Original Version

DEPARTMENT OF STATISTICS

University of Wisconsin
Madison, Wisconsin 53706

TECHNICAL REPORT NO. 313

October 1972

ON THE DISPERSION MATRIX OF THE MAXIMUM LIKELIHOOD ESTIMATORS OF THE PARAMETERS OF A DYNAMIC MODEL

by

Bernard Viort
University of Wisconsin

Typist: Bernice R. Weitzel

This report was supported by the Air Force Office of Scientific Research under Grant AFOSR-72-2363 and by the Wisconsin Alumni Research Foundation.

ON THE DISPERSION MATRIX OF THE MAXIMUM LIKELIHOOD ESTIMATORS OF THE PARAMETERS OF A DYNAMIC MODEL.

Dealing with multi-parameters problems, a central role is played by the dispersion matrix (variance-covariance matrix) of the estimators of the parameters: this matrix is an intrinsic part of the results on linear as well as non-linear models. It is also the starting point in the theory of the design of experiments: the particular form of the dispersion matrix presented here was derived in connection with this design problem (Viort [15]), but is believed to be an interesting result per se.

I. Introduction.

The purpose of the following is to derive the dispersion matrix of the maximum likelihood estimators of the parameters of the model

$$y_t = \delta^{-1}(B)\omega(B)x_{t-b} + e_t \quad t = \ldots, -1, 0, 1, \ldots \tag{1}$$

where $\omega(B)$, $\delta(B)$ are unknown polynomials in the backwardshift operator B defined by

$$Bx_t = x_{t-1} \tag{2}$$

the noise or error $\{e_t\}$ being a known stochastic process.

The situation considered here is then a special case of the "Transfer Function Models" or "Dynamic Models" of Box & Jenkins [1], who generalized (and gave a powerful treatment of) models frequently considered in the econometric (distributed lag) and engineering practice.

It is assumed that:

A1. (Stationarity Condition). All the roots of $\omega(z) = 0$, $\delta(z) = 0$ are outside the unit circle in the complex plane $\mathbb{C}$;

A2. (Irreducibility Condition). There exists no common root z_0: $\omega(z_0) = \delta(z_0) = 0$;

A3. $\{x_t\}$ is a stationary stochastic process with known normalized spectral density[1] $f_x(\theta)$ and variance σ_x^2;

A4. $\{e_t\}$ is a stationary stochastic process with known normalized spectral density[1] $f_e(\theta)$ bounded away from 0 and variance σ_e^2;

A5. $\{x_t\}$ and $\{e_t\}$ are independent.

In addition and without loss of generality because of A3, the unknown constant b in (1) can be taken as zero.

The estimation of the ω- and δ- parameters of (1) or of the similar model

$$\delta(B)y_t = \omega(B)x_t + \tilde{e}_t \tag{3}$$

has been considered by many authors (see Box & Jenkins [1], Fishman [4],

---

[1]The terminology of 'spectral density' is used here in the wide sense of 'generalized derivative of the spectral distribution function'. (Schwartz [13]).

Griliches [6], Hannan [7], Parzen [11], Theil [14], Wahba [16] for discussion, complete references and different approaches). There are two main difficulties:

1) The first one is that, as soon as $\delta(B) \not\equiv 1$, and $\{e_t\}$ is not white noise[2], the "regressors" $y_{t-1}, \ldots$ are correlated with the errors $e_t, \ldots$ violating one fundamental assumption of the standard linear regression theory (Malinvaud [10]);

2) The second difficulty is that the problem is non-linear, namely the dispersion matrix is a function of the unknown parameters: it is then necessary to use an iterative method and, in order to start it, to have reliable preliminary estimates of the unknown polynomials $\omega(B)$ and $\delta(B)$ (degrees as well as values of the coefficients).

The complete iterative procedure of estimation is not the purpose of this report (see Box & Lucas [2] for the theory, Box & Jenkins [1] for numerical methods), and it will be assumed that the exact values of the polynomials are known

$$\begin{cases} \omega(B) = \omega_0 + \omega_1 B + \ldots + \omega_p B^p \\ \delta(B) = 1 - \delta_1 B - \ldots - \delta_q B^q \end{cases} \qquad (4)$$

[2]Another wide sense notation for i.i.d r.v. (in discrete time).

II. Derivation of the dispersion matrix when $f_X(\theta)$ is absolutely continuous with respect to the Lebesgue measure on [0, 2Π].

II. 1. The results of Grenander & Szegö [5].

The first step is the approximation of the dispersion matrix of a stationary process by a circulant matrix.

Let $f(\theta)$ be a real-valued integrable function on $[0, 2\Pi]$.

Let

$$c_n = \frac{1}{2\Pi}\int_0^{2\Pi} e^{-in\theta}\, f(\theta)\, d\theta \qquad n = \ldots, -1, 0, 1, \ldots \tag{5}$$

be the coefficients of the Fourier series expansion of $f(\theta)$. The matrix

$$T_n(f) = \{c_{\mu\nu} = c_{\mu-\nu}\ ;\ \mu,\ \nu = 1, \ldots, n\} \tag{6}$$

is called the Toeplitz matrix (of dimension n) associated with f. It is clear that, when $f(\theta)$ is the spectral density of a stationary process, $M_n(f)$ is the correlation matrix of n successive observations. [Note that, when the process is real-valued, $f(\theta) = f(2\Pi-\theta)$ and $T_n(f)$ is real].

When n is large, the Toeplitz matrices are closely related to circulant matrices

$$C_n = \{\tilde{c}_{\mu\nu}:\ \tilde{c}_{\mu-\nu} = c_{\mu'-\nu'} \quad \text{for} \quad \mu-\nu \equiv \mu'-\nu'(n);\ \mu,\nu = 1, \ldots, n\} \tag{7}$$

whose spectral decomposition is very easy to find.

The approximation of Toeplitz matrices by circulant matrices can be viewed in two intuitive ways, but its proof is more involved.

