Serial correlation in the analysis of time series

Richard L. Anderson
Iowa State College

Follow this and additional works at: https://lib.dr.iastate.edu/rtd
Part of the Mathematics Commons

Recommended Citation
https://lib.dr.iastate.edu/rtd/12880
NOTE TO USERS

This reproduction is the best copy available.

UMI®
SERIAL CORRELATION IN THE ANALYSIS
OF TIME SERIES

by

Richard L. Anderson

A Thesis Submitted to the Graduate Faculty
for the Degree of

DOCTOR OF PHILOSOPHY

Major Subject Mathematics

Approved:

Signature was redacted for privacy.

In charge of Major work

Signature was redacted for privacy.

Head of Major Department

Signature was redacted for privacy.

Dean of Graduate College

Iowa State College
1941
INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTRODUCTION</td>
<td>5</td>
</tr>
<tr>
<td>THEORETICAL BACKGROUND</td>
<td>11</td>
</tr>
<tr>
<td>Large Sample Distributions for Lag 1</td>
<td>12</td>
</tr>
<tr>
<td>Serial Covariance</td>
<td>12</td>
</tr>
<tr>
<td>Serial Correlation Coefficient</td>
<td>15</td>
</tr>
<tr>
<td>Small Sample Distributions for Lag 1</td>
<td>17</td>
</tr>
<tr>
<td>Serial Covariance</td>
<td>17</td>
</tr>
<tr>
<td>Characteristic function method</td>
<td>17</td>
</tr>
<tr>
<td>$X^2$ method</td>
<td>23</td>
</tr>
<tr>
<td>General density and probability functions for odd $N$</td>
<td>25</td>
</tr>
<tr>
<td>Serial correlation coefficient</td>
<td>27</td>
</tr>
<tr>
<td>$N = 2$</td>
<td>27</td>
</tr>
<tr>
<td>$N = 3$</td>
<td>29</td>
</tr>
<tr>
<td>$N = 4$</td>
<td>29</td>
</tr>
<tr>
<td>$N = 5$</td>
<td>30</td>
</tr>
<tr>
<td>$N = 6$</td>
<td>31</td>
</tr>
<tr>
<td>$N = 7$</td>
<td>32</td>
</tr>
<tr>
<td>Density function for any odd $N$</td>
<td>33</td>
</tr>
<tr>
<td>Density functions for even $N &gt; 6$</td>
<td>25</td>
</tr>
</tbody>
</table>
Distributions of Serial Correlation Coefficients for General Lag, L.......................... 36
p=2(N-2L)........................................... 41
p=3................................................. 41
p=4................................................. 42
p>4................................................. 43
Distributions of $R_L$ for Data Corrected for Trend.................................................. 46

COMPUTATION AND USE OF SIGNIFICANCE LEVELS FOR SMALL

SAMPLES WITH MEANS ZERO........................................... 50
Serial Covariance for Lag 1........................................... 50
Serial Correlation Coefficient, $R_L$........................................... 52
Lag 1................................................. 52
General lag, L........................................... 59
Examples.................................................. 62

DISCUSSION OF RESULTS............................................... 67

SUMMARY.................................................. 72

ACKNOWLEDGMENTS.................................................. 74

BIBLIOGRAPHY.................................................. 75
# LIST OF FIGURES AND TABLES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1</td>
<td>Density functions of $R_1$</td>
<td>28</td>
</tr>
<tr>
<td>Figure 2</td>
<td>Density functions of $R_2$</td>
<td>40</td>
</tr>
<tr>
<td>Figure 3</td>
<td>Significance levels of $R_1$</td>
<td>54</td>
</tr>
<tr>
<td>Figure 4</td>
<td>Significance levels of $R_L$ for $p = 4$</td>
<td>61</td>
</tr>
</tbody>
</table>

| Table 1 | Significance levels of $R_4$                | 53   |
| Table 2 | Calculations of 5% significance levels of $R_4$, N = 25 | 58   |
| Table 3 | Significance levels of $R_L$ for $p = 2$ and 3 | 60   |
| Table 4 | Serial R's for annual hide prices            | 64   |
INTRODUCTION

The analysis of most time series demands more rigorous tests of significance than those used in ordinary regression problems (10, 20). As Bartlett has said,

"It has long been recognized that correlating time series is a matter where caution has to be the rule, and a legitimate use of the ordinary statistical tests for testing significance the exception. In economic and sociological problems....., the statistical requirements for exact tests to be possible are rarely, if ever, completely satisfied. ..... The test of significance implies that the observations are normally distributed and successive observations independent. ..... The first assumption is known to be reasonably elastic. ..... The second assumption is the more important one. In most time series, even if trends have first been fitted, deviations are to some extent the product of preceding events, which may have influenced many preceding observations." (3, p.536)

Yule treats this problem of significance in a similar manner. He states that the use of the standard error test assumes: (a) drawing from the same aggregate throughout, (b) all draws equally likely to be drawn from any part of the aggregate, (c) each value drawn independent of all other values drawn, and (d) for simplification, the values are distributed in a certain way (not necessary in most cases). In time series, conditions (b) and (c) obviously are not fulfilled (24).

Yule has adopted the term lagged serial correlation to denote the relationship between successive values of some
variable, \( x \). If we have a set of \( n \) observations, \( x_1, x_2, \ldots, x_n \), we will define the serial correlation coefficient for the first lag to be

\[
R_1 = \frac{x_1 x_2 + x_2 x_3 + \ldots + x_{n-1} x_n + x_n x_1}{\sum x_i^2}.
\]

Similarly the serial \( R \) for the second lag is

\[
R_2 = \frac{x_1 x_3 + x_2 x_4 + \ldots + x_{n-1} x_2 + x_n x_3}{\sum x_i^2},
\]

and for a general lag \( L \),

\[
R_L = \frac{x_1 x_{L+n} + x_2 x_{L+n-1} + \ldots + x_{n-L} x_{L-1} + x_n x_L}{\sum x_i^2}.
\]

Although previous definitions of the serial \( R \) have omitted all terms in the numerator beyond \( x_{n-L} x_n \), we have introduced the cyclic arrangement of the cross-products in order to simplify the distributions of \( R_L \) which will be presented later.

Most of the previous discussions on serial correlation have been concerned with the correlation between the items in two time series. In an article criticizing the variate-difference method, Yule mentioned the problem of determining the true correlation between two series when both of these series consisted of serially correlated variates (23). In a subsequent article, he demonstrated that the correlations between samples drawn at random from two serially correlated series tended to form a definite U-shaped distribution, with a majority of the
correlations near ±1. These correlations could be expected, because time series tend to be always rising or falling in a small segment of the total range; there are no random up-and-down swings (24).

Bartlett concluded that we could use the usual tests of significance as a preliminary measure. If these tests show that the correlation between two series was non-significant, further study would not be required; however, if the correlation was significant, it would be necessary to test for serial correlation. He concluded that there was no adequate test at hand if the serial correlation was large (3). Most other writers have tended to evade the problem of serial correlation entirely.

Because the ordinary tests of the significance of regression and harmonic coefficients do not hold for serially correlated variates, there is a demand for some new method of analyzing data which might be serially correlated. The writer faced this situation in a recent study of the harmonic components of the cyclical deviations of agricultural prices (1). The ordinary tests of significance of the harmonic terms derived by Fisher (9) and Walker (5) could not be used directly. The most obvious difficulty was the determination of the proper number of degrees of freedom in the price series considered. A tentative solution suggested by E. B. Wilson was adopted at that time (31). Wilson proposed that a sequence of serial correlation coefficients, \( R_1, R_2, \ldots \),
be computed until the first non-significant $R_L$ was determined. Since no significance levels for the serial correlation coefficient were available, we extended the computations until we found the $R_L$ which was closest to zero. The best estimate of the true number of degrees of freedom in the price series was taken to be the number of items in the original series divided by $L$.

A test of significance based on such a determination of the number of degrees could not be exact, since one must assume that the same harmonic terms would be obtained from the serially correlated as from uncorrelated variates. Above all, it could not be assumed that the ratio of the largest harmonic term to the total sum of squares would be distributed as the ratio, $g$, which Fisher used in his test of the significance of harmonic coefficients (9).

W. G. Cochran has suggested that the next step in the study of serially correlated variates should be devoted to the problem of the exact distribution of the harmonic components of a series of these variates. As a preliminary step towards solving this difficult problem, we have extended our research to include the distributions of the serial correlation coefficients, $R_L$. These distributions can be used to determine significance levels of $R_L$; thus, we can determine at what lag the serial correlation is not significantly greater than zero. This manuscript will deal with this two-
fold problem of the distribution of the serial correlation coefficient and of the significance levels for various lags.

In connection with the determination of the appropriate number of degrees of freedom in any serially correlated series, it might be mentioned that the amount of information in any such series is not materially increased by taking a larger sample over a certain time interval; that is, we do not learn much more concerning the structure of an economic time series from weekly than from monthly data and probably little more from monthly than from quarterly or even yearly data.

The utility of the distribution of the serial correlation coefficient is not limited to the problem of determining the number of independent items in a serially correlated series. The tests of significance of the serial correlation should also be useful in the determination of the proper method of studying a time series. H. Wold (22) has prepared an excellent treatise on this topic, but he includes no test of significance. He has shown that a graph of the various serial correlation coefficients, which he called the correlogram, can be used to determine whether the method of hidden periodicities, linear autoregression or moving averages should be used to analyze a time series. If the correlogram consists of a more or less regular sequence of harmonic terms with periods corresponding to the periods of the time series, and the amplitudes are not damped in successive swings, the method of hidden periodicities
or harmonic analysis should be used; if the amplitudes are
damped towards the horizontal axis as the serial correlations
for higher lags are computed, one of the other two methods
should be considered. The method of moving averages is better
if all the serial correlations are zero after a certain point,
while the linear auto-regression method is used when the
correlations merely damp slowly towards zero. Since we feel
that the harmonic analysis method is too restrictive for most
time series, one of the other more adaptable methods is pre-
ferred; therefore, we want to determine if the serial correla-
tions are all zero after a certain point. Our tables of the
significance levels of the serial R's should facilitate the
analysis of time series along the lines suggested by Wold.

Of course the utility of these tables is not limited to
the treatment of economic time series. Other data, especially
meteorological data, are presumed to be serially correlated.
In most biological and agricultural experiments, there is
probably no need for a test of significance of the serial
correlation, but certain high correlations even here might be
explained by a relationship between variates within classes.
It is evidently true that any ordered series should be studied
for serial correlation before any definite statements regard-
ing its relation to other ordered series are made.
THEORETICAL BACKGROUND

The serial correlation coefficient $R_l$ equals the serial covariance $C_l$ divided by the variance $V$. The first section will deal with the correlation for lag 1; the results then will be extended to include the general lag, $L$. We define our serial covariance and variance for lag 1 and $N$ observations, $X_1, X_2, \ldots, X_N$, as follows:

$$C_l = X_1^2 + X_2^2 + \ldots + X_{N-1}^2 + X_N^2 + X_N X_1$$

$$V = \frac{N}{N-1} X_1^2.$$  

Note that this definition of the covariance differs slightly from the ordinary serial covariance, because the $X_N X_1$ term is ordinarily omitted; also, we have used the term variance to denote the sum of squares. In addition, it is assumed that the values of $X$ are taken as deviates from their mean; thus, the $X$'s are assumed to have an expectation of 0.

If the population variance were known a priori, the variates could be transformed so that they would have a variance of 1; under such an unusual circumstance, the only distribution required would be that of the serial covariance. Since the population variance is seldom, if ever, known, the main emphasis in this paper will be placed on the use of the serial correlation coefficient, although the distribution of the covariance will be included as an introduction to the subject at hand.
Large Sample Distributions for Lag 1

Since much of the time series data consists of a large number of items, especially if weekly or monthly data is studied, it is advisable to consider the problem of our serial covariance and correlation on the large sample as well as the small sample basis.

Serial covariance.

We will consider the covariance \( C_x \) as defined on page 11, assuming that \( E(C_x) = 0 \) and that the \( X \)'s are normally and independently distributed with mean 0 and variance 1. The problem is to determine how large \( C_x \) must be to refute these assumptions at a given level of significance.* The density function of \( C_x \) is

\[
D(C_x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \phi(\theta) e^{-i\theta C_x} d\theta,
\]

where \( \phi(\theta) \) is the characteristic function for \( C_x \),

\[
\phi(\theta) = K \lim_{n \to \infty} \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2} \left\{ \sum_{i=1}^{n} x_i^2 - 2\theta \cdot (x_1 x_i + \cdots + x_n x_i) \right\}} dx_1 \cdots dx_n \ (n > 2).
\]

The solution of this integral for \( \phi(\theta) \) is usually denoted by \( \Delta^\frac{1}{2} \), where \( \Delta \) is the determinant of the

* This problem was suggested by Professor Harold Hotelling, whose methods have been followed in the succeeding pages.
coefficients of the \( X \)'s in the exponent of \( e \) in \( \phi (e) \). This \( \Delta \) is

\[
\Delta = \begin{vmatrix}
1 & -ie & 0 & \cdots & 0 & -ie \\
-ie & 1 & -ie & \cdots & 0 & 0 \\
0 & -ie & 1 & \cdots & 0 & 0 \\
& \ddots & \ddots & \ddots & \ddots & \ddots \\
& & \ddots & \ddots & \ddots & \ddots \\
& & & \ddots & \ddots & \ddots \\
0 & 0 & 0 & \cdots & -ie & 0 \\
0 & 0 & 0 & \cdots & 1 & -ie \\
-ie & 0 & 0 & \cdots & -ie & 1 \\
\end{vmatrix}
\]

This determinant can be evaluated by the method of circulants. Let

\[
P = \begin{vmatrix}
1 & \omega_1 & \omega_1^2 & \cdots & \omega_1^{N-1} \\
1 & \omega_2 & \omega_2^2 & \cdots & \omega_2^{N-1} \\
& \ddots & \ddots & \ddots & \ddots \\
& & \ddots & \ddots & \ddots \\
& & & \ddots & \ddots & \ddots \\
1 & \omega_N & \omega_N^2 & \cdots & \omega_N^{N-1} \\
\end{vmatrix}
\quad \text{and} \quad \Delta = \begin{vmatrix}
a_1 & a_2 & \cdots & a_N \\
a_N & a_1 & \cdots & a_{N-1} \\
& \ddots & \ddots & \ddots \\
& & \ddots & \ddots & \ddots \\
& & & a_3 & a_4 & \cdots & a_1 \\
\end{vmatrix}
\]

with \( \omega_k \) denoting the \( k^{\text{th}} \) root of unity. If we set \( A_k = \sum_{i=1}^{N} a_i \omega_k^{i-1} \),

\[
\Delta \cdot P = \begin{vmatrix}
A_2 & (\omega_2 A_2) & \cdots & (\omega_2^{N-1} A_2) \\
A_3 & (\omega_3 A_2) & \cdots & (\omega_3^{N-1} A_3) \\
& \ddots & \ddots & \ddots \\
& & \ddots & \ddots & \ddots \\
& & & A_N & (\omega_N A_N) & \cdots & (\omega_N^{N-1} A_N) \\
\end{vmatrix} = \prod_{k=1}^{N} A_k \cdot P.
\]

Therefore, \( \Delta = \prod_{k=1}^{N} A_k \).
In the problem for the distribution of $C_z$, $a_z = 1$, 
$a_z = -i\epsilon$, $a_n = -i\epsilon$ and all other $a^i$s = 0. Thus,
$$
\Delta^{\frac{1}{2}} = \prod_{k=1}^{N} \left\{1-i\epsilon \left(\omega_k + \omega_k^{-1}\right)\right\}^{\frac{1}{2}} = \prod_{k=1}^{N} \left\{1-i\epsilon \sigma_k\right\}^{\frac{1}{2}},
$$
where $\sigma_k = \omega_k + \omega_k^{-1} = 2 \cos 2\pi k/N$.

