

DECONVOLUTION AND ESTIMATION OF TRANSFER FUNCTION PHASE AND COEFFICIENTS FOR NONGAUSSIAN LINEAR PROCESSES¹

BY K. S. LIH AND M. ROSENBLATT

University of California, Riverside and San Diego

NonGaussian linear processes are considered. It is shown that the phase of the transfer function can be estimated under broad conditions. This is not true of Gaussian linear processes and in this sense Gaussian linear processes are atypical. The asymptotic behavior of a phase estimate is determined. The phase estimates make use of bispectral estimates. These ideas are applied to a problem of deconvolution which is effective even when the transfer function is not minimum phase. A number of computational illustrations are given.

1. Introduction. Most of the literature on finite parameter time series models is either centered on Gaussian models or the results are motivated by what one can do in the case of Gaussian models. Here we deal with stationary nonGaussian linear processes and show that under broad conditions, aspects of the structure that are completely nonidentifiable in the Gaussian case can be resolved in the nonGaussian case. Assume that the random variables v_t , $t = \dots, -1, 0, 1, \dots$ are independent and identically distributed with mean zero, $Ev_t = 0$, and variance one $Ev_t^2 = 1$. Let $\{\alpha_j\}$ be a sequence of real constants with

$$\sum_{j=-\infty}^{\infty} \alpha_j^2 < \infty.$$

Consider the linear process generated by $\{\alpha_j\}$ and $\{v_t\}$

$$(1) \quad x_t = \sum_{j=-\infty}^{\infty} \alpha_j v_{t-j}.$$

Let $\alpha(z) = \sum_j \alpha_j z^j$ be the z -transform corresponding to the process $\{x_t\}$. Then

$$\alpha(e^{-i\lambda}) = \sum_j \alpha_j e^{-ij\lambda}$$

is called the frequency response function or transfer function. We are concerned with the estimation of $\alpha(e^{-i\lambda})$ on the basis of observations only on the process $\{x_t\}$. Further, the linear nonGaussian process serves as a model for types of seismic exploration. The weights α_j can be regarded as the signature of a disturbance passing through a medium and the random weights v_j as reflectivity of slabs in a layered medium. It has been observed that in many geophysical contexts the observed data (the x_t 's) are distinctly nonGaussian and a basic object is to deconvolve, estimating the α_j 's and v_j 's in the process. A discussion of related questions in the geophysical context can be found in Donoho (1981), Godfrey and Rocca (1981), and Wiggins (1978).

The spectral density of $\{x_t\}$ is

$$f(\lambda) = \frac{1}{2\pi} |\alpha(e^{-i\lambda})|^2.$$

In the Gaussian case (when $\{x_t\}$ is normally distributed) the full probability structure of $\{x_t\}$ is determined by $f(\lambda)$ or equivalently by the modulus of $\alpha(e^{-i\lambda})$, $|\alpha(e^{-i\lambda})|$. The phase

Received November 1981; revised May 1982.

¹ Research supported in part by Office of Naval Research Contract N00014-81-K-0003.

AMS 1980 subject classification. Primary 62M15; secondary 62G99.

Key words and phrases. Asymptotics, bispectrum, deconvolution, linear process, non-Gaussian, phase.

information in $\alpha(e^{-i\lambda})$ is not identifiable in the Gaussian case.

If $\alpha(z)$ is a rational function

$$\alpha(z) = \frac{A(z)}{B(z)}$$

with $A(z), B(z)$ polynomials

$$A(z) = \sum_{k=0}^q a_k z^k, \quad a_0 \neq 0, \quad B(z) = \sum_{k=0}^p b_k z^k, \quad b_0 = 1,$$

the process $\{x_t\}$ is a finite parameter autoregressive moving average process; that is,

$$(2) \quad \sum_{j=0}^p b_j x_{t-j} = \sum_{k=0}^q a_k v_{t-k}.$$

If $\{x_t\}$ is a Gaussian process satisfying (2), then any real root $z_j \neq 0$ of $A(z)$ or $B(z)$ can be replaced by its inverse and pairs of nonzero conjugate roots by their paired conjugated inverses \bar{z}^{-1} without changing the probability structure of $\{x_t\}$ if the process is appropriately rescaled. This follows since $|e^{i\lambda} - z_0| = |z_0| |e^{-i\lambda} - z_0^{-1}|$. With real distinct roots there are 2^{p+q} ways of specifying the roots without changing the structure of $\{x_t\}$. To ensure unique determination of the coefficients a_k and b_j of (2) (since there is a different specification of these coefficients corresponding to each of the possible root specifications) in the Gaussian case, it is the custom to assume that all the roots of $A(z)$ and $B(z)$ are outside the unit circle $|z| \leq 1$ in the complex plane.

However, for a nonGaussian stationary process satisfying (2) (in which case the independent v_t 's are nonGaussian) the different 2^{p+q} specifications of roots mentioned above generally correspond to different probability structures and different stationary processes. As a simple example, consider the moving average

$$x_t = 6v_t - 5v_{t-1} + v_{t-2}$$

with the roots of $A(z)$ being 2 and 3, and the moving average

$$y_t = 3v_t - 7v_{t-1} + 2v_{t-2}$$

having a polynomial $A(z)$ with roots $\frac{1}{2}$ and 3. Both $\{x_t\}$ and $\{y_t\}$ have the same spectral density but if the independent random sequence $\{v_t\}$ is exponentially distributed, the marginal distributions of the $\{x_t\}$ and $\{y_t\}$ sequences are different. In the problem of deconvolution where one wishes to recover the process $\{v_t\}$ (assumed nonGaussian which is most often the case in applications) in some sense, the proper specification of nonzero roots (which are inside and which are outside) becomes crucial (see Rosenblatt, 1974). There is a discussion concerning the distribution of roots as related to prediction problems in Rosenblatt (1980). The assumption that all roots are outside the unit disc is often called the minimum phase condition.

There are results on the estimation of the coefficients a_j and b_k of (2) (corresponding to roots assumed outside the unit circle) in Box and Jenkins (1976). In the Gaussian case these are essentially equivalent asymptotically to maximum likelihood procedures. In the nonGaussian case the computations are carried out as if the process were Gaussian. One has a least squares but not a maximum likelihood solution in the nonGaussian case. The coefficients estimated are those corresponding to roots outside the unit circle even though the actual structure of the process may not be one with all the roots outside the unit circle. Thus one will typically not be able to resolve the actual structure using these procedures in the nonGaussian case. Of course, if one knows the actual nonGaussian distribution of the v_t , one can use the maximum likelihood estimate or an asymptotically equivalent procedure to estimate the coefficients in (2) even if the roots z_j are not all outside the unit circle (see Basawa, Feigen and Heyde, 1975). Higher order spectral methods discussed in the next section do not require this knowledge. Our discussion follows that of Rosenblatt (1980).

