papers were based upon the signs of the $X_j - X_i$ for $j > i$, $i=1, 2, \dots$. In [9] and [4], $j=i+1$, $i=1, 2, \dots, n$; in [8], $j=i+k$, $i=1, 2, \dots, n+1-k$ (k is a fixed positive integer); in [2], other pairs of i, j were used (e.g., $j=1+2(n+1)/3$, $i=1, 2, \dots, (n+1)/3$). If the sign of the difference $X_j - X_i$ is positive, then the "total movement" in X between time i and j is positive. Most of the tests presented in the present paper, will also be based on the signs of the differences (i.e., the signs of the movements) in the time series. Moore and Wallis [14] have pointed out that "in certain types of data the signs of differences are more accurate than the magnitudes of either the observations or the differences With economic

measures of the kind for which index numbers are ordinarily used, it may be certain that a change has been in a given direction (e.g., when all components of the index change in the same way), questionable how much the change has been (because of ambiguities in the weighting system), and meaningless to state an absolute standing". For these types of data and for data available only in ordinal form, the methods presented herein may be particularly well suited.

It was not necessary, in the earlier papers, to assume that the distributions from which the observations ^{come} had Λ had a specific form (e.g., it was not necessary to assume that the observations were normal variates) in order to derive the distribution of the appropriate test statistics. The tests presented there were distribution-free. It was, however, assumed in these papers that the observations were mutually independent and that the distribution from which they ^{come} had Λ was continuous. To derive the asymptotic distribution of the test statistics presented herein, the assumptions made will be somewhat more general.

The calculations required to perform the tests presented in the earlier papers and those presented herein are rather simple. While the modified tests presented here require more calculations than the earlier tests, these calculations remain simple.

Let $X = \{X_1, X_2, \dots, X_{n+1}\}$ and $Y = \{Y_1, Y_2, \dots, Y_{n+1}\}$ be two different observed time series of equal duration. Let $X_{i+1} - X_i = W_i$ and $Y_{i+1} - Y_i = Z_i$ for $i=1, 2, \dots, n$. We shall for convenience assume that the W_i and Z_i have continuous distributions. Let $U_i = 1$ if $W_i > 0$, and $U_i = 0$ if $W_i \leq 0$. Let $V_i = 1$ if $Z_i > 0$, and $V_i = 0$ if $Z_i \leq 0$. If $U_i = 1$, then a positive movement in X occurred between times i and $i+1$; if $V_i = 1$, then a positive movement in Y occurred between times i and $i+1$; if $U_i = V_i = 1$, then a positive comovement occurred; if $U_i = V_i = 0$, then a negative comovement occurred; if $U_i = 1, V_i = 0$, then a contramovement took place. (The term "comovement" used here is adopted from earlier work by Friedman [5]. It was also used by

Grumfeld in [12]. The term "contramovement" is adopted from Goodman and Grumfeld [9].) The observed distribution of the n pairs (U_i, V_i) (for $i=1,2,\dots,n$) can be summarized in the following 2x2 cross classification table:

$V_i \backslash U_i$	1	0
1	a	b
0	c	d

The number of positive comovements in the two time series is a ; the number of negative comovements is d ; the number of comovements is $a+d$; the number of contramovements is $b+c$; the number of positive movements in X is $a+b$; the number of positive movements in Y is $a+c$.

The usual test of independence in a 2x2 cross classification table (see e.g., [17], pp. 65-72) corresponds to a test of whether a differs significantly from its estimated expected value, $A=(a+b)(a+c)/n$, under the null hypothesis of independence. It tests whether $a-A=\Delta$ differs significantly from zero using the fact that the estimated variance of Δ is $s^2 = [(a+b)(c+d)(a+c)(b+d)/n^3]$ under the null hypothesis. Moore and Wallis [14] have suggested that the usual 2x2 table test applied to the table given above could be used as a test of the existence of correlation between the movements in X and Y , which would eliminate at least the primary effects of trends in the series. They pointed out that this test is appropriate "for the case of randomly arranged [signs of the] first differences" in X and in Y . In the case where X and Y are purely random processes (i.e., 0-dependent stationary time series), they noted in [14] that the signs of the first differences in the series would have a negative serial correlation, and that if the 2x2 table test, which was justified by them when the signs of the first differences are randomly arranged, leads to acceptance of the hypothesis that correlation is absent, then it

surely would lead to acceptance of this hypothesis if the "null hypothesis of random observations were used; but nothing more can be said definitely until the sampling distribution [of Δ] appropriate to the assumption of random observations is known." The problem raised in [14] led to the derivation by Goodman and Grunfeld [9] of the sampling distribution of Δ for purely random processes of long duration and also for a certain kind of generalization of such time series. (They showed in [9] that the usual estimate, s^2 , of the variance of Δ required modification.) For these time series, a simple test of whether Δ differed significantly from zero was obtained.

