The Annals of Statistics
1977, Vol. 5, No. 2, 408-413

A TEST FOR SERIAL CORRELATION IN MULTIVARIATE DATA[1]

BY WALTER S. LIGGETT, JR.

Tennessee Valley Authority

Consider a sample from a multiple time series that is stationary and Gaussian. A test is presented for independence among the multivariate observations that comprise this sample. The test is a generalization of the Kolmogorov–Smirnov test for serial correlation in a single time series. In the test, pairs of spectral-matrix estimates are compared using the largest-root statistic. The comparisons, which are tested simultaneously, are between estimates obtained from upper and lower parts of the frequency band. Under the null hypothesis, the joint distribution of the largest roots is obtained in a form suitable for computation of significance levels.

1. Introduction. A common assumption in multivariate analysis is that the data (or the residuals after removing the mean) are a sequence of random vectors that are independent and normally distributed with zero means and identical but unknown covariance matrices. In the test presented here, this is the null hypothesis. The alternative hypothesis for which this test is designed is that the data are a stationary Gaussian multiple time series in which observations at different times are correlated. The usual approach to detecting serial correlation is graphical presentation of either spectral-matrix estimates or covariance function estimates. For multiple time series, this approach has several difficulties including the need for several different graphs including ones that show dependence among the component series and the choice of spectral resolution. Thus, a way to test simultaneously for all the types of serial correlation that a stationary Gaussian time series can exhibit is needed.

A univariate time series can be tested for serial correlation by computing the periodogram, dividing the frequency band into two parts, and comparing the sum of the periodogram over the lower part with the sum over the whole band. The sums are compared for all divisions of the frequency band using Kolmogorov–Smirnov limits (Bartlett (1966), Durbin (1969)). The test presented here is a generalization of this. As with the univariate case, the multidimensional periodogram is computed, the frequency band is divided, and the sum over the lower part is compared with the sum over the whole band. These estimates are compared using the largest and smallest eigenvalues of one estimate with respect to the other. Some but not all divisions of the frequency band are considered simultaneously. As shown in Section 3, the computation of the distribution under the null hypothesis also has similarities to Kolmogorov–Smirnov tests.

Received July 1975; revised May 1976.

[1] This research was supported at Raytheon, Submarine Signal Division, by the Office of Naval Research under Contract N00014-72-C-0314 (NR 042-287).

AMS 1970 *subject classifications.* Primary 62M15; Secondary 62H15.

Key words and phrases. Multiple time series, spectral analysis, complex multivariate analysis.

Beyond the question of whether serial correlation exists is the problem of estimating the form of this dependence. Part of this problem is deciding how much smoothing to use. A comparison of spectral matrices like the one described here is used for this purpose in Liggett (1973a, b).

In the following three sections, we specify the test, derive its distribution under the null hypothesis, and discuss ways to choose divisions of the frequency band and a critical region so that an effective test with an easily computed significance level results.

2. The test. The test presented here results from the decision to transform the data into the frequency domain, the choice of spectral estimates to compare, the decision to use the largest-root statistic, and the choice of a critical region.

The data, which are denoted by X_t, $t = 1, 2, \cdots, T$, are a sample from a real, p-dimensional random sequence. The finite Fourier transform of this sequence is

$$\xi_f = (2\pi T)^{-\frac{1}{2}} \sum_{t=1}^{T} X_t e^{-i2\pi ft/T}, \qquad f = 0, 1, \cdots, [T/2]. \tag{1}$$

The rank-one matrices $\xi_f \xi_f^*$ (where * denotes conjugate transpose) are a multidimensional periodogram with frequency indexed by f. Under the null hypothesis, the random vectors ξ_f are independent, and for $0 < f < T/2$, ξ_f is complex-Gaussian distributed with zero mean and spectral matrix that does not depend on f (Goodman (1963), Hannan (1970)). The test is based on ξ_f for $1 \leqq f \leqq N_0$, where $N_0 = [(T-1)/2]$, and thus excludes ξ_0 and when T is even, $\xi_{T/2}$. These excluded coefficients are real, not complex. ξ_0 is proportional to the usual estimate of the mean.