2. Higher order spectral method. We prove the following lemma.

LEMMA 1. Let $\{x_t\}$ be a nonGaussian linear process (see (1)) with the independent random variables $\{v_t\}$ having all moments finite. Assume that

$$\sum |j| |a_j| < \infty$$

and $\alpha(e^{-i\lambda}) \neq 0$ for all λ . Then the function $\alpha(e^{-i\lambda})$ is identifiable on the basis of observations on $\{x_t\}$ alone up to the integer a in a factor $e^{ia\lambda}$ and the sign of $\alpha(1) = \sum \alpha_k$. Actually it is enough to have moments of order $k > 2$ finite with cumulant $\gamma_k \neq 0$.

Since the v_t 's are assumed nonGaussian with all moments finite, there must be a cumulant of v_t , $\gamma_k \neq 0$ of smallest subscript $k > 2$. The k th order cumulant spectral density of the process $\{x_t\}$ is given by

$$\begin{aligned} b_k(\lambda_1, \dots, \lambda_{k-1}) &= \frac{1}{(2\pi)^{k-1}} \sum_{j_1, \dots, j_{k-1}} \text{cum}(x_t, x_{t+j_1}, \dots, x_{t+j_{k-1}}) \exp(-\sum_{s=1}^{k-1} i j_s \lambda_s) \\ (3) \qquad \qquad \qquad &= \frac{\gamma_k}{(2\pi)^{k-1}} \alpha(e^{-i\lambda_1}) \dots \alpha(e^{-i\lambda_{k-1}}) \alpha(e^{i(\lambda_1 + \dots + \lambda_{k-1})}). \end{aligned}$$

Let

$$h(\lambda) = \arg \left\{ \alpha(e^{-i\lambda}) \frac{\alpha(1)}{|\alpha(1)|} \right\}.$$

Then

$$\left\{ \frac{\alpha(1)}{|\alpha(1)|} \right\}^k \gamma_k = (2\pi)^{k/2-1} b_k(0, \dots, 0) / \{f(0)\}^{k/2},$$

and

$$(4) \quad h(\lambda_1) + \dots + h(\lambda_{k-1}) - h(\lambda_1 + \dots + \lambda_{k-1}) = \arg \left[\left\{ \frac{\alpha(1)}{|\alpha(1)|} \right\}^k \gamma_k^{-1} b_k(\lambda_1, \dots, \lambda_{k-1}) \right]$$

since $h(-\lambda) = -h(\lambda)$. Further

$$(5) \quad h'(0) - h'(\lambda) = \lim_{\Delta \rightarrow 0} \frac{1}{(k-2)\Delta} \{h(\lambda) + (k-2)h(\Delta) - h(\lambda + (k-2)\Delta)\}.$$

Now

$$h(\lambda) = \int_0^\lambda \{h'(u) - h'(0)\} du + c\lambda = h_1(\lambda) + c\lambda$$

where $c = h'(0)$. In particular

$$h(\pi) = h_1(\pi) + c\pi.$$

Since the α_j 's are real we must have $h(\pi) = a\pi$ for some integer a . Set $h_1(\pi)/\pi = \delta$. Then $h(\pi) = a\pi = (\delta + c)\pi$, so that

$$c = a - \delta.$$

The integer a cannot be determined without further assumptions since it corresponds to reindexing or subscripting the v_t 's. The sign of $\alpha(1)$ is also intrinsically undecidable since one can multiply α_j 's and v_t 's by (-1) without changing the observed process $\{x_t\}$. Thus, under the conditions specified above for a nonGaussian linear process $\{x_t\}$, $\alpha(e^{-i\lambda})$ is identifiable up to the integer a and the sign on the basis of observations on $\{x_t\}$ only and is given by

$$\alpha(e^{-i\lambda}) = |2\pi f(\lambda)|^{1/2} \exp\{ih(\lambda)\}$$

with

$$(6) \quad h(\lambda) = \int_0^\lambda \{h'(u) - h'(0)\} du + c\lambda = h_1(\lambda) - \frac{h_1(\pi)}{\pi} \lambda + a\lambda.$$

Notice that $h_1(\lambda)$ can actually be computed. In Section 3 we will consider an estimate of the identifiable aspects of $\alpha(e^{-i\lambda})$ suggested by Lemma 1 and the convergence properties of this estimate. Comments on the computations using spectral methods follow in Section 4. A few remarks on other computational methods are made in Section 5. Examples of estimation and deconvolution are given in Section 6. It should be noted that in some seismic investigations, a can be estimated through knowledge of speed of wave (e.g. sound) propagation through the medium.

3. Phase estimation and convergence of estimates. There are many discussions concerned with the estimation of the second order spectral density $f(\lambda)$ (see Anderson 1971, or Jenkins and Watts, 1968). We will concentrate on the estimation of $h(\lambda)$. For simplicity of discussion we will assume that the third order cumulant γ_3 of v_t is nonzero. The program in the higher order case can be carried out in a similar manner. Equation (5) becomes

$$h'(0) - h'(\lambda) = \lim_{\Delta \rightarrow 0} \frac{1}{\Delta} \{h(\lambda) + h(\Delta) - h(\lambda + \Delta)\}$$

when $k = 3$. From (4) we find that up to a sign

$$h(\lambda) + h(\Delta) - h(\lambda + \Delta) = \arg\{b_3(\lambda, \Delta)\}.$$

From this point on we will drop the subscript and understand that we are dealing with the bispectral density $b(\lambda, \mu)$.

The following approximation is useful in deriving the theorem of this section.

