A LINEAR COMBINATION TEST FOR DETECTING SERIAL CORRELATION IN MULTIVARIATE SAMPLES

BY RICHARD A. JOHNSON[1] AND THORE LANGELAND[1]

University of Wisconsin and StatOil-Norway

If multivariate observations taken at adjacent times are correlated the quality of inferences, based on an independence assumption, can be seriously eroded. After illustrating these effects, we propose a new test for detecting dependence among adjacent observations. Our test statistic is the maximum absolute value of the lag 1 correlation obtainable from a linear combination of the observations. We express the statistic in terms of two eigenvalues and then obtain the asymptotic null distribution. Asymptotic power is examined for sequences of local alternatives in a multivariate normal autoregressive process. An explicit expression is obtained for the density of the limit distribution in the bivariate case. We then compare power with the likelihood ratio statistic.

1. Introduction. The presence of even a moderate autocorrelation, among univariate observations, can cause serious difficulties for procedures based on an assumption of independence. To illustrate, suppose normal observations are treated as independent but they actually follow a first order autoregressive (AR) model

$$X_t - \mu = \phi(X_{t-1} - \mu) + \varepsilon_t$$

where the ε_t are independent and identically distributed with mean 0 and variance σ_ε^2 and $|\phi| < 1$. It is well known that $\mathrm{corr}(X_t, X_{t-1}) = \phi$ and $\sqrt{n}(\bar{X} - \mu)/s \xrightarrow{\mathcal{L}} N(0, (1+\phi)(1-\phi)^{-1})$. The coverage of the large sample nominal 95% confidence interval $\bar{X} \pm 1.96s/\sqrt{n}$ depends rather dramatically on ϕ.

Table 1. Coverage Probability of the Interval $\bar{X} \pm 1.96s/\sqrt{n}$

ϕ	-0.3	0	0.3	0.5	0.7
Coverage probability	.992	.950	.849	.742	.590

[1]Sponsored by the Air Force Office of Scientific Research under Grant AFOSR–87–0256.
AMS 1980 subject classifications. 62H20; 62F05.
Key words and phrases. Serial correlation; test of independence; linear combination.

In the context of $\bar{X}$ charts, the value $\bar{X} = \Sigma_{t=1}^{n} X_t/n$ is plotted on a chart with control limits $2s/\sqrt{n}$ where s is based on a large number of observations. The $\bar{X}$ chart will produce an excessive number of false signals under positive correlation.

Table 2. False Signal on $\bar{X}$ Chart

ϕ	-0.3	0	0.3	0.5	0.7
$P[\bar{X}$ outside 2-sigma limits$]$	.01	.05	.14	.25	.40

Johnson and Bagshaw (1974) have shown a similar deterioration occurs for the distribution of time to signal with *CUSUM* charts.

In the multivariate setting, both inferences about the mean μ, and covariance matrix, $\boldsymbol{\Sigma}$, can be severely affected by serial correlation. Let the $k \times 1$ random vectors $\mathbf{X}_t$ follow the multivariate AR(1) model

$$\mathbf{X}_t - \boldsymbol{\mu} = \Phi(\mathbf{X}_{t-1} - \boldsymbol{\mu}) + \boldsymbol{\varepsilon}_t \tag{1}$$

where the $\boldsymbol{\varepsilon}_t$ are independent and identically distributed with $E(\boldsymbol{\varepsilon}_t) = \mathbf{0}$ and $\mathrm{Cov}(\boldsymbol{\varepsilon}_t) = \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}$ and all of the eigenvalues of Φ are between -1 and 1. Under this model $\mathrm{Cov}(\mathbf{X}_t, \mathbf{X}_{t-j}) = \Phi^j \boldsymbol{\Sigma}_{\mathbf{X}}$, where

$$\boldsymbol{\Sigma}_{\mathbf{X}} = \mathrm{Cov}(\mathbf{X}_t) = \sum_{j=0}^{\infty} \Phi^j \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}} \Phi'^j$$

As a consequence of the ergodic theorem

$$\bar{\mathbf{X}} \overset{a.s.}{\to} \boldsymbol{\mu} \text{ and } \mathbf{S} = \frac{1}{n-1}\sum_{t=1}^{n}(\mathbf{X}_t - \bar{\mathbf{X}})(\mathbf{X}_t - \bar{\mathbf{X}})' \overset{a.s.}{\to} \boldsymbol{\Sigma}_{\mathbf{X}} \tag{2}$$

Also,

$$\mathrm{Cov}(n^{-1/2}\sum_{t=1}^{n}\mathbf{X}_t) \overset{a.s.}{\to} (\mathbf{I} - \Phi)^{-1}\boldsymbol{\Sigma}_{\mathbf{X}} + \boldsymbol{\Sigma}_{\mathbf{X}}(\mathbf{I} - \Phi')^{-1} - \boldsymbol{\Sigma}_{\mathbf{X}}$$

and $\sqrt{n}(\bar{\mathbf{X}} - \boldsymbol{\mu})$ is asymptotically normal with this limiting covariance matrix. Suppose the underlying process has $\Phi = \phi\mathbf{I}_k$, $|\phi| < 1$, then the nominal 95% large sample confidence ellipsoid

$$\{\boldsymbol{\mu} : n(\bar{\mathbf{X}} - \boldsymbol{\mu})'\mathbf{S}^{-1}(\bar{\mathbf{X}} - \boldsymbol{\mu}) \le \chi_k^2(.05)\}$$

has coverage probability $P[\chi_k^2 \le (1-\phi)(1+\phi)^{-1}\chi_k^2(.05)]$.