The test proposed in [9] is appropriate when the X and Y are 0-dependent and the series $U = \{U_1, U_2, \dots, U_n\}$ and $V = \{V_1, V_2, \dots, V_n\}$ are stationary. For such time series, this test is a test of the existence of correlation between the movements in X and Y that takes into account (a) the time trends that may exist in X and in Y , and (b) the serial correlation between the signs of the first differences of 0-dependent time series. (The Moore-Wallis test takes into account (a) but not (b), and a related test by Stuart [18] takes into account (b) but not (a).) When X and Y are m -dependent ($m > 0$), the test in [9] does not fully take into account the serial correlation between the signs of the first differences. The test presented herein will take into account both (a) and the serial correlation between the signs of the first differences of m -dependent time series.

Tests of the existence of correlation between the movements in X and Y based on $\Delta = a - A$ compare the observed number, a , of positive comovements, given in the above 2×2 table, with its estimated expected value, A , when the null hypothesis of independence is true. Since A is based on $(a+b)/n$ and $(a+c)/n$, which are measures of an aspect of the "trend" in X and in Y , respectively, tests based on Δ test whether the observed number of positive comovements can be "explained" by the trends in X and in Y alone. If we

denote the estimated expected values of b, c, d , in the above 2×2 table, by $B=(a+b)(b+d)/n$, $C=(a+c)(c+d)/n$, $D=(c+d)(b+d)/n$, respectively, we see that $a-A=d-D=B-b=C-c = \frac{[(a+d)-(A+D)]}{2} = \frac{[(B+C)-(b+c)]}{2} = \left\{ \frac{[(a+d)-(b+c)]}{2} - \frac{[(a+b)-(c+d)] [(a+c)-(b+d)]}{n} \right\} / 4 = (ad-bc)/n$. Thus, a test of whether Δ differs significantly from zero is also a test of whether the number of comovements $(a+d)$, the number of contramovements $(b+c)$, the difference between the number of comovements and contramovements $([a+d] - [b+c])$, differ from their respective estimated expected values. It is a test of whether the observed values of $(a+d)$, $(b+c)$, $([a+d] - [b+c])$ can be "explained" by the trends in X and in Y alone.

The statistic Δ/n is the observed covariance between U_1 and V_1 . For certain types of data, of the kind referred to in [14], where the U_1 (and V_1) are more accurate than the W_1 (and Z_1), tests based on the observed covariance between U_1 and V_1 are to be preferred to tests based on the observed covariance between W_1 and Z_1 . The methods used here to obtain a test based on the observed covariance between U_1 and V_1 can be modified in a straightforward fashion in order to obtain a similar test based on the observed covariance between W_1 and Z_1 . The latter test will also be presented here. Tests based on the rank covariance (or the rank correlation) between W_1 and Z_1 could also be obtained, which test should be used in a particular situation will depend upon the type of data available, on the accuracy of the U_1 and V_1 as compared with the W_1 and Z_1 , and on the particular null and alternate hypotheses under consideration. For example, if the W_1 and Z_1 form 0-dependent stationary normal series, the usual test of the null hypothesis that the correlation between two normal variates is zero should be applied to the (W_1, Z_1) (see, e.g., [3]). In some situations, tests based on the rank correlation between the W_1 and Z_1 will recommend themselves (see, e.g., related remarks in [4]); in other kinds of situations, discussed earlier herein, tests based on the (U_1, V_1) will be more appropriate

Under the null hypothesis that the U_1 and V_1 (or the W_1 and Z_1) are independent, the covariance between U_1 and V_1 (or between W_1 and Z_1) will be zero. It will sometimes be of interest to test the null hypothesis that this covariance is some specified value other than zero. For certain kinds of m -dependent time series, methods of testing such hypotheses, of estimating the magnitude of this covariance (i.e., the covariance between U_1 and V_1 or that between W_1 and Z_1), and of obtaining confidence intervals for this covariance, will be presented in this article. These methods require more computation than the methods referred to earlier herein.

The U_1 and V_1 (or the W_1 and Z_1) are based on the differences $X_j - X_1$ and $Y_j - Y_1$ for $j=1+l$. In other words, movement is here defined in terms of the difference in the time series at successive time points. The results obtained in this case can be generalized to the case where movement is defined in terms of the difference in the time series at time points k units apart, where k is a fixed positive integer. These generalizations are presented in Section 3 herein.

The methods used to derive tests of the existence of correlation between the movements in m -dependent time series can also be used to derive tests of the existence of trend in such series. We present such a test in Section 4. Other tests for trend could also be readily derived.