A. The first way consists in writing, using (5) and (6)

$$T_n(f) = \{\frac{1}{2\Pi}\int_0^{2\Pi} e^{-i(\mu-\nu)\theta} f(\theta)d\theta \ ; \ \mu,\nu = 1, \ldots, n\} \tag{8}$$

and, for n large, the integral can be considered as the sum[3]

$$\frac{2\Pi}{n}\sum_{k=1}^{n} e^{-i(\mu-\nu)\frac{2k\Pi}{n}} f(\frac{2k\Pi}{n}) \tag{9}$$

suggesting that:

$$T_n(f) \simeq U_n D_n U_n^* \tag{10}$$

with

$$U_n = \left\{\frac{1}{\sqrt{n}} e^{2i\Pi\frac{\mu\nu}{n}} \ ; \ \mu,\nu = 1, \ldots, n\right\} \tag{11}$$

$$D_n = \text{Diagonal } \{d_{\nu\nu} = f(\frac{2\Pi\nu}{n}) \ ; \ \nu = 1, \ldots, n\} \tag{12}$$

where U_n is not only a circulant, but also a unitary matrix (see Hannan [7] for another interesting interpretation of U_n). D_n is a circulant,

---

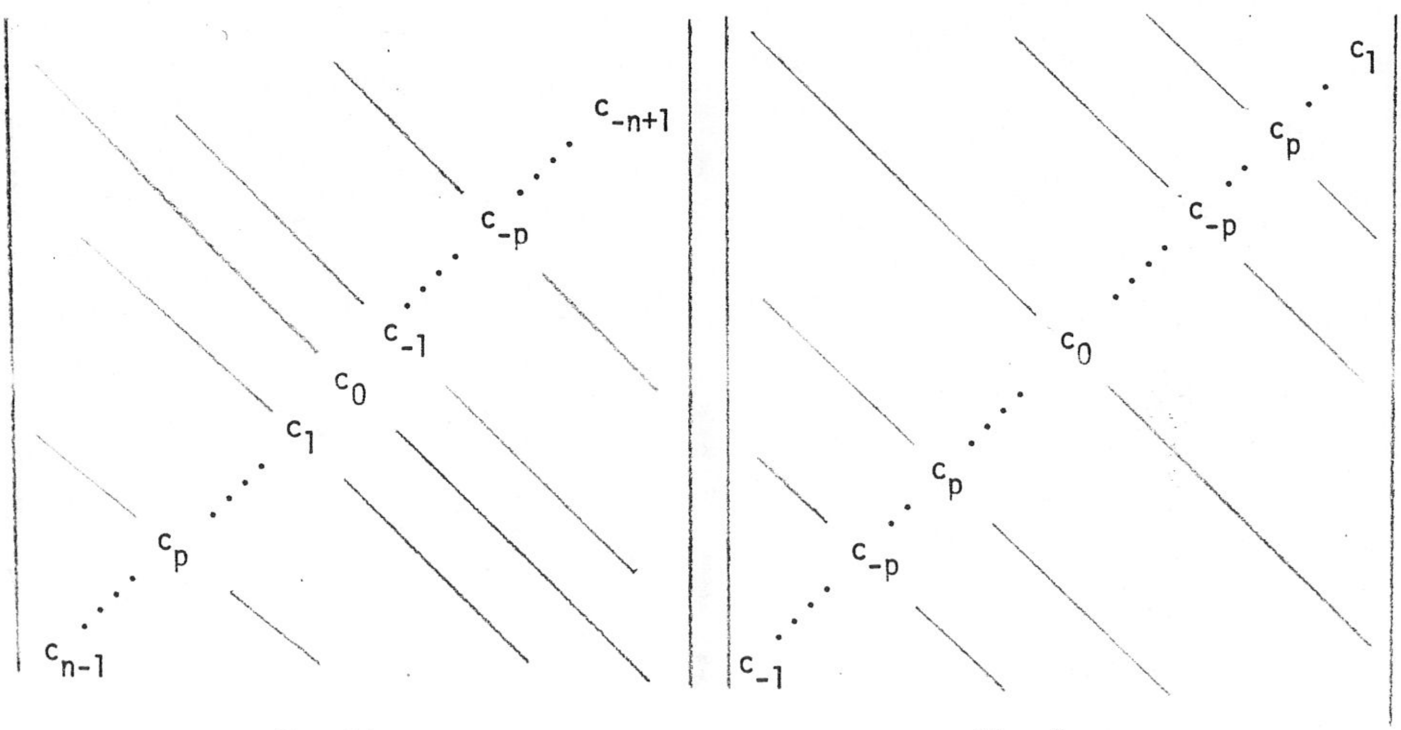
[3] If $f(\theta)$ is $\mathfrak{R}$-integrable, a sufficient condition being that $f(\theta)$ is a.e. [dx] continuous.

and the set of circulant being closed for matrix multiplication, (10) shows that $T_n(f)$ is in some sense approximated by a circulant. As n increases, it is dear that a fixed element of $U_n D_n U_n^*$ converges to the corresponding element of T_n, but it is also dear that this is not a proof since this type of convergence is not uniform on all the elements, their number being precisely n.

B. The second way is to consider the fact, that since the c_ν's are the Fourier coefficients of $f(\theta)$, one has by Parseval's inequality

$$\sum_{\nu=-\infty}^{\infty} |c_\nu| < \infty \tag{13}$$

so that the c_ν's are all negligeable for $|\nu| \geq p$, p large enough. One can consider the Toeplitz matrix and the circulant



Toeplitz   Circulant

or, if one considers only the 2^{nd} diagonal, and assuming n even $(> 2P)$:

Toeplitz:	$c_{n-1}\cdots$	$c_{n-p}\cdots$	$c_{\frac{n}{2}}\cdots$	$c_p\cdots$	c_1	c_0	$c_{-1}\cdots$	$c_{-p}\cdots$	$c_{-\frac{n}{2}}\cdots$	$c_{p-n}\cdots$	c_{1-n}
Circulant:	$c_{-1}\cdots$	$c_{-p}\cdots$	$c_{-\frac{n}{2}}\cdots$	$c_p\cdots$	c_1	c_0	$c_{-1}\cdots$	$c_{-p}\cdots$	$c_{\frac{n}{2}}\cdots$	$c_p\cdots$	c_1
Number of elements on the parallel to the first diagonal	1	p	$\frac{n}{2}$	n-p	n-1	n	n-1	n-p	$\frac{n}{2}$	p	1

and one can measure the distance of the two matrices by

$$d(T,C) = \left[\frac{1}{n} \sum_{\mu,\nu=1}^{n} |T_{\mu\nu} - C_{\mu\nu}|^2 \right]^{\frac{1}{2}} \tag{14}$$

where $(T_{\mu\nu})$ represent the Toeplitz and $(C_{\mu\nu})$ the circulant matrix.