LEMMA 2. *Let the assumptions of Lemma 1 be satisfied. Assume that ${}_n b(\lambda, \mu)$ is a consistent estimate of $b(\lambda, \mu)$ (as $n \rightarrow \infty$) based on a sample of size n . If one considers*

$$\theta_n(\lambda, \mu) = \arctan(\text{Im}_n b(\lambda, \mu) / \text{Re}_n b(\lambda, \mu))$$

as an estimate of

$$\theta(\lambda, \mu) = \arg b(\lambda, \mu)$$

then

$$(7) \quad \theta_n(\lambda, \mu) - \theta(\lambda, \mu) = -\frac{\text{Im } b(\lambda, \mu)}{|b(\lambda, \mu)|^2} \{\text{Re}_n b(\lambda, \mu) - \text{Re } b(\lambda, \mu)\} + \frac{\text{Re } b(\lambda, \mu)}{|b(\lambda, \mu)|^2} \{\text{Im}_n b(\lambda, \mu) - \text{Im } b(\lambda, \mu)\} + o_p({}_n b(\lambda, \mu) - b(\lambda, \mu)).$$

We note that for a complex number

$$z = x + iy = re^{i\theta}$$

with $r = |z|$ and $\theta = \arctan(y/x)$ a principal value determination, one has

$$\frac{\partial \theta}{\partial y} = \frac{x}{r^2}, \quad \frac{\partial \theta}{\partial x} = -\frac{y}{r^2}, \quad \frac{\partial^2 \theta}{\partial x^2} = \frac{2xy}{r^4}, \quad \frac{\partial^2 \theta}{\partial y^2} = -\frac{2xy}{r^4}$$

and

$$\frac{\partial^2 \theta}{\partial x \partial y} = \frac{1}{r^2} - \frac{2x^2}{r^4} = -\frac{1}{r^2} + \frac{2y^2}{r^2}.$$

The relation (7) follows on using these observations in a Taylor expansion of the arctan function.

Let us consider estimating

$$h_1(\lambda) - \frac{h_1(\pi)}{\pi} \lambda.$$

Set $\Delta = \Delta(n)$, $k\Delta = \lambda$, and let $\Delta = \Delta(n) \rightarrow 0$ as $n \rightarrow \infty$. Assume for convenience that $b(0, 0)$ is positive. A simple modification indicated later takes care of the case in which $b(0, 0)$ is negative. Now

$$\begin{aligned} h_1(\lambda) &= h(\lambda) - h'(0)\lambda \cong h(k\Delta) - \frac{h(\Delta)}{\Delta} k\Delta \\ &= -\sum_{j=1}^{k-1} \{h(j\Delta) + h(\Delta) - h((j+1)\Delta)\} = -\sum_{j=1}^{k-1} \arg b(j\Delta, \Delta). \end{aligned}$$

This suggests

$$H_n(\lambda) = -\sum_{j=1}^{k-1} \arg_n b(j\Delta, \Delta)$$

as an appropriate estimate of $h_1(\lambda)$.

THEOREM. Assume that $b(0, 0)$ is positive and that $EX_i^6 < \infty$. Consider

$$(8) \quad H_n(\lambda) = -\sum_{j=1}^{k-1} \arg_n b(j\Delta, \Delta),$$

$k\Delta = \lambda$, as an estimate of

$$h_1(\lambda) = h(\lambda) - h'(0)\lambda,$$

where it is understood that the bispectral estimates are weighted averages of 3rd order periodogram values. If $b(\lambda, \mu) \in \mathcal{C}^2$ and the weight function is symmetric and bandlimited with bandwidth Δ , then

$$H_n(\lambda) - h_1(\lambda) = R_n(\lambda) + o_p(H_n(\lambda) - h_1(\lambda))$$

with

$$\begin{aligned} ER_n(\lambda) &\sim \int_0^\lambda \frac{1}{2} \frac{\text{Im } b(u, 0)}{|b(u, 0)|^2} \{AD_u^2 \text{Re } b(u, 0) + 2BD_u D_v \text{Re } b(u, 0) + CD_v^2 \text{Re } b(u, 0)\} du \Delta \\ &\quad - \int_0^\lambda \frac{1}{2} \frac{\text{Re } b(u, 0)}{|b(u, 0)|^2} \{AD_u^2 \text{Im } b(u, 0) + 2BD_u D_v \text{Im } b(u, 0) + CD_v^2 \text{Im } b(u, 0)\} du \Delta + o(\Delta) \end{aligned}$$

and

$$(9) \quad \text{Cov}\{R_n(\lambda), R_n(\mu)\} \cong \frac{2\pi^2}{\Delta^3 n \gamma_3^2} \min(\lambda, \mu) \int w^2(u, v) du dv$$

for $\Delta(n) \rightarrow 0$, $\Delta^2 n \rightarrow \infty$ as $n \rightarrow \infty$. Here A , B , and C are the second moments of the weight function w of the bispectral estimates

$$A = \int u^2 w(u, v) du dv, \quad B = \int uvw(u, v) du dv, \quad C = \int v^2 w(u, v) du dv$$

and D_u and D_v are partial derivatives with respect to u and v respectively.

It has been shown in Brillinger and Rosenblatt (1967) under the assumptions of the theorem that

$$(10) \quad E\{ {}_n b(\lambda, \mu) \} - b(\lambda, \mu) \sim \frac{1}{2} \int \int (uD_\lambda + vD_\mu)^2 b(\lambda, \mu) w(u, v) du dv \Delta^2 + o(\Delta^2)$$

and

$$(11) \quad \text{Var}\{ {}_n b(\lambda, \mu) \} \sim \frac{f(\lambda)f(\mu)f(\lambda + \mu)}{\Delta^2 n} \int w^2(u, v) du dv, \quad 0 \neq \lambda \neq \mu \neq \pi,$$

if $\Delta^2 n \rightarrow \infty$ as $n \rightarrow \infty$, $\Delta(n) \rightarrow 0$. Further estimates ${}_n b(\lambda, \mu)$, ${}_n b(\lambda', \mu')$, $0 \leq \mu \leq \lambda$, $0 \leq \mu' \leq \lambda'$, are asymptotically uncorrelated as $n \rightarrow \infty$ if (λ, μ) and (λ', μ') are distinct. Using (7) and (8) we can write

$$(12) \quad H_n(\lambda) - h_1(\lambda) = R_n(\lambda) + o_p(H_n(\lambda) - h_1(\lambda)),$$

$$R_n(\lambda) = \sum \left[\frac{\text{Im } b(j\Delta, \Delta)}{|b(j\Delta, \Delta)|^2} \{ \text{Re } {}_n b(j\Delta, \Delta) - \text{Re } b(j\Delta, \Delta) \} \right. \\ \left. - \frac{\text{Re } b(j\Delta, \Delta)}{|b(j\Delta, \Delta)|^2} \{ \text{Im } {}_n b(j\Delta, \Delta) - \text{Im } b(j\Delta, \Delta) \} \right]$$

and show by employing (10) that the approximation for ER_n given in the statement of the theorem is valid. Further, by using (11) it follows that (9) holds.