Most of the tests presented here are based on a dichotomous classification of the movements in X (and in Y) indicating whether they were positive or negative. Some of the results presented herein can be modified in a straightforward fashion for situations where other fixed methods of classification, into two or more classes based on the magnitude of the movements, are used. The cross classification table describing the joint distribution of the classified movements in X and in Y can be analyzed in ways other than those discussed here. For example, the problem of measuring the extent of the correlation between the movements in X and Y can be considered, in part, a problem in the

measurement of association for this cross classification table. For a discussion of the latter problem, see [10], [11], and the literature cited there. The particular method of analysis that will be appropriate will depend on the purpose of the particular investigation at hand.

2. TESTS BASED ON $a-A$

We shall use the same notation as in the preceding section. We assume throughout that X and Y are m -dependent ($m \geq 0$). Thus, $U = \{U_1, U_2, \dots, U_n\}$ and $V = \{V_1, V_2, \dots, V_n\}$ are $(m+1)$ -dependent. For simplicity we assume also that U and V are stationary. (This assumption implies that trends in X and Y will be, in a certain sense, either constant or nonexistent; some of the results presented here will hold under more general conditions, but we shall not go into these details in the present paper (see [13]).) For such series, we shall now present a test of the hypothesis that U and V are independent.

Since the U and V series are $(m+1)$ -dependent, we know that the $\text{cov}\{U_1, U_{1+t}\} = \epsilon_t$ and $\text{cov}\{V_1, V_{1+t}\} = \phi_t$ will be zero for $t > m+1$. To estimate ϵ_t and ϕ_t for $t \leq m+1$, we first compute $g_t = \sum_{i=1}^{n-t} U_i U_{i+t}$ and $h_t = \sum_{i=1}^{n-t} V_i V_{i+t}$ for $t=1, 2, \dots, m+1$. The statistic g_t is the observed number of pairs of time points, t units apart, where there were positive movements in X (at both time points in the pair); h_t is the observed number of pairs of time points, t units apart, where there were positive movements in Y . These quantities can also be obtained from 2×2 tables of the following form describing the observed distribution of the $(n-t)$ pairs (U_i, U_{i+t}) and (V_i, V_{i+t}) for $i=1, 2, \dots, n-t$:

$U_i \backslash U_{i+t}$	1	0
1		
0		

$V_i \backslash V_{i+t}$	1	0
1		
0		

The entry in the upper-left cell of the first table will be g_t , and the entry in the corresponding cell of the second table will be h_t . (It is interesting to note that the usual 2×2 table test of independence applied to the pairs (U_1, U_{1+1}) and to the pairs (V_1, V_{1+1}) has been suggested as a test of the hypothesis of randomness in the U series and in the V series, respectively; i.e., the hypothesis that these series are 0-dependent (see [6], [7]).)

Having computed g_t and h_t , the following consistent estimators of ϵ_t and ϕ_t are obtained:

$$(2.1) \quad \hat{\epsilon}_t = g_t / (n-t) - [(a+b)/n]^2,$$

and

$$(2.2) \quad \hat{\phi}_t = h_t / (n-t) - [(a+c)/n]^2.$$

Writing

$$(2.3) \quad s^2 = (a+b)(c+d)(a+c)(b+d)/n^3,$$

and

$$(2.4) \quad \tilde{s}^2 = s^2 + 2n \sum_{t=1}^{m+1} \hat{\epsilon}_t \hat{\phi}_t,$$

we shall prove in the Appendix that, when U and V are independent, the asymptotic distribution ($n \rightarrow \infty$) of

$$(2.5) \quad (z-A)/\tilde{s}$$

will be normal with zero mean and unit variance. The statistic (2.5) can therefore be used to test whether U and V are independent.

The test proposed in [14] did not include the term $2n \sum_{t=1}^{m+1} \hat{\epsilon}_t \hat{\phi}_t$ in the variance computation; i.e., s^2 was used rather than $\tilde{s}^2$. The test proposed in [9] included the term $2n \sum_{t=1}^{m+1} \hat{\epsilon}_t \hat{\phi}_t$ but not $2n \sum_{t=2}^{m+1} \hat{\epsilon}_t \hat{\phi}_t$.

If time series X is m_1 -dependent and time series Y is m_2 -dependent, then $\epsilon_t = 0$ for $t > m_1 + 1$, $\phi_t = 0$ for $t > m_2 + 1$, $\epsilon_t \phi_t = 0$ for $t > \ell + 1$, where $\ell = \min[m_1, m_2]$. It follows from this that (2.4) can be replaced by

$$(2.6) \quad \tilde{s}^2 = s^2 + 2n \sum_{t=1}^{m+1} \hat{\epsilon}_t \hat{\phi}_t,$$

in the computation of (2.5). In particular, tests appropriate for the case where X and Y are 0-dependent will be appropriate also for the case where one of these m -dependent series is not 0-dependent and the other one is. It can also be seen that the test appropriate for the case where both U and V are 0-dependent (obtained by replacing $\tilde{s}^2$ by s^2) will be appropriate also for the case where one of these m -dependent series is not 0-dependent and the other one is.