The rigorous proof of the approximation is the justification of the intuitive feeling that $d(T,C)$ can be made less than any small fixed number for n large enough, since

$$c_{-j} = c_j \tag{15}$$

$$d^2(T,C) = \frac{2}{n} \sum_{\nu=1}^{\frac{n}{2}} \nu[c_\nu - c_{\nu-n}]^2 \tag{16}$$

$$= \frac{2}{n}\sum_{\nu=1}^{p} \nu[c_\nu - c_{\nu-n}]^2 + \frac{2}{n}\sum_{\nu=p+1}^{\frac{n}{2}} \nu[c_\nu - c_{\nu-n}]^2 \tag{17}$$

and, as will be shown below, as $n \to \infty$

$$p \text{ fixed} \qquad \frac{2}{n} \sum_{\nu=1}^{p} \nu[c_\nu - c_{n-\nu}]^2 \to 0 \tag{18}$$

while

$$\frac{2}{n} \sum_{p+1}^{\frac{n}{2}} \nu[c_\nu - c_{n-\nu}]^2 \to \varepsilon_p \tag{19}$$

where ε_p can be made arbitrarily small by choosing p large enough.

In order to prove the approximation it is useful to use a suitable approximation of $f(\theta)$. Let therefore

$$f_p(\theta) = \sum_{\nu=-p}^{p} (1 - \frac{|\nu|}{p}) c_\nu \, e^{-i\nu\theta} \tag{20}$$

and consider the diagonal matrix

$$D_n^p = \{d_{\nu\nu}^p = f_p(\frac{2\Pi\nu}{n}) \; ; \; \nu = 1, \ldots, n\} \tag{21}$$

Let then

$$T_n^p = U_n \, D_n^p \, U_n^* \tag{22}$$

and consider $d(T_n, T_n^p)$

Using the matrix K_n^p

$$K_n^p = \{(1 - \frac{|\mu-\nu|}{p}) c_{\mu-\nu} \; ; \; \mu,\nu = 1, \ldots, p+1; \; 0 \quad \text{elsewhere}\} \tag{23}$$

one has:

$$d(T_n, T_n^p) \leq d(T_n, K_n^p) + d(K_n^p, T_n^p) \tag{24}$$

The terms of the right hand side will now be considered separately:

A. $d(T_n, K_n^p)$.

$$d^2(T_n, K_n^p) \leq \frac{2}{n} \sum_{\nu=1} (p-\nu) \frac{\nu^2}{p^2} |c_\nu|^2 + \frac{2}{n} \sum_{\nu=p+1}^{n} (n-\nu) c_\nu^2 \qquad (25)$$

Since

$$0 \leq \frac{n-\nu}{n} < 1 \qquad \nu \varepsilon [p+1, n] \qquad (26)$$

and since $\sum |c_\nu|^2$ is a.c, it is possible, given $\varepsilon > 0$, to choose p such that $\sum_{\nu=p+1}^{n} |c_\nu| < \frac{\varepsilon}{2}$.

So, as $n \to \infty$

$$d^2(T_n, K_n^p) \leq \frac{kp}{n} + \varepsilon \qquad (27)$$

with

$$k = \sum_{-\infty}^{\infty} |c_\nu|^2 \qquad (28)$$

B. $d|K_n^p, T_n^p|$

$$d^2|K_n^p, T_n^p| \leq \frac{2}{n} \sum_{\nu=1}^{p} \nu(1 - \frac{|\nu|}{p})^2 |c_\nu|^2 \qquad (29)$$

$$\leq \frac{2p}{n} \sum_{\nu=1}^{p} |c_\nu|^2 \qquad (30)$$

$$\leq \frac{kp}{n} \qquad (31)$$

and now, using (27) and (31)

$$d(T_n, T_n^p) \leq \sqrt{\frac{kp}{n}} + \sqrt{\frac{kp}{n} + \varepsilon} \tag{32}$$

i.e. the Toeplitz matrix T_n can be approximated[4] to any degree of accuracy by a circulant matrix.

II. 2. The approximate inverse of $T_n(f)$.

The preceding approximation will now be used to derive the approximate inverse of a Toeplitz matrix. A sufficient condition for the existence of an inverse is that f is bounded away from zero. Assume therefore that there exists $\delta > 0$:

$$f(\theta) \geq \delta > 0 \qquad \theta \varepsilon [0, 2\Pi] \tag{33}$$

Since

$$f(\theta) = \sum_{\nu=-\infty}^{+\infty} c_\nu \, e^{-i\nu\theta} \tag{34}$$

$$f_p(\theta) = \sum_{\nu=-p}^{p} \left(1 - \frac{|\nu|}{p}\right) c_\nu \, e^{-i\nu\theta} \tag{35}$$

one has

$$|f(\theta) - f_p(\theta)|^2 \leq \sum_{\nu=-p}^{p} \frac{\nu^2}{p^2} |c_\nu|^2 + 2 \sum_{\nu=p+1}^{\infty} |c_\nu|^2 \tag{36}$$

---

[4] The approximation is valid in the distance $d(\cdot,\cdot)$: In order to use it, it is necessary to make sure that the topology is compatible with the usage considered.

Given $\varepsilon > 0$ it is possible to choose p_1, p_2 $(p_2 > p_1 > 0)$ such that

$$\frac{p_1^2}{p_2^2} \leq \frac{\varepsilon}{2k} \tag{37}$$

and

$$\sum_{\nu=p+1}^{\infty} |c_\nu|^2 \leq \frac{\varepsilon}{4} \tag{38}$$

and

$$p_2 > p \tag{39}$$

(where p, k are the same as in II. 1).