Table 3. Coverage Probability of the Nominal 95% Confidence Ellipsoid

		ϕ			
		-0.3	0	0.3	0.5
	2	.996	.950	.801	.632
k	5	.999	.950	.690	.405
	10	1.000	.950	.547	.193
	20	1.000	.950	.341	.041

In the context of principal component analysis, suppose we wish to analyze $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}$, which is the covariance matrix for $\mathbf{X}_t$ under independence, but that the AR(1) autocorrelation structure is introduced by selecting a sampling interval that is too short. The first principal component has coefficient vector $\mathbf{e}_1$ where $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}\mathbf{e}_1 = \lambda_1\mathbf{e}_1$ and $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}\mathbf{e}_i = \lambda_i\mathbf{e}_i$ with $\lambda_1 \geq \cdots \geq \lambda_k$. If the underlying process is an AR(1) process with $\Phi = c\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}$,

$$\boldsymbol{\Sigma}_{\mathbf{X}}\mathbf{e} = \frac{\lambda^3}{\lambda^2 - c^2}\mathbf{e}$$

so, if c is just smaller than λ_k, the ordering of the eigenvalues is reversed. That is, $\mathbf{e}_k$ is incorrectly identified as the coefficient of the first principal component.

The message is clear, a series of observations need to be checked for serial correlation.

Numerous tests have been proposed for the univariate case. The most common tests for independence among a collection of vectors $\{\mathbf{X}_t\}_{t=1}^{T}$ depend on the sample cross covariance matrix of lag j

$$\text{(3)} \qquad \mathbf{C}_j = \frac{1}{T}\sum_{t=1}^{T-j}(\mathbf{X}_t - \bar{\mathbf{X}})(\mathbf{X}_{t+j} - \bar{\mathbf{X}})' \text{ for } j = 1, 2, \cdots, T-1$$

where $\bar{\mathbf{X}} = T^{-1}\boldsymbol{\Sigma}_{t=1}^{T}\mathbf{X}_t$. The likelihood ratio test is derived by considering the multivariate autoregressive process of order p

$$\text{(4)} \qquad \mathbf{X}_t = \Phi_1\mathbf{X}_{t-1} + \cdots + \Phi_p\mathbf{X}_{t-p} + \boldsymbol{\theta} + \boldsymbol{\varepsilon}_t$$

where the $\boldsymbol{\varepsilon}_t$ are independent and identically distributed normal random vectors with mean $\mathbf{0}$, variance $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}$ and the roots of $|\mathbf{I} - \boldsymbol{\Phi}_1 \cdots - \boldsymbol{\Phi}_p| = 0$ lie outside the unit circle. The likelihood ratio test of $H_0 : [\boldsymbol{\Phi}_1, \cdots, \boldsymbol{\Phi}_p] = \mathbf{0}$ leads to the statistic that is asymptotically equivalent to

$$\text{(5)} \qquad S_L = -[T - p - 1 - \frac{1}{2}(kp + k + 1)]\log\left(\frac{|\mathbf{C}_0 - \hat{\Phi}\mathbf{C}(p)\hat{\Phi}'|}{|\mathbf{C}_0|}\right)$$

where $\hat{\Phi} = [\Phi_1, \cdots, \Phi_p]$ is the solution of the Yule-Walker equations

$$\mathbf{C}_m = \sum_{j=1}^{p} \hat{\Phi}_j \mathbf{C}_{m-j} \quad \text{for } m = 1, 2, \cdots, p$$

and

$$\mathbf{C}(p) = \begin{bmatrix} \mathbf{C}_0 & \mathbf{C}_1 & \mathbf{C}_2 & \cdots & \mathbf{C}_{p-1} \\ \mathbf{C}_1' & \mathbf{C}_0 & \mathbf{C}_1 & \cdots & \mathbf{C}_{p-2} \\ \vdots & \vdots & \vdots & & \vdots \\ \mathbf{C}_{p-1}' & \mathbf{C}_{p-2}' & \mathbf{C}_{p-3}' & \cdots & \mathbf{C}_0 \end{bmatrix}.$$

The test based on S_L is seemingly the most popular multivariate test.

Chitturi (1974) proposed testing the same hypothesis using the statistic

$$T \sum_{m=1}^{p} \sum_{u=1}^{k} \sum_{v=1}^{k} \hat{r}_{uv}(m) \hat{r}_{uv}(-m) \tag{6}$$

where $\hat{r}_{uv}(m)$ is the cross autocorrelation of lag m between the u-th and v-th components of $\mathbf{X}_t$.

The extension of Quennouille's test, due to Bartlett and Rajalaksham (1953), is based on the test statistic

$$\sum_{u=1}^{p} \operatorname{tr}(\mathbf{G}_u \mathbf{G}_u')$$

where $\mathbf{G}_u = \mathbf{A}_0^{-1}(\Sigma_{j=0}^{p} \hat{\Phi}_j \mathbf{C}_{u-j}')\hat{\mathbf{B}}_0'$, $\Phi_0 = \mathbf{I}$ and $\hat{\mathbf{A}}_0$ and $\hat{\mathbf{B}}_0$ are given by

$$\hat{\mathbf{A}}_0' \hat{\mathbf{A}}_0 = (\mathbf{C}_0 - \mathbf{C}_1' \mathbf{C}_0^{-1} \mathbf{C}_1 - \cdots - \mathbf{C}_p' \mathbf{C}_0^{-1} \mathbf{C}_p)^{-1}$$
$$\hat{\mathbf{B}}_0' \hat{\mathbf{B}}_0 = (\mathbf{C}_0 - \mathbf{C}_1 \mathbf{C}_0^{-1} \mathbf{C}_1' - \cdots - \mathbf{C}_p \mathbf{C}_0^{-1} \mathbf{C}_p')^{-1}.$$

Legget (1977) proposed a multivariate extension of the Bartlett periodogram test.

While this collection of tests, generalized from the univariate case, may be adequate for testing for serial dependence, we found it useful to take an alternative approach. In the next section, we introduce a statistic that concentrates the first order serial correlation into a single linear combination.