Under the alternative hypothesis, ξ_f for $f = 1, \cdots N_0$ are approximately independent and complex-Gaussian distributed. Wahba (1968) and Hannan (1970) provide asymptotic justification for these distributional properties. Thus, the two hypotheses are distinguished by whether or not the spectral matrix varies with frequency. For this reason, the test proposed is based on comparisons among estimates of the spectral matrix at different frequencies.

The type of frequency variation to which the test is most sensitive is determined by which estimates are compared. Let $0 < N_m < N_{m-1} < \cdots < N_0$ and let

$$A_k = \sum_{f=1}^{N_k} \xi_f \xi_f^*, \qquad k = 0, 1, \cdots, m. \tag{2}$$

The test presented here compares A_k and $A_0 - A_k$ $(k \neq 0)$ with A_0. As noted above, this is analogous to the Kolmogorov–Smirnov test for the univariate case. Under some circumstances such as a search for periodic components, a different set of comparisons might be more appropriate.

As with the choice of frequency intervals, the choice of the method for comparing the estimates determines the alternatives to which the test is most sensitive. At least four ways to compare two spectral matrices can be adapted from tests for comparing covariance matrices (Pillai and Jayachandran (1968)). Let λ_{k1}, $\lambda_{k2}, \cdots, \lambda_{kp}$ be the eigenvalues of $A_0^{-1}A_k$ in descending order. The test presented

here is based on λ_{k1}, the largest eigenvalue of $A_0^{-1}A_k$, and $1 - \lambda_{kp}$, the largest eigenvalue of $A_0^{-1}(A_0 - A_k)$.

The test is now specified except for m, the number of simultaneous comparisons, $N_k (k = 1, \cdots, m)$, the divisions of the frequency band, and the critical region for the eigenvalues $\lambda_{11}, \lambda_{21}, \cdots, \lambda_{m1}$ and $\lambda_{1p}, \lambda_{2p}, \cdots, \lambda_{mp}$. The choice of these parameters affects not only the effectiveness of the test but also the computation of the distribution under the null hypothesis. The result derived in the next section is exact but does not hold for all choices of these parameters. Further, computation of the distribution is easier for some parameter values. Section 4 discusses the choice of these parameters.

3. Distribution under the null hypothesis. Theorems 1 and 2 give the joint density of λ_{kj} for $1 \leqq k \leqq m$ and $1 \leqq j \leqq p$ and the joint distribution of λ_{k1} and λ_{kp} for $1 \leqq k \leqq m$. Khatri (1964) obtained this result for a single comparison, $m = 1$. Both the density and the distribution are valid for only part of the range of the eigenvalues. Nevertheless, these results are appropriate in most cases since the probability that the eigenvalues will occur in this range increases to 1 as $N_k - N_{k+1} (0 \leqq k \leqq m - 1)$ and N_m increase.

THEOREM 1. *Let* $N_m, N_{k-1} - N_k \geqq p$ *for* $k = 1, \cdots, m$. *Let* $\xi_f (1 \leqq f \leqq N_0)$ *be independent and complex-Gaussian distributed with zero means and identical second moments. Let* A_k *be defined by* (2) *and let the eigenvalues of* $A_0^{-1}A_k$ *in descending order be* $\lambda_{k1}, \lambda_{k2}, \cdots, \lambda_{kp}$. *Then, when* $\lambda_{kp} \geqq \lambda_{(k+1)1}$ *for* $k = 1, \cdots, m - 1$, *the joint density of the eigenvalues is*

$$
\begin{aligned}
p(\lambda_{11}, \cdots, \lambda_{mp}) = c_m \prod_{i=1}^{p} \{\lambda_{mi}^{N_m - p}(1 - \lambda_{1i})^{N_0 - N_1 - p}\} \\
\times \prod_{k=1}^{m-1} \{\det [(\lambda_{ki} - \lambda_{(k+1)j})^{N_k - N_{k+1} - 1}]\} \\
\times \prod_{i=1}^{p-1} \prod_{j=i+1}^{p} (\lambda_{1i} - \lambda_{1j})(\lambda_{mi} - \lambda_{mj}) ,
\end{aligned} \tag{3}
$$

where

$$
\begin{aligned}
c_m = \prod_{i=1}^{p} \left\{ \frac{\Gamma(N_0 - i + 1)}{\Gamma(N_0 - N_1 - i + 1)\Gamma(N_m - i + 1)\Gamma(p - i + 1)} \right. \\
\left. \times \prod_{k=1}^{m-1} \frac{1}{\Gamma(N_k - N_{k+1})} \right\} .
\end{aligned} \tag{4}
$$

