ON TESTS OF SIGNIFICANCE IN TIME SERIES

BY G. TINTNER

The purpose of this note is to give some tests of significance for problems connected with time series. H. Wold [1], in a recent book gives an excellent theoretical treatment of this subject (without treating, however, the important problem of the trend), but he does not give any tests of significance, [2]. These have proved extremely important in other fields, especially in biological applications. The method used in what follows may be found useful also for other problems in time series.

1. A Test of Significance for the Variances of Differences. The Variate Difference Method, [3, 4, 5], starts from the assumption that the time series w_i ($i = 1, 2 \cdots N$) consists of two additive parts: A "smooth" part m_i, the mathematical expectation of w_i, and a random part x_i, which we will assume to be normally and independently distributed with mean 0 and variance σ^2. Hence we have

$$w_i = m_i + x_i, \qquad i = 1, \cdots, N \tag{1}$$

if we have N items in our series.
We form the finite differences and get for the difference of order k

$$\Delta^k w_i = \Delta^k m_i + \Delta^k x_i. \tag{2}$$

But the smooth component or mathematical expectation can be eliminated to any desired degree by successive differencing. This would not be true of a "zig-zag" component or a periodic function with small period [6]. It will be remembered, that for instance the differences of order k of a polynomial of order k are constant and that differences of order $(k + 1)$ and higher are zero.

O. Anderson and others who worked in this field have tested the order of the difference, say k_0, beginning from where the component m_i is sufficiently eliminated, in the following way. They define the variance σ_k^2 of the k-th difference by

$$\sigma_k^2 = \sum_{i=1}^{N-k} (\Delta^k w_i)^2/((N - k)_{2k}C_k). \tag{3}$$

We note that all variances of the differences beginning from k_0 must be equal to each other, because they will contain only the component x_i, if the other component m_i has been eliminated through taking k_0 differences. O. Anderson and R. Zaykoff, [3, 7], give formulae for the standard errors of the difference between the variances of the differences k and $k + 1$. These formulae are valid only for large samples and suppose a knowledge of the true variance σ^2.

We propose a new method for testing the equality of the variances of two successive differences in order to find the order of the difference k_0 beginning from which we have

$$\sigma_{k_0}^2 = \sigma_{k_0+1}^2 = \sigma_{k_0+2}^2 = \cdots \tag{4}$$

This method is one of selection and it consists in selecting the items to be included in the variance of the k-th difference, σ_k^2, in such a way that they become independent of the items to be included in the calculation of the variance of the difference of order $k + 1$, σ_{k+1}^2. Then the ordinary test of significance, i.e. the one involved in the analysis of variance as developed by R. A. Fisher [8], becomes applicable.

Let us consider an example. Suppose we want to compare the variance of the first differences and of the second differences, in order to test the hypothesis that the component m_i has already been eliminated in the first difference. But the process of forming finite differences has introduced correlations, even if the original random elements x_i are independently distributed. Each item in the series of the first differences will be correlated with the next and the preceding item. Each item in the series of second differences will be correlated with the two preceding and the two following items of the same series. But each item of the series of the first differences will also be correlated with the two preceding, the corresponding and the following item of the series of second differences.

We can make a very simple valid comparison in spite of these correlations if we sacrifice some of the available information. We can for instance calculate $\sigma_1^{2\prime}$ by including only items number 1, 6, 11, 16 etc. of the series of first differences. And we calculate $\sigma_2^{2\prime}$ by including only items number 3, 8, 13, 18 etc. of the series of second differences. The two quantities $\sigma_1^{2\prime}$ and $\sigma_2^{2\prime}$ are independent and hence can be compared by using either Fisher's z test, [8], or Snedecor's F table [9]. The variances are

$$\sigma_1^{2\prime} = \sum{}' (\Delta^1 w_i)^2 \Big/ \left(\frac{(N-1)}{5} {}_2C_1\right) \quad \text{and} \tag{5}$$

$$\sigma_2^{2\prime} = \sum{}'' (\Delta^2 w_i)^2 \Big/ \left(\frac{(N-2)}{5} {}_4C_2\right) \tag{6}$$

where $\sum'$ and $\sum''$ denote summation over the selected items. Other selections which are possible are: Items number 2, 7, 12 etc. of the series of first differences and items number 4, 9, 14 etc. of the series of second differences. Or items number 3, 8, 13 of the series of the first differences and items number 5, 10, 15 etc. of the series of second differences. Or items number 4, 9, 14 etc. of the series of first differences and items number 6, 11, 16 etc. of the series of second differences. Finally, items number 5, 10, 15 etc. of the series of first differences and items number 7, 12, 17 etc. of the series of second differences. These 5 selections are of course not independent of each other. The comparison can always be made by calculating the variances according to formulae (5) and (6) and using either Fisher's z table, [8], or Snedecor's F table, [9], for $(N - 1)/5$ and $(N - 2)/5$ degrees of freedom. If N is large enough, these two numbers will be near enough together in order to use the property of the z distribution to become normal for equal degrees of freedom, [8]. Then we can assume that $z = (\log_e \sigma_1^{2'} - \log_e \sigma_2^{2'})/2$ is normally distributed with mean zero and standard error $\sqrt{5/(N - 2)}$.

Should the test turn out positive, i.e. if the difference between the variances is greater than permitted from the point of view of certain significance levels, then we have to compare the variance of the second and the third differences, by selecting items in a similar manner and so on.