2. A Linear Combination Test. Because first order autocorrelation is most common, it is worthwhile to develop a test for first order correlation that is both easy to apply and has a graphic interpretation. We reduce the problem to one dimension by considering linear combinations $\mathbf{a}'\mathbf{X}_t$, $t = 1, 2, \ldots, T$ and selecting $\mathbf{a}$ to maximize the lag 1 correlation

$$r_{\mathbf{a}}(1) = \frac{\Sigma_{t=1}^{T-1} \mathbf{a}'(\mathbf{X}_t - \bar{\mathbf{X}})(\mathbf{X}_{t+1} - \bar{\mathbf{X}})'\mathbf{a}}{\Sigma_{t=1}^{T} [\mathbf{a}'(\mathbf{X}_t - \bar{\mathbf{X}})]^2} = \frac{\mathbf{a}'\mathbf{C}_1\mathbf{a}}{\mathbf{a}'\mathbf{C}_0\mathbf{a}} \tag{7}$$

Our test statistic is then defined as the maximum attainable lag 1 correlation,

$$R_L = \sup_{\mathbf{a}\neq\mathbf{0}} |r_{\mathbf{a}}(1)|.$$

Setting $\mathbf{C}_s = 2^{-1}(\mathbf{C}_1+\mathbf{C}_1')$, $r_{\mathbf{a}}(1)$ can be expressed in terms of symmetric matrices as

$$R_L = \sup_{\mathbf{a}\neq\mathbf{o}} \frac{|\mathbf{a}'\mathbf{C}_s\mathbf{a}|}{\mathbf{a}'\mathbf{C}_0\mathbf{a}} = \max\{|\hat{\lambda}_1|, \hat{\lambda}_k\} \tag{8}$$

where $\hat{\lambda}_1 < \hat{\lambda}_2 < \cdots < \hat{\lambda}_k$ are the eigenvalues of $\mathbf{C}_0^{-1/2}\mathbf{C}_s\mathbf{C}_0^{-1/2}$ or $\mathbf{C}_0^{-1}\mathbf{C}_s$. One point of difficulty is that $\mathbf{C}_s$ is not necessarily non-negative definite.

Note that R_L has the properties

(i)

$$R_L > |r_i(1)|,\ r_i(1) = \frac{\Sigma_{t=1}^{T-1}(X_{ti} - \bar{X}_i)(X_{t+1,i} - \bar{X}_i)}{\Sigma_{t=1}^{T}(X_{ti} - \bar{X}_i)^2}.$$

(ii) R_L is invariant under

$$\mathbf{X}_t \to \mathbf{A}\mathbf{X}_t\mathbf{Q}$$

where $\mathbf{A}$ is non-singular and $\mathbf{Q}$ orthogonal.

A plot of $\hat{\mathbf{a}}'(\mathbf{X}_t - \bar{\mathbf{X}})$ versus $\hat{\mathbf{a}}'(\mathbf{X}_{t+1} - \bar{\mathbf{X}})$ displays the concentrated correlation estimated by R_L.

We now indicate the steps leading to the asymptotic null distribution for R_L leaving the more technical algebraic steps until Section 5. We say that the $k \times k$ matrix $\mathbf{B}$ is $N_{k^2}(0, \boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}^{-1})$ if $\mathrm{tr}(\mathbf{A}'\mathbf{B})$ is $N(0, \mathrm{tr}\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}'\boldsymbol{\Sigma}^{-1})$ for every $k \times k$ matrix $\mathbf{A}$. Mann and Wald (1943) showed that

$$T^{1/2}\mathbf{C}_0^{-1}\mathbf{C}_1 \xrightarrow{\mathcal{L}} N_{k^2}(0, \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}} \otimes \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1})$$

so $T^{1/2}\mathbf{C}_0^{-1/2}\mathbf{C}_s\mathbf{C}_0^{-1/2} \xrightarrow{\mathcal{L}} \mathbf{S}$ where, under the null hypothesis, $\mathbf{S}$ has pdf

$$f(\mathbf{S}) = \frac{1}{(2\pi)^{k(k+1)/4}} \cdot 2^{k(k-1)/4}\mathrm{etr}(-\frac{1}{2}\mathbf{S}\mathbf{S}'), \tag{9}$$

with respect to $k(k+1)/2$ dimensional Lebesgue measure.

Hsu (1939) encountered the same asymptotic distribution while studying a normal theory one-way MANOVA problem. He established that, if $\mathbf{S}$ is distributed as (9), the distribution of its eigenvalues $\lambda_1 < \cdots < \lambda_k$ has pdf

$$g(\lambda_1, \lambda_2, \ldots, \lambda_k) = [2^{k/2}\Pi_{i=1}^{k}\Gamma(i/2)]^{-1}\Pi_{i<j}^{k}(\lambda_j - \lambda_i) \cdot e^{-\Sigma_{i=1}^{k}\lambda_i^2/2}. \tag{10}$$

Since $T^{1/2}R_L$ is a continuous function of $T^{1/2}\mathbf{C}_0^{-1/2}\mathbf{C}_s\mathbf{C}_0^{-1/2}$,

$$\sqrt{T}R_L \xrightarrow{\mathcal{L}} \max(|\lambda_1|, \hat{\lambda}_k). \tag{11}$$

For $k = 2$, the limit distribution is easy to evaluate

$$P[T^{1/2}R_L < x] \to P[-x < \Lambda_1 < \Lambda_2 < x] = \sqrt{2}\int_{-x}^{x} ue^{-u^2/2}\Phi(u)du = F(x). \tag{12}$$

It is considerably more difficult to present expressions for the general case. Set

$$G_j(t) = \int_{-x}^{t} u^j e^{-u^2/2}du, \quad j = 0, 1, 2, \ldots, k, \tag{13}$$

$$G_{j,\ell}(x) = \int_{-x}^{x} G_j(t)t^{\ell}e^{-t^2/2}dt, \quad 0 < j,\ \ell < k \tag{14}$$

where it can be shown (see Mehta (1960), p. 399, eqn. (13))

$$G_{j,\ell}(x) = (-1)^{\ell+j}G_{\ell,j}(x). \tag{15}$$

In Section 5, we establish

THEOREM 2.1. *For k even, the asymptotic cdf of the LCT statistic $T^{1/2}R_L$, under the null hypothesis of independence, is*

$$F(x) = (\Pi_{i=1}^{k}\Gamma(i/2))^{-1}\det(\{G_{j,\ell}(x)\})$$

for $j = 0, 2, 4, \ldots, k-2$ and $\ell = 1, 3, 5, \ldots, k-1$, where $G_{j,\ell}(x)$ is defined in (14).