PROOF. The matrices $A_{k-1} - A_k$ for $k = 1, \cdots, m$ and A_m are independent and complex-Wishart distributed with spectral matrix $(2\pi)^{-1}EX_tX_t'$ (Hannan (1970)). Khatri (1965) gives the Jacobian of the transformation

$$
B_k = A_0^{-\frac{1}{2}} A_k A_0^{-\frac{1}{2}} , \qquad k = 1, \cdots, m . \tag{5}
$$

The joint distribution of B_k, $k = 1, \cdots, m$ which is obtained is

$$
\begin{aligned}
p(B_1, B_2, \cdots, B_m) \\
= \frac{\tilde{\Gamma}_p(N_0)(\det B_m)^{N_m - p}(\det (I - B_1))^{N_0 - N_1 - p}}{\tilde{\Gamma}_p(N_m)\tilde{\Gamma}_p(N_0 - N_1)} \\
\times \prod_{k=1}^{m-1} \left\{ \frac{(\det (B_k - B_{k+1}))^{N_k - N_{k+1} - p}}{\tilde{\Gamma}_p(N_k - N_{k+1})} \right\} ,
\end{aligned} \tag{6}
$$

where $\tilde{\Gamma}_p(N)$ is the complex multivariate gamma function,

$$\tilde{\Gamma}_p(N) = \pi^{p(p-1)/2} \prod_{i=1}^{p} \Gamma(N - i + 1) . \tag{7}$$

Using the results in Khatri (1965), make the following sequence of transformations. For $k = 1, 2, \cdots, m$, let $\Lambda_k = \operatorname{diag}(\lambda_{k1}, \lambda_{k2}, \cdots, \lambda_{kp})$ and let U_k be unitary matrices. First, transform B_1 to $U_1 \Lambda_1 U_1^*$ and then transform $U_1^* B_k U_1$ to $B_k^{(1)}$ for $k = 2, \cdots, m$. Second, transform $B_2^{(1)}$ to $U_2 \Lambda_2 U_2^*$ and then transform $U_2^* B_k^{(1)} U_2$ to $B_k^{(2)}$ for $k = 3, \cdots, m$. Continue this until $B_m^{(m-1)}$ is transformed to $U_m \Lambda_m U_m^*$. Finally, recalling that $\lambda_{kp} \geqq \lambda_{(k+1)1}$, integrate over U_k to obtain

$$\begin{aligned} p(\lambda_{11}, \cdots, \lambda_{mp}) &= c_m (\det \Lambda_m)^{N_m - p} (\det (I - \Lambda_1))^{N_0 - N_1 - p} \\ &\times \prod_{k=1}^{m} \left\{ \frac{\int_{U(p)} (\det (\Lambda_k - U \Lambda_{k+1} U^*)^{N_k - N_{k+1} - p} (dU)}{\prod_{i=1}^{p} [\Gamma(N_k - N_{k+1} - i + 1)\Gamma(p - i + 1)/\Gamma(N_k - N_{k+1})]} \right\} \\ &\times \prod_{k=1}^{m} \prod_{i=1}^{p-1} \prod_{j=i+1}^{p} (\lambda_{ki} - \lambda_{kj})^2 , \end{aligned} \tag{8}$$

where (dU) is the invariant measure on the unitary group $U(p)$ normalized to make the total measure unity (James (1964)). Note that when $\lambda_{kp} < \lambda_{(k+1)1}$ the range of integration of U is limited by the requirement that $\Lambda_k - U \Lambda_{k+1} U^*$ be positive semidefinite.

The integrals over U are given by

$$\int_{U(p)} [\det (I - AUBU^*)]^{n-p} (dU) = {}_1\tilde{F}_0(-n + p; A, B) , \tag{9}$$

where $A = \operatorname{diag}(\alpha_1, \cdots, \alpha_p)$, $B = \operatorname{diag}(\beta_1, \cdots, \beta_p)$, and ${}_1\tilde{F}_0$ is a hypergeometric function of matrix argument defined by James (1964). Khatri's result (1970, Lemma 3)

$$\begin{aligned} &(\det (\alpha_i^{p-j}) \det (\beta_i^{p-j}))\, {}_1\tilde{F}_0(-n + p; A, B) \\ &\quad = (-1)^{p(p-1)/2} \prod_{i=1}^{p} [\Gamma(p - i + 1)\Gamma(n - i + 1)/\Gamma(n)] \det ((1 - \alpha_j \beta_i)^{n-1}) \end{aligned} \tag{10}$$

completes the the proof.