COROLLARY. *Assume that all moments of $\{x_t\}$ exist and that the assumptions of the theorem are satisfied. Then $EH_n(\lambda) \rightarrow h_1(\lambda)$ and $H_n(\lambda)$ is asymptotically normal with variance given by (9). The mean square error of $R_n(\lambda)$ is of order*

$$C_1 \Delta^2 + \frac{C_2}{\Delta^3 n}.$$

The optimal rate of convergence is $n^{-2/5}$ when $\Delta(n) \sim n^{-1/5}$. If $b(\lambda, \mu) \in \mathcal{C}^3$ and the weight function is bandlimited with first and second moments zero, the mean square error of $R_n(\lambda)$ is of the order

$$C_1 \Delta^4 + \frac{C_2}{\Delta^3 n}.$$

The optimal rate of convergence is then $n^{-4/7}$ with $\Delta(n) \sim n^{-1/7}$.

The asymptotic normality with variance given by (9) follows on using (12).

Generally we will estimate $h_1(\lambda)$ and hence $h(\lambda)$ for a whole range of λ values. The sign of $b(0, 0)$ may not be positive. We estimate it by noting the real part of ${}_n b(\Delta, \Delta)$. If it is negative we multiply all ${}_n b(j\Delta, \Delta)$ with a minus sign. The estimate $H_n(\lambda)$ is then given by

$$H_n(\lambda) = - \sum_{j=1}^{k-1} \arg\{- {}_n b(j\Delta, \Delta)\}.$$

4. Computations using spectral methods. We remark on the computational aspect of phase estimation of $\alpha(e^{-i\lambda})$ and give a few illustrative examples to indicate its effectiveness.

Given a sample $\{x_t\}$ of size $n = kN$, we center and normalize it so that it has mean zero and variance one. Break up the sample into k disjoint subsections of equal length N so that the variance of the bispectral estimate from each section is not too large. Then choose a grid of points $\lambda_j = j\Delta$ in $(0, 2\pi)$, $j = 1, \dots, M$, $\Delta = 2\pi L/N$ for an appropriate integer L . Though the symmetry condition $h(\lambda) = -h(-\lambda)$ implies that one need only deal with λ in $(0, \pi)$, there may be some advantage in considering $\lambda \in (0, 2\pi)$. We will comment on this point later on. Form the bispectral estimate ${}_n b(j\Delta, \Delta)$ of the type discussed above with a weight function of bandwidth Δ from each subsection. Average the estimates from the different subsections so as to arrive at a final estimate ${}_n b(j\Delta, \Delta)$. A detailed discussion of this kind of algorithm can be found in Lii and Helland (1981). Compute $\theta_n(j) = \arg\{{}_n b(j\Delta, \Delta)\} + 2k\pi$ where the integer k is chosen to ensure continuity of $H_n(\ell\Delta) = H_n(\lambda_\ell) = -\sum_{j=1}^{\ell-1} \theta_n(j)$, $\ell = 2, \dots, M + 1$ (neighboring values are as close to each other as

possible). Since the upper index is $\ell - 1$ we start with $\ell = 2$. Since $h(0) = 0$ one sets $H_n(0) = 0$ and estimates $H_n(\Delta) = H_n(\lambda_1)$ by an interpolation between 0 and $H_n(\lambda_2)$, $\lambda_2 = 2\Delta$. $H_n(\pi)$ is also computed by an interpolation procedure. This amounts to a complete procedure for estimating $h(\lambda)$.

Since

$$\alpha_k = \frac{1}{2\pi} \int_0^{2\pi} \alpha(e^{-i\lambda}) e^{ik\lambda} d\lambda$$

an estimate $\hat{\alpha}_k$ of α_k is given by

$$(13) \quad \hat{\alpha}_k = \frac{1}{2\pi} \int_0^{2\pi} \hat{\alpha}(e^{-i\lambda}) e^{ik\lambda} d\lambda \\ \cong \frac{1}{(M+2)} \sum_{j=0}^{M+1} \sqrt{2\pi f_n(\lambda_j)} \exp\left\{i\left(H_n(\lambda_j) - \frac{H_n(\pi)}{\pi} \lambda_j + k\lambda_j\right)\right\}$$

and this computation can be carried out by using the fast Fourier transform.

The α_k 's are real numbers and so the $\hat{\alpha}_k$'s may or may not be real. If the symmetric property of $f(\lambda)$ and $h(\lambda)$ is used and the integration is carried out from $-\pi$ to π almost real $\hat{\alpha}_k$'s will be obtained. The imaginary part of the α_k 's will only be the size of rounding errors. In practice there is no indication of how good or bad the estimates are apart from asymptotic results. In actual practice $\{j\Delta\}_{j=0}^{M+1}$ may not be symmetric about π . If the estimates $H_n(\lambda_j)$ are reasonably good the estimated $\hat{\alpha}_j$'s from (13) should still be almost real. The size of the imaginary part reflects the level of noise. When the estimates $H_n(\lambda_j)$ are not good the imaginary part of the $\hat{\alpha}_k$'s becomes comparable to (or larger than) its real part. This can serve as a direct indication of the quality of the estimation.

If the linear process is one-sided with a finite number of parameters one has a moving average of order q

$$x_t = \sum_{j=0}^q \alpha_j v_{t-j}, \quad \alpha_0 \neq 0.$$

We could estimate $\alpha(z) = \sum_{j=0}^q \alpha_j z^j$ by $\hat{\alpha}(z) = \sum_{j=0}^q \hat{\alpha}_j z^j$. In deconvolution we try to recover the process $\{v_t\}$, $v_t = \frac{1}{\alpha(B)} x_t$ (B is the backward shift operator so that $B^j x_t = x_{t-j}$) by

computing the approximation $\hat{v}_t = \frac{1}{\hat{\alpha}(B)} x_t$. If all the roots of $\alpha(z)$ (and $\hat{\alpha}(z)$) are outside the unit circle (the frequency function is minimum delay) then $\hat{\alpha}^{-1}(z)$ has a one-sided expansion $\sum_{j=0}^{\infty} \alpha'_j B^j$. In the computation, the series is truncated after a certain number of terms. If some of the roots of $\hat{\alpha}(z)$ have modulus less than one we can still expand $\hat{\alpha}^{-1}(B)$ with a Laurent series expansion. Once the roots of $\hat{\alpha}(z)$ are computed, one can easily get the Laurent series expansion of $\hat{\alpha}^{-1}(B)$ by partial fractions as described in Rosenblatt (1974) or Henrici (1974).

