TECHNICAL REPORT NO. 157

May 1968

A FORM OF OPTIMUM WEIGHTING SYSTEM
FOR SPECTRUM ESTIMATES

by

Henry R. Neave

(University of Nottingham
and
University of Wisconsin)

This research was supported in part by the University of Wisconsin Computing Center, the National Science Foundation, and the Science Research Council of Great Britain.
SUMMARY

A previous paper has demonstrated some weaknesses of the usual class of estimates of the spectrum of a stationary time series, where the weights are dependent solely on a "lag window generator" and choice of truncation point. To pursue this matter further, we here calculate an expression for the optimum weights according to the criterion of minimum average mean square error (although a fairly natural choice, some bad aspects of this criterion are demonstrated). This expression, which depends on the spectrum being estimated, is rather complicated, and two approximations to it are devised and examined. Finally the optimum weights in certain autoregression situations are computed and compared with the weights in the standard types of spectrum estimate.
§1. Background

Let \( \{X_t; t = \ldots -2, -1, 0, 1, 2, \ldots\} \) be a real-valued, weakly stationary, discrete stochastic process (or time series) with zero mean and covariance function:

\[
R(v) = E\left[X_t X_{t+v}\right] = R(-v). \tag{1.1}
\]

The power spectrum or spectral density \( f(\cdot) \) of the process is the Fourier transform of the covariance function, i.e.

\[
f(\omega) = \frac{1}{2\pi} \sum_{v=-\infty}^{\infty} R(v) \cos v\omega \tag{1.2}
\]

\[
R(v) = \int_{-\pi}^{\pi} f(\omega) \cos v\omega \, d\omega. \tag{1.3}
\]

The process itself has a spectral representation:

\[
X_t = \int_{-\pi}^{\pi} e^{i\omega t} \, dZ(\omega) \tag{1.4}
\]

where \( Z(\cdot) \) is a complex-valued stochastic function with uncorrelated increments, and

\[
E \left[ |dZ(\omega)|^2 \right] = f(\omega) \, d\omega \tag{1.5}
\]

(Cramer (1940)). This expresses \( X_t \) as a superposition of harmonics with random amplitudes, the expectation of whose square is proportional to the spectral density. The fundamental importance of the spectrum in the analysis of a stationary stochastic process is therefore that the tendency of the process to oscillate with a frequency \( \omega \) is reflected directly in the value of \( f(\omega) \).
The apparently obvious way to estimate the spectrum using a finite sample \( \{X_t; \ t = 1, 2, \ldots, T\} \) is to form estimates \( R_T(v) \) of the covariance and, following (1.2), use

\[
f_0(\omega) = \frac{1}{2\pi} \sum_{v=-T+1}^{T-1} R_T(v) \cos v\omega.
\] (1.6)

The limits of summation here have been written \( \pm (T-1) \) because products \( X_i X_j \) can be formed only for \( |i-j| \leq T-1 \), and therefore meaningful estimates of \( R(v) \) can be formed only for \( |v| \leq T-1 \). Unbiased estimates for \( R(v) \) are

\[
R_T(v) = \frac{1}{T-|v|} \sum_{t=1}^{T-|v|} X_t X_{t+|v|}, \quad |v| < T
\] (1.7)

but it is more convenient computationally to replace the divisor \( T-|v| \) by \( T \); this is also justifiable through mean square error and other considerations (Schaerf (1964)). The resulting estimate is the periodogram:

\[
f_T(\omega) = \frac{1}{2\pi} \sum_{v=-T+1}^{T-1} R_T(v) \cos v\omega
\] (1.8)

where \( R_T(v) \) is the sample covariance function:

\[
R_T(v) = \frac{1}{T} \sum_{t=1}^{T-|v|} X_t X_{t+|v|}, \quad |v| < T.
\] (1.9)

However \( f_T(\omega) \) is unsatisfactory as an estimate because of excessive variance; in fact it is not consistent, i.e., its variance does not tend to zero as \( T \) tends to infinity.