Then

$$|f(\theta) - f_{p_2}(\theta)|^2 \leq \sum_{\nu=-p_1}^{p_1} \frac{\nu^2}{p_1^2} |c_\nu|^2 + 2 \sum_{\nu=p_1+1}^{\infty} |c_\nu|^2 \leq \varepsilon \tag{40}$$

For $0 < \varepsilon < \frac{\delta}{2}$ one has simultaneously

$$f_{p_2} > \frac{\delta}{2} > 0 \tag{41}$$

and $d(T_n(f), T_n(f_{p_2})) = d(T_n(f), T_n^{p_2}(f)) \leq \varepsilon$ for n large enough (42)

Since

$$D_n^{p_2} = \text{Diagonal } \{d_{\nu\nu}^{p_2} = f_{p_2}(\tfrac{2\Pi\nu}{n}) \ ; \ \nu = 1, \ldots, n\} \tag{43}$$

is invertible, a classical theorem on linear operators (see Schwartz [12]) says that T_n is invertible and has the approximate inverse (in the distance $d(\cdot,\cdot)$)

$$T_n^{-1} \simeq U_n^*[D_n^{p_2}]^{-1}U_n \tag{44}$$

or, using the definitions and (40), for $f(\theta)$ bourded away how 0:

$$\left\{\frac{1}{2\Pi}\int_0^{2\Pi} e^{-i(\mu-\nu)\theta} f(\theta)d\theta\right\}^{-1} \simeq U_n^* [\text{Diagonal}\{\frac{1}{d_{\nu\nu}} = \frac{1}{f(\frac{2\Pi\nu}{n})}\}]U_n \tag{45}$$

II. 3. Linearization of the model (1).

Consider now the ratio $\frac{\omega(B)}{\delta(B)}$. Because of assumption A1, this ratio can be uniformly approximated on $[0,2\Pi]$ to any desired degree of accuracy by a polynomial in B

$$\frac{\omega(B)}{\delta(B)} \simeq \Pi_0 + \Pi_1 B + \dots + \Pi_r B^r \tag{46}$$

and the corresponding model is:

$$y_t = \Pi_0 x_t + \dots + \Pi_r x_{t-r} + e_t \qquad t = \dots, -1, 0, 1, \dots \tag{47}$$

The dispersion matrix of the maximum likelihood estimators of

$$\Pi_{(r)} = (\Pi_0, \dots, \Pi_r)' \tag{48}$$

is, because of assumptions A4, A5:

$$\sum_{\Pi(r)} = [X'[\sum_e]^{-1}X]^{-1} \tag{49}$$

for n observations one has:

$$X_n^r = \{x_{j-k} \ ; \ j = 1, \ldots, n \ ; \ k = 1, \ldots, r\} \tag{50}$$

and

$$\sum_e^n = \left\{ \frac{\sigma_e^2}{2\Pi} \int_0^{2\Pi} e^{-i(\mu-\nu)\theta} f_e(\theta)d\theta \ ; \ \mu,\nu = 1, \ldots, n \right\} \tag{51}$$

Using (45), $\sum_{\Pi(r)}^{-1}$ can be approximated:

$$\sum_{\Pi(r)}^{-1} \simeq X_n^{r'} U_n^* \quad \text{Diagonal} \left\{ \frac{1}{\sigma_e^2 f_e(\frac{2\Pi\nu}{n})} \right\} U_n X_n^r \tag{52}$$

and since

$$U_n X_n^r = \left\{ \frac{1}{\sqrt{n}} \sum_{\nu=1}^n e^{2i\Pi\frac{\mu\nu}{n}} x_{\nu-k} \ ; \ \mu = 1, \ldots, n \ ; \ k = 1, \ldots, r \right\} \tag{53}$$

one has:

$$\left[\sum_{\Pi(r)}^{-1} \right]_{jk} \simeq \left\{ \frac{1}{n} \sum_{\mu=1}^n \left[\sum_{\nu=1}^n e^{-2i\Pi\frac{\mu\nu}{n}} x_{\nu-j} \right] \left[\sum_{\nu'=1}^n e^{2i\Pi\frac{\mu\nu'}{n}} x_{\nu'-k} \right] \Big/ \sigma_e^2 f_e(\tfrac{2\Pi\mu}{n}) \right\} \tag{54}$$

The attention will now be focused on the quantity:

$$S_\mu = \frac{1}{n} \sum_{\nu=1}^{n} e^{-2i\Pi\frac{\mu\nu}{n}} x_{\nu-j} \sum_{\nu'=1}^{n} e^{2i\Pi\frac{\mu\nu'}{n}} x_{\nu'-k} \tag{55}$$

$$= \frac{1}{n} \sum_{\nu=1}^{n} \sum_{\nu'=1}^{n} e^{2i\Pi(\nu'-\nu)\frac{\mu}{n}} x_{\nu-j} x_{\nu'-k} \tag{56}$$

the double sum in (56) represents the sum of all the elements of a square: one can perform first the partial sums of the elements on a same parallel to the main diagonal, and then the sum over all such parallels:

$$S_\mu = \frac{1}{n} \left\{ \sum_{d=-(n-1)}^{-1} e^{2i\pi d\frac{\mu}{n}} \sum_{\nu=1}^{n-|d|} x_{\nu-j} x_{\nu+d-k} + \sum_{\nu=1}^{n} x_{\nu-j} x_{\nu-k} + \sum_{d=1}^{n-1} e^{2i\pi d\frac{\mu}{n}} \sum_{\nu'=1}^{n-|d|} x_{\nu'-d-j} x_{\nu'-k} \right\} \tag{57}$$

and, since $\{x_t\}$ is a stationary process

$$S_\mu = \frac{1}{n} \sum_{d=-(n-1)}^{n-1} e^{2i\pi d\frac{\mu}{n}} \sum_{\nu=1}^{n-|d|} x_{\nu-j} x_{\nu+d-k} \tag{58}$$

$$= \frac{1}{n} e^{-2i\pi(j-k)\frac{\mu}{n}} \sum_{d=-(n-1)+j-k}^{n-1+j-k} e^{2i\pi d\frac{\mu}{n}} \sum_{\nu=1}^{n-|d-(j-k)|} x_\nu x_{\nu+d} \tag{59}$$

and, as $n \to \infty$, since $|j-k| < r$

$$S_\mu = \frac{1}{n} e^{-2i\pi(j-k)\frac{\mu}{n}} \sum_{d=-(n-1)}^{n-1} e^{2i\pi d\frac{\mu}{n}} \sum_{\nu=1}^{n-|d|} x_\nu x_{\nu+d} \tag{60}$$

It is well known (Jenkins & Watts [8]) that, as $n \to \infty$

$$E[S_\mu] = e^{-2i\pi(j-k)\frac{\mu}{n}} \sigma_x^2 f_x(2\pi\tfrac{\mu}{n}) \tag{61}$$

but that

$$\mathrm{Var}(S_\mu) \not\to 0 \tag{62}$$

(which is the main problem in the estimation of the spectral density).