If X is m_1 -dependent and Y is m_2 -dependent, then X and Y are m -dependent where $m = \max[m_1, m_2]$. In what follows we take $m = \max[m_1, m_2]$.

The test described above was based on the U_i and V_i . A similar test could be based on the W_i and Z_i . Assuming that $W = \{W_1, W_2, \dots, W_n\}$ and $Z = \{Z_1, Z_2, \dots, Z_n\}$ form stationary series, the covariances $\text{cov}\{W_i, W_{i+t}\} = \epsilon_t^*$ and $\text{cov}\{Z_i, Z_{i+t}\} = \phi_t^*$ can be estimated consistently by

$$(2.7) \quad \hat{\epsilon}_t^* = g_t^* / (n-t) \bar{W}^2,$$

$$(2.8) \quad \hat{\phi}_t^* = h_t^* / (n-t) \bar{Z}^2,$$

where $g_t^* = \sum_{i=1}^{n-t} W_i W_{i+t}$, $h_t^* = \sum_{i=1}^{n-t} Z_i Z_{i+t}$, $\bar{W} = \sum_{i=1}^n W_i / n$, $\bar{Z} = \sum_{i=1}^n Z_i / n$, for $t=1, 2, \dots, m+1$. Writing

$$(2.9) \quad s_W^2 = \sum_{i=1}^n (W_i - \bar{W})^2 / (n-1),$$

$$(2.10) \quad s_Z^2 = \sum_{i=1}^n (Z_i - \bar{Z})^2 / (n-1),$$

$$(2.11) \quad s^{*2} = s_W^2 s_Z^2,$$

$$(2.12) \quad \tilde{s}^{*2} = s^{*2} + 2n \sum_{t=1}^{m+1} \hat{\epsilon}_t^* \hat{\phi}_t^*,$$

it follows from the results in the Appendix that, when W and Z are

independent the asymptotic distribution ($n \rightarrow \infty$) of

$$(2.13) \quad \left(\sum_{i=1}^n W_i Z_i - n \bar{W} \bar{Z} \right) / S^*$$

will be normal with zero mean and unit variance. (To apply the Hoeffding-Robbins form of the central limit theorem [13], we assume here that the third absolute moments of W_i and Z_i are finite.) The statistic (2.13) can therefore be used to test whether W and Z are independent. As earlier herein, the statistic (2.12) can be replaced by

$$(2.14) \quad \hat{S}^{*2} = S^{*2} + 2n \sum_{t=1}^{m+1} \hat{\psi}_t^* \hat{\phi}_t^*$$

in the computation of (2.13),

The preceding asymptotic results were derived under the null hypothesis that W and Z (or U and V) were independent. We shall now consider the case where W and Z need not be independent. Let θ^* denote the covariance between W_i and Z_i . Under the null hypothesis considered in the preceding paragraph, $\theta^* = 0$. We shall assume that the $W_i Z_i$ form an $(m+1)$ -dependent stationary series; this will follow as a consequence, in the case where W and Z are independent, of the fact that the W and Z are $(m+1)$ -dependent stationary series. The statistic

$$(2.15) \quad \hat{\theta}^* = \sum_{i=1}^n W_i Z_i / n \bar{W} \bar{Z},$$

is a consistent estimator of θ^* . Writing

$$(2.16) \quad \hat{\psi}_t^* = \sum_{i=1}^{n-t} (W_i - \bar{W})(W_{i+t} - \bar{W})(Z_i - \bar{Z})(Z_{i+t} - \bar{Z}) / (n-t) \hat{\theta}^{*2},$$

for $t=0, 1, 2, \dots, m+1$, and

$$(2.17) \quad \hat{S}^{*2} = n \left[\hat{\psi}_0^{*2} + 2 \sum_{t=1}^{m+1} \hat{\psi}_t^* \hat{\phi}_t^* \right],$$

it follows from the results in the Appendix that the asymptotic distribution ($n \rightarrow \infty$) of

$$(2.18) \quad (\hat{\theta}^* - \theta^*)n/\hat{s}^*$$

will be normal with zero mean and unit variance. The statistic (2.18) can therefore be used to obtain approximate confidence intervals for θ^* and to test the hypothesis that θ^* is some specified value, say, θ_0^* . When the null hypothesis is that W_i and Z_i are independent, then $\theta_0^*=0$ and the value of $\hat{s}^*$ can be replaced by $\hat{s}_0^*$.

Let θ denote the covariance between U_i and V_i . A method analogous to that presented in the preceding paragraph could be used to estimate θ , obtain approximate confidence intervals for θ , and test hypothesis concerning θ . The details are given in the Appendix herein. When the null hypothesis is that $\theta=0$, the test presented at the beginning of this section will be simpler to apply than will the test corresponding to the one presented in the preceding paragraph.