If we let $K = \log \phi$, then
$$
K = \frac{1}{2} \left\{i\epsilon \sum_{k=1}^{N} \sigma_k + \frac{1}{2} (i\epsilon)^2 \sum_{k=1}^{N} \sigma_k^2 + \ldots \right\}
$$
$$
= \frac{N}{2} \left\{(i\epsilon)^2 + \frac{k_3}{3} (i\epsilon)^3 + \frac{k_4}{4} (i\epsilon)^4 + \ldots \right\},
$$

If we let $C_z = \sqrt{N} C_z'$, then for $N > 3$
$$
K' = K \left(\epsilon/\sqrt{N}\right) = -\frac{1}{2} \epsilon^2 + O(1/N).
$$

For $N \to \sigma$, $K' = -\frac{1}{2} \epsilon^2$; thus, $C_z'$ is normally distributed with mean 0 and variance 1 for large $N$. Therefore, the positive 1% and 5% significance levels for $C_z'$ are 2.327 and 1.645 respectively. If it is desired to consider both the positive and negative tails of the distribution, the levels will be 2.575 and 1.96. To find the significance levels for $C_z$, multiply these values by $\sqrt{N}$.

$$
N k_p = \sum_{k=1}^{N} \sigma_k^p = \sum_{k=1}^{N} \left(\omega_k + \omega_k^{-1}\right)^p
$$
$$
= \sum_{k=1}^{N} (\omega_k^p + p \omega_k^{p-2} + \ldots + p \omega_k^{q-2} + \omega_k^{N-p}).
$$

$$
\sum_{k=1}^{N} \omega_k^q = \begin{cases} 
N \text{ for } q \text{ a multiple of } N \\
0 \text{ for } q \text{ not a multiple of } N
\end{cases}
$$

Thus, $k_1 = 0$, $k_2 = 2$, $k_3 = 0 \text{ (for } N > 3) \text{ and } 2 \text{ (for } N = 3)$,
$$
k_4 = \begin{cases} 
6 \text{ for } N \neq 4 \\
8 \text{ for } N = 4
\end{cases} \text{ and } k_5 = \begin{cases} 
0 \text{ for } N > 5 \\
2 \text{ for } N = 5 \\
10 \text{ for } N = 3
\end{cases}.$$
**Serial correlation coefficient**

For the more important distribution of the serial correlation coefficient $R_z$, we must consider the simultaneous distribution of $C_z$ and $V$. The density function for $C_z$ and $V$ is

$$D(C_z, V) = (2\pi)^{-\frac{3}{2}} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \phi(s, t) e^{-iC_z - iV} \, ds \, dt.$$ 

The characteristic function $\phi(s, t)$ is

$$\phi(s, t) = K \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} e^{-i \left\{ \sum X_i^2 - z \ln \sum X_i^2 - z \ln \left( \prod_{i=1}^{n} (X_i - X_i') \right) \right\}} \, dX_1 \cdots dX_n.$$ 

$$= (1-2it)^{-\frac{N}{2}} \prod_{k=1}^{N} \left( 1 - \frac{is}{1-2it} \right)^{-\frac{1}{2}}.$$ 

As before, we set $\lambda = \log \phi(s, t)$. Then

$$\lambda = -N/2 \log(1-2it) - \frac{1}{2} \sum_{k=1}^{N} \log(1 - \frac{is}{1-2it} \cdot c_k)$$

$$- N \left\{ \left( \frac{1}{4} + \frac{1}{6} \right) \sum \frac{ik^2}{1-2it} + \cdots \right\}$$

$$+ \left\{ \left( \frac{1}{2} \sum \frac{ik^3}{1-2it} \right) + \left( \frac{1}{3} \sum \frac{ik^4}{1-2it} \right) + \cdots \right\},$$

where the $k$'s are the same as defined above for the covariance.

If we make the substitutions $C_z' = C_z / \sqrt{N}$ and $V' = \sqrt{N} / \sqrt{2} \cdot (1 - \frac{V}{N})$, the new semi-invariant $\lambda'$ has the following form:

$$\lambda'(s, t) = \lambda \left( \frac{s}{\sqrt{N}}, \frac{-t}{\sqrt{2N}} \right) + it \sqrt{\frac{N}{2}} = - \frac{1}{2} (t^2 + s^2) + e,$$
where $\theta$ is of order $N^{-\frac{1}{2}}$. We infer that $G_1^i$ and $V'$ are normally and independently distributed with means 0 and variance 1 for large $N$.

If we replace $G_1^i$ and $V'$ by their equivalent values in terms of $G_1$ and $V$, the asymptotic density function for $G_1$ and $V$ would be

$$D(G_1, V) = \left(2\pi N \sqrt{2}\right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \left[\frac{G_1^2}{N} + \frac{V^2}{2N} - \frac{1}{2} N\right]\right).$$

The density function for $R_1$ and $V$ is found by substituting $G_1 = R_1 V$ in $D(G_1, V)$. The resultant function is

$$\left(2\pi N \sqrt{2}\right)^{-\frac{1}{2}} V \exp\left(-\left[\frac{V^2(1+2R_1^2)}{4N} - \frac{V}{2} + \frac{N}{4}\right]\right).$$

Now let $x = \left[V - \frac{N}{1+2R_1^2}\right] \sqrt{\frac{1+2R_1^2}{4N}}$ and integrate with respect to $x$ from $-\frac{1}{2} \sqrt{\frac{N}{1+2R_1^2}}$ to $\infty$. The resultant density function for $R_1$ is

$$D(R_1) = \sqrt{\frac{N}{2\pi}} (1+2R_1^2)^{-\frac{3}{2}} e^{-\frac{NR_1^2}{1+2R_1^2}} + \frac{1}{2N} e^{-\frac{N}{4}} \left[\frac{2}{N} - \frac{1}{3} \cdot \frac{4(1+2R_1^2)}{N} + \ldots\right].$$

For large $N$, the second part can be neglected. If we let $y^2 = \frac{NR_1^2}{1+2R_1^2}$, $D(y) = (2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2} y^2}$. Of course, the positive 1% and 5% significance values for $y$ are 2.327 and 1.645 respectively. The corresponding significance levels for $R_1$ can be found by substituting these values for $y$ in the equation $R_1 = y(N-2y^2)^{-\frac{1}{2}}$. 
Small Sample Distributions for Lag 1

Serial covariance

Characteristic function method. A detailed account of the many methods available for the solution of the small sample distributions would necessitate a volume in itself, but we feel that the method of characteristic functions should be included, although it was not the one finally used to derive the distribution of the correlation coefficient. As noted on page 14, the characteristic function for the covariance is

$$\phi(\theta) = \prod_{k=1}^{n} (1 - i\theta c_k)^{-\frac{1}{2}},$$

where $c_k = 2 \cos 2\pi k/N$. The density function of $C_z$ is

$$D(C_z) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \phi(\theta) e^{-i\theta C_z} d\theta.$$

Certain simplifications of $\phi(\theta)$ have been used. Since $\cos(-\theta) = \cos \theta$, we can eliminate all of the radical terms except those involving $\cos 2\pi$ for odd $N$ and $\cos 2\pi$ and $\cos \pi$ for even $N$. When $N$ also is divisible by 4, it should be noted that the factors involving $\cos \pi/2$ or $\cos 3\pi/2$ can be omitted since they equal 1. Our density functions take these forms:

\[
\frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-i\theta C_z}}{(1-\iota c_z)(1-\iota c_z)^2 \cdots (1-\iota c_{z n-1})} \sqrt{1-2\iota \theta} \, d\theta \quad \text{for odd } N
\]
\[
\frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{e^{-i\theta c_i}}{(1-ic_i) \ldots (1-ic_i x^2) / (1+4\theta^2)} \, d\theta \quad \text{for } N \text{ even but not divisible by 4.}
\]

\[
\frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{e^{-i\theta c_i}}{(1-ic_i) \ldots (1-ic_i x^2 / 4) / (1+4\theta^2)} \, d\theta \quad \text{for } N \text{ divisible by 4.}
\]

Since the integrands are fundamentally different for odd and even \(N\), different procedures are required in evaluating each type. These integrands involve contour integration with poles at \(\theta = -1/c_i\), \(-1/c_i\), etc., plus a branch point at \(\theta = -\frac{1}{2}i\) for odd \(N\) and two branch points at \(\theta = -\frac{1}{2}i\) and \(+\frac{1}{2}i\) for even \(N\). In order to evaluate these integrals, we have utilized the formulas presented in Campbell's Fourier Integrals for Practical Applications (5). Two basic integrals needed in our work are:

\[
(1) \int_{-\infty}^{+\infty} F_1(f) e^{2\pi ifg} \, df = G_1(g)
\]

\[
(2) \int_{-\infty}^{+\infty} F_1 F_2 \, e^{2\pi ifg} \, df = \int_{-\infty}^{+\infty} G_1(x) G_2(g-x) \, dx.
\]

The denominator in our density functions is the \(F(f)\) mentioned above. The particular \(F\)'s which we need and the corresponding solutions follow:
in which \( p = 2\pi i \alpha \) and \( \alpha, \beta, \) and \( \rho \) are complex numbers whose real parts are \( > 0 \). We have set \( p = -i\alpha \) and \( g = C_1 \) for \( C_1 > 0 \) and \( p = i\alpha \) and \( g = -C_1 \) for \( C_1 < 0 \).

Let us consider an example of the methods of evaluation for \( N = 5 \). The density function for \( C_1 > 0 \) is

\[
D(C_1) = 2^{-\frac{3}{2}} \int_{-\infty}^{\infty} \frac{e^{\pi C_1}}{(a_1+p)(a_2-p)\sqrt{\frac{1}{\alpha} + p}} \, df,
\]

where \( a_k = \left| \left( 2 \cos 2\pi k/5 \right)^{-1} \right| \). Using formulas 3 and 5, we have

\[
G_1(x) = \left( \pi x \right)^{-\frac{1}{2}} e^{-\frac{1}{2} x} \quad \text{for } x > 0
\]

\[
G_2(C-x) = \begin{cases} (a_1+a_2)^{-1} e^{a_2(C-x)} & \text{for } x < C_1 \\ (a_1+a_2)^{-1} e^{-a_2(C_1-x)} & \text{for } x > C_1 \end{cases}
\]

The density function of \( C_2 \) for \( C_1 > 0 \) is found by substituting these \( G \)'s in formula 2, so that

\[
D(C_2) = \frac{1}{2\pi} \left( \frac{1}{(a_1+a_2)} \right) \int_{-\infty}^{C_1} \left( e^{-a_1 C_1} \int_{0}^{C_1} x^{-\frac{1}{2}} e^{x(a_1^{-\frac{1}{2}})} \, dx + \int_{C_1}^{\infty} x^{-\frac{1}{2}} e^{-x(a_2+C_1)} \, dx \right) \, dx.
\]
The probability that \( C_1 > C_0 \) for \( C_1 > 0 \) is

\[
P(C_1 > C_0) = \frac{1}{\sqrt{2\pi} (a_1 + a_2)} \left\{ \int_{C_1 > C_0} x^{-\frac{1}{2}} e^{+x (a_1 - \frac{1}{2} C_1)} dx \int_{C_1 > C_0} e^{a_1 C_1} dx \right. \\
+ \left. \int_{C_1 > C_0} x^{-\frac{1}{2}} e^{-x (a_2 + \frac{1}{2} C_1)} dx \int_{C_1 > C_0} e^{a_2 C_1} dx \right\} dc_1.
\]

These integrals can be simplified by interchanging the order of integration, as follows:

\[
P(C_1 > C_0) = \frac{1}{\sqrt{2\pi} (a_1 + a_2)} \left\{ \int_{X = C_0}^{\infty} x^{-\frac{1}{2}} e^{+X (a_1 - \frac{1}{2} X)} dx \int_{C_1 > X} e^{a_1 C_1} dc_1 \\
+ \int_{X = 0}^{\infty} x^{-\frac{1}{2}} e^{+X (a_1 - \frac{1}{2} X)} dx \int_{C_1 > X} e^{a_2 C_1} df_1 \\
+ \int_{X = C_0}^{\infty} x^{-\frac{1}{2}} e^{-X (a_2 + \frac{1}{2} X)} dx \int_{C_1 > X} e^{a_2 C_1} dc_1 \right\}
\]

\[
= \frac{1}{\sqrt{2\pi} (a_1 + a_2)} \left\{ \frac{1}{a_2} \left\{ \int_{x = C_1}^{\infty} x^{-\frac{1}{2}} e^{-\frac{1}{2} x} dx + \int_{x = 0}^{C_1} x^{-\frac{1}{2}} e^{x (a_2 - \frac{1}{2} C_1)} dx \right. \\
+ \frac{1}{a_2} \left\{ \int_{x = C_0}^{\infty} x^{\frac{1}{2}} (e^{-\frac{1}{2} x} - e^{a_2 (x - a_1 x + \frac{1}{2} C_0)}) dx \right\}\right\}.
\]

The first and third integrals in the last equation are merely incomplete gamma functions, whose values have been tabulated (12). The second integral is a more difficult one to evaluate. We have expanded the integrand in a binomial series, as follows:
\[
I' = \int_{x=0}^{c_0} x^{-\frac{1}{2}} e^{-\frac{1}{2} x} u_1(C_0-x) \, dx = \frac{2e^{-\frac{1}{2} c_0}}{\sqrt{C_0} (2a_1-1)} \int_{y=0}^{\frac{c_0}{2a_1-1}} \left[ 1 - \frac{2y}{C_0 (2a_1-1)} \right]^{-\frac{1}{2}} e^{-y} \, dy
\]

after substituting \( y = \frac{1}{2} (C_0-x)(2a_1-1) \). If we let \( k = \frac{1}{C_0 (2a_1-1)} \) and expand the integrand in a binomial series, we have

\[
I' = 2C_0^{\frac{1}{2}} k e^{-\frac{1}{2} C_0} \int_{y=0}^{\frac{1}{2}} \left[ 1 + ky + \frac{1}{2} k^2 y^2 + \frac{1}{3} k^3 y^3 + \ldots \right] e^{-y} \, dy.
\]

We have here a series of incomplete gamma functions. A detailed account of the methods of computation will be presented in the section on computations of significance levels.