The almost universally adopted form of spectrum estimate is

\[
f_T^*(\omega) = \frac{1}{2\pi} \sum_{v=-T+1}^{T-1} k_T^*(v) R_T(v) \cos v\omega \quad (-\pi \leq \omega \leq \pi)
\] (1.10)
where \( k^*(\cdot) \) is a weighting function called the \textit{covariance averaging kernel} (Parzen (1961a)) or the \textit{lag window} (Blackman and Tukey (1958)). The lag window is usually formed from

\[
k_T^*(v) = k^* \left( \frac{v}{M_T} \right)
\]

where \( k^*(\cdot) \) is a \textit{lag window generator}, which is a bounded function having the properties:

\[
\begin{align}
(i) & \quad k^*(0) = 1 \\
(ii) & \quad k^*(\theta) = k^*(-\theta) \\
(iii) & \quad k^*(\theta) = 0, \quad |\theta| > 1.
\end{align}
\]

and \( M_T \) is a function of \( T \) satisfying

\[
M_T \to \infty, \quad \frac{M_T}{T} \to 0 \text{ as } T \to \infty.
\]

It is well-known (Parzen (1957), Grenander and Rosenblatt (1957)) that, under these conditions, the variance of the estimate \( f_T^*(\cdot) \) is given by

\[
\lim_{T \to \infty} \frac{T}{M_T} \operatorname{var}(f_T^*(\omega)) = 2(1 + \delta_{0,\omega} + \delta_{\pi,\omega}) \int_0^1 k^*(\theta)^2 \, d\theta
\]

where \( \delta_{a,b} \) is the Kronecker delta function, equal to 1 if \( a = b \) and 0 otherwise. It is shown further in Neave (1968a, 1968b) that in practice when \( M_T/T \) is not necessarily close to zero, a better approximation is given by

\[
\frac{T}{M_T} \operatorname{var}(f_T^*(\omega)) = 2(1 + \delta_{0,\omega} + \delta_{\pi,\omega}) \int_0^1 k^*(\theta)^2 \left(1 - \frac{M_T}{T} \theta\right) d\theta.
\]
It will be convenient in most of the subsequent work to use the unbiased sample covariance function \( \hat{R}_T(\cdot) \) in our spectrum estimates. Accordingly we define

\[
\hat{S}_T(\omega) = \frac{1}{2\pi} \sum_{\nu=-T+1}^{T-1} k_T(\nu) \hat{R}_T(\nu) \cos \nu \omega
\]

(1.16)

which differs in form from (1.10) only in the use of \( \hat{R}_T(\cdot) \) instead of \( R_T(\cdot) \).

It is apparent that

\[
\hat{S}_T^*(\cdot) \equiv \hat{S}_T(\cdot)
\]

(1.17)

if and only if the weighting functions \( k_T^*(\cdot) \) and \( k_T(\cdot) \) satisfy the relation

\[
k_T(\nu) = (1 - |\nu|) k_T^*(\nu)
\]

(1.18)

where \( \gamma = M_0 / T \). It \( k_T^*(\cdot) \) is defined by (1.11) and (1.12), \( k_T(\cdot) \) may be defined similarly by using the modified lag window generator:

\[
k(\theta) = k_1(\gamma, \theta) = (1 - |\theta|) k^*(\theta).
\]

(1.19)

Assuming consistency of the estimate, which requires that \( \gamma \to 0 \) as \( T \to \infty \), the two generators \( k^*(\cdot) \) and \( k(\cdot) \) are asymptotically identical.

Clearly there are infinitely many functions satisfying (1.12) which can be used as lag window generators. Various methods of comparing generators are known of which perhaps the most familiar is the study of their Fourier transforms, the spectral window generators. The spectrum estimate (1.10) can be expressed as a weighted or smoothed periodogram:

\[
\hat{S}_T^*(\omega) = \int_{-\pi}^{\pi} k_T^*(\lambda) \hat{S}_T^{\pi}(\omega - \lambda) d\lambda
\]

(1.20)
where the spectral window $K_T^*(\cdot)$ is the Fourier transform of the sequence 
$\{k_T^*(v)\}$:

$$K_T^*(\lambda) = \frac{1}{2\pi} \sum_{v=-T}^{T} k_T^*(v) \cos v\lambda. \quad (1.21)$$

Defining $K^*(\cdot)$ as the spectral window generator:

$$K^*(\lambda) = \frac{1}{2\pi} \int_{-1}^{1} k^*(\theta) \cos \omega \theta \, d\theta, \quad (1.22)$$

it is shown in Parzen (1963) that for large $M_T$,

$$K^*(\lambda) = M_T \cdot K^*(M_T\lambda), \quad \omega \neq 0 \quad (1.23)$$

and so it is really only necessary to study $K^*(\cdot)$ to examine how $f_T(\cdot)$ is weighted and thus to draw conclusions on the properties of the estimate $f_T^*(\omega)$. Indeed, instead of defining the class of estimates by (1.10) and stipulating the form of $k^*(\cdot)$, we can use (1.20) as the defining formula of the class and specify \{K_T^*(\cdot)\} for a particular case. The very first consistent estimates of the spectrum considered (Daniell (1946)) were defined in this way with

$$K_T^*(\lambda) = 1, \quad |\lambda| < B_T$$

$$= 0, \quad \text{otherwise} \quad (1.24)$$

where $B_T \to 0$ as $T \to \infty$.