Another way to find the inverse weights in deconvolution is to use a least squares criterion as described in Wiggins (1978). Another general method of deconvolution will be mentioned in the section on computation. A discussion of related questions for autoregressive schemes can be found in Huzii (1981).

5. Other possible computational methods. We briefly discuss two other possible methods of estimating the coefficients of a nonGaussian moving average process of order q

$$(14) \quad x_t = \sum_{j=0}^q \alpha_j v_{t-j}.$$

As noted earlier, second order moments will not allow us to determine the location of the roots of

$$(15) \quad \alpha(z) = \sum_{j=0}^q \alpha_j z^j.$$

Higher order moments will be used in the first method which makes use of a *least squares procedure*. Assume $Ev_t = 0, Ev_t^3 = \gamma \neq 0$. Consider

$$(16) \quad c_k = Ex_t x_{t+k}^2 = \gamma \sum_j \alpha_j \alpha_{j+k}^2, \quad k = -q, -q + 1, \dots, q.$$

Estimate c_k by

$$\hat{c}_k = \frac{1}{n} \sum_{k=1}^n x_t x_{t+k}^2$$

and solve the extremal problem

$$(17) \quad \min_{\alpha_j} \sum_{k=-q}^q (\hat{c}_k - \gamma \sum_j \alpha_j \alpha_{j+k}^2)^2.$$

There are $q + 2$ unknowns $\alpha_0, \dots, \alpha_q$ and γ in (17). Due to the homogeneity of the α_j 's we have to normalize the problem appropriately; all the α_j 's can be multiplied by a constant $c \neq 0$ and γ can be divided by c^3 without changing (17). There are a number of ways of carrying out such a normalization. One could set $Ev_t^3 = \gamma = 1$. Alternatively $\alpha_0 = 1$ could be the normalization condition. Some comments on the asymptotic distribution of the \hat{c}_k 's are given in Appendix 2.

The second method is a searching procedure. One uses a typical second order method to estimate the roots of $\alpha(z), r_j, j = 1, \dots, q$, assuming all the roots have modulus greater than one. An accurate estimate of the distribution of roots is obtained by taking the conjugated inverse of an appropriate number of the r_j 's. Suppose all of the r_j 's are real and distinct. Then there are 2^q possible sets of roots that give the same second order structure. Each of these sets yields a distinct set of $\hat{\alpha}_j$'s which in turn lead to a distinct set of the c_k 's. Choose the set of $\hat{\alpha}_j$'s which determine the set of c_k 's minimizing

$$\sum_{k=-q}^q (\hat{c}_k - c_k)^2$$

among all the possible sets of $\{\hat{\alpha}_j\}$. If some of the roots r_j are complex, the inverse complex conjugates are taken in pairs. If there are multiple roots, the solution of roots in terms of coefficients is unstable. Some comments on this question are made in Appendix 1. The initial set of coefficient estimates corresponding to roots all outside the unit circle can be obtained by the method described in Box and Jenkins (1976). Alternatively, one could try to obtain the roots directly by solving for the roots of the polynomial

$$p(z) = z^q g(z)$$

where

$$g(z) = \alpha(z)\alpha(z^{-1}) = \sum_{j=-q}^q \beta_j z^j$$

with

$$\beta_{|j|} = Ex_t x_{t+|j|} = \sum_j \alpha_j \alpha_{j+|j|}.$$

We estimate β_j by

$$\hat{\beta}_j = \frac{1}{n} \sum_{t=1}^{n-|j|} x_t x_{t+|j|}.$$

The roots of $p(z)$ with modulus greater than one are the initial set of roots.

6. Examples. We will consider a few simple examples generated by Monte Carlo simulation to illustrate the computation and to give a qualitative feeling of the effectiveness of the theory. Details and possible "fine tuning" of the computational method will be considered elsewhere.

We generate $x_t = v_t + \alpha_1 v_{t-1} + \alpha_2 v_{t-2}, t = 1, \dots, 640$ where $v_t = v'_t - 1$ and v'_t 's are independent exponentially distributed random deviates with mean one obtained from the GGEXN subroutine in the International Mathematical and Statistical Library (IMSL).

Then

$$E v_t = 0, \quad \text{var}(v_t) = E v_t^2 = 1, \quad E v_t^3 = 2.$$

We partition $\{x_t\}_{t=1}^{640}$ into five sections, each of which has 128 points. Compute the bispectrum estimate ${}_{128}b^{(i)}(j\Delta, \Delta), j = 1, \dots, 13; i = 1, \dots, 5$ by the algorithm described in Lii and Helland (1981). Here we set $\Delta = 18\pi/128 = 0.442$. Our final bispectrum estimate is

$${}_{640}b(j\Delta, \Delta) = 1/5 \sum_{i=1}^5 {}_{128}b^{(i)}(j\Delta, \Delta).$$

Compute

$$\theta_n(j) = \arg\{{}_n\hat{b}(j\Delta, \Delta)\} = \arctan\{\text{Im } {}_n\hat{b}(j\Delta, \Delta)/\text{Re } {}_n\hat{b}(j\Delta, \Delta)\}$$

by taking the principal value as well as

$$-H_n(j\Delta) \equiv -H_n(\lambda_j) = \sum_{i=1}^{j-1} \theta_n(i), \quad j = 2, \dots, 14.$$

Let

$$H_n(0) \equiv H_n(\lambda_0) = 0,$$

$$H_n(\Delta) \equiv H_n(\lambda_1) = \frac{1}{2}H_n(\lambda_2),$$

$$H_n(\pi) = \frac{1}{2}\{H_n(7\Delta) + H_n(8\Delta)\} = \frac{1}{2}\{H_n(3.093) + H_n(3.534)\}$$

and

$$\hat{c} = -H_n(\pi)/\pi = -\delta.$$

Recall that \hat{c} is an estimate of

$$c = \lim_{\Delta \rightarrow 0} \frac{h(\Delta)}{\Delta} = h'(0)$$

up to an integer. From formula (13) we compute the α_k 's. A standard smoothed periodogram with uniform weights and bandwidth Δ was used to compute $f_n(\lambda)$ as an estimate of the spectrum $f(\lambda)$ of $\{x_t\}$; $f_n(0)$ is obtained by a linear extrapolation. These examples are as follows:

MODEL: $x_t = v_t + \alpha_1 v_{t-1} + \alpha_2 v_{t-2}$ with four cases specified in Table 1.