For \( C_x < 0 \), the density function of \( C_x \) is much simpler, since we need consider only the case for \( x < -C_x \) with \( C_x \), and \( x \) is integrated from 0 to \( \infty \). Thus the density function is

\[
D(C_x) = \frac{e^{a_x C_x}}{(a_x + a_x) \sqrt{2a_x + 1}}.
\]

The probability that \( C_x < C_0 \) is given by the equation

\[
P(C_x < C_0) = \frac{e^{a_x C_0}}{a_x (a_x + a_x) \sqrt{2a_x + 1}}.
\]

When \( N \) is greater than 5, the denominator in the integrand of \( D(C_x) \) contains more than two factors of the type \( (a_1 + p) \). In order to use Campbell's formulas, it is necessary to separate the integrand into a series of partial fractions of the type of formula 5 on page 19. As an example, the fractions for \( N = 9 \) would be
\[
\frac{1}{(p+a_1)(p+a_2)(p-a_3)(p-a_4)} = k \left[ \frac{1}{p+a_1} - \frac{1}{p+a_2} \right] \left[ \frac{1}{p-a_3} - \frac{1}{p-a_4} \right],
\]
where \(k = 1/(a_2-a_1)(a_3-a_4)\). Of course the rest of the problem is merely a repetition of the methods used for \(N = 5\). If we let
\[
P(a_1) = e^{-a_1C_1} \int_{x=0}^{e^{-x(a_1^{-\frac{1}{2}})}} \frac{1}{\sqrt{2\pi(a_2-a_3)(a_3-a_4)}} \left\{ \frac{1}{a_2+a_3} \left[ F(a_1) + G(a_3) \right] - \frac{1}{a_2+a_3} \left[ F(a_1) + G(a_4) \right] \right. \\
\left. - \frac{1}{a_3+a_4} \left[ F(a_3) + G(a_3) \right] + \frac{1}{a_3+a_4} \left[ F(a_3) + G(a_4) \right] \right\} = \\
\left(2\pi\right)^{-\frac{1}{2}} \left[ F(a_1)/k_2 + F(a_3)/k_3 + G(a_3)/k_3 + G(a_4)/k_4 \right],
\]
with \(k_2 = (a_2-a_1)(a_3+a_2)(a_4+a_2)\), \(k_3 = (a_3-a_1)(a_3+a_3)(a_4+a_3)\), \(k_3 = (a_3+a_3)(a_4+a_3)(a_4-a_3)\) and \(k_4 = (a_3+a_4)(a_3+a_4)(a_3-a_4)\).

The problem for even \(N\) is very complicated for \(N > 4\). The distribution for \(N = 4\) is not so difficult. In this case, \(D(C_1) = (2\pi)^{-\frac{1}{2}} K_0\left( \frac{1}{2} |C_1| \right)\), and the probability that \(C_1 > C_0\) can be reduced to
\[
P(C_1 > C_0) = \frac{1}{n} \int_{C_0}^{\infty} K_0(x) \, dx.
\]

The values of this integral have been tabulated (4).

For \(N > 4\), the probability function involves the definite integral of such functions as \(e^x K_0(x)\), whose values have not
been tabulated. Since we have not seen any real need for the significance levels of $C_1$, these integrals have not been evaluated. No general density function has been set up for even $N$.

χ² method. W. G. Cochran has suggested that we utilize a result given in his article on quadratic forms (7). If $x_1, x_2 \ldots, x_n$ are normally and independently distributed with variances 1 and means 0, then

"Every quadratic form $\sum a_{ij} x_i x_j$ is distributed like $\sum \lambda_i u_1$, where $r$ is the rank of the matrix, $A$, of the quadratic form, the $u_i$'s are independent and each is distributed as $\chi^2$ with 1 d.f., and the $\lambda_i$'s are the non-zero latent roots of the characteristic equation of $A$" (p. 179).

If each $\lambda_i$ appears $k_i$ times as a latent root, $u_1$ will be distributed as $\chi^2$ with $k_1$ degrees of freedom.

We wish to consider the distribution of $C_1 = X_1X_2 + X_3X_5 + X_2X_4 + \ldots + X_nX_1$. Referring to the solution of the characteristic function by the circulant determinant on page 13, we note that the characteristic or $\lambda$ determinant of $C_1$ has the same general solution with $a_1 = -\lambda$, $a_2 = \frac{1}{2}$, $a_3 = \frac{1}{2}$, and all other $a$'s = 0. Thus the characteristic equation for lag 1 and $N$ observations is

$$F_{1, N}(\lambda) = \prod_{i=1}^{N} (-\lambda + \cos \frac{2\pi i}{N}) = 0.$$ 

The solutions of this equation are $\lambda_1 = \cos \frac{2\pi i}{N}$. Again, we
have a different set of solutions for odd and even \( N \). For odd \( N \) all of the \( \lambda_i \)s are repeated except for \( \lambda_N \), while \( \lambda_N \) and \( \lambda_{n/2} \) are not repeated for even \( N \). Thus \( C_2 \) is distributed as follows:

\[
C_2 = \sum_{i=1}^{\frac{N}{2}} \lambda_i u_i + u \\
C_2 = \sum_{i=1}^{\frac{N}{2}} \lambda_i u_i + u_N - u_{n/2}
\]

for odd \( N \)

for even \( N \).

Note that for \( N \) divisible by 4, \( \lambda_{\frac{N}{4}} \) and \( \lambda_{\frac{3N}{4}} \) are 0.

As with the characteristic function method, let us consider an example for \( N = 5 \). Here \( C_2 = \lambda_1 u_1 + \lambda_2 u_2 + u_3 \), where \( \lambda_1 = 0.30902 \) and \( \lambda_2 = -0.80902 \), \( u_1 \) and \( u_2 \) are distributed as \( \chi^2 \) with 2 d.f. each and \( u \) is distributed as \( \chi^2 \) with 1 d.f.

The density function of \( u_1 \) and \( u \) is

\[
\frac{1}{2 \sqrt{2\pi}} u_{1/2} e^{-\frac{1}{2}(u+u_2)}
\]

while that for \( u_2 \) is \( \frac{1}{2} e^{-\frac{1}{2}u} \). If we let \( u_2 = (x-u)/\lambda_2 \) so that \( C_2 = x + \lambda_2 u_2 \), the density function of \( x \) and \( u \) would be

\[
\frac{1}{2 \sqrt{2\pi}} u_{1/2} e^{-\frac{1}{2} \frac{1-\lambda_2}{\lambda_2}} u e^{-x/\lambda_2},
\]

Since \( u_2 \) and \( u_3 \) are distributed as \( \chi^2 \), they can not be negative; therefore, \( u \) varies between 0 and \( x \). From the equation \( C_2 = x + \lambda_2 u_2 \), it is apparent that \( u_2 = -(x-C_2)/\lambda_2 \) and that \( x \) varies between \( C_2 \) and \( \infty \) for \( C_2 > 0 \) and between 0 and \( C_2 \) for \( C_2 < 0 \). For \( C_2 > 0 \), the density function of \( C_2 \) is
\[- \frac{e^{-C_1/2 \lambda_2}}{4 \lambda_1 \lambda_3 \sqrt{2\pi}} \int_{x = \lambda_1}^{\infty} e^{-x \left( \frac{1}{2 \lambda_1} - \frac{1}{2 \lambda_2} \right)} \, dx \int_{u = 0}^{\infty} u^{-\frac{1}{2}} e^{\frac{1}{2 \lambda_1} - \frac{1}{2 \lambda_3}} \, du.\]

By interchanging the order of integration, we have exactly the same distribution as obtained by the characteristic function method (page 19), since \(|2\lambda_1| = 1/\alpha_1\). Similar results hold for \(C_1 < 0\).

**General density and probability functions for odd \(N\).** If we extend the method of partial fractions illustrated for \(N = 9\) on pages 21 and 22, we can derive a general density function for odd \(N\). The density function for positive \(C_1\) in terms of the \(\lambda_s\) instead of the \(\alpha_i\)'s would be

\[
\frac{1}{2} \frac{\lambda}{\sqrt{2\pi}} \sum_{i=m}^{n-1} \frac{-S}{\beta_i} e^{-C_1/2 \lambda_i} G(\lambda_i) - \frac{1}{2} \frac{\lambda}{\sqrt{2\pi}} \sum_{i=1}^{n-1} \frac{-S}{\beta_i} e^{-C_1/2 \lambda_i} F(\lambda_i),
\]

where \(\lambda_m\) is the first negative \(\lambda_1\), \(\beta_1 = \lim_{j \to 1} (\lambda_j - \lambda_1)\) for \(j \neq 1\), \(F\) is the integral of

\[
u^{-\frac{1}{2}} e^{\frac{1}{2 \lambda_i} - \frac{1}{\lambda_i}}
\]

between the limits 0 and \(C_1\) and \(G\) is the same integral between the limits \(C_1\) and \(\infty\).

Similarly for \(C_1 < 0\), the density function of \(C_1\) is

\[
\frac{1}{2} \sum_{i=m}^{n-1} \frac{-S}{\beta_i} e^{-C_1/2 \lambda_i} \sqrt{\frac{\lambda_i}{\lambda_1 - \lambda_i}}.
\]
Corresponding to the given density function of $C_1$, the probability that $C_1 > C_0$ when $C_1 > 0$ is (for $n = \frac{N-1}{2}$)

$$
\sum_{i=1}^{\infty} (-1)^i \beta_i^t \frac{\lambda_1^i}{\lambda_1^i - 1} e^{-\frac{C_0}{\lambda_1^i - 1}} \left[ 1 - I\left(\mu_1, \frac{\lambda_1^i}{\sqrt{2}}, -\frac{i}{2}\right) \right]
\quad + \left[ 1 - I\left(C_0, \frac{\lambda_1^i}{\sqrt{2}}, -\frac{i}{2}\right) \right] + \left(2\pi\right)^{-\frac{1}{2}} \sum_{i=1}^{n-1} (-1)^{n-i} \beta_i^i I_i^i \left(\lambda_1^i \right),
$$

where $\beta_i^i = \frac{\lambda_1^{n-i}}{\beta_1^i}$ and $\mu_1 = \frac{C_0}{\sqrt{2} \lambda_1^i}$. The notation adopted for the incomplete gamma function is the same as that given by Pearson (12).

$$
\frac{1}{\Gamma(p+1)} \int_{x=L}^{\infty} x^p e^{-x} \, dx = \left[ 1 - I(L, p) \right].
$$

Pearson has published his tables in terms of $I(\mu, p)$, where $\mu = L/\sqrt{p+1}$. As indicated on page 21, we have set

$$
I_i^i = \int_{x=0}^{C_0} x^{-\frac{i}{2}} e^{-\frac{x}{2}} x^{-\frac{1}{2} \lambda_1^i} (C_0-x)^{\lambda_1^i-1} \, dx,
$$

which can be set up as the sum of terms of the form $I(\mu, p)$:

$$
I_i^i = 2 \sqrt{C_0} \sum_{k=0}^{C_0} k \frac{\lambda_1^i}{2} \sum_{k=0}^{C_0} (k^{i})^p z_p c_p I(\mu_1^{(p)}, p),
$$

where $k_1 = \frac{\lambda_1^i}{(1-\lambda_1^i) C_0}$ and $\mu_1^{(p)} = \frac{1}{2k_1 \sqrt{1+p}}$. 
Serial correlation coefficient.

The $X^2$ method has been used exclusively to derive the density functions for the serial correlation coefficient $R_1$. The general procedure to be followed can be summarized as follows: We determine the joint density function of the $u$'s which form the distribution of $C_1 (=R_1 V)$ and $V$. The $u$'s are integrated out, leaving the joint density function of $R_1$ and $V$. The distribution of $R_1$ is then obtained by integrating with respect to $V$ from 0 to $\infty$. We have adopted the same notation here as was used for the covariance. The graphs of the density functions for $N = 2$ to $N = 7$ inclusive plus those for $N = 9$ and 11 have been drawn (Figure 1). These distributions approach normality as $N$ increases; for $N = 11$, the density function is nearly normal except at the tails of the distribution.

The probability that $R_1 > R_0$ is found by integrating $D(R_1)$ from $R_0$ to 1. For simplicity of notation, we have denoted by $P(R_0)$ the indefinite integral of $D(R_1)$ evaluated at $R_1 = R_0$. The probability that $R_1 > R_0$ can be found by computing $1 - P(R_0)$ for odd $N$ and $1/2 - P(R_0)$ for even $N$. Although the density functions of $R_1$ for very small $N$ have no practical use, they have been derived as illustrations of the methods used.

$N = 2$. For $N = 2$, I have let $C_1 = R_1 V = 2X_1X_2$ in order to be consistent with the definition of $C_1$ for other $N$. 
Therefore, \( R_2 V = u_2 - u_2 \) and \( V = u_4 + u_2 \), where each \( u \) is distributed as \( \chi^2 \) with 1 d.f. The density function for \( u_2 \) and \( u_2 \) is

\[
D(u_2, u_2) = (2\pi)^{-1} u_2^{-\frac{1}{2}} u_2^{-\frac{1}{2}} e^{-\frac{1}{2} V}.
\]

After substituting \( u_2 = V(1+R_2)/2 \) and \( u_2 = V(1-R_2)/2 \) and integrating with respect to \( V \) from 0 to \( \infty \), we have

\[
D(R_2) = (1-R_2^2)^{-\frac{1}{2}}/\pi \text{ and } P(R_0) = \sin^{-1} R_0/\pi.
\]

**N=3.** \( C_3 = R_3 V = -u_3/2 + u \) and \( V = u_3 + u_3 \), where \( u_3 \) and \( u \) are distributed as \( \chi^2 \) with 2 and 1 d.f. respectively. The density function for the \( u \)’s is

\[
D(u_2, u) = (2\sqrt{2\pi})^{-1} u^{-\frac{1}{2}} e^{-\frac{1}{2} V}.
\]

After substituting \( u = V(2R_3 + 1)/3 \) and \( u_3 = 2V(1-R_3)/3 \) in this equation and integrating out \( V \), we have

\[
D(R_3) = (1+2R_3)^{-\frac{1}{2}} / \sqrt{3} \text{ and } P(R_0) = (1+2R_0)^{\frac{1}{2}} / \sqrt{3}.
\]

Since the \( u \)’s must be positive, \( R_3 \) can vary only from \(-\frac{1}{2}\) to 1.