A further method of comparing generators was presented in Neave (1968 c).

It was based on the following easily-proved result. Defining

$$S_V(\omega) = \frac{1}{2\pi} R(0) + \frac{1}{\pi} \sum_{u=1}^{V} R(u) \cos u\omega, \quad (1.25)$$
which is a truncation of the Fourier series representation of the spectrum (1.2), an unbiased estimate of \( S^*_v(\omega) \) is given by

\[
S^*_v(\omega) = \frac{1}{2\pi} \hat{R}_T(0) + \frac{1}{\pi} \sum_{u=1}^\infty \hat{R}_T(u) \cos u\omega.
\] (1.26)

Then the spectrum estimate \( \hat{f}^*_T(\omega) \), given by (1.16), may be expressed in terms of the \( S^*_v(\omega) \)'s as follows:

\[
\hat{f}^*_T(\omega) = \sum_{v=0}^{M_T} W_T(v) S^*_v(\omega)
\] (1.27)

where

\[
W_T(v) = k\left(\frac{v}{M_T}\right) - k\left(\frac{v+1}{M_T}\right), \quad v = 0, 1, \ldots, M_T-1
\]

\[
W_T(M_T) = k(1).
\] (1.28)

The use of the individual \( S^*_v(\omega) \)'s as estimates of the spectrum were considered in terms of bias and variance, and it was concluded that, as a general rule, the weight function \( W_T(v) \) should be very small for a short range of small \( v \), then rise fairly quickly to a peak, and then damp down to zero at a somewhat slower rate. Now from (1.28), it is easy to see that \( W_T(v) \) is approximately proportional to \(-k'(\frac{v}{M_T})\) for \( v = 0, 1, \ldots, M_T-1 \), and is equal to \( k(1) \) at \( v = M_T \). Consequently we may compare lag window generators by studying the derivatives of the corresponding modified lag window generators (1.19), and seeing how well they fit into this pattern.

The conclusions made in that paper were that, of the twelve generators compared there, the well-known Parzen generator:
\[ k^*(\theta) = 1 - 6 \theta^2 (1 - |\theta|), \quad 0 \leq |\theta| \leq \frac{1}{2}, \]
\[ = 2(1 - |\theta|)^3, \quad \frac{1}{2} \leq \theta \leq 1 \quad (1.29) \]

and one due to Bohman:

\[ k^*(\theta) = (1 - \theta) \cos \pi \theta + \frac{\sin \pi \theta}{\pi}, \quad 0 \leq |\theta| \leq 1 \quad (1.30) \]

were the most satisfactory. Nevertheless it was shown that the corresponding weighting functions \( W_T(*) \) still had weaknesses, which were concluded to be mainly due to the restricted nature of the class to which they belong—i.e.,

weighting functions obtained essentially by differentiating a continuous lag window generator, and accounting for different sample sizes \( T \) by merely changing the value of \( M_T \).

These conclusions led us to carry to the further stage presented in this paper the investigation of the formation of good weighting functions for spectrum estimation.
§2. A Form of Optimum Weighting System

In Watts (1964) an attempt was made to find covariance weighting functions which were optimum according to some criterion. This involved the minimisation of certain rather arbitrarily chosen integrals by methods of the calculus of variations. An earlier attack on this problem was by Parzen (1958), based on considerations of minimum integrated mean square error. However as the author admitted, 'in this paper, no conclusions are drawn or principles enunciated as to how to proceed in practice to estimate the spectral density.' Jones (1960) produced some optimisation theory, again in terms of mean square error, on a restricted class of generators.

The contribution to this topic given here is essentially an extension of the ideas of Lomnicki and Zaremba (1957). The criterion used in this paper is again the minimum integrated mean square error, although it will be indicated that this may not be a universally wise choice, although it is certainly a natural one mathematically. The considerations here are not merely asymptotic: the weights obtained are expressed in terms of the covariance function and the sample size $T$, i.e. for a given spectrum and sample size, the optimum weights are immediately derivable. Finally some particular autoregressive processes are chosen, and the resulting weighting systems calculated for various values of $T$, and discussed with reference to standard practices.