THEOREM 2.2. *For k odd, the asymptotic cdf of the LCT statistic $T^{1/2}R_L$, under the null hypothesis of independence, is*

$$F(x) = [2^{1/2}\Pi_{i=1}^{k}\Gamma(i/2)]^{-1}\Sigma_{j=0}^{(k-1)/2}(-1)^{(k-1)/2+j}G_{2j}(x)\det(\mathbf{B}_j)$$

where $G_j(x)$ is defined in (13),

$$\mathbf{B}_j = \begin{bmatrix} G_{0,1}(x) & G_{0,3}(x) & \cdots & G_{0,k-2}(x) \\ G_{2,1}(x) & G_{2,3}(x) & \cdots & G_{2,k-2}(x) \\ \vdots & \vdots & & \vdots \\ G_{2j-2,1}(x) & G_{2j-2,3}(x) & \ldots & G_{2j-2,k-2}(x) \\ G_{2j+2,1}(x) & G_{2j+2,3}(x) & \ldots & G_{2j+2,k-2}(x) \\ \vdots & \vdots & & \vdots \\ G_{k-1,1}(x) & G_{k-1,3}(x) & \ldots & G_{k-1,k-2}(x) \end{bmatrix}$$

for $j = 0, 1, 2, \ldots, (k-1)/2$, and $G_{j,\ell}(x)$ is defined in (14).

A table of 1-st, 5-th, and 10-th percentiles, for $k = 2(1)20$ were calculated using double precision arithmetic (see Langeland (1980)).

3. Some Competing Tests and Power Considerations. Most tests for independence are motivated from consideration of autoregressive alternatives. Let

$$\mathbf{X}_t - \boldsymbol{\mu} = \boldsymbol{\Phi}(\mathbf{X}_{t-1} - \boldsymbol{\mu}) + \boldsymbol{\varepsilon}_t$$

for $t = 1, 2, \ldots, T$. The hypothesis of independence is then

$$H : \boldsymbol{\Phi} = \mathbf{0}. \tag{16}$$

A natural test statistic to use is

$$S_L = -[N - \frac{1}{2}(k + k + 1)] \log[|\mathbf{C}_0 - \hat{\boldsymbol{\Phi}}\mathbf{C}_0\hat{\boldsymbol{\Phi}}'|/|\mathbf{C}_0|) \tag{17}$$

where $N = T - 1 - 1$ and $\hat{\boldsymbol{\Phi}} = \mathbf{C}_1\mathbf{C}_0^{-1}$. If the $\{\boldsymbol{\varepsilon}_t\}_{t=1}^{T}$ are i.i.d. multivariate normal, then the test statistic in (17) has the same asymptotic distribution as the logarithm of the likelihood ratio test statistic. See Hannan (1970, pp. 338–341).

THEOREM 3.1. *Under the null hypothesis of independence (16), the asymptotic distribution of the test statistic (17) is a $\chi^2_{k^2}$-distribution.*

In order to obtain an indication of asymptotic power, we introduce the normal theory AR(1) model (16) where the $\boldsymbol{\varepsilon}_t$ are independent $N(\mathbf{0}, \boldsymbol{\Sigma}_{\varepsilon})$. Let $\{\boldsymbol{\Phi}_T\}$ be a sequence of alternatives to independence, where $T^{1/2}\boldsymbol{\Phi}_T \to \mathbf{H}$, and let $P_{T,\boldsymbol{\Phi}_T}$ denote the distribution of $\mathbf{X}_1, \ldots, \mathbf{X}_T$. Let P_T be the distribution of $\mathbf{X}_1, \ldots, \mathbf{X}_T$ under independence.

THEOREM 3.2. *Under $\{P_T\}$*

$$\Lambda_T = \ln \frac{dP_{T,\boldsymbol{\Phi}_T}}{dP_T} = tr[\boldsymbol{\Sigma}_{\varepsilon}^{-1}T^{1/2}\boldsymbol{\Phi}_T T^{1/2}\mathbf{C}_1] - \frac{1}{2}tr[\boldsymbol{\Sigma}_{\varepsilon}T^{1/2}\boldsymbol{\Phi}_T\mathbf{C}_0T^{1/2}\boldsymbol{\Phi}_T'] + o_{P_n}(1)$$
$$\xrightarrow{\mathcal{L}} N(-\frac{1}{2}\sigma^2, \sigma^2)$$

so $\{P_T\}$ and $\{P_{T,\boldsymbol{\Phi}_T}\}$ are contiguous.

It can then be shown that $(\Lambda_T, T^{1/2}\mathbf{C}_0^{-1/2}\mathbf{C}_s\mathbf{C}_0^{-1/2})$ is asymptotically normal under P_T so that we can obtain the limiting distribution of the linear combination statistic, R_L, under $\{P_{T,\boldsymbol{\Phi}_T}\}$. Even the bivariate case is complicated. The limit distribution for $T^{1/2}R_L$ is

$$f(x) = 4e^{-(\lambda+\eta)/2} \sum_{j=0}^{\infty} \frac{(\lambda/2)^j}{(j!)^2} \sum_{i=0}^{\infty} \frac{(\eta/2)^i}{i!} \frac{x^{2(j+i+1)}}{\Gamma(\frac{2i+1}{2})}$$
$$\cdot \int_0^1 (1-u)^{2j+1} u^{2i} e^{-x^2[u^2+(1-u)^2]} du$$

for $x > 0$, where $\eta = (\mu_1 + \mu_3)^2/2$, $\lambda = [(\mu_1 - \mu_3)^2 + 4\mu_2^2]/2$ and