**N=4.** \( R_4 V = u_4 - u_3 \) and \( V = u_3 + u_3 + u_4 \), where \( u_3 \) and \( u_4 \) are distributed as \( \chi^2 \) with 1 d.f. and \( u_2 \) with 2 d.f. Therefore, \( u_3 = [V(1-R_4) - u_3]/2 \) and \( u_4 = [V(1+R_4) - u_3]/2 \). The density function of the \( u \)’s is

\[
D(u_2, u_3, u_4) = (4\pi)^{-1} u_2^{-\frac{1}{2}} u_3^{-\frac{1}{2}} u_4^{-\frac{1}{2}} e^{-\frac{1}{2} V}.
\]
For $R_2 \geq 0$, $u_3$ varies from 0 to $V(1-R_2)$, so that

$$D(R_2, V) = (4\pi)^{-1} V \int_{u_3=0}^{\sqrt{(1-R_2)}} e^{-\frac{1}{2} V \left[(V-u_3)^2-V^2R_2\right]} \, du_3.$$ 

After substituting $V-u_3 = x$ and integrating with respect to $x$ from $R_2 V$ to $V$ and with respect to $V$ from 0 to $\infty$, we have

$$D(R_2) = (\pi)^{-1} \log \left[\frac{1+\sqrt{1-R_2^2}}{R_2}\right] \quad \text{and}$$

$$P(R_0) = (\pi)^{-1} \left\{ R_0 \log \left[\frac{1+\sqrt{1-R_2^2}}{R_0}\right] + \sin^{-1} R_0 \right\}.$$ 

For $R_2 < 0$, replace $R_2$ by $-R_2$ in the above functions.

$N = 5$, $R_2 V = \lambda_1 u_2 + \lambda_3 u_3 + u$ and $V = u_1 + u_2 + u$. Therefore,

$$u_2 = \frac{V(R_2-\lambda_2)-u(1-\lambda_2)}{(\lambda_2-\lambda_3)} \quad \text{and} \quad u_3 = \frac{u(1-\lambda_3)-V(R_2-\lambda_3)}{(\lambda_2-\lambda_3)}.$$ 

The density function for the $u$'s is

$$D(u_2, u_3, u) = (4 \sqrt{2\pi})^{-1} u^{-\frac{1}{2}} e^{-\frac{1}{2} V \cdot u}.$$ 

The Jacobian of the transformation to the $R_2$, $V$, $u$ system is $V/(\lambda_2-\lambda_3)$. Therefore, the new density function would be

$$D(R_2, V, u) = \frac{u^{-\frac{1}{2}} V \cdot e^{-\frac{1}{2} V \cdot u}}{4 \sqrt{2\pi} \cdot (\lambda_2-\lambda_3)}.$$ 

For $R_2 \geq \lambda_2$, $u$ varies between $\frac{V(R_2-\lambda_2)}{1-\lambda_2}$ and $\frac{V(R_2-\lambda_2)}{1-\lambda_2}$, so that the density function of $R_2$ in this region is
\[
\frac{3}{2(\lambda_2 - \lambda_1)} \left[ \sqrt{\frac{R_2 - \lambda_1}{1 - \lambda_2}} - \sqrt{\frac{R_1 - \lambda_1}{1 - \lambda_2}} \right].
\]

For this same region,

\[
P(R_0) = \frac{1}{\lambda_2 - \lambda_1} \left[ (R_0 - \lambda_2) \sqrt{\frac{R_0 - \lambda_1}{1 - \lambda_2}} - (R_0 - \lambda_1) \sqrt{\frac{R_0 - \lambda_2}{1 - \lambda_2}} \right].
\]

For \( \lambda_2 \leq R_1 \leq \lambda_1 \), \( u \) varies between 0 and \( V(R_1 - \lambda_1) \); therefore the factor involving \( R_1 - \lambda_1 \) is not included in the density function of \( R_1 \) for this region.

\[N = 6, R_1 V = \frac{1}{2} u_1 - \frac{1}{2} u_2 - u_3 + u_6 \text{ and } V = u_4 + u_5 + u_3 + u_6,\]

where \( u_1 \) and \( u_4 \) are distributed as \( \chi^2 \) with 2 d.f. and \( u_3 \) and \( u_6 \) with 1 d.f.

\[u_1 = \frac{1}{2} \left[ V(2R_1 + 1) + u_3 - 3u_6 \right] \text{ and } u_4 = \frac{1}{2} \left[ V(1 - 2R_1) + u_6 - 3u_5 \right].\]

The density function for \( u_3, u_6, V \) and \( R_1 \) is

\[(8u)^{-1} V^{\frac{1}{2}} u_3^{-\frac{1}{2}} u_6^{-\frac{1}{2}}.\]

To determine the limits of integration, we have used a graph of \( u_6 \) in terms of \( u_3 \). From this graph, it was found that for \( R_1 \geq \frac{1}{2} \), \( u_6 \) varies between the two lines \([3u_3 - V(1 - 2R)]\) and \([V(1 + 2R_1) + u_3]/3 \); and \( u_3 \) varies from 0 to \( V(1 - R_1)/2 \). For \( R_1 \leq -\frac{1}{2} \), replace \( R_1 \) by \(-R_1\) in these limits. Finally when \( R_1 \) falls in the middle region, \( |R_1| \leq \frac{1}{2} \), the problem is complicated by the necessity of computing two integrals.

This problem has been simplified by drawing the line connecting
the origin with the intersection of the two boundary lines
for \( R_x > \frac{1}{2} \) and then adding the integrals taken over the two
regions created by this line. The equation of this line is
\( u_3 = u_0(1-R_x)/(1+R_x) \).

The density function of \( R_x \) is
\[
\frac{4}{\pi^{1/2}} \left[ (2R_x+1) \log \frac{1-R_x + \sqrt{3(1+R_x)}}{\sqrt{2(2R_x+1)}} + (1-2R_x) \log \frac{1+R_x + \sqrt{3(1-R_x)}}{\sqrt{2(2R_x-1)}} \right]
\]
for \( R_x \geq \frac{1}{2} \). The same distribution holds for \( |R_x| \leq \frac{1}{2} \) if
\((2R_x-1)\) in the second log is replaced by \((1-2R_x)\) and for
\( R_x \leq -\frac{1}{2} \) if \( R_x \) simply is replaced by \(-R_x\).

The probability function, \( P(R_o) \) is similar to \( D(R_o) \),
extcept that the factors \((2R_o+1)\) and \((1-2R_o)\) are squared, the
entire function is divided by 4 and \( \sin^{-1} R_o / \pi \) is added to \( D(R_o) \).

\( \text{N = 7.} \quad R_x V = \lambda_4 u_1 + \lambda_5 u_4 + \lambda_7 u_3 + u_2 \) and \( V = u_4 + u_3 + u_2 \). 
\[
\frac{V(R_x-\lambda_3)-u_4(\lambda_3-\lambda_4)-u(1-\lambda_3)}{\lambda_3-\lambda_4}
\]
and
\[
\frac{u_3 = u(1-\lambda_3)+u_4(\lambda_3-\lambda_4)-V(R_x-\lambda_3)}{\lambda_3-\lambda_4}
\]
From these two equations, it is evident that \( u \) varies from
\[
L = \frac{V(R_x-\lambda_3)-u_4(\lambda_3-\lambda_4)}{1-\lambda_4} \quad \text{to} \quad U = \frac{V(R_x-\lambda_3)-u_4(\lambda_3-\lambda_4)}{1-\lambda_3}
\]
These two lines intersect at \( u = V(R_x-\lambda_3)/(1-\lambda_3) \) and
\( u_3 = V(1-R_x)/(1-\lambda_3) \). When \( R_x \geq \lambda_3 \), \( u \) is integrated between
the limits \( L \) and \( U \) and \( u_3 \) from 0 to this point of intersection.
For \( \lambda_a \leq R_a \leq \lambda_z \), \( u \) is negative at the intersection point; since the \( u \)'s must be positive, it is necessary to subtract the integral taken over that part of the region for which \( u \) is negative. For \( R_a \leq \lambda_a \), the intersection of \( L \) on both the \( u \) and \( u_z \) axes is negative; therefore, \( u \) must be integrated from 0 to \( U \) and \( u_z \) from 0 to the intersection of \( U \) on the \( u_z \) axis, \( V(R_a-\lambda_z)/(\lambda_z-\lambda_2) \).

The density function of \( R_a \) for \( R_a \geq \lambda_a \) is

\[
\frac{5}{2} \left[ \frac{(R_a-\lambda_a)^{2+}}{(\lambda_a-\lambda_2)(\lambda_a-\lambda_z)\sqrt{1-\lambda_z}} + \frac{(R_a-\lambda_a)^{2+}}{(\lambda_2-\lambda_z)(\lambda_a-\lambda_z)\sqrt{1-\lambda_z}} \right].
\]

For \( \lambda_a \leq R_a \leq \lambda_z \), drop the term in \( (R_a-\lambda_a) \) and for \( \lambda_z \leq R_a \leq \lambda_a \), also drop the term in \( (R_a-\lambda_z) \).

**Density function for any odd \( N \).** If we change our notation to let \( R_n \) and \( V_n \) be \( R_a \) and \( V \) for \( N \) observations, then it would appear that the density function for \( R_n \) and \( V_n \) is

\[
D(R_n, V_n) = K V_n^N \exp\left(-\frac{1}{2} V_n \sum_{i=1}^{N-2} (R_i-\lambda_1)^2 \right) \text{ for } R_n \geq \lambda_1,
\]

where \( \alpha_1 = \sum_{j=1}^{N-2} (\lambda_j-\lambda_1) \sqrt{1-\lambda_1} \) \((i \neq j)\) and \( 1/K = 2^{N+1} \Gamma \left(\frac{N+1}{2}\right) \).

This distribution holds for \( N = 3, 5, \) and 7; we will show that it holds for \( N+2 \), assuming it true for \( N \). We see that

\[
R_{n+2} V_{n+2} = R_n V_n + \lambda_k u_k \text{ and } V_{n+2} = V_n + u_k,
\]

where \( k = \frac{(N+1)}{2} \). From these equations we obtain
\[ R_N = \frac{(R_{N+2} V_{N+2} - \lambda_k u_k)}{(V_{N+2} - u_k)} \quad \text{and} \quad V_N = V_{N+2} - u_k. \]

The density function for \( u_k, V_{N+2} \), and \( R_{N+2} \) is

\[
\frac{K V_{N+2}}{2} e^{-\frac{1}{2} V_{N+2}} \int_{0}^{N-2} \left[ V_{N+2} \frac{R_{N+2} - \lambda_1}{\lambda_k - \lambda_1} \right]^{\frac{n-1}{2}} \frac{\lambda_k - \lambda_1}{\lambda_k - \lambda_1}. \]

Since \( R_n \leq 1, u_k \leq V_{N+2} (1-R_{N+2})/(1-\lambda_k) \). For the region \( R_{N+2} \geq \lambda_1 \), the lower limit of \( u_k \) is 0. In this case, the density function for \( R_{N+2} \) and \( V_{N+2} \) would be

\[
\frac{K V_{N+2}^{\frac{n-1}{2}}}{2} e^{-\frac{1}{2} V_{N+2}} \int_{0}^{n-2} \left[ \frac{R_{N+2} - \lambda_1}{\lambda_k - \lambda_1} \right]^{\frac{n-1}{2}} \theta, \]

where \( \theta = \sum_{i=1}^{n-2} \frac{n-2}{\lambda_k - \lambda_1}. \) This density function is the same as that found by substituting \( N+2 \) for \( N \) in the density function of \( R_n \) and \( V_n \), if the coefficient of \( (R_{N+2} - \lambda_k) \) equals the reciprocal of \( \prod_{j=1}^{k-1} (\lambda_j - \lambda_k) \sqrt{1-\lambda_k} \). The proof of this equality is suggested by Lagrange's interpolation formula (1519). Let \( f(x) \) be a polynomial of degree \( (k-1) \), which for \( x=\lambda_i \) has the value \( f(\lambda_i) \), \( i=1, 2, \ldots, k \). Then Lagrange's formula is

\[
\frac{f(\lambda_1)}{(\lambda_1-x)} + \frac{f(\lambda_2)}{(x-\lambda_2)} + \frac{f(\lambda_3)}{(x-\lambda_2)(\lambda_1-\lambda_2)} + \frac{f(\lambda_4)}{(x-\lambda_3)(\lambda_1-\lambda_2)(\lambda_1-\lambda_3)} + \ldots + \frac{f(\lambda_k)}{(x-\lambda_k)(\lambda_1-\lambda_2) \cdots (\lambda_1-\lambda_k)} = 0. \]
where \( \| \) indicates that the zero factor is omitted. If we let 
\[
f(x) = (1-x)^{k-1},
\]
then \( f(1) = 0 \), and
\[
\frac{(1-\lambda_k)^{k-\frac{3}{2}}}{\prod_{j=1}^{k-1} (\lambda_j-\lambda_k) \sqrt{1-\lambda_k}} = -\sum_{l=1}^{k-1} \frac{(1-\lambda_l)^{k-\frac{3}{2}}}{\prod_{j=1}^{l-1} (\lambda_j-\lambda_l) \sqrt{1-\lambda_l}}.
\]

If we divide both sides of this equation by \((1-\lambda_k)^{k-\frac{3}{2}}\), we have exactly the desired equality, since \( k = (N+1)/2 \).

For \( \lambda_a \leq R_{N+2} \leq \lambda_a \), we must consider the cases when \( R_N \geq \lambda_a \) and also when \( R_N \leq \lambda_a \). For the first case, \( u_k \) varies from \( V_{N+2} (\lambda_a - R_{N+2})/(\lambda_a - \lambda_k) \) to \( V_{N+2} (1-R_{N+2})/(1-\lambda_k) \). In the second case, the limits are 0 and \( V_{N+2} (\lambda_a - R_{N+2})/(\lambda_a - \lambda_k) \).

Similar limits must be set up for the other regions. It can be shown that all terms for which \( \lambda_i \geq \lambda_m \) must be omitted for any region in which \( R_{N+2} \leq \lambda_m \); in other words, we omit from the density function any term which is imaginary.

Likewise the general form of the probability function is
\[
P(R_o) = \frac{N-1}{a_1} \left( R_o - \lambda_1 \right)^{N-2} /a_1 \text{ for } \lambda_m \leq R_o \leq \lambda_{m-1},
\]

where \( a_1 = \prod_{j=1}^{N-1} (\lambda_j-\lambda_1) \sqrt{1-\lambda_1} \). As mentioned in the introduction to this section, \( P(R_n > R_o) = 1 - P(R_o) \).

**Density Functions for even** \( N > 6 \). We have not attempted to determine any distributions of \( R_a \) for \( N > 6 \), because we have been able to determine the significance levels by interpolating between the levels for the adjacent odd \( N \). We checked the accuracy of this interpolation for \( N = 6 \).
Distributions of Serial Correlation Coefficients
for General Lag, L

Although the distribution of the serial correlation coefficient for lag 1 is fundamental, we must be able to find the significance levels for other lags in order to determine which lag gives us a non-significant serial correlation. This problem is a much more difficult one and has not been solved completely, although the procedure is fairly definite. Time does not permit the solution of all the distributions required to publish a complete table of significance levels for all N and all lags, but the procedure for computing such levels will be indicated.