THEOREM 2. *If the conditions of Theorem* 1 *hold and if* $v_1 \geqq u_1 \geqq v_2 \geqq u_2 \geqq \cdots \geqq v_m \geqq u_m$, *then*

$$\Pr\{v_k \geqq \lambda_{k1}, \cdots, \lambda_{kp} > u_k; k = 1, \cdots, m\} = c_m \det (a_{ij}) \tag{11}$$

where

$$\begin{aligned} a_{ij} = \int_{u_1}^{v_1} \int_{u_2}^{v_2} \cdots \int_{u_m}^{v_m} &(\lambda_m - \alpha_1)^{p-i} (\lambda_1 - \alpha_2)^{p-j} \lambda_m^{N_m - p} \\ &\times (1 - \lambda_1)^{N_0 - N_1 - p} \prod_{k=1}^{m-1} (\lambda_k - \lambda_{k+1})^{N_k - N_{k+1} - 1} d\lambda_m \cdots d\lambda_1 . \end{aligned} \tag{12}$$

The constants α_1 *and* α_2 *can be chosen arbitrarily.*

PROOF. The theorem follows from repeated application of Khatri's result (1969, 1970, Lemma 2)

$$\begin{aligned} \int_{u}^{v} \int_{u}^{w_1} \cdots \int_{u}^{w_{p-1}} &\det (a_j(w_i)) \det (b_j(w_i))\, dw_p \cdots dw_1 \\ &= \det \left(\int_{u}^{v} a_j(y) b_i(y)\, dy \right) . \end{aligned} \tag{13}$$

First substitute

$$\begin{aligned} a_j(w_i) &= (\lambda_{mi} - \alpha_1)^{p-j} \\ b_j(w_i) &= \lambda_{mi}^{N_m - p} (\lambda_{(m-1)j} - \lambda_{mi})^{N_{m-1} - N_m - 1} . \end{aligned} \tag{14}$$

Next substitute

$$a_j(w_i) = (\lambda_{(m-2)j} - \lambda_{(m-1)i})^{N_{m-2}-N_{m-1}-1}$$
$$b_j(w_i) = \int_{u_m}^{v_m} (\lambda_m - \alpha_1)^{p-j}\lambda_m^{N_m-p}(\lambda_{(m-1)i} - \lambda_m)^{N_{m-1}-N_m-1}\, d\lambda_m \,. \tag{15}$$

The rest follows similarly.

Equation (12) is familiar. If $\alpha_1 = 0$ and $\alpha_2 = 1$, a_{ij} is proportional to the joint distribution of order statistics from the uniform distribution (David (1970)). If $\alpha_1 = N_m/N_0$ and $\alpha_2 = N_1/N_0$, then an asymptotic approximation to a_{ij} can be derived as the asymptotic distribution of order statistics is derived (David (1970)). This gives an asymptotic distribution for the largest and smallest eigenvalues.

4. Computations. If N_0, p, and the desired significance level are regarded as given, then the parameters yet to be specified are m, N_k, u_k, and v_k for $k = 1, \cdots, m$. The critical region is given by $\lambda_{k1} > v_k$ and $\lambda_{kp} < u_k$, $k = 1, \cdots, m$. The computational methods presented by Durbin (1973) can be used to evaluate a_{ij} for any of the parameter values permitted by Theorems 1 and 2. However, this section just presents an algorithm for a special case.

The case to be considered arises from the following. First, the computation of a_{ij} is simplified if

$$u_k = v_{k+1} \qquad k = 1, \cdots, m-1 \,. \tag{16}$$

Second, since the population values of λ_{kj} are N_k/N_0, letting

$$N_k - N_{k+1} = [N_0/(m+1)]$$
$$u_k = v_{k+1} = (N_k + N_{k+1})/2N_0 \,, \qquad k = 1, \cdots, m-1 \tag{17}$$

creates a nearly symmetrical critical region.