In Neave (1968 d), the variance of the spectrum estimate $\hat{f}_T(\omega)$ given by (1.16) was found to be of the form (putting $M_T = T$):

$$\begin{align*}
\text{var}(\hat{f}_T(\omega)) &= \frac{1}{\pi^2} \left\{ A k_T(0)^2 + k_T(0) \sum_{v=1}^{T} B(v) k_T(v) \cos v\omega \\
+ \sum_{v=1}^{T-1} C(v) \frac{k_T(v)^2 \cos^2 v\omega}{T-v} + 2 \sum_{1 \leq v < u \leq T} \frac{D(u, v) k_T(u) k_T(v) \cos v\omega \cos u\omega}{T-v} \right\}
\end{align*}$$
where
\[ A = \frac{1}{2T^2} \left\{ T \, R(0)^2 + 2 \sum_{x=1}^{T} (T-x) \, R(x)^2 \right\} , \]  
\[ C(v) = R(0)^2 + R(v)^2 + \frac{2}{T-v} \left\{ \sum_{x=1}^{T-v} (T-v-x) \left\{ R(x)^2 + R(v+x) \, R(|v-x|) \right\} \right\} \]  

and \( B(\cdot) \) and \( D(\cdot) \) are functions whose forms are irrelevant for the present purposes. Integrating the variance, the second and fourth terms vanish, thus
\[ \int_0^\pi \text{var}(\tilde{f}_T(\omega)) \, d\omega = \frac{1}{\pi} \left\{ A \, k_T(0)^2 + \frac{1}{2} \sum_{v=1}^{T-1} \frac{C(v) \, k_T(v)^2}{T-v} \right\} . \]

The bias \( b_T(\omega) \) of \( \tilde{f}_T(\omega) \) is given by
\[ b_T(\omega) = E \left[ \tilde{f}_T(\omega) \right] - f(\omega) \]
\[ = - \frac{1}{2\pi} \left\{ \sum_{|v| < T} (1-k_T(v)) \, R(v) \cos v\omega \right. \]
\[ + \left. \sum_{|v| \geq T} R(v) \cos v\omega \right\} \]
\[ = - \frac{1}{2\pi} \sum_{v=-\infty}^{\infty} (1-k_T(v)) \, R(v) \cos v\omega \]
if we define \( k_T(v) = 0 \) for \( |v| \geq T \). Thus the integrated squared bias is
\[ \int_0^\pi b_T(\omega)^2 \, d\omega \]
\[ = \frac{1}{2\pi} \sum_{v=1}^{\infty} R(v)^2 (1-k_T(v))^2 + \frac{1}{4\pi} \, R(0)^2 (1-k_T(0))^2 . \]

Consequently, the integrated mean square error of \( \tilde{f}_T(\omega) \) is
\[ \text{IMSE} = \int_0^\pi \left\{ \text{var} \left( \mathbf{w}_T(\omega) \right) + b_T(\omega)^2 \right\} \, d\omega \]

\[ = \frac{1}{4\pi} \left\{ 4 \, A \, k_T(0)^2 + 2 \sum_{v=1}^{T-1} \frac{C(v) \, k_T(v)^2}{T-v} \right. \]

\[ + \left. 2 \sum_{v=1}^\infty R(v)^2 \left( 1 - k_T(v) \right)^2 + R(0)^2 \left( 1 - k_T(0) \right)^2 \right\} . \]

Regarding \( k_T(0), k_T(1), \ldots, k_T(T-1) \) as variables, then partially differentiating this expression w.r.t. \( k_T(0) \):

\[ \frac{\partial}{\partial k_T(0)} \text{IMSE} = \frac{1}{4\pi} \left\{ 8 \, A \, k_T(0) - 2 \, R(0)^2 \left( 1 - k_T(0) \right) \right\} . \]

Also for \( 0 < v < T \),

\[ \frac{\partial}{\partial k_T(v)} \text{IMSE} = \frac{1}{4\pi} \left\{ 4 \, k_T(v) \, \frac{C(v)}{T-v} - 4 \, R(v)^2 \left( 1 - k_T(v) \right) \right\} . \]

For minimisation of IMSE, these partial derivatives should all vanish, giving the solutions:

\[ k_T^0(0) = \frac{R(0)^2}{R(0)^2 + 4 \, A} \tag{2.3} \]

\[ k_T^0(v) = \frac{R(v)^2}{R(v)^2 + \frac{C(v)}{T-v}}, \quad 0 < v < T. \tag{2.4} \]