For a general lag L, the covariance $C_L$ is given by

$$C_L = X_1X_{L+1} + X_2X_{L+2} + \cdots + X_NX_L.$$  

The constants in the characteristic equation, referred to on page 13, are $a_1 = -\lambda$, $a_{L+1} = \frac{1}{2}$, $a_{N-L+1} = \frac{1}{2}$ and all other $a$'s = 0. Hence, the characteristic equation is

$$F_{L,N} = \prod_{i=1}^{N} \left[ -\lambda + \frac{1}{2} (\omega_i^L + \omega_i^{-L}) \right] = \prod_{i=1}^{N} \left[ -\lambda + \cos \frac{2\pi i L}{N} \right] = 0.$$

The solutions of this equation are $\lambda_1 = \cos \frac{2\pi i L}{N}$.

Certain important generalizations concerning $F_{L,N}$ may be set down:
(1) When \( L \) is not a factor of \( N \) or has no common factor with \( N \), \( F_{L,N} = F_{1,N} \).

(2) When \( L \) and \( N \) have a common factor, \( \alpha \),

\[
F_{L,N} = \left[ \frac{F_{1,N}}{\alpha} \right]^a.
\]

(2a) If \( L \) is a factor of \( N \), \( F_{L,N} = \left[ \frac{F_{1,N}}{L} \right]^L \).

The proof of the first statement was suggested by Professor Cochran. We first note that \( \cos \frac{2\pi NL}{N} = \cos \frac{2\pi N}{N} = 1 \).

Since \( \cos (2\pi k/N + 2\pi m) = \cos 2\pi k/N \), we must prove in addition that the series of numbers

\[ L, 2L, 3L, \ldots, (N-1)L, \]

when reduced to modulus \( N \), can be arranged to form the series

\[ 1, 2, 3, \ldots, (N-1), \]

where \( L \) and \( N \) are relatively prime to each other. In other words, we must show that all terms of the reduced series are different, since it is obvious that there is one term which will reduce to 1 and that the maximum value for any reduced term is \( (N-1) \). Suppose it is possible, for some \( r \) and \( s (r > s) \), that

\[ rL/N = a + k/N; \quad sL/N = b + k/N, \]

where \( a \) and \( b \) are integers and \( 1 \leq r, s \leq (N-1) \). Then

\[ (r-s)L/N = (a-b) \text{ or } L/N = (a-b)/(r-s). \]

This last equation can be true only if \( a = b \), so that \( L = 0 \) (which is nonsensical), or \( L/N \) has a common factor, since
(a-b) is an integer and \((r-s) < N\). Hence, if \(L/N\) is a prime fraction, each term of the sequence \(\{\cos \frac{2\pi Li}{N}\}\) reduces uniquely to one of the sequence \(\{\cos \frac{2\pi i}{N}\}\), for \(i = 1, \ldots, N\).

If \(L\) and \(N\) have a common factor \(\alpha\), \(L = q\alpha\) and \(N = p\alpha\), where \(p\) and \(q\) are integers prime to each other. Hence,

\[
F_{L,N} = \prod_{t=1}^{p} \left\{-\lambda + \cos \frac{2\pi q i}{p}\right\} = (F_{q,p})^\alpha = (F_{1,p})^\alpha.
\]

If \(\alpha = L\), \(F_{L,N} = (F_{1,p})^L\), where \(p = N/L\).

When these results are applied to the asymptotic large sample distribution, we find that the latter is independent of \(L\). If \(N\) and \(L\) have a common factor \(\alpha\),

\[
K_L = -\alpha/2 \log \Delta_{1,p} \text{ instead of } K_1 = -\frac{1}{2} \log \Delta_{1,n}.
\]

Since the asymptotic distribution is linear in \(N\), the \(\alpha\)'s cancel; hence, we need consider only the distributions given for lag \(L\). Of course it should be noted that even though \(N\) is large enough to permit one to use the normal approximation for lag \(L\), \(p\) may be too small to make the approximation accurate for a lag greater than one.

The results which we have obtained for the general lag \(L\) seem to indicate that one should study the small sample distributions of \(R_L\) in terms of \(N/L\), which we shall indicate by \(p\), instead of considering each \(N\) and the various lags for
that \( N \). In the following distributions of \( R_L \), we shall call lag \( L \) the primary lag, as contrasted to any other lag \( a \), for which the distribution is the same as that for lag \( L \). That is, for \( N=6 \), \( R_a \) and \( R_L \) have the same density function, but we will denote the distribution of \( R_a \) as the primary distribution. All of these distributions have been derived by the \( \chi^2 \) method. The principal change from the methods presented for lag 1 is that the \( u \)'s are now distributed as \( \chi^2 \) with \( 2L \) and \( L \) degrees of freedom instead of 2 and 1 d.f. This difference will become apparent for the various distributions which follow; in each case, the distribution of \( R_a \) is simply the one presented in the previous sections.

Only the general distribution formulas will be set up here. Some graphs of the density functions for \( L=2 \) are presented in Figure 2. The general density functions for \( p=2, 3 \) and 4 have been derived. In addition, the distributions for \( L=2 \) have been set up for \( p = 5, 6, 7 \) and 9. For \( N>9 \), we must consider regions bounded by figures of more than two dimensions; we have not been able to simplify the complicated integrals which must be solved in order to determine these density functions for \( p > 9 \). Any suggestions on methods of simplifying these integrations will be appreciated.

In order to condense the general formulas which follow, the numerical constant times the factors involving \( V \) will be denoted by \( E_L(V) \). In those cases for which the complete dis-
tribution is derived, the value of \( K \) will not be stated specifically.

\[ p=2(N-2L) \]

\[ \mathbf{R}_L \mathbf{V} = \mathbf{u}_1 - \mathbf{u}_2 \] and \( \mathbf{V} = \mathbf{u}_1 + \mathbf{u}_2 \), where both \( \mathbf{u}'s \) are distributed as \( \chi^2 \) with \( L \) d.f. Hence,

\[ \mathbf{D}_L(u_1, u_2) = K_L(V) \left( u_1 u_2 \right)^{1/2L-1} \]

After substituting \( u_1 = V(1+R_L) \) and \( u_2 = V(1-R_L) \) and integrating with respect to \( V \) from 0 to \( \infty \), we have

\[ \mathbf{D}(R_L) = \frac{(1-R_L)^{1/2L-1}}{2^{1/2} \beta(L/2, 1/2)} \]

If we set \( 1-R_L = 2y \), then

\[ \mathbf{P}(R_L > R_0) = \frac{1}{\beta(1/2, L/2)} \int_{y=0}^{1/2(1-R_0)} y^{1/2L-1} (1-y)^{1/2L-1} \, dy. \]

Pearson has tabulated the values of these incomplete Beta functions (13).

\[ p=3 \]

\[ \mathbf{R}_L \mathbf{V} = -\mathbf{i} \mathbf{u}_1 \mathbf{u} \] and \( \mathbf{V} = \mathbf{u}_1 + \mathbf{u}_2 \), where \( \mathbf{u}_2 \) is distributed as \( \chi^2 \) with \( 2L \) d.f. and \( \mathbf{u} \) as \( \chi^2 \) with \( L \) d.f. Therefore,

\[ \mathbf{D}_L(u_2 u) = K_L(V) u_2^{L-1} u^{1/2L-1} \]

Proceeding as before, \( \mathbf{D}(R_L) \) for \( R_L < -\frac{1}{2} \) is

\[ \frac{2^L}{3^{1/2(3L-2)}} \frac{1}{\beta(L/2, L/2)} \frac{(1-R_L)^{L-1}}{(1+2R_L)^{1/2(L-2)}}. \]
If we set \( R_L = 1 - 1.5y \),

\[
P(R_L > R_0) = \frac{1}{\beta(L, \frac{1}{2})} \int_{y=0}^{\frac{3}{2}(1-\kappa_0)} y^{L-1} \ (1-y)^{\frac{L}{2}-1} \ dy.
\]

\( p=4 \). \( R_L V = u_4 - u_3 \) and \( V = u_3 + u_3 + u_4 \), where \( u_3 \) is distributed as \( \chi^2 \) with \( 2L \) d.f., and \( u_3 \) and \( u_4 \) as \( \chi^2 \) with \( L \) d.f. Hence,

\[
D(R_L, V) = k_L(V) \int_{u_3 > 0} u_3^{L-1} \left\{ u_3^2 - 2u_3V + V^2(1-R_L^2) \right\}^{\frac{L}{2}-1} \ du_3 \ dV \ dR_L.
\]

If we substitute \( u_3 = Vx \), we can separate the parts involving \( V \) from those in \( x \) and integrate with respect to \( V \); hence,

\[
D(R_L) = k_L \int_{x=0}^{(1-\kappa_1)} x^{L-1} \ x^{\frac{L}{2}-1} \ dx \ dR_L,
\]

where \( k_L = \frac{1}{2^{L-1}} \frac{\Gamma((2L)/2)}{(L/2)!} \Gamma(L) \) and \( X = \left[ x^2 - 2x + (1-R_L^2) \right] \).

For even \( L \), let \( m = (L-1) \) and \( n = (L-2)/2 \). Then \( D(R_L) \) can be represented as \( I_{m, n}^{(1-\kappa_1)} \), where \( I_{m, n} \) is the integral

\[
\int x^m x^n \ dx.
\]

A recursion formula can now be set up to simplify the calculations:

\[
I_{m, n} = (m+1)^{-1} x^n x^{m+1} + \frac{2n}{m+1} \left[ I_{m+1, n-1} - I_{m+2, n-1} \right].
\]

When evaluating these integrals, it is noted that \( x=0 \) at the upper limit and \( (1-R_L^2) \) at the lower limit and that \( x=0 \) and \( (1-R_L) \) at the lower and upper limits respectively, so that any
term which contains both \( X \) and \( x \) will vanish when evaluated at both the upper and lower limits. Hence, \( D(R_L) \) will be a polynomial in \((1-R_L)\) only. \( P(R_L > R_o) \) must be derived for each \( L \).

Since the density function is much more complicated for \( L \) odd, we have interpolated for most of the significance levels of \( R_L \) for such cases.

Some examples will now be given. If we let \( y = (1-R_o) \), \( x = \sqrt{1-R_o^2} \) and \( P(R_L > R_o) = P_L \),

\[
P_2 = \frac{y^3}{2}
\]

\[
P_3 = \frac{1}{2} - \left[ 8 \sin^{-1} R_o - 65R_o x^3 + 97R_o x + 3R_o^3 x
\right.
\]

\[
+ (80R_o^3 + 12R_o^5) \log_e \left( \frac{(1-x)/R_o}{3} \right) \bigg] /8\pi.
\]

\[
P_6 = y^6 \left( 5R_o + 2 \right)/4.
\]

\[
P_6 = y^7 \left( 110 - 165y + 63y^2 \right)/16.
\]

\[
P_o = y^{15} \left( 31008 - 96,900y + 114,570y^8 - 60,705y^5
\right.
\]

\[
+ 12,155y^7 \bigg) /256.
\]

\[
P_6 = y^{27} \left( 42,073,200 - 235,609,920y + 568,182,384y^3
\right.
\]

\[
- 764,725,360y^5 + 620,283,030y^7 - 303,155,820y^7
\]

\[
+ 82,649,379y^9 - 9,694,845y^7 \right) /4096.
\]

\( p > 4 \). For \( p > 4 \), we will present only the integrands which must be evaluated in solving for \( D(R_L) \). The integrand for each \( p \) will be denoted by \( I_p \). The limits of integration for \( p=5,6 \) and \( 7 \) have been included in the derivations of \( D(R_L) \), pages 30-33.
\[ I_6 = K_L(V) \left( \frac{1}{2} \right) (yz)^{-1}, \text{ where} \]
\[ y = V(R_L - \lambda_3) - u(1 - \lambda_3), \quad z = u(1 - \lambda_3) - V(R_L - \lambda_3) \]
\[ K_L(V) = \frac{Ve}{2^{\frac{3}{2}L} \cdot \Gamma(L/2) \cdot \Gamma(3L)} \]

\[ I_6 = K_L(V)(u_5u_6)^{\frac{1}{2} - 1} (yz)^{-1}, \text{ where} \]
\[ y = V(2RL + 1) + u_3 - 3u_6, \quad z = V(1 - 2RL) - 3u_3 + u_6 \text{ and} \]
\[ K_L(V) = \frac{Ve}{2^{\frac{3}{2}L} \cdot \Gamma(L/2) \cdot \Gamma(3L)} \]

\[ I_7 = K_L(V) \left( \frac{1}{2} \right) (u_5yz)^{-1}, \text{ where} \]
\[ y = V(R_L - \lambda_3) - u(1 - \lambda_3) - u_3 (\lambda_2 - \lambda_3), \]
\[ z = u(1 - \lambda_3) + u_3 (\lambda_2 - \lambda_3) - V(R_L - \lambda_2) \text{ and} \]
\[ K_L(V) = \frac{Ve}{2^{\frac{3}{2}L} \cdot \Gamma(L/2) \cdot \Gamma(3L)} \]

\[ I_8 = K_L(V) \left( \frac{1}{2} \right) (u_5u_6yz)^{-1}, \text{ where} \]
\[ y = V(R_L - \lambda_3) - u(1 - \lambda_3) - u_3 (\lambda_2 - \lambda_3) - u_3 (\lambda_3 - \lambda_3), \]
\[ z = u(1 - \lambda_3) + u_3 (\lambda_2 - \lambda_3) + u_3 (\lambda_3 - \lambda_3) - V(R_L - \lambda_3) \text{ and} \]
\[ K_L(V) = \frac{Ve}{2^{\frac{3}{2}L} \cdot \Gamma(L/2) \cdot \Gamma(3L)} \]

For \( I_8 \), \( u_5 \) varies from
\[ S = \frac{V(R_L - \lambda_3) - u_3 (\lambda_2 - \lambda_3) - u(1 - \lambda_3)}{\lambda_3 - \lambda_3} \]
\[ U = \frac{V(R_L - \lambda_3) - u_3 (\lambda_2 - \lambda_3) - u(1 - \lambda_3)}{\lambda_3 - \lambda_3} \]

These two planes intersect in the straight line whose projection on the plane \( u_5 = 0 \) is given by the equation
\[ u_5 (\lambda_2 - \lambda_3) + u(1 - \lambda_3) = V(R_L - \lambda_3). \]
This line of intersection meets the plane \( u = 0 \) at the point
\[
u = V(1-R_L)/(1-\lambda_1), \quad u = V(R_L-\lambda_2)/(1-\lambda_2).
\]
The integration consists of two parts: (1) the integral between the plane \( U \) and the plane perpendicular to \( u = 0 \) and through the intersection of \( U \) and \( S \) minus (2) the integral between the plane \( S \) and this perpendicular plane. When \( R_L \geq \lambda_2 \), the limits of integration for (1) are: \( u_z \) from 0 to \( U \),
\[
u \text{ from } \frac{V(R_l-\lambda_2)-u_z(\lambda_2-\lambda_4)}{1-\lambda_4} \text{ to } \frac{V(R_l-\lambda_4)-u_z(\lambda_4-\lambda_4)}{1-\lambda_4} \quad \text{and}
\]
\( u_z \) from 0 to \( V(1-R_L)/(1-\lambda_1) \). The limits for (2) are the same except that \( \lambda_4 \) is replaced by \( \lambda_3 \). For \( R_L \leq \lambda_2 \), \( u \) is negative at the point of intersection; hence, we must subtract the value of the integral between \( u = 0 \) and \( u = V(R_L-\lambda_2)/(1-\lambda_2) \).