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1/2 & 1/2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} (\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{1/2} \otimes \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1/2})\mathrm{vec}(\boldsymbol{\Phi}).$$

It also follows directly that (Λ_T, S_L) are each jointly normal under $\{P_T\}$. From the contiguity, we then obtain

THEOREM 3.3. *Under $\{P_{T,\boldsymbol{\Phi}_T}\}$, the asymptotic distribution of S_L is noncentral $\chi^2_{k^2}$ with noncentrality parameter $tr[\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}\mathbf{H}\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}\mathbf{H}']$.*

It is well-known that the likelihood ratio test has several large sample optimal properties. However, a calculation of asymptotic power in Table 3.1 with $k = 2$, $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}} = \mathbf{I}$ shows that the linear combination test has higher power than the others when $T^{1/2}\boldsymbol{\Phi}_T \to \mathrm{diag}(h_{11}, 0)$. In the other cases considered, where $\mathbf{H}$ is of full rank or $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}$ is not proportional to $\mathbf{I}$, the likelihood ratio test has higher power. The superiority of the likelihood ratio test prevailed in a number of other cases that are not given in Table 3.1.

4. Example. We consider some data reported by Simon (see Duncan (1959), pp. 626–630) consisting of burning times of 30 fuses as recorded by three observers. Since there is one missing observation for the second observer, we first confine ourselves to the data given by observers one and three. Let $\mathbf{X}_t = (X_{t,1}, X_{t,2})'$, $t = 1, 2, \ldots, 30$ denote the observations. The plot of $X_{t,i}$ versus $X_{t+1,i}$ for $i = 1$ is given below in Figure 4.1. The plot for $i = 2$ is similar. Neither exhibits clear signs of first order serial dependence. The LCT statistic $\sqrt{30}R_L = 2.40$ and it is significant at the 10 percent level. The value of the corresponding eigenvector is $\hat{\mathbf{a}} = (1.0, -.99)'$. The plot of $\hat{\mathbf{a}}'\mathbf{X}_t$ versus $\hat{\mathbf{a}}'\mathbf{X}_{t+1}$ given in Figure 4.2 gives an indication of serial dependence in the two series of data. If the missing observation is estimated, the evidence for dependence with three observers is much stronger. The statistic becomes significant at the 3% level.

5. Derivation of Limiting Null Distribution. The asymptotic cdf of $T^{1/2}R_L$ is given by

$$F(x) = P[-x < \Lambda_1 < \Lambda_k < x] = \int_{Q(-x,x)} \cdots \int g(\lambda_1, \ldots, \lambda_k)d\lambda_1 \ldots d\lambda_k \tag{18}$$

where $g(\cdot)$ is defined in (10) and $Q(a,b) = \{a < \lambda_1 < \lambda_2 \cdots < \lambda_k < b\}$. Since

$$\Pi_{1<i<j<k}(\lambda_j - \lambda_i) = \det \begin{bmatrix} 1 & 1 & \ldots & 1 \\ \lambda_1 & \lambda_2 & \ldots & \lambda_k \\ \vdots & \vdots & & \vdots \\ \lambda_1^{k-1} & \lambda_2^{k-1} & \ldots & \lambda_k^{k-1} \end{bmatrix}$$

Table 3.1
Asymptotic Power

Σ_ε	$\mathbf{H}$	R_L	S_L
$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 0.1 & 0 \\ 0 & 0 \end{bmatrix}$	.0513	0.505
$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 0.5 & 0 \\ 0 & 0 \end{bmatrix}$	.0849	.0627
$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$	.1769	.1055
$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}$	.4666	.3201
$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 3 & 0 \\ 0 & 0 \end{bmatrix}$	.7714	.6635
$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 5 & 0 \\ 0 & 0 \end{bmatrix}$	.9952	.9894
$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 0.4 & -0.2 \\ 0.2 & 0.4 \end{bmatrix}$	.0731	.0707
$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 2 & -1 \\ 1 & 2 \end{bmatrix}$	.6338	.7160
$\begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$	$\begin{bmatrix} 2 & -1 \\ 1 & 2 \end{bmatrix}$	.6890	.7763
$\begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$	$\begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}$	.0834	.0956