These expressions thus define the optimum weighting function on the criterion of minimum integrated mean square error. Lomnicki and Zaremba obtained this function in the form

\[ k_T^0(v) = \frac{R(v)^2}{R(v)^2 + \text{var} \left( R_T(v) \right)} = \frac{R(v)^2}{E \left[ \frac{\mathbf{w}_T(v)}{R_T(v)^2} \right]} . \]
Certain points are apparent. Firstly, contrary to the usual practice, (2.3) gives a value of $k_T^0(0)$ which is less than one. Now if a lag window generator does not satisfy

$$k(0) = 1$$

then asymptotically the spectrum estimate is biased, and thus is of course unacceptable. However it is normally the case that $k_T^0(0) \to 1$ as $T \to \infty$ since

$$0 < A = \frac{1}{2T^2} \left\{ TR(0)^2 + 2 \sum_{x=1}^{T} (T-x) R(x)^2 \right\}$$

$$< \frac{1}{T} \sum_{x=0}^{T} R(x)^2 ,$$

and this tends to zero under fairly general conditions, namely those of ergodicity—see Parzen (1961 b).

Secondly, if $R(v) = 0$, then $k_T^0(v) = 0$. For in this case, the covariance estimate

$$k_T^0(v) \hat{R}_T(v)$$

is unbiased and has zero variance; if $k_T^0(v)$ were non-zero, this estimate would still be unbiased but would have positive variance. Obviously then, if the process were white noise:

$$f(\omega) = \frac{R(0)}{2\pi}$$

$$R(v) = 0 , \quad v \neq 0 ,$$

the best truncation point to use is zero! Even then, the weight applied to $\hat{R}_T(0)$, i.e. $k_T^0(v)$, is still not 1 but
\[
\frac{R(0)^2}{R(0)^2 + \frac{1}{2T^2} \cdot TR(0)^2} = \frac{T}{T + 2}.
\]

That is to say the estimate of a white noise spectrum having least integrated mean square error is

\[
\frac{1}{2\pi (T+2)} \sum_{t=1}^{T} X_t^2.
\]

An important conclusion from (2.4) concerns the vexed problem of choice of \(M_T\) in the standard type of spectrum estimate. From (2.6), \(R(v)\) must damp down as \(v\) increases. One fairly often encounters spectra for which \(R(v)\) is effectively negligible for \(v > v_0\), where \(v_0\) is quite small. From the above work, nothing is gained (though much may be lost) by including estimates of \(R(v)\) for \(v > v_0\). Most generators \(k^*(\theta)\), and therefore the modified generators \(k(\theta)\), are very small for a range of \(\theta \leq 1\), and therefore to obtain sufficient weight for \(v < v_0\), \(M_T\) should exceed \(v_0\), possibly by quite a considerable factor. However, \(M_T\) may well have to be less than the figure arrived at from such deliberations due to considerations of variance, but then one must expect the spectrum estimate to be significantly biased. On the other hand, if the sample size \(T\) can be increased, there is still no point in giving very much weight to covariance estimates for \(v > v_0\). The advice of Parzen (1965), who suggests specific ratios of \(M_T/T\), seems therefore rather dubious. It would seem advantageous to always plot \(R_T(\cdot)\) or \(R_T'(\cdot)\) and hence obtain a rough estimate of \(v_0\). Having regard to the shape of the generator being used, a maximum useful value of \(M_T\) could then be estimated.
§ 3. Optimum Weights for First Order Autoregressive Processes

A mathematically convenient and useful class of processes to consider is the first order autoregressive class, which consists of the processes whose covariance function is of the form:

\[ R(v) = R(0) \alpha^{|v|}, \quad 0 < \alpha < 1. \]  

(3.1)

Expressions for the optimum weights will be calculated for general \( \alpha \), and then some particular values will be considered in more detail. Without loss of generality, \( R(0) \) may be taken equal to 1, since the expressions for the optimum weights are homogeneous in the covariance function. Writing

\[ s(T) = \sum_{x=1}^{T} \alpha^{2x}, \quad u(T) = \sum_{x=1}^{T} x \alpha^{2x}, \]

then, from (2.1) and (2.2),

\[ A = \frac{1}{2T} + \frac{1}{T} s(T) - \frac{1}{T^2} u(T), \]

and

\[ C(v) = 1 + \alpha^{2v} + \frac{2}{T-v} \left\{ \sum_{x=1}^{T-v} (T-v-x) (\alpha^{2x} + \alpha^{v+x} + v - x) \right\} \]

which for \( v < T/2 \) is equal to

\[
1 + \alpha^{2v} + 2s(T-v) - \frac{2}{T-v} u(T-v) + \frac{2}{T-v} \alpha^{2v} \sum_{x=1}^{T-v} (T-v-x) \\
+ 2 \left\{ s(T-v) - s(v) \right\} - \frac{2}{T-v} \left\{ u(T-v) - u(v) \right\}
\]
and for \( v \geq T/2 \) is equal to
\[
1 + \alpha^{2v} + 2s(T-v) - \frac{2}{T-v} u(T-v) + \frac{2}{T-v} \alpha^{2v} \sum_{x=1}^{T-v-1} x.
\]