In this case, the limits for (1) are: \( u_z = 0 \) to \( U \),
\[
u = \frac{V(R_l-\lambda_2)-u_z(1-\lambda_2)}{\lambda_2-\lambda_4} \text{ to } \frac{V(R_l-\lambda_4)-u_z(1-\lambda_4)}{\lambda_4-\lambda_4} \quad \text{and}
\]
\( u = V(R_l-\lambda_2)/(1-\lambda_2) \) to 0. Replace \( \lambda_4 \) with \( \lambda_3 \) for (2).

Similar corrections must be made for \( R_L \leq \lambda_2 \) and \( R_L \leq \lambda_3 \).

In evaluating these integrals for lag 2, a few general formulas were useful. If we let
\[
I_{r,n,m} = \int S \int_{-\infty}^{\infty} x^n y^m z^m dx,
\]
for \( y = a^xbz \), \( z = a^xb'x \) and \( S \) and \( U \) the lower and upper limits of integration,
\[
I_{n, m}^{(s)} = x I_{n, m}^{(s)} - \frac{1}{(m+n+1)b} \left[ I_{n+1, m}^{(s)} \right.
- \frac{m}{(m+n)b} I_{n+1, m-1}^{(s)}
+ \left. \frac{m(m-1)p^2}{(m+n)(m+n-1)b} I_{n+1, m-2}^{(s)} - \cdots \right]^{(s)} \]

where \( I_{n, m} = I_{0, n, m} \) and \( p = ab^* - a^*b \). Also

\[
I_{n, m}^{(s)} = \frac{1}{(m+n+1)b} \left[ y^{n+1} \left( z^m - \frac{m}{(m+n)b} z^{m-1} \right) + \cdots \right.
+ \left. \frac{(-m^m m^m b^m)}{(m+n)b^m} \right]^{(s)} .
\]

If \( y(U) = 0, z(S) = 0 \) and \( y(S) = p/b^* \),

\[
I_{n, m}^{(s)} = \frac{(-m^m m^m b^m)}{(m+n+1)b^m} .
\]

**Distributions of \( R_L \) for Data Corrected for Trend**

We have assumed that the observations were normally and independently distributed about a zero mean. It will be shown in some examples which follow that most economic data should be corrected for a trend effect before they are analyzed for any other component, such as a cyclical movement. Also, Professor Cochran has mentioned the serial correlation problem in connection with meteorological data, which must be corrected for a non-zero mean \( (8) \). Since the correction for
the mean is simply a correction for a constant trend, a
general trend equation would include this correction as a
special case.

Some general distribution formulas have been derived for
$R_L$ when the variance and covariance have been corrected for
the mean, but no significance levels have been tabulated. We
hope to extend the distributions to include a correlation
which has been corrected for a linear trend. If the variance
and covariance are corrected for the mean, they have the
following form:

$$C_L = X_1 X_{1+1} + X_2 X_{1+2} + \cdots + X_n X_n - (\Sigma X_1)^2/N$$
$$V = \Sigma X_1^2 - (\Sigma X_1)^2/N.$$

For $C_L$, we note that the constants in our circulant
determinant are $a_1 = -\left[\lambda + \frac{1}{N}\right]$ , $a_2 = a_N = \left[\frac{1}{2} - \frac{1}{N}\right]$ and all other
a's = $\frac{1}{N}$. Hence,

$$A_k = \left[ -(\lambda + \frac{1}{N}) + \left( \frac{1}{2} - \frac{1}{N}\right)(\omega_k^{1} + \omega_k^{-1}) - \frac{1}{N} \Sigma_{i=3}^{N-1} \omega_k^{-i} \right].$$

Since $\Sigma_{i=3}^{N-1} \omega_k^{-i} = -(1 + \omega_k^{1} + \omega_k^{-1})$ for $k \neq N$ and $N-3$ for $k = N$,

$$A_k = \begin{cases} -\lambda_k + \cos \frac{2\pi k}{N} & \text{for } k \neq N \\ -\lambda_k & \text{for } k = N. \end{cases}$$

Therefore,

$$C_L = \begin{cases} \sum_{k=1}^{N/2} \lambda_k u_k & \text{for } N \text{ odd} \\ \sum_{k=1}^{N/2} \lambda_k u_k - u_{\frac{N}{2}} & \text{for } N \text{ even}, \end{cases}$$

where $\lambda_k = \cos \frac{2\pi k}{N}$ and $u_k$ is distributed as $\chi^2$ with 2 d.f.
and $u_{\frac{N}{2}}$ with 1 d.f.
The general density function of \( R_z \) for \( N \) odd is

\[
D(R_z) = k \sum_{l=m}^{k+1} (R_z - \lambda_1)^{k-1} / a_1 \quad \text{for} \quad \lambda_m \leq R_z \leq \lambda_{m-1},
\]

where \( k = (N-3)/2 \), \( a_1 = \prod_{j=1}^{k+1} (\lambda_j - \lambda_1) \) for \( i \neq j \) and \( m \geq 2 \).

Comparing this function with that for the uncorrected data (page 33), we note that the factor \( \sqrt{(R_z - \lambda_1)/(1-\lambda_1)} \) is omitted and that \( D(R_z) \) vanishes at \( R_z = \lambda_z \) instead of \( R_z = 1 \).

The new probability function is

\[
P(R_z > R_0) = 1 - \sum_{l=m}^{k+1} (R_0 - \lambda_1)^{k-1} / a_1 .
\]

It appears that a simple formula for \( N \) even also can be determined. In fact we have found that for \( N \) even, the distribution is of the same form as that found for \( N \) odd before.

The general probability formula is

\[
P(R_z > R_0) = \sum_{l=1}^{m} (\lambda_1 - R_0)^{k-1} / a_1, \quad \text{for} \quad \lambda_{m+1} \leq R_z \leq \lambda_m,
\]

where \( k = (N-3)/2 \) and \( a_1 = \prod_{j=1}^{k+1} (\lambda_j - \lambda_1) \sqrt{1+\lambda_1} \), if \( j \).

The distributions of \( R_L \) in terms of \( p (=N/L) \) are also similar to those for the data with a zero mean. The covariance and variance are still distributed as the sums of terms individually distributed as \( \chi^2 \):

\[
\sigma_L = \sum_{k=1}^{p-1} \lambda_k u_k + u \quad \text{and} \quad V = \sum_{k=1}^{p-1} u_k + u,
\]
where $\lambda_k = \cos \frac{2\pi k}{p}$ and the $u_k$'s are distributed as $\chi^2$ with $L$ d.f. and $u$ as $\chi^2$ with $L-1$ d.f.

For $p = 2$, the probability function is

$$P(R_L > R_o) = \Phi \left( \frac{1}{2}, \frac{L-1}{2} \right) \int_{y=0}^{\lambda} y^{L-1} (1-y)^{\frac{L-3}{2}} \, dy,$$

where $x = (1-R_o)/2$.

For $p = 3$, where $x = 2(1-R_o)/3$,

$$P(R_L > R_o) = \Phi \left( L, \frac{L-1}{2} \right) \int_{y=0}^{\lambda} y^{L-1} (1-y)^{\frac{L-3}{2}} \, dy.$$
\[
\begin{align*}
\text{if } x &= 0.06, \quad (x_1, x_2, x_3, x_4) = (0, 0, 0, 0) \\
\text{if } x &= 0.06, \quad \left\{ \left( \frac{Z}{\sqrt{V}} \right) \left( I - \frac{1}{X} \right) \right\} = (0, 0, 0, 0) \\
\text{and } \delta \text{ will be presented. For the distribution of } \delta \text{, refer to Table I for the covariance for } N = 2. \\
to illustrate the methods of computing the significance levels. For the covariance of } \delta \text{ a detailed account of the skewness with means zero.}
\end{align*}
\]
\[
\begin{align*}
0.0500 &= \left( \frac{2}{3} - \frac{\pi}{6} \right) + \\
&\quad \left( \left( \frac{2}{3} - \frac{\pi}{6} \right) \frac{\pi}{3} \right) + \\
&\quad \left( \left( \frac{2}{3} - \frac{\pi}{6} \right) \frac{\pi}{3} \right)
\end{align*}
\]

\[
T = \frac{1}{(\pi/3)^2} = \frac{1}{(0.406)^2} = 0.0191
\]

\[
\begin{array}{cccccc}
9.90 & 20.00 & 39.40 & 6 & 9 & 9.96 \ 
6.60 & 1.40 & 5.90 & 4 & 3 & 6.06 \ 
12.1 & 9.90 & 8.40 & 2 & 2 & 12.14 \ 
7.92 & 9.82 & 0.00 & 1 & 1 & 7.92 \ 
5.04 & 39.40 & 49.50 & 0 & 0 & 5.04 \ 
0.00 & 0.00 & 0.00 & 0 & 0 & 0.00
\end{array}
\]

Let \( \theta = 4.06 \). Then \( \frac{\theta}{3} = 0.406 \), \( r = \frac{\theta}{6} \), and \( e = \frac{\theta}{6} \). Then, \( N = 0.13124 \).

\[
N = \frac{1}{3} \left( \frac{\pi}{3} \right) = \frac{1}{3} \left( \frac{\pi}{3} \right)
\]

In this case, \( \chi = 0.06042 \) and \( \chi^2 = 0.6042 \).
Serial Correlation Coefficient, $R_L$

Lag 1

A complete table of both the positive 5% and 1% probability levels of $R_L$ not corrected for mean for $N$ up to 45 has been compiled (Table 1). This table includes the significance levels using both the exact small sample distributions and the approximate normal distribution for large samples. These results are shown graphically in Figure 3. It would appear that the approximate distribution wouldn't vary much from the exact distribution with the 5% significance levels for $N$ greater than 50; a somewhat larger $N$ would be required for the approximation to be accurate for the 1% significance levels.

In case it is desired to test a negative $R_L$ for significance, it might be important to note that the density functions are symmetrical within three-decimal accuracy for all $N$ equal to or greater than 11. The negative significance levels for $N = 5, 7, \text{ and } 9$ respectively are $-0.631, -0.553, \text{ and } -0.496$. The corresponding 1% points are $-0.752, -0.718, \text{ and } -0.660$.

A brief note on the methods used in estimating these significance levels might be useful. We have derived the exact significance levels for $N = 2, 3, 4, 5, 6, 7, 9, 11, 13, 15, 25 \text{ and } 45$. The levels for intermediate $N$'s have been determined by graphical interpolation. This interpolation was
<table>
<thead>
<tr>
<th>H₀</th>
<th>H₀⁺</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.000</td>
<td>*0.021</td>
</tr>
<tr>
<td>0.020</td>
<td>*0.060</td>
</tr>
<tr>
<td>0.040</td>
<td>*0.125</td>
</tr>
<tr>
<td>0.060</td>
<td>*0.210</td>
</tr>
<tr>
<td>0.080</td>
<td>*0.315</td>
</tr>
<tr>
<td>0.100</td>
<td>*0.430</td>
</tr>
<tr>
<td>0.120</td>
<td>*0.565</td>
</tr>
<tr>
<td>0.140</td>
<td>*0.715</td>
</tr>
<tr>
<td>0.160</td>
<td>*0.885</td>
</tr>
<tr>
<td>0.180</td>
<td>*1.075</td>
</tr>
<tr>
<td>0.200</td>
<td>*1.285</td>
</tr>
<tr>
<td>0.220</td>
<td>*1.515</td>
</tr>
<tr>
<td>0.240</td>
<td>*1.765</td>
</tr>
<tr>
<td>0.260</td>
<td>*2.045</td>
</tr>
<tr>
<td>0.280</td>
<td>*2.355</td>
</tr>
<tr>
<td>0.300</td>
<td>*2.695</td>
</tr>
<tr>
<td>0.320</td>
<td>*3.065</td>
</tr>
<tr>
<td>0.340</td>
<td>*3.465</td>
</tr>
<tr>
<td>0.360</td>
<td>*3.895</td>
</tr>
<tr>
<td>0.380</td>
<td>*4.365</td>
</tr>
<tr>
<td>0.400</td>
<td>*4.875</td>
</tr>
<tr>
<td>0.420</td>
<td>*5.435</td>
</tr>
<tr>
<td>0.440</td>
<td>*6.045</td>
</tr>
<tr>
<td>0.460</td>
<td>*6.715</td>
</tr>
<tr>
<td>0.480</td>
<td>*7.445</td>
</tr>
<tr>
<td>0.500</td>
<td>*8.235</td>
</tr>
<tr>
<td>0.520</td>
<td>*9.105</td>
</tr>
<tr>
<td>0.540</td>
<td>*10.045</td>
</tr>
<tr>
<td>0.560</td>
<td>*11.065</td>
</tr>
<tr>
<td>0.580</td>
<td>*12.165</td>
</tr>
<tr>
<td>0.600</td>
<td>*13.355</td>
</tr>
<tr>
<td>0.620</td>
<td>*14.645</td>
</tr>
<tr>
<td>0.640</td>
<td>*16.045</td>
</tr>
<tr>
<td>0.660</td>
<td>*17.545</td>
</tr>
<tr>
<td>0.680</td>
<td>*19.165</td>
</tr>
<tr>
<td>0.700</td>
<td>*20.945</td>
</tr>
<tr>
<td>0.720</td>
<td>*22.835</td>
</tr>
<tr>
<td>0.740</td>
<td>*24.875</td>
</tr>
<tr>
<td>0.760</td>
<td>*27.065</td>
</tr>
<tr>
<td>0.780</td>
<td>*29.415</td>
</tr>
<tr>
<td>0.800</td>
<td>*31.945</td>
</tr>
<tr>
<td>0.820</td>
<td>*34.665</td>
</tr>
<tr>
<td>0.840</td>
<td>*37.575</td>
</tr>
<tr>
<td>0.860</td>
<td>*40.705</td>
</tr>
<tr>
<td>0.880</td>
<td>*44.125</td>
</tr>
<tr>
<td>0.900</td>
<td>*47.845</td>
</tr>
<tr>
<td>0.920</td>
<td>*51.945</td>
</tr>
<tr>
<td>0.940</td>
<td>*56.445</td>
</tr>
<tr>
<td>0.960</td>
<td>*61.345</td>
</tr>
<tr>
<td>0.980</td>
<td>*66.745</td>
</tr>
</tbody>
</table>

Table I: Significance Levels of H₀
FIGURE 3 - SIGNIFICANCE LEVELS OF \( R \)
carried out as follows: We plotted the significance levels which we had computed for $N$ up to 15 and then drew a smooth curve through these points to estimate the significance levels for $N = 8, 10, 12$ and 14 and also extrapolated to determine an approximate value of $R_0$ for $N = 25$. The exact $R_0$ for $N = 25$ was then determined and the smooth curve extended to this point to determine the significance levels for $N$ between 15 and 25. We found that our extrapolated value of $R_0$ for $N = 25$ was nearly exact; so, we have concluded that the graphical interpolation gives a very good approximation to the true significance levels. The next key value used was $N = 45$. The same procedure was used to estimate the $R_0$ for this $N$ and then to determine the significance levels for the $N$'s between 25 and 45.