3. TESTS BASED ON a_k, b_k

We now present a generalization of the tests described in the preceding section. It is also a generalization of the tests described by Goodman in [8].

Let $X_{i+k} = X_i = W_{ik}$ and $Y_{i+k} = Y_i = Z_{ik}$ for $i=1, 2, \dots, n+1-k$, where k is a fixed integer. We shall for convenience assume that the W_{ik} and Z_{ik} have continuous distributions. Let $U_{ik}=1$ if $W_{ik} > 0$, and $U_{ik}=0$ if $W_{ik} \leq 0$. Let $V_{ik}=1$ if $Z_{ik} > 0$, and $V_{ik}=0$ if $Z_{ik} \leq 0$. The observed distribution of the $n+1-k$ pairs (U_{ik}, V_{ik}) , for $i=1, 2, \dots, n+1-k$, can be summarized in the following 2×2 cross classification table:

$U_{ik} \backslash V_{ik}$	1	0
1	a_k	b_k
0	c_k	d_k

For $k=1$, $a_k=a$, $b_k=b$, $c_k=c$, $d_k=d$. For $k \geq 1$ the definitions of a_k, b_k, c_k, d_k are analogous to those of a, b, c, d , where now "movement" is defined by considering pairs of time points k units apart, rather than pairs of successive time points.

A test of the existence of correlation between movements in X and Y , which eliminates at least the primary effects of trends in the series, can be based upon a modification of the usual test of independence in a 2×2 table applied to the table given above. This would test whether a_k differs significantly from its estimated expected value, $A_k = (a_k + b_k)(a_k + c_k) / (n+1-k)$, under the null hypothesis of independence.

Let $U^{(k)} = \{U_{1k}, U_{2k}, \dots, U_{n+1-k,k}\}$ $V^{(k)} = \{V_{1k}, V_{2k}, \dots, V_{n+1-k,k}\}$. Assuming that X and Y are m -dependent, then $U^{(k)}$ and $V^{(k)}$ will be $(m+k)$ -dependent. For simplicity we also assume that $U^{(k)}$ and $V^{(k)}$ are stationary.

Let $g_{tk} = \sum_{i=1}^{n+1-k-t} U_{ik} U_{i+t,k}$ $h_{tk} = \sum_{i=1}^{n+1-k-t} V_{ik} V_{i+t,k}$. Writing

$$(3.1) \quad \hat{\epsilon}_{tk} = g_{tk} / (n+1-k-t) - [(a_k + b_k) / n]^2,$$

$$(3.2) \quad \hat{\phi}_{tk} = h_{tk} / (n+1-k-t) - [(a_k + c_k) / n]^2,$$

$$(3.3) \quad s_k^2 = (a_k + b_k)(c_k + d_k)(a_k + c_k)(b_k + d_k) / (n+1-k)^3,$$

$$(3.4) \quad \tilde{s}_k^2 = s_k^2 + 2(n+1-k) \sum_{t=1}^{m+k} \hat{\epsilon}_{tk} \hat{\phi}_{tk},$$

we shall prove in the Appendix that, when $U^{(k)}$ and $V^{(k)}$ are independent, the asymptotic distribution ($n \rightarrow \infty$) of

$$(3.5) \quad (a_k - A_k) / \tilde{s}_k$$

is normal with zero mean and unit variance. The statistic (3.5) can therefore be used to test whether $U^{(k)}$ and $V^{(k)}$ are independent.

The test proposed in [8] was also based on the $a_k - A_k$, but that test was derived under the assumption that $m=0$. In this special case, the serial covariance, ϵ_{tk} , between U_{ik} and $U_{i+t,k}$ is zero for $t < k$; the serial covariance, ϕ_{tk} , between V_{ik} and $V_{i+t,k}$ is zero for $t < k$. Thus, when $m=0$, the variance formula (3.4) can be replaced by

$$(3.6) \quad s_k^2 = s_k^2 + 2(n-k+1)\hat{\epsilon}_{kk}\hat{\phi}_{kk},$$

which corresponds to the variance formula given in [8].

The test in Section 2, based on the (U_1, V_1) , is a special case of the test given in the present section based on (U_{1k}, V_{1k}) . Similarly, a test analogous to the one based on (U_{1k}, V_{1k}) could be presented, based on the (W_{1k}, Z_{1k}) rather than on the (U_{1k}, V_{1k}) , which would be a direct generalization of the test in Section 2 based on the (W_1, Z_1) . Writing θ_k (or θ_k^*) for the covariance between U_{1k} and V_{1k} (or between W_{1k} and Z_{1k}), a test analogous to the test in Section 2 of whether θ (or θ^*) is equal to some specified value (not necessarily zero) could also be obtained. These tests are given in the Appendix herein.