Now,
\[
\sum_{x=1}^{V} (T-v-x) = \sum_{s=T-2v}^{T-v-1} s
\]
\[
= v(T-\frac{1}{2}(3v+1))
\]
and
\[
\sum_{x=1}^{T-v-1} x = \frac{1}{2} (T-v-1) (T-v).
\]

Also
\[
s(T) = \frac{\alpha^{2} (1-\alpha^{2T})}{1-\alpha^{2}}
\]
(3.2)

using the formula for summation of a geometric series. Differentiating \( s(T) \),
\[
\frac{ds(t)}{d\alpha} = 2 \sum_{x=1}^{T} x\alpha^{2x-1}
\]
so that
\[
u(T) = \frac{\alpha}{2} \frac{ds(T)}{d\alpha}
\]
\[
= \frac{\alpha}{2} \left\{ \frac{(1-\alpha^{2}) (2\alpha^{2} - 2(T+1) \alpha^{2T+1}) + 2 \alpha^{4} (1-\alpha^{2T})}{(1-\alpha^{2})^{2}} \right\}
\]
\[
= \frac{\alpha^{2} (1-(T+1) \alpha^{2T})}{1-\alpha^{2}} + \frac{\alpha^{4} (1-\alpha^{2T})}{(1-\alpha^{2})^{2}}.
\]
(3.3)
Consequently for $v < T/2$,

$$C(v) = 1 + \alpha^{2v} + 2 \left\{ 2s(T-v) - s(v) + \frac{v\alpha^{2v}(T - \frac{3}{2}v - \frac{1}{2}) + u(v) - 2u(T-v)}{T-v} \right\}$$  \hspace{1cm} (3.4)$$

and for $v \geq T/2$,

$$C(v) = 1 + 2 \left\{ s(T-v) - \frac{u(T-v)}{T-v} \right\} + (T-v)\alpha^{2v}. \hspace{1cm} (3.5)$$

On substituting these expressions into (2.3) and (2.4), the required optimum weights are obtained.

Because of the complicated nature of these expressions, it is of interest to investigate possible approximations to $C(\ast)$ and therefore to $k_{0T}^0(\ast)$. First, fixing $v$, the limiting behavior of $C(v)$ as $T \to \infty$ will be examined. The terms in (3.4) will be considered separately. We have the following results:

$$4s(T-v) = \frac{4\alpha^2(1-\alpha^2(T-v))}{1-\alpha^2} = \frac{4\alpha^2}{1-\alpha^2} + 0(\alpha^2T).$$

$$\frac{2v\alpha^{2v}(T - \frac{3}{2}v - \frac{1}{2})}{T-v} = 2v\alpha^{2v} + 0(\frac{1}{T}).$$

$$\frac{2u(v)}{T-v} = 0(\frac{1}{T}).$$

$$\frac{4u(T-v)}{T-v} = \frac{4}{T-v} \left\{ \frac{\alpha^2(1-(T-v+1))\alpha^2(T-v)}{1-\alpha^2} + \frac{\alpha^4(1-\alpha^2(T-v))}{(1-\alpha^2)^2} \right\}$$

$$= 0(\frac{1}{T}).$$
The remaining part, \(1 + \alpha^{2\nu} - 2s(\nu)\), is independent of \(T\). Consequently

\[
C(\nu) = 1 + \alpha^{2\nu} + \frac{4\alpha^{2}}{1-\alpha^{2}} \left( 1 - \frac{2\alpha^{2}(1-\alpha^{2\nu})}{1-\alpha^{2}} \right) + 2\nu \alpha^{2\nu} + O\left(\frac{1}{T}\right)
\]

\[
= \frac{1 + \alpha^{2}}{1 - \alpha^{2}} \left( 1 + \alpha^{2\nu} \right) + 2\nu \alpha^{2\nu} + O\left(\frac{1}{T}\right)
\]

Then using (2.4),

\[
k_{T}^{0}(\nu) \sim (1 + \frac{1}{T\alpha^{2\nu}}) \left\{ \frac{1 + \alpha^{2}}{1 - \alpha^{2}} \left( 1 + \alpha^{2\nu} \right) + 2\nu \alpha^{2\nu} \right\}^{-1}
\]  \hspace{1cm} (3.6)

There is no point in approximating (3.5) in this way since that equation refers to \(\nu \geq T/2\) whereas these considerations assume \(T\) to be large and \(\nu\) fixed, i.e. small in comparison with \(T\).