At this point, it is important to remember that, according to (54), one is not interested in the individual values of S_μ, as it is the case in the estimation of the spectral density, but in the average

$$\sum_{\mu=1}^{n} \frac{S_\mu}{\sigma_e^2 f_e(\frac{2\pi\mu}{n})} \tag{63}$$

This situation is extremely favorable since averaging is exactly what is suggested in spectral analysis in order to obtain convergent estimators [the counterpart of the reduction of the variance being a bias of the expectation and some correlation between the estimators].

In order to see how this works here, lets assume first that f_x and f_e are continuous functions of θ on $[0,2\pi]$.

Given $\varepsilon > 0$, let $\delta > 0$ be such that

$$|\theta_1-\theta_2| < \delta \Rightarrow \begin{cases} |f_x(\theta_1) - f_x(\theta_2)| < \varepsilon \\ |f_e(\theta_1) - f_e(\theta_2)| < \varepsilon \end{cases} \tag{64}$$

let N be an integer such that

$$N > \frac{1}{\delta} \tag{65}$$

and for h integer given, take

$$n = (2h+1)N \tag{66}$$

then

$$\sum_{\mu=1}^{n} \frac{S_\mu}{\sigma_e^2 f_e(\frac{2\pi\mu}{n})} = \sum_{\nu=1}^{N} \sum_{\mu=\nu(2h+1)-2h}^{\nu(2h+1)} \frac{S_\mu}{\sigma_e^2 f_e(2\pi\frac{\mu}{n})} \tag{67}$$

Now for $\mu \in [\nu(2h+1)-2h, \nu(2h+1)]$

$$(64) \Rightarrow f_e(2\pi\tfrac{\mu}{n}) \simeq f_e(2\pi \frac{\nu(2h+1)-h}{n}) \tag{68}$$

and with

$$\tilde{S}_\mu = e^{2i\pi(j-k)\frac{\mu}{n}} S_\mu \tag{69}$$

$$E[\tilde{S}_\mu] \simeq \sigma_x^2 f_x(2\pi\frac{\nu(2h+1)-h}{n}) \tag{70}$$

$$\mathrm{Var}[\tilde{S}_\mu] \simeq V_\nu \tag{71}$$

The different S_μ being independent,

$$R_\nu = \sum_{\mu=\nu(2h+1)-2h}^{\nu(2h+1)} \frac{\tilde{S}_\mu}{\sigma_e^2 f_e(2\pi\frac{\mu}{n})} \tag{72}$$

is now a random variable with expectation

$$E[R_\nu] \simeq (2h+1) \; \frac{\sigma_x^2 f_x(\frac{2\pi}{n}[\nu(2h+1)-h])}{\sigma_e^2 f_e(\frac{2\pi}{n}[\nu(2h+1)-h])} \tag{73}$$

and variance

$$\mathrm{Var}(R_\nu) = (2h+1)V_\nu \tag{74}$$

so that

$$\frac{1}{2h+1} \sum_{\mu=\nu(2h+1)-2h}^{\nu(2h+1)} \frac{\tilde{S}_\mu}{\sigma_e^2 f_e(2\pi\frac{\mu}{n})} \tag{75}$$

is a r.v. with expectation

$$\frac{\sigma_x^2}{\sigma_e^2} \; \frac{f_x(\frac{2\pi}{n}[\nu(2h+1)-h])}{f_e(\frac{2\pi}{n}[\nu(2h+1)-h])} \tag{76}$$

and variance

$$\frac{V_\nu}{2h+1} \tag{77}$$

and, for different ν's, the R_ν's are uncorrelated.

By choosing h large enough, and accordingly n large enough, it is possible to show (Jenkins & Watts [8]) that:

$$\sum_{\mu=1}^{n} \tilde{S}_\mu \Big/ \sigma_e^2 f_e(2\pi\tfrac{\mu}{n}) \xrightarrow{P} (2h+1) \sum_{\nu=1}^{N} \frac{\sigma_x^2 f_x(\frac{2\pi}{n}(\nu(2h+1)-h)}{\sigma_e^2 f_e(\frac{2\pi}{n}(\nu(2h+1)-h)} \tag{78}$$

or since all the functions are assumed to be continuous,

$$[\Sigma_{\pi(r)}^{-1}]_{jk} \xrightarrow{P} (2h+1) \sum_{\nu=1}^{N} \frac{\sigma_x^2 f_x(\frac{2\pi\nu}{N})}{\sigma_e^2 f_e(\frac{2\pi\nu}{N})} \tag{79}$$

or, using the argument of II. 1. A, since now r is finite and f_x/f_e is $\mathfrak{R}$-integrable.

$$[\Sigma_{\pi(r)}^{-1}]_{jk} \xrightarrow{P} \frac{N(2h+1)}{2\pi} \frac{\sigma_x^2}{\sigma_e^2} \int_0^{2\pi} e^{-2i\pi(j-k)\theta} \frac{f_x(\theta)}{f_e(\theta)} d\theta \tag{80}$$

or, if one is interested in the dispersion matrix "for one observation"

$$\Sigma_{\pi(r)}^{-1} = \left\{ \frac{1}{2\pi} \frac{\sigma_x^2}{\sigma_e^2} \int_0^{2\pi} e^{-2i\pi(j-k)\theta} \frac{f_x(\theta)}{f_e(\theta)} d\theta \ ; \ j,k = 1, \ldots, r \right\} \tag{81}$$

Note that, since r is now finite, it is not possible to use here the results of II. 1, II. 2 to derive the exact form of the dispersion matrix. Now it is important to remember that this result is derived with the implicit assumption that the approximation in (52) is good = this point will be verified now.