Here and from this point on all $\hat{\alpha}_k$'s are adjusted by sign and index shift.

Table 2 gives the details of the computations for Case 2. Figure 1 compares the theoretical h with the estimated h .

As an indication of discretization error, notice that if we use the exact $h(\lambda)$ and $|\alpha(e^{-i\lambda})|$ instead of estimated ones, we get

$$\hat{\alpha}_0 = 0.9136, \quad \hat{\alpha}_1 = -2.247, \quad \hat{\alpha}_3 = 0.5977.$$

When we increase the number of parameters to four, we observe qualitatively the same type of result as in the three parameter case. But generally speaking, the instability is increased. We give one example. The model is

$$x_t = v_t - 4.25v_{t-1} + 4.75v_{t-2} - 0.938v_{t-3}$$

with roots $\frac{2}{3}, \frac{2}{5}, 4$ and we get

$$\hat{\alpha}_0 = 1.106, \quad \hat{\alpha}_1 = -3.572, \quad \hat{\alpha}_2 = 4.293, \quad \hat{\alpha}_3 = -0.9762.$$

These simple examples indicate that one can estimate the unknown coefficients reasonably well and one is able to discriminate between different models even though they have the same spectral structure.

TABLE 1
Coefficients, roots and estimated coefficients for four cases

Case	Coefficients			Roots		Estimated coefficients		
	α_0	α_1	α_2	r_1	r_2	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\alpha}_2$
1	1.0	-0.833	0.167	2.0	3.0	0.9593	-0.5816	0.1158
2	1.0	-2.333	0.667	0.5	3.0	0.7164	-2.175	0.7605
3	1.0	-3.50	1.50	2.0	0.333	0.7561	-3.334	1.778
4	1.0	-5.0	6.0	0.5	0.333	0.9603	-3.904	4.966

TABLE 2
Computations for Case 2

λ	Length $ \alpha(e^{-\lambda}) $	Est. Length	Argument $h(\lambda)$	Argument by Sum $-H_n(\lambda)$	Argument at J $\theta_n(\lambda)$	Adjusted $-H_n(\lambda) +$ $\hat{c}\lambda$	$-H_n(\lambda)$ $+ (\hat{c} - 1)\lambda$
0	0.6667	0.473	0.0000	0	0	0	0
0.442	0.8389	0.894	-0.6125	-0.2713	0	0.0904	-0.352
0.884	1.3018	1.315	-1.0828	-0.5425	-0.5425	0.1808	-0.703
1.325	1.9553	1.888	-1.4915	-0.4851	0.0574	-0.0575	-1.382
1.767	2.6785	2.748	-1.8895	-0.6904	-0.2053	-0.0331	-1.80
2.209	3.3369	3.348	-2.2893	-0.6481	0.0423	-0.2563	-2.465
2.651	3.8064	3.520	-2.6920	-1.0097	-0.3615	-0.0756	-2.727
3.093	3.9980	3.809	-3.0966	-1.0884	-0.0787	-0.1778	-3.27
3.534	3.8752	3.951	-3.5014	-1.4842	-0.3958	0.0371	-3.497
3.976	3.4614	3.297	-3.9047	-1.6797	-0.1956	0.0518	-3.924
4.418	2.8351	2.802	-4.3051	-1.5538	0.1259	-0.2550	-4.693
4.860	2.1146	2.168	-4.7031	-1.8428	-0.2890	-0.1469	-5.007
5.301	1.4344	1.429	-5.1070	-1.8785	-0.0357	-0.2921	-5.593
5.743	0.9199	0.936	-5.5558	-2.3478	-0.4693	-0.0037	-5.747
6.185	0.6755	0.443	-6.1366	-2.6185	-0.2706	0.0861	-6.099

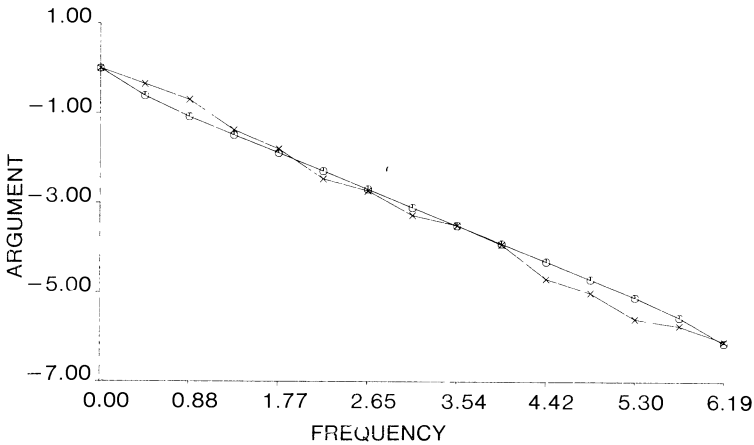


FIG. 1. Theoretical h and estimated h for case 2 denoted by o and x .