Figure 4.1
PLOT OF DATA OF OBSERVER ONE VERSUS
THESE DATA LAGGED ONE UNIT

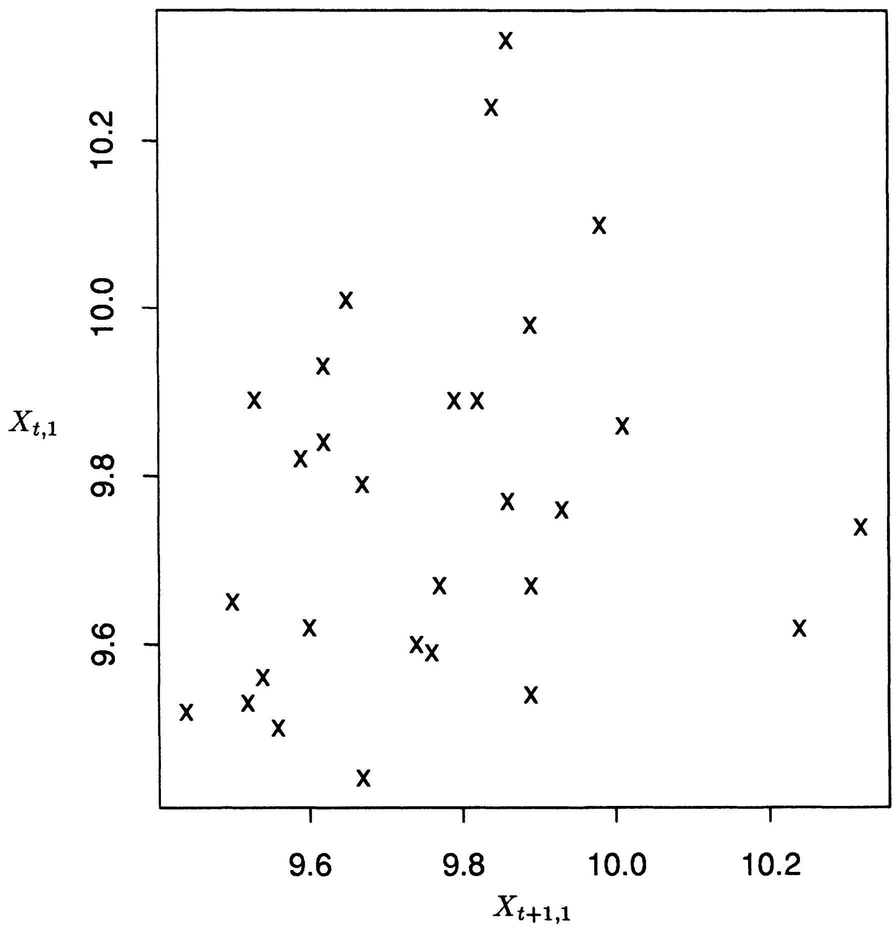
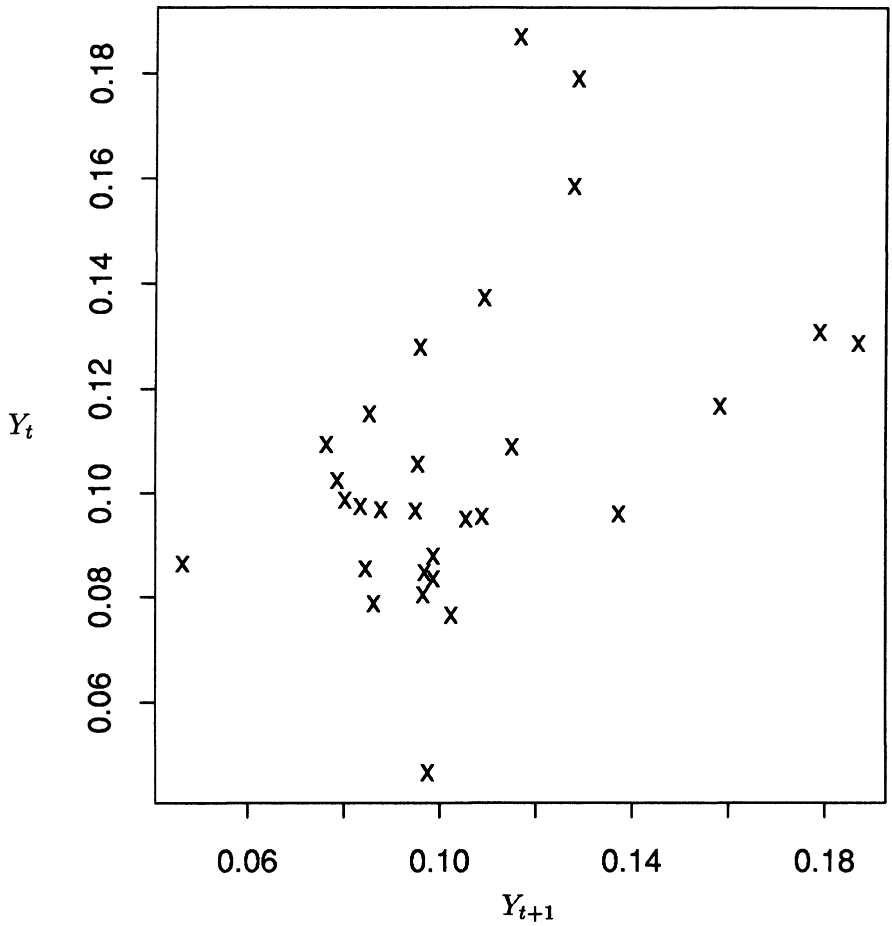


Figure 4.2
PLOT OF $\hat{\mathbf{a}}'\mathbf{X}_t$ VERSUS $\hat{\mathbf{a}}'\mathbf{X}_{t+1}$ FOR DATA FOR OBSERVERS ONE AND THREE ($Y = \hat{\mathbf{a}}'\mathbf{X}$).



(the Vandermonde determinant), (18) can be rewritten as

$$\int_{Q(-x,x)} \cdots \int c_k \det \begin{bmatrix} e^{-\lambda_1^2/2} & e^{-\lambda_2^2/2} & \dots & e^{-\lambda_k^2/2} \\ \lambda_1 e^{-\lambda_1^2/2} & \lambda_2 e^{-\lambda_2^2/2} & \dots & \lambda_k e^{-\lambda_k^2/2} \\ \vdots & \vdots & & \vdots \\ \lambda_1^{k-1} e^{-\lambda_1^2/2} & \lambda_2^{k-1} e^{-\lambda_2^2/2} & \dots & \lambda_k^{k-1} e^{-\lambda_k^2/2} \end{bmatrix}$$

$$\cdot d\lambda_1 \cdot d\lambda_2 \cdots d\lambda_k$$

where $c_k = [2^{k/2} \Sigma_{j=1}^k \Gamma(j/2)]^{-1}$.

In order to obtain an explicit expression for the densities we need some additional concepts and lemmas (see Aitken (1939), pp. 50 and 111).

The *signature function* $E(x_1, x_2, \dots, x_k)$ is defined as

$$E(x_1, x_2, \dots, x_k) = \Pi_{1<i<j<k} \operatorname{sign}(x_j - x_i) \tag{19}$$

for $x = (x_1, x_2, \dots, x_k)' \in R^k$, $E(x_1, x_2, \dots, x_k) = 0$ if $x_i = x_j$ for some $i \neq j$, $i, j = 1, 2, \dots, k$, and $E(x_1) = 1$ for all $x_1 \in R$.