Considering (49) and (52), one has the square of the distance between the true inverse of the dispersion matrix and the approximate one

$$\begin{aligned} d_n^2(X'\Sigma_e^{-1}X,\ X'U_n^* \ \text{Diagonal}\{\frac{1}{\sigma_e^2 f_e(\frac{2\pi\mu}{n})}\} U_n X) &= d_n(-) \\ = \frac{1}{r} \sum_{j,k=1}^{r} \sum_{\mu,\nu=1}^{n} | x_{\mu-j} (\Sigma_e^{-1} - U_n^* D\{\frac{1}{\sigma_e^2 f_e(\frac{2\pi\mu}{n})}\} U_n)_{\mu\nu} \, x_{\nu-k}|^2 & \end{aligned} \tag{82}$$

$$\leq \frac{1}{r} \sum_{\mu,\nu=1}^{n} \left\{ |(\Sigma_e^{-1} - U_n^* D.\{\frac{1}{\sigma_e^2 f_e(\frac{2\pi\mu}{n})}\} U_n)_{\mu\nu}| \sum_{j,k=1}^{r} |x_{\mu-j} x_{\nu-k}| \right\}^2 \quad (83)$$

$$\leq \frac{1}{r} \sum_{\mu,\nu=1}^{n} \left\{ |(\Sigma_e^{-1} - U_n^* D.\{\frac{1}{\sigma_e^2 f_e(\frac{2\pi\mu}{n})}\} U_n)_{\mu\nu}|^2 (\sum_{j,k=1}^{r} |x_{\mu-j} x_{\nu-k}|)^2 \right\} \quad (84)$$

Now $\sum_{j,k=1}^{r} |x_{\mu-j} x_{\nu-k}|$ is bounded in probability, so that

$$d_n^2(-) = O(n) \quad \text{as} \quad n \to \infty \quad (85)$$

$$d_n(-) = O(\sqrt{n}) \quad (86)$$

and, for the dispersion matrix for one observation

$$d = \frac{d_n(-)}{n} = O(\frac{1}{\sqrt{n}}). \quad (87)$$

i.e. $d \xrightarrow{P} 0$, which is the desired result since only the convergence in probability was considered in (78), (79), (80).

The assumptions on f_x and f_e can be now weakened since it is obvious that the same demonstration, with easy modifications, holds when f_x and f_e are continuous but for a finite number of points, since it holds for each subinterval where f_x and f_e are continuous and there is a finite number of such intervals.

II. 4. The dispersion matrix for the original parameters.

The final step is to return to the original parameters.

Let
$$m = p+q+1 \tag{88}$$

be the number of parameters and

$$\underline{P} = (P_1, \ldots, P_{p+1}, P_{p+2}, \ldots, p_m)' = (\omega_0, \ldots, \omega_p, \delta_1, \ldots, \delta_q)' \tag{89}$$

represent the vector of "unknown" parameters.

In the preceding it was tacitly assumed that $r \geq m$.

The Π-parameters are non-linear functions of the P-parameters but, using now the assumption that the exact values of the $\underline{P}$-parameters are known, one has the approximation in the neighborhood of this known value

$$\begin{cases} d\Pi_0 = \dfrac{\partial\Pi_0}{\partial p_1} dp_0 + \ldots + \dfrac{\partial\Pi_0}{\partial p_m} dp_m + \varepsilon_0 \\ \ldots \\ d\Pi_r = \dfrac{\partial\Pi_r}{\partial p_1} dp_0 + \ldots + \dfrac{\partial\Pi_r}{\partial p_m} dp_m + \varepsilon_r \end{cases} \tag{90}$$

Since the dispersion matrix of $\Pi_{(r)}$ is $\sum_{\Pi_{(r)}}$, the dispersion matrix of $\underline{P}$ is immediately

$$\sum_p = \left[\left\{\frac{\partial\Pi}{\partial \underline{p}}\right\} \sum_{\Pi_{(r)}}^{-1} \left\{\frac{\partial\Pi}{\partial \underline{p}}\right\}' \right]^{-1} \tag{91}$$

with the notation

$$\left\{\frac{\partial \Pi}{\partial \underline{p}}\right\} = \left\{\frac{\partial \Pi_j}{\partial p_k} \; ; \; j = 1, \ldots, r \; ; \; k = 1, \ldots, m\right\} \tag{92}$$

Define now

$$G(\theta) = \frac{\omega(e^{-i\theta})}{\delta(e^{-i\theta})} \tag{93}$$

which, because of A1, is continuous and bounded on $[0,2\pi]$. In addition it is easy to check that all the partial derivatives $\frac{\partial G}{\partial p_j}(\theta)$ are continuous and bounded on $[0,2\pi]$ so that using Loéve [9], the derivative with respect to p_j of the expansion of G is the same as the expansion of the derivative of G with respect to p_j:

$$G(\theta) = \Pi_0 + \Pi_1 e^{-i\theta} + \ldots \tag{94}$$

$$\frac{\partial}{\partial p_j} G(\theta) = \frac{\partial \Pi_0}{\partial p_j} + \frac{\partial \Pi_1}{\partial p_j} e^{-i\theta} + \ldots \tag{95}$$

and, since only a finite number m of parameters are involved, it is easy to choose r large enough so that, to any degree of accuracy, the $m+1$ approximations are satisfied:

$$\begin{cases} G(\theta) \cong \Pi_0 + \Pi_1 \, e^{-i\theta} + \ldots + \Pi_r \, e^{-ri\theta} \\ \frac{\partial}{\partial p_j} G(\theta) \cong \frac{\partial \Pi_0}{\partial p_j} + \frac{\partial \Pi_1}{\partial P_J} e^{-i\theta} + \ldots + \frac{\partial \Pi_r}{\partial p_j} e^{-ir\theta} \quad j = 1, \ldots, m \end{cases} \tag{96}$$