The tests proposed in this and the preceding section were based, in part, on the observed distribution of the $n+1-k$ pairs (U_{ik}, V_{ik}) , $i=1, 2, \dots, n+1-k$, or on the pairs (W_{ik}, Z_{ik}) , when k is a fixed positive integer. Considering all possible values of k , there are a total of $n(n+1)/2$ pairs (U_{ik}, V_{ik}) . The observed distribution of these $n(n+1)/2$ pairs (U_{ik}, V_{ik}) can be summarized in a 2×2 table; the usual measure of association computed for this table will be equal to the partial rank correlation coefficient, $T_{XY.T}$, between X and Y with "time T held constant" (see [8], [9]). The situation considered here is, in a sense, a special case of the usual situation where partial rank correlation may be of interest; here the actual observations on the variable to be "held constant", T , are the integers $1, 2, \dots, n+1$ (or a linear transformation of them). The tests of the existence of correlation (based on the U_{ik}, V_{ik})

presented in the present paper are tests based on modified partial rank correlation coefficients where the $(n+1-k)$ pairs of (U_{1k}, V_{1k}) are used rather than the total $n(n+1)/2$ pairs. This follows from the fact that a test of whether $a_k - A_k$ differs significantly from zero is also a test of whether the usual measure of association for the corresponding 2×2 table differs significantly from zero (see [8]).

4. A TEST FOR TREND

For a 0-dependent time series, Cox and Stuart [2] have presented a very simple test for trend that has rather high asymptotic relative efficiency against the alternative hypothesis that the series consists of 0-dependent normal variates having a linear trend. For m -dependent time series ($m > 0$), this test requires modification. A modified test for trend suitable for m -dependent time series will be presented here. The method used to derive this test is similar to that used to derive the tests presented in the preceding sections. This method can be used to derive other tests for trend. Presentation herein of the modification of the Cox-Stuart test will make clear how other tests for 0-dependent series can be modified in order to obtain tests for trend in time series that are m -dependent.

Let $U_1^0 = 1$ if $X_{2n^0+1} - X_1 > 0$, and $U_1^0 = 0$ if $X_{2n^0+1} - X_1 \leq 0$, for $i=1, 2, \dots, n^0$, $n^0 = (n+1)/3$, where $(n+1)$ is divisible by 3. (We assume that $X_{2n^0+1} - X_1$ has a continuous distribution.) The test proposed in [2] is based on the fact that $U_0^0 = \sum_{i=1}^{n^0} U_i^0$ will have a binomial distribution with mean $n^0/2$ and variance $n^0/4$ when X is a 0-dependent stationary time series. A simple sign test of whether U_0^0 differs significantly from $n^0/2$ can serve as a test for trend in X (see [2]). When X is 0-dependent and stationary, the asymptotic distribution of

$$(4.1) \quad (U_0^0 - n^0/2) / \sqrt{n^0/4}$$

is normal with zero mean and unit variance. This fact also provides us with a simple test for trend for 0-dependent series.

If X is m -dependent, the time series $U^0 = \{U_1^0, U_2^0, \dots, U_{n^0}^0\}$ will also be m -dependent (for $n^0 \geq m$). If the X series is stationary, the U^0 series will also be stationary. Writing

$$(4.2) \quad f_t = \sum_{i=1}^{n^0-t} U_i^0 U_{i+t}^0$$

$$(4.3) \quad \sigma_U^2 = n^0 \left\{ 2 \sum_{t=1}^m f_t / (n^0-t) - (2m-1)/4 \right\}$$

we find, using a method of derivation similar to that used in the Appendix, that when X is an m -dependent stationary series the asymptotic distribution ($n^0 \rightarrow \infty$) of

$$(4.4) \quad (U_{n^0}^0 - n^0/2)/\sigma_U$$

is normal with zero mean and unit variance. The statistic (4.4) can therefore be used to test for trend in X . This test is a modification of the sign test in [2] where the variance $n^0/4$ is now replaced by (4.3). When n^0 is large,

the term $2 \sum_{t=1}^m f_t / (n^0-t)$ in (4.3) can be replaced by $2 \sum_{i=1}^{n^0-m} e_i / (n^0-m)$, where e_i is the number of values of j such that $U_j^0 = 1$ for $i < j \leq i+m$ when $U_i^0 = 1$, and $e_i = 0$ when $U_i^0 = 0$.

APPENDIX

We shall prove here ^a ~~some~~ ^{result} ~~what~~ more general ^{than} those presented in Sections 2 and 3 herein. The results presented in those sections will then be viewed in terms of this more general result. We shall use here the terminology of Section 3. The method of proof presented here can also be used to prove the result presented in Section 4.