Let $k = 2m$ and $m = 1, 2, \dots$, and let $\mathbf{A} = \{a_{ij}\}$ be a skew $(k \times k)$ matrix, then the *Pfaffian* of $\mathbf{A}$, $Pf(\mathbf{A})$, is defined as

$$\begin{aligned} Pf(\mathbf{A}) &= (2^m m!)^{-1} \Sigma_{j_1=1}^k \Sigma_{j_2=1}^k \cdots \Sigma_{j_k=1}^k E(j_1, j_2, \dots, j_k) \\ &\quad \cdot \; a_{j_1 j_2} \cdot a_{j_3 j_4} \cdots a_{j_{k-1} j_k} \cdot \end{aligned}$$

It is well-known that $[Pf(\mathbf{A})]^2 = \det \mathbf{A}$.

de Bruijn (1955) has established the following expression for k even.

LEMMA 5.1. *Assume* $\det(\{\phi_j(x_j)\}) \in L(R^k)$ *and let* $k = 2m$ *and* $m = 1, 2, \dots$, *then*

$$\begin{aligned} &\int_{Q(a,b)} \cdots \int \det(\{\phi_j(x_i)\}) dx_1 dx_2 \cdots dx_k \\ &= Pf(\{a_{ij} = \int_a^b \int_a^b \phi_i(x)\phi_j(y) sign(y - x) dx dy\}). \end{aligned} \tag{20}$$

REMARK. de Bruijn (1955) gives a somewhat unusual definition of the Pfaffian and his derivation of the integral on the left-hand side of (20), for k odd, is only valid in a very special case. However, Krishnaiah and Chang (1971, equation 2.6) give a general solution to the odd case. In their notation $\phi_j(x) = x^{r+j-1}\psi(x)$ for $r > 0$ and some function $\psi(x)$ satisfying the integrability conditions. We restate their results as Lemma 5.2 (an alternative proof is given in Langeland (1980)).

LEMMA 5.2. *Assume* $\det(\{\phi_j(x_i)\}) \in L(R^k)$ *and let* k *be odd, then*

$$\int_{Q(a,b)} \cdots \int \det(\{\phi_j(x_i)\}) dx_1 dx_2 \cdots dx_k = \Sigma_{j=1}^k (-1)^{j-1} \psi_j(b) Pf(\mathbf{A}_j)$$

where

$$\psi_j(b) = \int_a^b \phi_j(t) dt \;\; for \;\; j = 1, 2, \ldots, k,$$

and

$$\mathbf{A}_j = \begin{bmatrix} 0 & a_{12} & \cdots & a_{1,j-1} & a_{1,j+1} & \cdots & a_{1,k} \\ a_{21} & 0 & \cdots & a_{2,j-1} & a_{2,j+1} & \cdots & a_{2,k} \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ a_{j-1,1} & a_{j-1,2} & \cdots & 0 & a_{j-1,j+1} & \cdots & a_{j-1,k} \\ a_{j+1,1} & a_{j+1,2} & \cdots & a_{j+1,j-1} & 0 & \cdots & a_{j+1,k} \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ a_{k,1} & a_{k,2} & \cdots & a_{k,j-1} & a_{k,j+1} & \cdots & 0 \end{bmatrix}$$

for $j = 1, 2, \ldots, k$, *and* $Q(a,b)$ *and* a_{ij} *are as in Lemma 5.1.*

We can now establish Theorem 2.1.

PROOF OF THEOREM 2.1. First we notice that

$$\begin{aligned} &\int_{-x}^{x} [\int_{-x}^{x} u^j e^{-u^2/2} t^\ell e^{-t^2/2} \text{sign}(t-u) du] dt \\ &= \int_{-x}^{x} t^\ell e^{-t^2/2} [\int_{-x}^{t} u^j e^{-u^2/2} du - \int_{t}^{x} u^j e^{-u^2/2} du] dt \\ &= G_{j,\ell}(x) - \int_{-x}^{x} t^\ell e^{-t^2/2} [\int_{t}^{x} u^j e^{-u^2/2} du] dt \\ &= G_{j,\ell}(x) - \int_{-x}^{x} u^j e^{-u^2/2} [\int_{-x}^{u} t^\ell e^{-t^2/2} dt] du \\ &= G_{j,\ell}(x) - G_{\ell,j}(x) \text{ for } 0 < j,\ \ell < k - \ell. \end{aligned}$$

By (15), the last quantity equals 0 or $\pm 2G_{j,\ell}(x)$. Lemma 5.1 then gives

$$F(x) = [2^{k/2} \Pi_{j=1}^k \Gamma(j/2)]^{-1} \tag{21}$$

$$\cdot Pf \begin{bmatrix} 0 & 2G_{0,1}(x) & 0 & \ldots & 2G_{0,k-1}(x) \\ 2G_{1,0}(x) & 0 & 2G_{1,2}(x) & \ldots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 2G_{k-2,1}(x) & 0 & \ldots & 2G_{k-2,k-1}(x) \\ 2G_{k-1,0}(x) & 0 & 2G_{k-1,2}(x) & \ldots & 0 \end{bmatrix}$$

Let $k = 2m$, then, according to definition of the Pfaffian and the relation for signature functions

$$\begin{aligned} E(x_1, x_2, \ldots, x_k) &= (2^m m!)^{-1} \sum_{j_1=1}^{k} \sum_{j_2=1}^{k} \cdots \sum_{j_k=1}^{k} E(j_1, j_2, \ldots, j_k) \\ &\quad \cdot E(x_{j_1} x_{j_2}) \cdot E(x_{j_3} x_{j_4}) \cdots E(x_{j_{k-1}} x_{j_k}) \end{aligned}$$

established in de Bruijn (1955), the Pfaffian in (21) can be reduced to

$$\begin{aligned} &2^m \sum_{j_1=1}^{m} \sum_{j_2=1}^{m} \cdots \sum_{j_m=1}^{m} E(j_1, j_2, \ldots, j_m) \cdot \\ &\quad \cdot G_{0,2j_1-1}(x)\, G_{2,2j_2-1}(x) \ldots G_{k-2,2j_m-1}(x). \end{aligned}$$

But this is nothing but 2^m times the determinant in Theorem 2.1. The proof is complete.