Consider now the matrix product

$$M = \left\{\frac{\partial \Pi}{\partial \underline{P}}\right\} \left\{e^{-i(j-k)\theta} \;;\; j,k = 1, \ldots, r\right\} \left\{\frac{\partial \Pi}{\partial \underline{P}}\right\} \tag{97}$$

since

$$\left\{\frac{\partial \Pi}{\partial \underline{P}}\right\} \left\{e^{-i(j-k)\theta}\right\} = \left\{\frac{\partial \bar{G}}{\partial p_\mu} e^{-i(\nu-1)\theta} \;;\; \mu = 1, \ldots, m \;;\; \nu = 1, \ldots, r\right\} \tag{98}$$

one has

$$M = \left\{\frac{\partial \bar{G}}{\partial p_\mu} \frac{\partial G}{\partial p_\nu}\right\} \tag{99}$$

and finally:

$$\sum_{\underline{p}}^{-1} = \left\{\frac{1}{2\pi} \frac{\sigma_x^2}{\sigma_e^2} \int_0^{2\pi} \frac{\partial G(\theta)}{\partial p_\mu} \frac{\partial \bar{G}(\theta)}{\partial p_\nu} \frac{f_x(\theta)}{f_e(\theta)} d\theta \;;\; \mu,\nu = 1, \ldots, m\right\} \tag{100}$$

III. A Transformation.

When ω_0, which can be considered as a scale factor, is known $m = p+q$ and it is possible to find an expression of the matrix $\sum_{\underline{P}}^{-1}$ which is sometimes more convenient than (100).

Consider the matrix M defined in (99). Returning to the parameters ω and δ, it is possible to partition M:

$$M = \left[\begin{array}{c|c} M_{11} & M_{12} \\ \hline M_{21} & M_{22} \end{array}\right] \tag{101}$$

where, with the notations:

$$\begin{cases} \delta = \delta(e^{-i\theta}) \; ; \; |\delta|^2 = \delta\bar{\delta} \\ \omega = \omega(e^{-i\theta}) \; ; \; |\omega|^2 = \omega\bar{\omega} \end{cases} \tag{102}$$

$$\begin{cases} M_{11} = \left\{ \frac{1}{|\delta|^2} e^{-i(j-k)\theta} \; ; \; j,k = 1, \ldots, p \right\} \\ M_{12} = \left\{ \frac{\omega}{\delta|\delta|^2} e^{-i(j-k)\theta} \; ; \; j = 1, \ldots, p \; ; \; k = 1, \ldots, q \right\} \\ M_{22} = \left\{ \frac{|\omega|^2}{|\delta|^4} e^{-i(j-k)} \; ; \; j,k = 1, \ldots, q \right\} \end{cases} \tag{103}$$

so that

$$M = \frac{1}{|\delta|^4} \left| \begin{array}{c|c} |\delta|^2 e^{-i(j-k)\theta} & \omega\bar{\delta} e^{-i(j-k)\theta} \\ \hline \bar{\omega}\delta e^{-i(j-k)\theta} & |\omega|^2 e^{-i(j-k)\theta} \end{array} \right| \tag{104}$$

$$= \frac{1}{|\delta|^4} \tilde{M}$$

It is easy to check that each element of $\tilde{M}$ is a quadratic form in the parameters (ω,δ), and that each of the four matrices $\tilde{M}_{11}$, $\tilde{M}_{12}$, $\tilde{M}_{21}$, $\tilde{M}_{22}$ has the same elements or parallels to the "main diagonal". It turns out that:

$$\tilde{M} = \mathcal{P} \mathcal{U} \mathcal{P}' \tag{105}$$

with

$$= \begin{vmatrix} 1 & -\delta_1 & \cdots & -\delta_q & & 0 \\ 0 & \ddots & 1 & -\delta_1 & \cdots & -\delta_q \\ \omega_0 & \cdots & \omega_p & & & \\ & 0 & \ddots & & \ddots & 0 \\ & & \omega_0 & \cdots & & \omega_p \end{vmatrix} \tag{106}$$

and

$$\mathcal{M} = \{e^{-i(\mu-\nu)\theta} \;;\; \mu,\nu = 1, \ldots m\} \tag{107}$$

The proof is by direct matrix multiplication:

A. Take $m_{jk} \varepsilon \tilde{M}_{11}$

$$m_{jk} = \sum_{\mu,\nu=1}^{m} \mathcal{P}_{j\mu}\, e^{-i(\mu-\nu)\theta}\, \mathcal{P}_{k\nu} \tag{108}$$

$$= \sum_{\nu=1}^{m} [e^{-i(j-\nu)\theta} - \delta_1 e^{-i(j+1-\nu)\theta} - \ldots - \delta_q e^{-i(j+q-\nu)\theta}]\, \mathcal{P}_{k\nu} \tag{109}$$

$$= [e^{-i(j-k)\theta} - \delta_1 e^{-i(j+1-k)\theta} - \ldots - \delta_q e^{-i(j+q-k)\theta}] - \ldots$$

$$- \delta_q [e^{-i(j-k-q)\theta} - \delta_1 e^{-i(j+1-k-q)\theta} - \ldots - \delta_q e^{-i(j+q-k-q)\theta}] \tag{110}$$

$$= e^{-i(j-k)\theta}\{(1 - \delta_1 e^{-i\theta} - \dots - \delta_q e^{-iq\theta}] - \delta_1 e^{i\theta}[1 - \delta_1 e^{-i\theta} - \dots - \delta_q e^{-iq\theta}]$$

$$- \dots - \delta_q e^{iq\theta}[1 - \delta_1 e^{-i\theta} - \dots - \delta_q e^{-iq\theta}]\} \qquad (111)$$

$$= |\delta|^2 e^{-i(j-k)\theta} \qquad (112)$$

B. Take $m_{jk} \varepsilon \tilde{M}_{12}$

$$m_{jk} = \sum_{\nu=1}^{m} [e^{-i(j-\nu)\theta} - \delta_1 e^{-i(j+1-\nu)\theta} - \dots - \delta_q e^{-i(j+q-\nu)\theta}] \mathcal{P}_{k\nu} \qquad (113)$$

$$= \omega_0 [e^{-i(j-k)\theta} - \delta_1 e^{-i(j+1-k)\theta} - \dots - \delta_q e^{-i(j+q-k)\theta}] + \dots$$

$$+ \omega_p [e^{-i(j-p-k)\theta} - \delta_1 e^{-i(j+1-p-k)\theta} - \dots - \delta_q e^{-i(j+1-p-k)\theta}] \qquad (114)$$

$$= e^{-i(j-k)\theta}\{[\omega_0 [1 - \delta_1 e^{-i\theta} - \dots - \delta_q e^{-iq\theta}] + \dots$$

$$+ \omega_p e^{ip\theta}[1 - \delta_1 e^{-i\theta} - \dots - \delta_q e^{-q\theta}]\} \qquad (115)$$

$$= \bar{\omega}\delta \; e^{-i(j-k)\theta} \qquad (116)$$

For the elements of $\tilde{M}_{22}$, the computations go as for $\tilde{M}_{11}$ replacing the δ's by the ω'; and $\tilde{M}_{21} = \bar{\tilde{M}}_{12}$.