It may be observed that if \(\nu\) is put equal to 0 in (3.6), the resulting value is

\[
\frac{1}{1 + \frac{2}{T} \left( \frac{1 + \alpha^{2}}{1 - \alpha^{2}} \right)}
\]  \hspace{1cm} (3.7)

From (2.3), the true value of \(k_{T}^{0}(0)\) is

\[
\frac{1}{1 + 4\Lambda}
\]

where

\[
A = \frac{1}{2T^{2}} \left\{ T + 2 \sum_{x=1}^{T} (T - x) \alpha^{2x} \right\}
\]

\[
= \frac{1}{2T^{2}} \left\{ T + 2Ts(T) - 2u(T) \right\}
\]
\[ k_0^T(0) \approx \frac{1}{1 + \frac{2(1 + \alpha^2)}{T(1 - \alpha^2)}}. \]

agreeing with (3.7). Consequently the expression (3.6) is valid for all non-negative \( v \) which are small compared with \( T \).

A further stage of approximation is possible. For any value of \( c \) in \( 0 < c < 1 \), let \( T \to \infty \) and consider the corresponding function of \( T \) and \( c \), \( V_c(T) \) satisfying

\[ k_0^T(V_c(T)) = c. \]

Abbreviating \( V_c(T) \) by \( V \), it follows from (3.6) that

\[ \left(1 + \frac{1}{T\alpha^{2V}} \left[ \frac{1 + \alpha^2}{1 - \alpha^2} - (1 + \alpha^2V) + 2V\alpha^{2V} \right]\right)^{-1} \to c \]

or

\[ \frac{1}{T\alpha^{2V}} \left[ \frac{1 + \alpha^2}{1 - \alpha^2} - (1 + \alpha^2V) + 2V\alpha^{2V} \right] \to D \]

where

\[ c = \frac{1}{1 + D} ; \quad D = \frac{1}{c} - 1 \]

and \( D \) lies in the range \((0, \infty)\). Therefore

\[ \frac{1 + \alpha^2}{1 - \alpha^2} \left. \frac{1}{T\alpha^{2V}} \right. + \frac{1 + \alpha^2}{1 - \alpha^2} \frac{1}{T} + \frac{2V}{T} \to D. \]
The centre term clearly tends to zero. It is apparent also that \( \frac{V}{T} \to 0 \); otherwise if there were a sequence of \( T \)'s for which \( V > \mu T \) for some \( \mu > 0 \), then the first term would exceed

\[
\frac{1 + a^2}{1 - a^2} \cdot \frac{1}{T a^2 \mu T} \to \infty.
\]

Therefore the third term may be neglected, leaving

\[
\frac{1 + a^2}{1 - a^2} \cdot \frac{1}{T a^2 V} \to D.
\]

Taking logarithms

\[
\log\left(\frac{1 + a^2}{1 - a^2}\right) - \log T + 2V \log\left(\frac{1}{a}\right) \to \log\left(\frac{1}{c} - 1\right) \quad (3.8)
\]

For all values of \( c \) then (except the extremes, 0 and 1), \( V \) and \( T \) are connected by a relation of the form

\[
V = x \log T + y.
\]

This approximation may of course be poor in finite cases where \( \frac{V}{T} \) is not small and \( T \) is not large. In practice it is usually found to overestimate the values of \( k_T^0(v) \) near \( v = 0 \), but otherwise to be fairly satisfactory.
§4. Some Particular Cases

The autoregressive form is a good choice of spectrum both because it allows the easy mathematical treatment of §3, and because it has an important practical interpretation (see e.g. Neave (1968 d)). In addition, the fact that its covariance function steadily decreases results in the optimum weights having the same convenient property; with more complicated spectra, although the general tendency is to decrease, the optimum weights used need not be strictly decreasing. The arguments of §2 of Neave (1968 c) however show that a strictly decreasing set of weights should be used when estimating an unknown spectrum, and this again makes the autoregressive form an appropriate choice here.