From the remaining parameters, the significance level depends most strongly on m, the number of divisions. In fact, by the proper choice of m (and in some cases, also N_m, u_m and v_1), a significance level near the desired one can be obtained. Note that some large values of m that might be desired are prohibited by the condition $u_k \geqq v_{k+1}$ introduced to allow computation of the significance level.

The following formulas for a_{ij} apply when $v_1 = 1$ and $u_m = 0$. They are easily verified by integrating (12) by parts. Let

$$w = v_k - u_k$$
$$n = N_{k+1} - N_k \qquad \text{for} \quad k = 1, 2, \cdots, m-1$$
$$h_{rs} = (n-1)!/(n-s+r)! \tag{18}$$
$$g_{1r} = \frac{(N_0 - N_1 - j)!}{(N_0 - N_1 - j + n - r + 1)!}\left(\frac{1-u_1}{w}\right)^{N_0-N_1-j+n-r+1}$$
$$g_{2s} = \frac{(N_m - i)!}{(N_m - i + s)!}\left(\frac{v_m}{w}\right)^{N_m-i+s} \,.$$

If the $n \times n$ matrix (h_{rs}) and the n-vectors (g_{1r}) and (g_{2s}) are denoted by H, G_1

and G_2, respectively, then a_{ij} is given by

$$a_{ij} = (-1)^{p-j}(n-1)!\, w^{N_0-i-j+1} G_1 * H^{m-2} G_2 . \tag{19}$$

As an example, consider 256 bivariate observations. For this case, $p = 2$ and $N_0 = 127$, which by (17) implies that $n = 31$ and $w = 31/127$. When $m = 3$, $N_m = 33$, $u_m = 0$ and $v_1 = 1$, the significance level is 0.085.

REFERENCES

[1] BARTLETT, M. S. (1966). *An Introduction to Stochastic Processes.* Cambridge Univ. Press.

[2] DAVID, H. A. (1970). *Order Statistics.* Wiley, New York.

[3] DURBIN, J. (1969). Tests for serial correlation in regression analysis based on the periodogram of least-squares residuals. *Biometrika* **56** 1-15.

[4] DURBIN, J. (1973). *Distribution Theory for Tests Based on the Sample Distribution Function.* SIAM, Philadelphia.

[5] GOODMAN, N. R. (1963). Statistical analysis based on a certain multivariate complex Gaussian distribution (an introduction). *Ann. Math. Statist.* **34** 152-177.

[6] HANNAN, E. J. (1970). *Multiple Time Series.* Wiley, New York.

[7] JAMES, A. T. (1964). Distribution of matrix variates and latent roots derived from normal samples. *Ann. Math. Statist.* **35** 475-501.

[8] KHATRI, C. G. (1964). Distribution of the largest or the smallest characteristic root under null hypotheses concerning complex multivariate normal populations. *Ann. Math. Statist.* **35** 1807-1810.

[9] KHATRI, C. G. (1965). Classical statistical analysis based on certain multivariate complex Gaussian distribution. *Ann. Math. Statist.* **36** 98-114.

[10] KHATRI, C. G. (1969). Noncentral distributions of ith largest characteristic roots of three matrices concerning complex multivariate normal populations. *Ann. Inst. Statist. Math.* **21** 23-32.

[11] KHATRI, C. G. (1970). On the moments of traces of two matrices in three situations for complex multivariate normal populations. *Sankhyā Ser. A* **32** 65-80.

[12] LIGGETT, W. S., JR. (1973). Passive sonar: fitting models to multiple time series. In *Signal Processing* (J. W. R. Griffiths, P. L. Stocklin, and C. Van Schooneveld, eds.) 327-345. Academic Press, London.

[13] LIGGETT, W. S., JR. (1973). Determination of smoothing for spectral-matrix estimation. Raytheon Company Report, Portsmouth, R. I.

[14] PILLAI, K. C. S. and Jayachandran, K. (1968). Power comparisons of tests of equality of two covariance matrices based on four criteria. *Biometrika* **55** 335-342.

[15] WAHBA, G. (1968). On the distribution of some statistics useful in the analysis of jointly stationary time series. *Ann. Math. Statist.* **38** 1849-1862.

STATISTICIAN, DIVISION OF ENVIRONMENTAL PLANNING
TENNESSEE VALLEY AUTHORITY
CHATTANOOGA, TENNESSEE 37401