We assume that X and Y are m -dependent. Thus, $U^{(k)}$ and $V^{(k)}$ are $(m+k)$ -dependent. We also assume that $U^{(k)}$ and $V^{(k)}$ are stationary. Let $E\{U_{ik}\} = p_k$, $E\{V_{ik}\} = r_k$, for $i=1,2,\dots,n+1-k$. Writing $G_{ik} = U_{ik} - p_k$, $H_{ik} = V_{ik} - r_k$, $F_{ik} = G_{ik} H_{ik}$, $Q_{ik} = U_{ik} V_{ik}$, we see that

$$(A.1) \quad \bar{Q} - \bar{U}\bar{V} = \bar{F} - \bar{G}\bar{H},$$

where $\bar{Q} = \frac{1}{n+1-k} \sum_{i=1}^{n+1-k} Q_{ik}$, and $\bar{U}, \bar{V}, \bar{F}, \bar{G}, \bar{H}$ are defined similarly. (Since k is fixed, we shall for the sake of brevity omit the subscript k associated with $\bar{Q}, \bar{U}, \bar{V}$, etc.) We shall assume that the Q_{ik} also form an $(m+k)$ -dependent stationary series; this will follow as a consequence, in the case where U_{ik} and V_{ik} are independent, of the fact that $U^{(k)}$ and $V^{(k)}$ are $(m+k)$ -dependent and stationary. Let $E\{F_{1k}\} = \theta$. Applying the Hoeffding-Robbins form of the central limit theorem [13], we find that the asymptotic distribution of $(\bar{F} - \theta)\sqrt{n+1-k}$ is normal with zero mean and variance

$$(A.2) \quad E\{G_{1k}^2 H_{1k}^2\} + 2 \sum_{t=1}^{m+k} E\{G_{1k} H_{1k} G_{1+t,k} H_{1+t,k}\} = [2(m+k) + 1] \sigma^2.$$

Similarly the asymptotic distribution of $\bar{G}\sqrt{n+1-k}$ is normal with zero mean and variance

$$(A.3) \quad E\{G_{1k}^2\} + 2 \sum_{t=1}^{m+k} E\{G_{1k} G_{1+t,k}\}$$

the asymptotic distribution of $\bar{H}\sqrt{n+1-k}$ is normal with zero mean and variance

$$(A.4) \quad E\{H_{1k}^2\} + 2 \sum_{t=1}^{m+k} E\{H_{1k} H_{1+t,k}\}.$$

Since $\bar{G}$ converges in probability to zero, $\bar{G}\bar{H}\sqrt{n+1-k}$ will also converge in probability to zero (see [3], p.254). Thus, the asymptotic distribution of $(\bar{F} - \bar{G}\bar{H} - \theta)\sqrt{n+1-k}$ is also the asymptotic distribution of $(\bar{F} - \theta)\sqrt{n+1-k}$. We have therefore shown that the asymptotic distribution of $(\bar{Q} - \bar{U}\bar{V} - \theta)\sqrt{n+1-k}$ is normal with mean zero and variance (A.2). Writing

$$(A.5) \quad \hat{\theta} = \bar{Q} - \bar{U}\bar{V},$$

$$(A.6) \quad \hat{\psi}_t = \frac{1}{n+1-k-t} \sum_{i=1}^{n+1-k-t} \frac{(U_{ik} - \bar{U})(V_{ik} - \bar{V})(U_{1+t,k} - \bar{U})(V_{1+t,k} - \bar{V})}{n+1-k-t} = \hat{\theta}^2,$$

we see that (A.2) can be estimated consistently by

$$(A.7) \quad \hat{\psi}_0 + 2 \sum_{t=1}^{m+k} \hat{\psi}_t,$$

which we denote by $\hat{s}^2/(n+1-k)$. Thus, the asymptotic distribution of

$$(A.8) \quad [\bar{Q} - \bar{U}\bar{V} - \theta] (n+1-k)/\hat{s}$$

will be normal with zero mean and unit variance. A test of the hypothesis that θ is equal to some specified value θ_0 can be based on the asymptotic distribution of (A.8).

The preceding results did not make use of the fact that U_{ik} and V_{ik} take on only the values 0 or 1. The results will therefore apply as well, if U_{ik} and V_{ik} are replaced by W_{ik} and Z_{ik} , respectively. (It will be necessary to assume that the third moments of W_{ik} and Z_{ik} are finite in order to apply the Hoeffding-Robbins theorem [13].) We thus have obtained a generalization of the result in Section 2 concerning the asymptotic distribution of (2.18). The result presented there was for the special case where $k=1$.