Since we wanted to determine the values of the density function as well as the probability function for the smaller $N$, we calculated the values of $D(R_0)$ and then multiplied by the factors $(R_0 - \lambda_1)$ to determine $P(R_0)$. The values of $\alpha_1$ were calculated by simply multiplying the factors together. This procedure becomes quite tedious for large $N$ and is not recommended if only $P(R_0)$ is required.

The computations were materially simplified by using logarithms for the significance levels for $N \geq 15$. Certain trigonometric relations were used to simplify these results:
\[ a_i = \prod_{j=1}^{N-1} (\lambda_j - \lambda_1)^{1/(1-\lambda_1)} (i=1, \ldots, \frac{N-1}{2}) \]

\[ = 2^{\frac{N-2}{2}} \prod_{j=1}^{\frac{N-2}{2}} \sin \frac{\pi}{N}(i+j) \sin \frac{\pi}{N}(i-j) \sin \frac{\pi}{N}i, \]

where \( \prod' \) indicates that the factor for \( i=j \) is omitted.

Let \( a_i' = |a_i|/2^{\frac{N-2}{2}} \). Since \( \sin \frac{\pi}{N}(i+j) \) and \( \sin \frac{\pi}{N}i/N \) are always positive,

\[ a_i' = \prod_{j=1}^{\frac{N-2}{2}} \sin \frac{\pi}{N}(i+j) \sin \frac{\pi}{N}i \sin \frac{\pi}{N}|i-j|. \]

If we let \( (N-1)/2 = n \), then

\[ \log a_i' = \sum_{j=0}^{n} \log \sin \frac{\pi}{N}(i+j) + \sum_{j=1}^{n} \log \sin \frac{\pi}{N}|i-j| \]

\[ - \log \sin \frac{2\pi i}{N} \]

\[ = 2 \sum_{k=1}^{n} \log \sin \frac{\pi i}{N} + \sum_{k=1}^{n} \log \sin \frac{\pi i}{N} - \log \sin \frac{2\pi i}{N}. \]

Since \( \sin(n-\theta) = \sin \theta \) and \( N = 2n+1 \), \( \sin \frac{n+1}{N} = \sin \frac{n-1}{N} \pi \).

Therefore,

\[ \log a_i' = 2 \sum_{k=1}^{n} \log \sin \frac{\pi i}{N} - \log \sin \frac{2\pi i}{N}. \]

We have converted \( (R_0-\lambda_1) \) into the product of two sines and used the log sine tables here also. Let \( R_0 = \cos x \). Then

\[ \log (R_0-\lambda_1) = \log 2^{\frac{N-2}{2}} + \frac{N-2}{2} \log \sin \left( \frac{\pi i}{N} + \frac{x}{2} \right) \]

\[ + \frac{N-2}{2} \log \sin \left( \frac{\pi i}{N} - \frac{x}{2} \right). \]
\[ R_0 = \cos 64^\circ 04' = 0.4376. \]

end the positive 1% strength at

\[ R_0 = \cos 71^\circ 20' = 0.3193. \]

5% strength tests were located at

the values of \( z \) for \( t > 6 \) have not been computed. The positive

equation for \( z \) is

\[ z \geq \frac{1}{2} \sin \frac{N}{2}. \]

since the \( z \)'s for \( t > 6 \) are less than 0.001.

\[ \sin \frac{N}{2} \leq \frac{1}{2}. \]

since or multiples of 180/2 degrees. For \( N = 25^\circ \), 180/2/25

in this table. In the equation for \( z = \frac{1}{2} \), we need the log

\[
\frac{T_p}{T_N} (R_0 - 1/2 \sin \frac{N}{2}) = T_p
\]

positive 5% strength tests for \( N = 25^\circ \). We have let

Table 2 presents the calculations used to locate the

distribution by \( N \), interpolation is necessary to obtain the values

distributable by \( N \) at 1/2 degree intervals when \( 90^\circ \) is not

since 90 is distributable by each of these \( 90^\circ \) when \( 90^\circ \) is not

\( N = 15^\circ, 25^\circ \) and \( 35^\circ \) have been selected as the check solutions.

The distribution of \( R_0 \) for

- Another tables also indicate a very satisfactory method of Inter-

The (16) and for \( N \) = 45 the 14 place tables of \( z = \) negative (16) for

For \( N = 15^\circ \) and \( 25^\circ \) we have used 16 place log tables

\[
\begin{align*}
\log T_p &= \log \left( \frac{\frac{T_p}{T_N} (R_0 - \frac{1}{2} \sin \frac{N}{2})}{\log \sin \frac{N}{2}} \right) = \frac{T_p}{T_N} \log \left( \frac{T_p}{T_N} R_0 \right) + \frac{1}{2} \frac{T_p}{T_N} \log \sin \frac{N}{2} \\
&= \frac{T_p}{T_N} \log \left( \frac{T_p}{T_N} R_0 \right) + \frac{1}{2} \frac{T_p}{T_N} \log \sin \frac{N}{2}
\end{align*}
\]

Since

- 59 -
Table 2. Calculation of 5% significance level of $R_1$, $N = 25$

<table>
<thead>
<tr>
<th>$k$</th>
<th>$180k/N$</th>
<th>$\log \sin \left[ \frac{180k}{N} \right]^\circ$</th>
<th>$i$</th>
<th>$\log \alpha_{i}^\prime$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7.2</td>
<td>9.0980662445-10</td>
<td>12</td>
<td>5.0751538677-10</td>
</tr>
<tr>
<td>2</td>
<td>14.4</td>
<td>9.39565681233</td>
<td>11</td>
<td>4.6072253278</td>
</tr>
<tr>
<td>3</td>
<td>21.6</td>
<td>9.5659947644</td>
<td>10</td>
<td>4.4040014270</td>
</tr>
<tr>
<td>4</td>
<td>28.8</td>
<td>9.662250355</td>
<td>9</td>
<td>4.2864000049</td>
</tr>
<tr>
<td>5</td>
<td>36.0</td>
<td>9.7692168852</td>
<td>8</td>
<td>4.166545356</td>
</tr>
<tr>
<td>6</td>
<td>43.2</td>
<td>9.8345033387</td>
<td>7</td>
<td>4.109816049</td>
</tr>
<tr>
<td>7</td>
<td>50.4</td>
<td>9.868781073</td>
<td>6</td>
<td>4.174079398</td>
</tr>
<tr>
<td>8</td>
<td>57.6</td>
<td>9.926511534</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>64.8</td>
<td>9.9565655766</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>72.0</td>
<td>9.9782062355</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>79.2</td>
<td>9.992235073</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>86.4</td>
<td>9.9991421724</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$x = 71.5^\circ$, $R_0 = \cos x = 0.3173$

<table>
<thead>
<tr>
<th>$i$</th>
<th>$\frac{180i}{N} + \frac{x}{2}$</th>
<th>$\log \sin \left( \frac{180i}{N} + \frac{x}{2} \right)^\circ$</th>
<th>$\frac{180i}{N} - \frac{x}{2}$</th>
<th>$\log \sin \left( \frac{180i}{N} - \frac{x}{2} \right)^\circ$</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>122.15</td>
<td>9.9277079476-10</td>
<td>50.65</td>
<td>9.8633408136-10</td>
</tr>
<tr>
<td>11</td>
<td>114.35</td>
<td>9.9574522377</td>
<td>43.45</td>
<td>9.8374124787</td>
</tr>
<tr>
<td>10</td>
<td>107.75</td>
<td>9.9783174132</td>
<td>36.25</td>
<td>9.7718149654</td>
</tr>
<tr>
<td>9</td>
<td>100.55</td>
<td>9.9925957093</td>
<td>29.05</td>
<td>9.6862542484</td>
</tr>
<tr>
<td>8</td>
<td>93.35</td>
<td>9.9992572438</td>
<td>21.85</td>
<td>9.5707506402</td>
</tr>
<tr>
<td>7</td>
<td>86.15</td>
<td>9.9990187998</td>
<td>14.65</td>
<td>9.4029724370</td>
</tr>
</tbody>
</table>

| $i$ | $\log |P_1|$ | $P_1$ |
|-----|-----------|-------|
| 12  | 2.8094068861 | 644.7731 |
| 11  | 3.0237189108 | -1080.7343 |
| 10  | 2.7282715950 | 554.8988  |
| 9   | 2.0203245086 | -104.7936 |
| 8   | 0.9284231504 | 6.8925   |
| 7   | 3.9419176183-10 | 0.0675 |

$P(R_0) = \prod P_1 = 0.9500$

$\log |P_1| = (11.5) \left[ \log \sin \left( \frac{180i}{N} + \frac{x}{2} \right)^\circ + \log \sin \left( \frac{180i}{N} - \frac{x}{2} \right)^\circ \right] - \log \alpha_{i}^\prime$
General Lag L

The positive 5% and 1% significance levels of $R_L$ have been calculated for different $P(=\frac{N}{L})$, where $L$, a factor of $N$, is the primary lag. Referring to page 41, we note that the probability functions are merely incomplete Beta functions for $P = 2$ and $3$. Pearson (13) has tabulated the ratio

$$I_x(m,n) = \frac{\beta_x(m,n)}{\beta(m,n)}, \text{ where } \beta_x(m,n) = \int_0^y y^{m-1}(1-y)^{n-1} \, dy^*.$$

For $P = 2$, the probability integral in Pearson's notation is

$$P(R_L > R_0) = I_x\left(\frac{L}{2}, \frac{L}{2}\right),$$

where $R_0 = 1-2x$. We set $P = 0.05$ and 0.01 for the positive 5% and 1% significance levels respectively and used Pearson's tables to locate the values of $x$ which satisfied these equations.

Similarly for $P = 3$, $P(R_L > R_0) = I_x(L,\frac{L}{2})$, where $R_0 = 1-1.5x$.

The significance levels of $R_L$ for $P = 2$ and $3$ have been tabulated for all lags up to 20 and for lags 22, 24, 26, 30, 40 and 50 (Table 3).

As noted on pages 42-43, the probability functions $P(R_L > R_0)$ are much more complicated for $P > 3$. For $P = 4$, we computed the significance levels for $L = 1, 2, 3, 4, 6, 10$ and 16 and interpolated for the intermediate lags. These significance levels have been plotted in Figure 4.

* Pearson uses $p$ and $q$ instead of $m$ and $n.$
Table 3. Significance levels of $R_L$ for $p = 2$ and 3.

<table>
<thead>
<tr>
<th>Lag</th>
<th>$R_{.05}$</th>
<th>$R_{.01}$</th>
<th>$R_{.05}$</th>
<th>$R_{.01}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.988</td>
<td></td>
<td>0.854</td>
<td>0.970</td>
</tr>
<tr>
<td>2</td>
<td>0.900</td>
<td>0.980</td>
<td>0.685</td>
<td>0.850</td>
</tr>
<tr>
<td>3</td>
<td>0.805</td>
<td>0.924</td>
<td>0.556</td>
<td>0.745</td>
</tr>
<tr>
<td>4</td>
<td>0.729</td>
<td>0.882</td>
<td>0.486</td>
<td>0.667</td>
</tr>
<tr>
<td>5</td>
<td>0.670</td>
<td>0.833</td>
<td>0.427</td>
<td>0.607</td>
</tr>
<tr>
<td>6</td>
<td>0.622</td>
<td>0.769</td>
<td>0.400</td>
<td>0.560</td>
</tr>
<tr>
<td>7</td>
<td>0.582</td>
<td>0.750</td>
<td>0.370</td>
<td>0.522</td>
</tr>
<tr>
<td>8</td>
<td>0.549</td>
<td>0.718</td>
<td>0.347</td>
<td>0.491</td>
</tr>
<tr>
<td>9</td>
<td>0.521</td>
<td>0.685</td>
<td>0.327</td>
<td>0.464</td>
</tr>
<tr>
<td>10</td>
<td>0.497</td>
<td>0.658</td>
<td>0.310</td>
<td>0.441</td>
</tr>
<tr>
<td>11</td>
<td>0.476</td>
<td>0.634</td>
<td>0.296</td>
<td>0.421</td>
</tr>
<tr>
<td>12</td>
<td>0.458</td>
<td>0.612</td>
<td>0.283</td>
<td>0.404</td>
</tr>
<tr>
<td>13</td>
<td>0.441</td>
<td>0.593</td>
<td>0.272</td>
<td>0.389</td>
</tr>
<tr>
<td>14</td>
<td>0.426</td>
<td>0.575</td>
<td>0.262</td>
<td>0.375</td>
</tr>
<tr>
<td>15</td>
<td>0.413</td>
<td>0.558</td>
<td>0.253</td>
<td>0.362</td>
</tr>
<tr>
<td>16</td>
<td>0.400</td>
<td>0.543</td>
<td>0.245</td>
<td>0.350</td>
</tr>
<tr>
<td>17</td>
<td>0.389</td>
<td>0.529</td>
<td>0.238</td>
<td>0.340</td>
</tr>
<tr>
<td>18</td>
<td>0.378</td>
<td>0.516</td>
<td>0.231</td>
<td>0.331</td>
</tr>
<tr>
<td>19</td>
<td>0.369</td>
<td>0.504</td>
<td>0.225</td>
<td>0.322</td>
</tr>
<tr>
<td>20</td>
<td>0.260</td>
<td>0.493</td>
<td>0.219</td>
<td>0.314</td>
</tr>
<tr>
<td>22</td>
<td>0.344</td>
<td>0.472</td>
<td>0.209</td>
<td>0.299</td>
</tr>
<tr>
<td>24</td>
<td>0.330</td>
<td>0.454</td>
<td>0.200</td>
<td>0.286</td>
</tr>
<tr>
<td>26</td>
<td>0.317</td>
<td>0.437</td>
<td>0.192</td>
<td>0.275</td>
</tr>
<tr>
<td>30</td>
<td>0.296</td>
<td>0.410</td>
<td>0.178</td>
<td>0.256</td>
</tr>
<tr>
<td>40</td>
<td>0.258</td>
<td>0.368</td>
<td>0.154</td>
<td>0.221</td>
</tr>
<tr>
<td>50</td>
<td>0.231</td>
<td>0.322</td>
<td>0.138</td>
<td>0.198</td>
</tr>
</tbody>
</table>
The density functions of only $R_s$ have been derived for $P = 5, 6, 7$ and 9. The corresponding significance levels are (interpolating for $P = 8$):

<table>
<thead>
<tr>
<th>P</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R_{.05}$</td>
<td>0.471</td>
<td>0.437</td>
<td>0.412</td>
<td>0.390</td>
<td>0.369</td>
</tr>
<tr>
<td>$R_{.01}$</td>
<td>0.646</td>
<td>0.592</td>
<td>0.556</td>
<td>0.527</td>
<td>0.503</td>
</tr>
</tbody>
</table>

**Examples.**

I have derived the various serial correlations for part of the data used in my recent study of agricultural prices (1). Since the means of these price series were not zero, we should use significance levels which have been derived for a correlation coefficient corrected for the mean, but we will present these examples to illustrate the methods of deriving the serial $R$'s, of testing them for significance and of analyzing the results obtained. Our general formulas indicate that the true significance levels for data corrected for the mean will be below those presented in this thesis; therefore, any $R_L$ found to be significant using our significance levels would be adjudged at least significant at the true significance levels.