As a final result, one has:

The dispersion matrix per observation for the parameters $P = (\omega,\delta)$ of the model (1) is $\sum_{\underline{P}}$ with:

$$\sum_{\underline{P}}^{-1} = \left\{ \frac{1}{2\pi} \frac{\sigma_x^2}{\sigma_e^2} \int_0^{2\pi} e^{-i(j-k)} \frac{f_x(\theta)}{|\delta(\theta)|^4 f_e(\theta)} d\theta \right\} \tag{117}$$

IV. Generalization to the case where $f_x(\theta)$ is a generalized function.

In order to make the result (117) applicable to a wider class of situations it is necessary to weaken the conditions on $f_x(\theta)$. This is also a good occasion to show how a rigorous treatment of the special densities should proceed.

The spectral density is introduced as a consequence of the spectral representation of a stationary stochastic process continuous in quadratic mean: the central role is played by the spectral distribution function $F(\theta)$. (see Cramer & Leadbetter [3]).

$F(\theta)$ can be decomposed in three parts:

$$F = F_1 + F_2 + F_3 \tag{118}$$

where F_1 is a.c. with derivative f_1 called spectral density,

F_2 is a jump function with a finite number of jumps,

F_3 is continuous but has no derivative.

In most of the usual processes, $F = F_1 + F_2$, and it is possible to consider the generalized derivative of F as the sum of the derivative of its a.c. part and a finite linear complication of Dirac functions. The corresponding Rieman-Stieltjes integral is then the sum of the Rieman-Stieltjes integral corresponding to the derivative of the a.c. part and the linear combination of the value of the function at the jumps of F.

Now the jumps of F correspond to deterministic components in the process, so that for these components the variance in (62) is equal to zero: they can be treated separately and exactly, and II. 3 shows then that the inverse of the covariance matrix is the sum of two matrices, one (81) corresponding to the derivative of the a.c. part of the spectral distribution function, one corresponding to the deterministic part of the process, which is the linear combination of the values of the matrices

$$\left\{ \frac{1}{2\pi} \frac{\sigma_x^2}{\sigma_e^2} e^{-i(j-k)\theta} \frac{f_x(\theta)}{f_e(\theta)} \right\}$$

at the jumps of F, so that the extended definition of the integral makes (117) a general result for

- $f_x(\theta)$ generalized function with a finite number of discontinuities,
- $f_e(\theta)$ function with a finite number of discontinuities, bounded away from 0,

conditions which include a very large class of actual situations.

References.

[1] G. E. P. Box & G. M. Jenkins: Time Series Analysis: Forecasting and Control. Holden-Day, San Francisco (1970).

[2] G. E. P. Box & H. L. Lucas. Design of Experiments in Non-linear Situations. Biometrika, 46, (1959).

[3] H. Cramer & M. R. Leadbetter. Stationary and Related Stochastic Processes. J. Wiley and Sons, New York (1967).

[4] G. Fishman: Spectral Methods in Econometrics. Rand Corporation Santa Monica (1970).

[5] U. Grenander & G. Szegö: Toeplitz Forms and their Applications. University of California Press, Berkeley and Los Angeles (1958).

[6] Z. Griliches. Distributed Lag: A Survey. Econometrica, Vol. 35 (1967).

[7] E. J. Hannan. Time Series Analysis. Methuen. London (1966).

[8] G. M. Jenkins & D. Watts. Spectral Analysis and its Applications. Holden-Day, San Francisco (1969).

[9] M. Loéve. Probability Theory. Van Nostrand. Princeton (1955).

[10] E. Malinvaud. Methods Statistiques de l' Econometrie. Dunod. Paris, (1969).

[11] E. Parzen. Time Series Papers. Holden-Day, San Francisco (1969).

[12] L. Schwartz. Cours d' Analyse. Ecole Polytechnique, Paris (1963-64).

[13] L. Schwartz. Les Distributions. Hermann, Paris (1959).

[14] H. Theil. Principles of Econometrics. J. Wiley and Son, New York (1971)

[15] B. L. Viort. On the Construction of D-optimal Designs for Dynamic models. Technical Report No.314. Department of Statistics, University of Wisconsin, Madison (1972).

[16] G. Wahba. Estimation of the Coefficients in a Multidimensional Distributed Lag Model. Econometrica, Vol. 37, 3. (July 1969).

Security Classification

DOCUMENT CONTROL DATA – R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) University of Wisconsin Department of Statistics Madison, Wisconsin 53706	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3. REPORT TITLE

ON THE DISPERSION MATRIX OF THE MAXIMUM LIKELIHOOD ESTIMATORS OF THE PARAMETERS OF A DYNAMIC MODEL

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Scientific Interim

5. AUTHOR(S) (First name, middle initial, last name)

Bernard Viort

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
October 1972	28	16

8a. CONTRACT OR GRANT NO. AFOSR-72-2363 b. PROJECT NO. c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) Technical Report No. 313 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

10. DISTRIBUTION STATEMENT

This document has been approved for public release and sale; its distribution is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY Air Force Office of Scientific Research 1400 Wilson Boulevard Arlington, Virginia

13. ABSTRACT

The approximation of a Toeplitz matrix by a circulant matrix is used to derive the dispersion matrix of the parameters of a dynamic model when the error process is known and when one has an approximation of the value of the parameters.

DD FORM 1 NOV 65 1473

Security Classification

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
dispersion matrix						
dynamic model						
maximum likelihood						

Security Classification