PROOF OF THEOREM 2.2.

$$F(x) = [2^{k/2} \Sigma_{i=1}^{k} \Gamma(i/2)]^{-1} \Sigma_{j=0}^{k-1} (-1)^j G_j(x) Pf(\mathbf{A}_j)$$

where $\mathbf{A}_j = \{a_{pq}\}$ is a $(k-1) \times (k-1)$ matrix with entries $a_{pq} = G_{p,q} - G_{q,p}$ for $p, q = 0, 1, \ldots, j-1, j+1, \ldots, k-1$. Next, by (13)

$$G_j(x) = 0$$

for j odd. (It can also be shown that $Pf(\mathbf{A}_j) = 0$ for j odd.) According to (15), for j even, $\mathbf{A}_j$ is

$$\mathbf{A}_j = 2^{(k-1)} \left[\begin{matrix} 0 & G_{0,1}(x) & 0 & \ldots & G_{0,j+1}(x) \\ G_{1,0}(x) & 0 & G_{1,2}(x) & \ldots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ G_{j-1,0}(x) & 0 & G_{j-1,2}(x) & \ldots & 0 \\ G_{j+1,0}(x) & 0 & G_{j+1,2}(x) & \ldots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ G_{k-2,0}(x) & 0 & G_{k-2,2}(x) & \ldots & 0 \\ 0 & G_{k-1,1}(x) & 0 & & G_{k-1,j-1}(x) \end{matrix} \right.$$

$$\left. \begin{matrix} G_{0,j+1}(x) & \ldots & G_{0,k-2}(x) & 0 \\ 0 & \ldots & 0 & G_{0,k-1}(x) \\ \vdots & & \vdots & \vdots \\ 0 & \ldots & 0 & G_{j-1,k-1}(x) \\ 0 & \ldots & 0 & G_{j+1,k-1}(x) \\ \vdots & & \vdots & \vdots \\ 0 & \ldots & 0 & G_{k-2,k-1}(x) \\ G_{k-1,j+1}(x) & \ldots & G_{k-1,k-2}(x) & 0 \end{matrix} \right]$$

All entries containing j as a first or as a second index vanish, i.e., all $G_{\ell,j}(x)$ and $G_{j,\ell}(x)$ for $\ell = 1, 3, \ldots, k-2$ vanish. The remaining number of terms $G_{p,q}(x)$, with p even, is exactly $(k-1)/2$. Thus, the Pfaffian of A_j reduces to

$$2^{(k-1)/2}\Sigma_{j_1=1}^{m}\Sigma_{j_2=1}^{m}\cdots\Sigma_{j_m=1}^{m}E(j_1,j_2,\ldots,j_m)G_{0,2j_1-1}(x)\;\cdot$$
$$\cdot\, G_{2,2j_2-1}(x)\cdots G_{j-2,2j_{(j-2)/2}-1}(x)\cdot G_{j+2,2j_{(j+2)/2}-1}(x)\cdots G_{k-1,2j_m-1}(x)$$

where $m = (k-1)/2$. Except for a possible sign change this is nothing but $2^{(k-1)/2}$ times the determinant of the matrix $B_{(j/2)}$ appearing in the statement of Theorem 2.2. By inspection, the sign is given by $(-1)^{(k-1)/2+j}$. The proof is complete.

We remark that nuclear physicists (e.g. Mehta (1967), Wigner (1967)) are interested in distributions of the eigenvalues of $\mathbf{S}$.

REFERENCES

AITKEN, A. (1939). *Determinants and Matrices.* Oliver and Boyd, Edinburg.

BARTLETT, S. and RAJALAKSHMAN, D.V. (1953). Goodness-of-fit tests for simultaneous autoregressive series. *J. Roy. Statist. Soc. B* **15** 107–124.

CHITTURI, R.V. (1974). Distribution of residuals autocorrelations in multiple autoregressive schemes. *J. Amer. Statist. Assoc.* **69** 928–934.

DUNCAN, A. (1959). *Quality Control and Industrial Statistics.* Irwin, Homewood, Ill.

DE BRUIJN, N. (1955). On some multiple integrals involving determinants. *J. Indian Math. Soc.* **19** 133–152.

HANNAN, E.J. (1970). *Multiple Time Series.* John Wiley and Sons, New York.

HSU, P.L. (1939). On the distribution of the roots of certain determental equations. *Ann. Eugen.* **9** 250–258.

JOHNSON, R.A. and BAGSHAW, M. (1974). The effect of serial correlation on the performance of cusum tests. *Technometrics* **16** 103–112.

KRISHNAIAH, P. and CHANG, T. (1971). On the exact distributions of the extreme roots of the Wishart and MANOVA matrix. *J. Multiv. Anal.* **1** 108–117.

LANGELAND, T. (1980). Tests for Dependence in Multivariate Observations. Ph.D. Thesis, University of Wisconsin.

LIGGET, W.S. (1977). A test for serial correlation in multivariate data. *Ann. Statist.* **5** 408–413.

MANN, H. and WALD, A. (1943). On the statistical treatment of linear stochastic difference equations. *Econometrika* **11** 173–220.

MEHTA, M. (1967). *Random Matrices and Statistical Theory of Energy Levels.* Academic Press, New York.

SRIVASTAVA, M.S. and KHATRI, C.G. (1979). *Introduction to Multivariate Statistics.* North Holland, New York.

WIGNER, E. (1967). Random matrices in physics. *SIAM Review* **9** 1–23.

DEPARTMENT OF STATISTICS
UNIVERSITY OF WISCONSIN
1210 W. DAYTON STREET
MADISON, WI 53706

STATOIL RESGEO FORUS
STAVANGER 4001
NORWAY