The first example consists of sixteen annual rye prices for the period 1890-1915. Our tables of significance levels can be used to test for significance all the serial $R$'s for
the series, although the moving average method might be
intended to use the method of linear regression to analyze
than the previous high values for \( R^2 \) and \( R^e \) we would be
zero. Since \( R^2 \) and \( R^e \) were almost statistically different from
zero and all other \( R^s \) were not statistically different from
not quite statistically, \( R^s \) and \( R^e \) were statistically less than
the results were meteorally expected. In this case, \( R^e \) was
expected price and \( P \) the year, 1980 before detrended as (1)
the simple linear trend \( (P_e = 66.5 + 4.45t) \) where \( P \) is the
same linear trend but when the prices were corrected for \( R^e \) of the
seasonal price for the period 1980-1984 (Table 4). Most
annual prices for the period 1980-1984 (Table 4). I have selected twenty-five
In order to compare data which had been corrected for
from this data, since they were not corrected for trend.
longest statistically. Few different conclusions should be drawn
be statistically different from 0, although \( R^2 \) and \( R^e \) are
for these various tests are 0.0.266, 0.368, 0.265, 0.365, 0.0.259, 0.0.269, 0.0.259, 0.0.265, 0.0.259, 0.0.266,
For these various tests are 0.407, 0.407, 0.407. Since the %
significance levels
the 6% respectively are: 0.407, 0.407, 0.407. The linear correlation
coefficient for Table 1 to
end 71.4. The series correlation coefficients for Table 1 to
46.2, 36.2, 37.4, 38.6, 39.6, 40.6, 41.6, 42.6, 43.6, 44.6.45.6, 46.6.
the price for 1980, were: 54.5, 55.2, 56.2, 57.2, 58.2, 59.2, 60.2, 61.6, 62.6, 63.6.
the sample. The prices in cents per bushel start with

- 62 -
Table 4. Serial R's for annual hide prices, 1890-1914.

<table>
<thead>
<tr>
<th>Year</th>
<th>P</th>
<th>PE</th>
<th>D</th>
<th>lag</th>
<th>ΣPiPj</th>
<th>R(P)</th>
<th>Eμd₁d₁</th>
<th>R(D)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1890</td>
<td>93</td>
<td>73</td>
<td>20</td>
<td>0</td>
<td>30054</td>
<td>4332</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1891</td>
<td>95</td>
<td>77</td>
<td>18</td>
<td>1</td>
<td>21414</td>
<td>0.713**</td>
<td>1200</td>
<td>0.28</td>
</tr>
<tr>
<td>1892</td>
<td>97</td>
<td>82</td>
<td>5</td>
<td>2</td>
<td>14423</td>
<td>0.480**</td>
<td>-184</td>
<td>-0.04</td>
</tr>
<tr>
<td>1893</td>
<td>75</td>
<td>68</td>
<td>-11</td>
<td>3</td>
<td>9660</td>
<td>0.322*</td>
<td>-32</td>
<td>0.00</td>
</tr>
<tr>
<td>1894</td>
<td>64</td>
<td>91</td>
<td>-27</td>
<td>4</td>
<td>5362</td>
<td>0.178</td>
<td>659</td>
<td>-0.15</td>
</tr>
<tr>
<td>1895</td>
<td>103</td>
<td>95</td>
<td>8</td>
<td>5</td>
<td>-2442</td>
<td>-0.081</td>
<td>-2304</td>
<td>-0.33**</td>
</tr>
<tr>
<td>1896</td>
<td>81</td>
<td>100</td>
<td>-19</td>
<td>6</td>
<td>-5272</td>
<td>-0.175</td>
<td>-1745</td>
<td>-0.40*</td>
</tr>
<tr>
<td>1897</td>
<td>100</td>
<td>104</td>
<td>-4</td>
<td>7</td>
<td>-5722</td>
<td>-0.190</td>
<td>750</td>
<td>0.17</td>
</tr>
<tr>
<td>1898</td>
<td>115</td>
<td>109</td>
<td>6</td>
<td>8</td>
<td>-8613</td>
<td>-0.287</td>
<td>445</td>
<td>0.10</td>
</tr>
<tr>
<td>1899</td>
<td>124</td>
<td>113</td>
<td>11</td>
<td>9</td>
<td>-11978</td>
<td>-0.399*</td>
<td>-476</td>
<td>-0.11</td>
</tr>
<tr>
<td>1900</td>
<td>119</td>
<td>117</td>
<td>2</td>
<td>10</td>
<td>-11311</td>
<td>-0.376*</td>
<td>802</td>
<td>0.19</td>
</tr>
<tr>
<td>1901</td>
<td>124</td>
<td>122</td>
<td>2</td>
<td>11</td>
<td>-9796</td>
<td>-0.326*</td>
<td>1133</td>
<td>0.26</td>
</tr>
<tr>
<td>1902</td>
<td>134</td>
<td>126</td>
<td>8</td>
<td>12</td>
<td>-10769</td>
<td>-0.358*</td>
<td>-1098</td>
<td>-0.25</td>
</tr>
<tr>
<td>1903</td>
<td>117</td>
<td>131</td>
<td>-14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1904</td>
<td>117</td>
<td>135</td>
<td>-13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1905</td>
<td>143</td>
<td>140</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1906</td>
<td>154</td>
<td>144</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1907</td>
<td>146</td>
<td>149</td>
<td>-3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1908</td>
<td>134</td>
<td>153</td>
<td>-19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1909</td>
<td>165</td>
<td>158</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1910</td>
<td>155</td>
<td>162</td>
<td>-7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1911</td>
<td>148</td>
<td>166</td>
<td>-13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1912</td>
<td>176</td>
<td>171</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1913</td>
<td>194</td>
<td>175</td>
<td>19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1914</td>
<td>196</td>
<td>180</td>
<td>16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

F = prices in mills per pound
PE = prices estimated by linear regression
D = P - PE
ΣPiPj = sum of cross products of P's in terms of deviations from mean; Σμd₁d₁ is the same for the D's.
R(P) and R(D) = serial R's for F's and D's.
* Significant at 5% point.
** Significant at 1% point.
considered. It should be noted also that the distribution of $R_L$ for variates corrected for a linear trend is merely approximated by our distribution of $R_L$.

A series of forty-seven monthly rye prices which had been corrected for trend by a variable moving average exhibited a decided periodic movement. In my previous study the method of hidden periodicities was successfully used to analyze this data. The various serial correlation coefficients for this series were:

<table>
<thead>
<tr>
<th>lag</th>
<th>$R_L$</th>
<th>lag</th>
<th>$R_L$</th>
<th>lag</th>
<th>$R_L$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.957</td>
<td>9</td>
<td>0.295</td>
<td>17</td>
<td>-0.563</td>
</tr>
<tr>
<td>2</td>
<td>0.918</td>
<td>10</td>
<td>0.179</td>
<td>18</td>
<td>-0.350</td>
</tr>
<tr>
<td>3</td>
<td>0.857</td>
<td>11</td>
<td>0.063</td>
<td>19</td>
<td>-0.750</td>
</tr>
<tr>
<td>4</td>
<td>0.783</td>
<td>12</td>
<td>-0.053</td>
<td>20</td>
<td>-0.800</td>
</tr>
<tr>
<td>5</td>
<td>0.698</td>
<td>13</td>
<td>-0.164</td>
<td>21</td>
<td>-0.858</td>
</tr>
<tr>
<td>6</td>
<td>0.607</td>
<td>14</td>
<td>-0.273</td>
<td>22</td>
<td>-0.898</td>
</tr>
<tr>
<td>7</td>
<td>0.509</td>
<td>15</td>
<td>-0.373</td>
<td>23</td>
<td>-0.920</td>
</tr>
<tr>
<td>8</td>
<td>0.405</td>
<td>16</td>
<td>-0.473</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

All of the correlations in the first and third columns are highly significant (1% point), while only those for lags 10-13 are not significant. It would appear that the use of the variable moving average tends to accentuate the size of the serial correlation coefficient; therefore, we probably need a more critical test of significance than that given on our results. Only the first half of the series of $R$'s has been tabulated since these $R$'s will be reproduced in reverse order in the second half. It should be noted that only one-tenth of the monthly items would be adjudged independent by our criterion, while at least one-half of the yearly items
were independent. In order to compare the serial R's obtained by our method with those obtained by dropping one term for each lag, I computed \( R_{z_1} \) and \( R_{z_2} \) by the latter method. These R's were -0.963 and -0.952, not decidedly different from those calculated by our method.
DISCUSSION OF RESULTS

In the introduction, it was pointed out that there is need to improve the techniques of analyzing data which are ordered in time, especially economic and meteorological data. Such data usually consist of items which are dependent on one another; we have denoted this dependence by the term serial correlation. This condition of dependence is not confined to data ordered in time, although such data have received most of the attention to date; perhaps, some high correlations between variates in other fields have been caused by this interdependence. In addition, Tschuprow has pointed out that our ordinary tests of the significance of regression coefficients must be revised if the variance or any other distribution statistic of the dependent variable does not remain constant for all values of the independent variables (17).

Most of our examples of serial correlation have been taken from economic time series, because we have been concerned with such data in the past. Nevertheless, meteorological series, such as those studied by Schuster (14), obviously concern serially correlated items. Yule also has indicated that studies of the serial correlation should be encouraged with reference to the periodicities of sun-spots (25).

This manuscript is the result of an attempt to solve some of the complicated problems connected with serial
correlations by presenting the distributions of the serial R's for various lags. As noted before, we have used a simpler definition of this serial R than has been used heretofore. As an introduction to the problem, we have derived a few distributions and significance levels of the serial covariance. These distributions are rather useless from a practical viewpoint, since we seldom, if ever, know the true population variance.

We then proceeded to study the problem of the serial correlation coefficient, $R_L$. Since many economic and meteorological series consist of a large number of items, the large sample distribution of $R_L$ was deemed essential in formulating a test of significance for such series. Unfortunately we do not have any estimate of the error involved in using the large sample significance levels; hence, we are not certain what value of $N$ is needed to permit the use of these significance levels. For $R_2$, $N$ apparently must be somewhat larger than fifty at the 5% point and much larger at the 1% point; however, the large sample approximation is fairly accurate to two decimal places for the 5% point with $N$ greater than fifty. Although the distribution of $R_L$ for large $N$ is independent of $L$, the value of this large $N$ must be increased as $L$ is increased in order to have an accurate approximation of the significance levels.

The general density function of $R_4$ for $N$ odd has been set up. Most of the significance levels for $N$ greater than
Improvement as present conditions seem to indicate, the study of the susceptibility of a serial correlation is es
sential to the solution of the problem. If the need for these
measures, not factors of N. This study is more an intro-
in for all N, which have some I as a factor, but which are, in
each N. The susceptibility levels of any N will be the same
instead of counteracting the susceptibility levels of all N for
or Ha in terms of different primary bases I for each N/L.
It was found that the best method of studying the distri-
bution is to use a non-susceptible correlation to find a
series before a non-susceptible correlation is found; we
have found that the co-susceptible of the serial correlation for most

- 69 -
can be extended to include more distributions. In the meantime, we will attempt to simplify the complicated integrals which have been set up for \( N \) greater than four. The general distribution formulas have been tabulated for \( N/L = 2, 3 \) and 4.

In all of these distribution problems, we have neglected the so-called type II error \((11, 20)\), which is defined as the probability that our statistic really differs from the postulated parametric value even though it has been adjudged not significantly different from this parameter by our test of significance. To be strictly accurate, it must be shown that our particular distributions minimize this type II error.

Again we should emphasize that all our significance levels have been derived for data with a zero mean. The distributions of \( R_L \), when \( C_L \) and \( V \) are corrected for the sample mean have been indicated. We hope to extend this research to include distribution of the serial correlation corrected for at least a linear trend.

In conclusion, it should be emphasized that this paper is merely an introduction to the problem of serial correlation. Much has been written in the past concerning serially correlated variates, but little has been done about estimating the importance and the distribution of the serial correlation. Time series analyses have been made, disregarding all known hypotheses about independent observations; yet, few people have tried to rectify these mistakes. Dr. Tintner has indicated
a method of selection which avoids some of these difficulties in the use of the variate-difference method (16), and Dr. Wold also has emphasized new methods of handling time series, but even Wold avoided the question of testing hypotheses. If we can start a trend towards a more rational study of data ordered in time or space, much will have been accomplished, even though the methods presented here are soon replaced by simpler ones. Although economic and meteorological time series may be the best example of such data, I suspect that some aspect of the serial correlation problem might be applied to such problems as those of competition in field crop experiments.
SUMMARY

(1) Items ordered in time usually are not independent of one another; thus, some modifications must be made in the usual regression or harmonic analysis of such data.

(2) The number of independent events in such a time series can be approximated by dividing the total number of observations by the number of the first lag which will produce a non-significant serial correlation coefficient.

(3) The distributions and significance levels of the serial covariance have been indicated.

(4) If the series consists of a large number of items, the large sample normal distribution of the serial R can be used.

(5) For small samples, we have calculated what we purport to call the general density and probability functions for N odd and lag 1. The 1% and 5% significance levels for N up to 45 have been tabulated.

(6) It has been indicated that the same distributions hold for RL as for R when N is a prime number or when L and N have no common factor. Some of the distributions of RL for L a factor of N have been derived and directions presented for extensions.
(7) Several examples of the calculations of the serial correlation coefficients have been set up, illustrating applications to Wold's analysis of time series and to the problem of estimating the amount of information contained in various series.

(8) Points to be considered in the future are: discussion of type II error, extension to other fields than that of economic time series, a more vigorous test of the significance of the regression and harmonic coefficients in a serially correlated series, an extension of the distributions to include more lags of higher order than those considered in this manuscript and, above all, distributions and significance levels of $R_L$ when the data are corrected for trend effects.
ACKNOWLEDGMENTS

Sincere thanks are due the author's major professor, Dr. Gerhard Tintner, whose suggestions on the formulation and subsequent analysis of the problem of serial correlation have been indispensable. In addition, no one can devote four years to research under Dr. Tintner without acquiring some insight into the solutions of the many complex problems in present-day economics.

Much of the credit for the mathematical dodges used to simplify the results obtained should go to Professor W. G. Cochran, whose patience in the teaching of the elements of mathematical statistics can not be overemphasized.

Special thanks are due Professor E. R. Smith, who has rendered invaluable service in editing the manuscript.

Finally, sincere appreciation is expressed to all those who have helped with the computational parts of the thesis.
BIBLIOGRAPHY

Literature Cited


Supplementary Literature Not Cited


Morrison, J. T. Note on the correlation of time series. Phil. Mag., Ser. 7. 18:545-554. 1934.