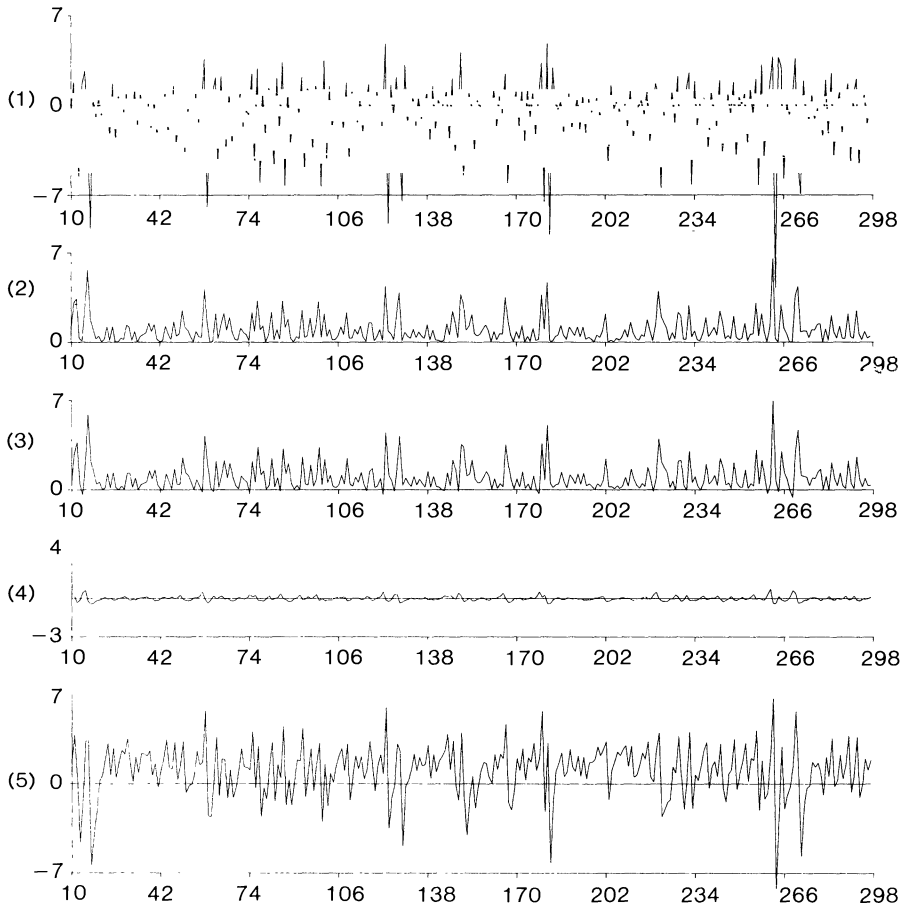


FIG. 2. Deconvolution for case 2. (1) Moving average process x_t , $t = 10$ to $t = 298$, (2) exponential v_t sequence that generates x_t , (3) estimated \hat{v}_t as deconvolved by our procedure, (4) $v_t - \hat{v}_t$, (5) deconvolution of x_t using minimum phase assumption.

To illustrate our deconvolution procedure we consider the moving average

$$x_t = v_t - 2.333v_{t-1} + 0.667v_{t-2}$$

with the v_t 's independent exponential variables with mean one. The v_t 's are generated as pseudo-random variates. The estimated model (using our bispectral techniques) is

$$x_t = 0.7164v_t - 2.175v_{t-1} + 0.7605v_{t-2}.$$

Let $\hat{\alpha}(B) = 0.7164 - 2.175B - 0.7605B^2$. Set

$$(18) \quad \hat{v}_t = \sum_{j=-9}^9 \hat{\alpha}_j x_{t-j} \cong \frac{1}{\hat{\alpha}(B)} x_t.$$

The first line of Figure 2 is the moving average process x_t as generated for $t = 10$ to $t = 298$. The second line is the independent exponential v_t sequence that generates x_t for the same t range. Line 3 shows the estimated \hat{v}_t as deconvolved by our procedure. The fourth line is the difference $v_t - \hat{v}_t$. The last line is a deconvolution of x_t using the assumption of minimum phase. The sample mean of $v_t - \hat{v}_t$ is .00032 and the sample variance .030. Notice

how our deconvolution has generally preserved the shape of the original v_t sequence while the minimum phase deconvolution has no resemblance to the original v_t sequence.

A general way to find the deconvolution weights can be described as follows. We have an estimate

$$\hat{\alpha}(e^{-i\lambda}) = \sqrt{2\pi\hat{f}_n(\lambda)} \exp\{i(H_n(\lambda) + \hat{c}\lambda)\}.$$

We compute $b(e^{-i\lambda}) = \hat{\alpha}^{-1}(e^{-i\lambda})$. Then

$$\hat{\omega}_k = \frac{1}{2\pi} \int_0^{2\pi} b(e^{-i\lambda}) e^{ik\lambda} d\lambda$$

give the coefficients of the series expansion of $\hat{\alpha}^{-1}(e^{-i\lambda})$ which are the weights desired in deconvolution. This method is general. We do not require knowledge of the order of the moving average process and there is no need to compute the roots from the estimated coefficients and how the roots are distributed is irrelevant.

APPENDIX 1

In this appendix we consider the relationship between the coefficients of a polynomial and the roots of the polynomial, at least locally. The polynomial is

$$\sum_{j=0}^p a_j z^j = \prod_{j=1}^p (z - z_j)$$

where the roots are $z_j, j = 1, \dots, p$, and the coefficients $a_j, j = 0, 1, \dots, p$, with $a_p = 1$. It is well known that

$$\begin{aligned} a_{p-1} &= -\sum_j z_j, & a_{p-2} &= \sum_{j \neq k} z_j z_k, \\ a_{p-3} &= -\sum_{j \neq k \neq \ell} z_j z_k z_\ell, \dots, & a_0 &= (-1)^p z_1 \dots z_p. \end{aligned}$$

Let us consider the relationship between the differentials of the coefficients a_j and the differentials of the roots z_k . Now

$$\begin{aligned} \frac{\partial a_{p-1}}{\partial z_\ell} &= -1, & \frac{\partial a_{p-2}}{\partial z_\ell} &= \sum_{j \neq \ell} z_j = -a_{p-1} - z_\ell, \\ \frac{\partial a_{p-3}}{\partial z_\ell} &= -\sum_{j \neq k \neq \ell} z_j z_k = -a_{p-2} - 2a_{p-1} z_\ell - 2z_\ell^2, \end{aligned}$$

$\ell = 1, \dots, p$. Thus

$$\begin{aligned} da_{p-1} &= -\sum_j dz_j, & da_{p-2} &= -a_{p-1} \sum_j dz_j - \sum_j z_j dz_j, \\ da_{p-3} &= -a_{p-2} \sum_j dz_j - 2a_{p-1} \sum_j z_j dz_j - 2 \sum_j z_j^2 dz_j, \dots \end{aligned}$$

and this can be written in matrix notation as

$$\begin{pmatrix} da_{p-1} \\ da_{p-2} \\ da_{p-3} \\ \vdots \end{pmatrix} = UV \begin{pmatrix} dz_1 \\ dz_2 \\ dz_3 \\ \vdots \end{pmatrix}$$

where

$$U = \begin{pmatrix} -1 & 0 & 0 & 0 \\ -a_{p-1} & -1 & 0 & 0 \\ & -2a_{p-1} & -2 & 0 \\ & & \dots & \end{pmatrix}$$

and

$$V = \begin{pmatrix} 1 & 1 & 1 & \dots & 1 \\ z_1 & z_2 & z_3 & \dots & z_n \\ z_1^2 & z_2^2 & z_3^2 & \dots & z_n^2 \\ & & \dots & & \\ & & & & \dots \end{pmatrix}.$$

U is a triangular nonsingular matrix and V is the Vander Monde matrix. V is nonsingular as long as the roots z_i are distinct.