The general procedure is as follows: If we want to compare the variance of the difference number k and the difference number $k + 1$, we find that we can only use a part of our available series, because we must make a selection in order to get two independent estimates. We can make $2k + 3$ different selections, which are not independent but each give two unbiased, independent estimates of the variances of the differences k and $k + 1$. The selections consist in taking items number $j, j + (2k + 3), j + 2(2k + 3), j + 3(2k + 3)$ etc. of the series of k-th differences and items number $j + k + 1, j + k + 1 + (2k + 3), j + k + 1 + 2(2k + 3), j + k + 1 + 3(2k + 3)$ etc. of the series of $(k + 1)$ difference. j is here equal to $1, 2, 3 \cdots 2k + 3$, giving $2k + 3$ possible selections for the comparison.

The variances of the difference number k and $k + 1$ are calculated according to the formulae

$$\sigma_k^{2'} = \sum{}'(\Delta^k w_i)^2 \Big/ \left(\frac{(N - k)}{2k + 3}\, {}_{2k}C_k\right) \tag{7}$$

$$\sigma_{k+1}^{2'} = \sum{}''(\Delta^{k+1} w_i)^2 \Big/ \left(\frac{(N - k - 1)}{2k + 3}\, {}_{2k+2}C_{k+1}\right). \tag{8}$$

The summations are again taken over the selected items and we can make an ordinary analysis of variance with Fisher's z table or Snedecor's F table entering it for $(N - k)/(2k + 3)$ and $(N - k - 1)/(2k + 3)$ degrees of freedom. If N is appreciably large, we can assume the number of degrees of freedom as equal and $z = (\log_e \sigma_k^{2'} - \log_e \sigma_{k+1}^{2'})/2$ is normally distributed about zero with a standard error of $\sqrt{(2k + 3)/(N - k - 1)}$.

2. The Distribution of the Serial Co-variance. A similar method yields the distribution of the serial covariance, i.e., the product of a random series with itself if lagged by a lag L. We assume that x_i, $i = 1, \cdots, N$, is a series of N terms which are normally and independently distributed with mean zero and variance one. We form the serial covariance w by lagging it by L terms and make a selection. We only include the products number 1, $1 + (L + 1)$, $1 + 2(L + 1)$, $1 + 3(L + 1)$ etc. The following formulae are exact only if N is a multiple of $L + 1$, otherwise they have to be regarded as approximations. The serial covariance w is

$$w = (x_1x_{L+1} + x_{L+2}x_{2L+2} + x_{2L+3}x_{3L+3} + \cdots) \Big/ \left(\frac{N}{L+1}\right) \tag{9}$$

We shall use the method of characteristic functions, [10], in order to establish the distribution of w. The characteristic function is in our case

$$E(e^{iwy}) = g(y) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} e^{iwy} f \, dx_1 \cdots dx_N \tag{10}$$

where f is the distribution function of the x_i, i.e., a distribution of N normal and independent variates with zero means and unit variances.

An orthogonal transformation of the quadratic form in the exponent yields a determinant, which consists of $\frac{N}{L+1}$ steps each of the form

$$\begin{vmatrix} 1 & 0 & \cdots & -i(L+1)y/N \\ 0 & 1 & & 0 \\ \vdots & \vdots & & \vdots \\ -i(L+1)y/N & 0 & & 1 \end{vmatrix} = 1 + \frac{(L+1)^2y^2}{N^2}. \tag{11}$$

The characteristic function is therefore given by

$$g(y) = \left[1 + \frac{(L+1)^2y^2}{N^2}\right]^{-\frac{1}{2}N(L+1)} \tag{12}$$

and the distribution of w, say $D(w)$, is given, [11], by

$$\begin{aligned} D(w) &= \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-iwy} g(y) \, dy \\ &= \frac{2^{\frac{1}{2}\left(2-\frac{N}{L+1}\right)}\left(\dfrac{N}{L+1}\right)^{\frac{1}{2}\left(1+\frac{N}{L+1}\right)}}{\sqrt{2\pi}\,\Gamma\left(\dfrac{N}{2L+2}\right)} \, w^{\frac{1}{2}\left(\frac{N}{L+1}-1\right)} K_{\frac{1}{2}\left(\frac{N}{L+1}-1\right)}\left(\frac{Nw}{L+1}\right) \end{aligned} \tag{13}$$

where K is a Bessel function of the second kind for a purely imaginary argument, [12].

We can also get from (12) an asymptotic formula for large N. In this case w is distributed normally about zero with a variance of $(L + 1)/N$.

REFERENCES

[1] H. Wold, *Analysis of Stationary Time Series*, Upsala, 1938.

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

[3] O. Anderson, "On the Logic of Decomposition of Statistical Series into Separate Components," *Journal of the Royal Statistical Society*, vol. 90, 1927, pp. 548 ff.
[4] O. Anderson, *Die Korrelationrechnung in der Konjunkturforschung*, Bonn, 1929.
[5] G. Tintner, *Prices in the Trade Cycle*, Vienna, 1935.
[6] A. Wald, *Berechnung und Ausschaltung von Saisonschwankungen*, Vienna, 1936, pp. 37 ff.
[7] R. Zaykoff, "Ueber die Ausschaltung der zufaellingen Komponente nach der Variate Difference Methods," *Publications of the Statistical Institute for Economic Research*, State University of Sofia, 1937, no. 1, pp. 75 ff.
[8] R. A. Fisher, *Statistical Methods for Research Workers*, 4th ed., London, 1932, pp. 202 ff.
[9] G. W. Snedecor, *Statistical Methods*, Ames, Iowa, 1938, pp. 184 ff.
[10] S. S. Wilks, *Statistical Inference*, Princeton, 1937, pp. 34 ff.
[11] G. N. Watson, *The Theory of Bessel Functions*, Cambridge, 1922, p. 172.
[12] M. S. Bartlett and J. Wishart, "The Distribution of Second Order Moment Statistics in a Normal System," *Proceedings Cambridge Philosophical Society*, vol. 28, 1931, pp. 455 ff.

Iowa State College,
Ames, Iowa.