When $U^{(k)}$ and $V^{(k)}$ are independent, then $\theta=0$, and the variance (A.2) can be replaced by

$$(A.9) \quad E \{G_{1k}^2\} E \{H_{1k}^2\} + 2 \sum_{t=1}^{m+k} E \{G_{1k} G_{1+t,k}\} E \{H_{1k} H_{1+t,k}\}.$$

(It should be mentioned that formula (A.9) is similar to, but different from, the usual formula for the variance of the observed correlation between two independent linear autoregressive series (see, e.g., [1], [15], [16]).) Writing

$$(A.10) \quad \hat{\epsilon}_{tk}^+ = g_{tk}/(n+1-k-t)\bar{U}^2,$$

$$(A.11) \quad \hat{\phi}_{tk}^+ = h_{tk}/(n+1-k-t)\bar{V}^2,$$

$$(A.12) \quad s_U^2 = \sum_{i=1}^{n+1-k} (U_{ik} - \bar{U})^2 / (n+1-k),$$

$$(A.13) \quad s_V^2 = \frac{n+1-k}{\sum_{i=1}^{n+1-k}} (V_{ik} - \bar{V})^2 / (n+1-k),$$

we see that (A.9) can be estimated consistently by

$$(A.14) \quad s_U^2 s_V^2 + 2 \sum_{t=1}^{n+k} \hat{\epsilon}_{tk}^+ \hat{\phi}_{tk}^+,$$

which we denote by $s^2/(n+1-k)$. Thus, when $U^{(k)}$ and $V^{(k)}$ are independent, the asymptotic distribution of

$$(A.15) \quad (\bar{Q}_n - \bar{UV})(n+1-k)/s^2$$

will be asymptotically normal with zero mean and unit variance. This result will also apply as well, if U_{ik} and V_{ik} are replaced by W_{ik} and Z_{ik} respectively. We thus have obtained a generalization of the result in Section 2 concerning the asymptotic distribution of (2.13).

Using the fact that U_{ik} and V_{ik} were defined in Section 3 to take on only the values 0 or 1, we see that $\bar{Q}_n - \bar{UV} = (a_k - A_k)/(n+1-k)$, and that the asymptotic distribution of

$$(A.16) \quad (a_k - A_k) / \tilde{s}_k,$$

where $\tilde{s}_k$ is given by (3.4), is normal with zero mean and unit variance.

REFERENCES

- [1] Bartlett, M.S., An Introduction to Stochastic Processes, Cambridge: Cambridge University Press, 1960.
- [2] Cox, D.R., and Stuart, A., "Some quick sign tests for trend in location and dispersion", Biometrika, 42 (1955), 80-95.
- [3] Cramér, Harald, Mathematical Methods of Statistics, Princeton: Princeton University Press, 1946.
- [4] Daniels, H.E., "Rank correlation and population models", Journal of the Royal Statistical Society, Series B, 12 (1950), 171-181.

- [5] Friedman, Milton, "An empirical study of the relationship between railroad stock prices and railroad earnings for the period 1921-1931", Master's Thesis, Department of Economics, University of Chicago, 1933.
- [6] Goodman, Leo A., "Simplified runs tests and likelihood ratio tests for Markov chains", Biometrika, 45 (1958), 181-97.
- [7] Goodman, Leo A., "Exact probabilities and asymptotic relationships for some statistics from m-th order Markov chains", Annals of Mathematical Statistics, 29 (1958), 476-90.
- [8] Goodman, Leo A. "Partial tests for partial taus", Biometrika, 46 (1959), 525-32.
- [9] Goodman, Leo A., and Grunfeld, Yehuda, "Some nonparametric tests for comovements between time series", Journal of the American Statistical Association, 56 (1961), 11-26.
- [10] Goodman, Leo A., and Kruskal, William H., "Measures of association for cross classifications", Journal of the American Statistical Association, 49 (1954), 732-64.
- [11] Goodman, Leo A., and Kruskal, William H., "Measures of association for cross classifications. II: Further discussion and references", Journal of the American Statistical Association, 54 (1959), 124-63.
- [12] Grunfeld, Yehuda, "The determinants of corporate investment", Doctoral Thesis, Department of Economics, University of Chicago, 1958.
- [13] Hoeffding, Wassily, and Robbins, Herbert, "The central limit theorem for dependent random variables", Duke Mathematical Journal, 15 (1948), 773-80.
- [14] Moore, Geoffrey H., and Wallis, W. Allen, "Time series significance tests based on signs of differences," Journal of the American Statistical Association, 38 (1943), 153-64.
- [15] Moran, P.A.P., "Some theorems on time series. I", Biometrika, 34 (1947), 281-291.

- [16] Orcutt, G.H., and James, S.F., "Testing the significance of correlation between time studies", Biometrika, 35 (1948), 397-413.
- [17] Pearson, E.S., and Hartley, H.O., Biometrika Tables for Statisticians, Vol. 1, Cambridge: Cambridge University Press, 1954.
- [18] Stuart, Alan, "The power of two difference-sign tests", Journal of the American Statistical Association, 47 (1952), 416-24.
- [19] Stuart, Alan, "The asymptotic relative efficiencies of distribution-free tests of randomness against normal alternatives", Journal of the American Statistical Association, 47 (1952), 416-24.
- [20] Wold, Herman, A Study in the Analysis of Stationary Time Series, Stockholm: Almqvist and Wiksell, 1954.