APPENDIX 2

The object is to remark on some aspects of the asymptotic behavior of estimates of the c_k 's in the context of a general linear process. Let x_t be a nonGaussian linear process (1) with

$$E v_t \equiv 0, \quad E v_t^2 \equiv 1, \quad E v_t^3 \equiv \gamma_3$$

$$\text{cum}(v_t^4) = \gamma_4, \quad \text{cum}(v_t^6) = \gamma_6.$$

Set

$$y_t = x_t^2.$$

For convenience we introduce

$$g_u = \text{Cov}(x_t, y_{t-u}), \quad r_u = \text{Cov}(x_t, x_{t-u}), \quad h_u = \text{Cov}(y_t, y_{t-u}).$$

Consider the estimates

$$\hat{g}_a = \frac{1}{N} \sum_{t=1}^N x_t y_{t+a}$$

of $E x_t y_{t+a}$. Then

$$\text{Cov}(\hat{g}_a, \hat{g}_b) = \frac{1}{N^2} \sum_{t,\tau=1}^N \text{Cov}(x_t y_{t+a}, x_\tau y_{\tau+b})$$

where

$$\text{Cov}(x_t y_{t+a}, x_\tau y_{\tau+b}) = r_{t-\tau} c_{t-\tau+a-b} + g_{t-\tau-b} g_{t-\tau+a} + \text{cum}(x_t, x_\tau, y_{t+b}) \text{cum}(y_{t+a})$$

$$+ \text{cum}(x_t, x_\tau, y_{t+a}) \text{cum}(y_{t+b}) + \text{cum}(x_t, y_{t+a}, x_\tau, y_{\tau+b}).$$

Now

$$r_u = \sum_j \alpha_j \alpha_{j-u}, \quad g_u = \gamma_3 \sum_j \alpha_j \alpha_{j-u}^2, \quad h_u = \gamma_4 \sum \alpha_j^2 \alpha_{j-u}^2 + 2 (\sum_j \alpha_j \alpha_{j-u})^2.$$

Further

$$\text{cum}(y_{t+a}) = \sum_j \alpha_j^2,$$

$$\text{cum}(x_t, x_\tau, y_{t+b}) = \gamma_4 \sum_j \alpha_j \alpha_{j+b}^2 \alpha_{j+(t-\tau)} = k(b, t - \tau),$$

$$\text{cum}(x_t, x_\tau, y_{t+a}) = \gamma_4 \sum_j \alpha_j \alpha_{j+a}^2 \alpha_{j+(t-\tau)} = k(a, \tau - t),$$

$$\text{cum}(x_t, y_{t+a}, x_\tau, y_{\tau+b}) = \gamma_6 \sum_u \alpha_u \alpha_{u+(t-\tau)} \alpha_{u+a}^2 \alpha_{u+b+(t-\tau)}^2$$

$$+ 4\gamma_4 (\sum_u \alpha_u \alpha_{u+(t-\tau)} \alpha_{u+a} \alpha_{u+b+(t-\tau)}) (\sum_u \alpha_u \alpha_{u+(t-\tau)+b-a})$$

$$+ 2\gamma_3^2 (\sum_u \alpha_u \alpha_{u+(t-\tau)} \alpha_{u+a}) (\sum \alpha_u^2 \alpha_{u+(t-\tau)+a-b})$$

$$+ 2\gamma_3^2 (\sum_n \alpha_n \alpha_{n+(t-\tau)} \alpha_{n+b}) (\sum \alpha_u^2 \alpha_{u+(t-\tau)+b-2})$$

$$= k(a, b, t - \tau).$$

It is then clear that

$$\lim_{N \rightarrow \infty} N \text{Cov}(\hat{g}_a, \hat{g}_b) = \sum_s r_s c_{s+a-b} + \sum_s g_{s-b} g_{s+a} + r_0 \sum_s \{k(b, s) + k(a, s)\} + \sum_s k(a, b, s).$$

Under the assumption that $\sum |\alpha_r| < \infty$, with a truncation argument like that employed in Anderson (1970), one can show that

$$\sqrt{N} (\hat{g}_a - g_a)$$

are asymptotically normally distributed with covariance structure given by the preceding formula.

Acknowledgment. Research of the second author was carried out in part in the University of Manchester Institute of Science and Technology.

REFERENCES

- ANDERSON, T. W. (1971). *The Spectral Analysis of Time Series*. Wiley, New York.
- BASAWA, I. V., FEIGEN, P. D. and HEYDE, C. C. (1975). Asymptotic properties of maximum likelihood estimates for stochastic processes. *Sankhya, Ser. A* **38** 259–270.
- BOX, G. E. P. and JENKINS, G. M. (1976). *Time Series Analysis, Forecasting and Control*. Holden-Day, San Francisco.
- BRILLINGER, D. and ROSENBLATT, M. (1967). Asymptotic theory of estimates of k th order spectra. In *Spectral Analysis of Time Series* (ed. B. Harris) 153–188. Wiley, New York.
- DONOHO, D. (1981) On minimum entropy deconvolution. In *Applied Time Series Analysis II* (ed. D. F. Findley) Academic, New York. 565–608.
- GODFREY, R. and ROCCA, F. (1981) Zero memory nonlinear deconvolution. *Geophys. Prospect.* **29** 189–228.
- HUZII, M. (1981) Estimation of coefficients of an autoregressive process by using a higher order moment. *J. Time Series Analysis* **2(2)**, 87–93.
- HENRICI, P. (1974). *Applied and Computational Complex Analysis*, vol. 1. Wiley, New York.
- JENKINS, G. M. and WATTS, D. G. (1968). *Spectral Analysis and Its Applications*. Holden-Day, San Francisco.
- LII, K. S. and HELLAND, K. N. (1981). Cross-bispectrum computation and variance estimation. *ACM Trans. Math. Software* **7** No. 3.
- ROSENBLATT, M. (1974). *Random Processes*. Springer, Berlin.
- ROSENBLATT, M. (1980). Linear processes and bispectra. *J. Appl. Probab.* **17** 265–270.
- WIGGINS, R. A. (1978). Minimum entropy deconvolution. *Geoexploration* **17**.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA
RIVERSIDE, CALIFORNIA 92521

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF CALIFORNIA, SAN DIEGO
LA JOLLA, CALIFORNIA 92093