In Figure 1(a-c) are shown graphs of the optimum weights, and approximations to them, in the case of a .9 first order autoregressive process; the sample sizes considered are 50, 200, and 1000. Figure 1(a) shows the exact weights, computed from (3.4), (2.3) and (2.4). Figure 1(b) shows the first approximations (3.6) and Figure 1(c) shows the cruder approximations derived from (3.8). It can be seen that the first approximations are practically indistinguishable from the true results when T is 200 and 1000, although they present a slightly flattened version of the true case for T = 50. The cruder approximations consistently overestimate the true values, especially for small v.

There are a number of points worth observing from the shapes of these optimum weighting functions. Firstly, even ignoring scale factors, these shapes vary greatly with T. For small T, the variance of the sample covariance function is comparatively so large that approximately equal weights
are given to the first 15 or so $S_v^*(',*)'$s (using the notation of (1.26)), including the very biased $S_0^*, S_1^*, etc.$ But for large $T$, the optimum weighting function becomes very flat near $v = 0$ so that practically no weight is given to these first few $S_v^*(',*)'$s: the majority of the weight, nearly 80% in fact, goes to the $S_v^*(',*)'$s for $15 \leq v \leq 35$ when $T = 1000$. Also, even for $T = 200$, $k_T^0(0)$ is considerably less than 1, with the result, pointed out previously, of increased bias and decreased variance. Mean square error criteria treat the squared bias and the variance as equally undesirable; in practice, large bias is generally worse in effect than large variance, thus possibly explaining some of the discrepancies between the accepted methods and these. Finally it should be noticed how slowly the significant range of $v$ increases with $T$. If the optimum weights were being used except for being ignored if less than .0001, which can then be regarded as the truncation point, it turns out that

$$M_{50} = 46; \quad M_{200} = 57; \quad M_{1000} = 66,$$

i.e. the values of $M_T/T$ are respectively

$$0.920; \quad 0.285; \quad 0.006.$$  

This tremendous decrease in the ratio is perhaps the most striking result of this exercise.

Taking into account these values of $M_T/T$, it is apparent from Figure 1 of Neave (1968 c) that these optimum weighting functions bear marked resemblance to the Parzen, Tukey and Bohman generators, which came most favourably out of the discussion in that paper. When $T$ is large, it would appear that the form of optimum weights is very like the Tukey generator for small $v$, and very like the Parzen for large $v$ -- i.e. a rather extended 'tail'. It is also particularly encouraging that the different shapes obtained for the
optimum weights with different values of $T$ also closely resemble the shapes of these generators according to the appropriate values of $M_T/T$ throughout almost the whole range $(0,1)$.

In Figure 2(a-c) are shown the optimum weighting functions for the same three values of $T$, and autoregressive parameter $\alpha = .8, .95$ and $.99$ respectively. As $\alpha$ approaches $1$, the spectral peak centered on the zero frequency becomes higher and narrower, and certainly for $\alpha = .99$ the spectrum is beyond the class discussed in Neave (1968 d). This latter case really emphasizes the way in which the optimum weights attempt to reduce the variance by biasing the estimate to what might seem an absurd extent, e.g. the total weight applied, $k_T^0(0)$, is only $.4$ when $T = 50$. Another point of interest is that the weights do not necessarily reduce to 0 as $v$ nears $T$ if the corresponding $R(v)$'s are still large. In such cases, comparatively large weight is given to these covariances, thus apparently encouraging a high variance, but the effect is reversed for small $v$ where the weights are almost linear (instead of flattening out), thus increasing the bias but decreasing the variance.

Not all of the features apparent in these graphs have been satisfactorily explained. It would appear that where the sample size makes it possible to obtain good spectrum estimates, then the current procedures are not far removed from the optimum if the truncation point happens to be well chosen. In particular, the extended 'tail', questioned in Neave (1968 c), is very much in evidence. But it would seem that much work could still be carried out on 'crude estimation' of spectra from small samples. This is particularly relevant if one is to follow the advice of Blackman and Tukey (1958, p. 45)
(supported by the evidence in Neave (1968 d)) to use a pilot analysis to find the rough location of outstanding peaks, and then prewhiten the data before carrying out a standard estimation procedure.
FIGURE 1: OPTIMUM WEIGHTS AND APPROXIMATIONS FOR .9 AUTOREGRESSIVE SPECTRUM.
(a) $\alpha = .8$

(b) $\alpha = .95$

(c) $\alpha = .99$

FIGURE 2: OPTIMUM WEIGHTS FOR THREE AUTOREGRESSIVE CASES.
REFERENCES


