The Annals of Statistics
1984, Vol. 12, No. 2, 646–664

SIGNAL EXTRACTION FOR NONSTATIONARY TIME SERIES

BY WILLIAM BELL[1]

Bureau of the Census

The nonstationary signal extraction problem is to estimate s_t given observations on $z_t = s_t + n_t$ (signal plus noise) when either s_t or n_t or both is nonstationary. Homogeneous or explosive nonstationary time series described by models of the form $\delta(B)z_t = w_t$ where $\delta(B)$ has zeroes on or inside the unit circle and w_t is stationary are considered. For certain cases, approximate solutions to the nonstationary signal extraction problem have been given by Hannan (1967), Sobel (1967), and Cleveland and Tiao (1976). The paper gives exact solutions in the forms of expressions for $E(s_t \mid \{z_t\})$ and $\text{Var}(s_t \mid \{z_t\})$ (assuming normality) under two sets of alternative assumptions regarding the generation of z_t, s_t, and n_t. Extensions to signal extraction with a finite number of observations, to the nonGaussian case, and to the multivariate case are discussed.

1. Introduction.

Suppose that

$$z_t = s_t + n_t, \quad t = 0, \pm 1, \pm 2, \cdots \tag{1.1}$$

where z_t is an observable time series and s_t and n_t are unobservable signal and noise time series. The signal extraction problem is to find the best (e.g., minimum mean squared error) estimate of s_t for any fixed t given the observed data. The problem for the case where s_t and n_t are independent and stationary was solved independently by Kolmogorov (1939, 1941) and Wiener (1949). This paper deals with the signal extraction problem when either s_t or n_t or both are nonstationary.

NOTATION. If y_t is a time series we shall write $\{y_t\}$ for its entire doubly infinite realization. The segment of the time series between and including any two time points $i < j$, shall be denoted by $\mathbf{y}_{(j)}^{(i)} = (y_i, y_{i+1}, \cdots, y_j)'$, where prime denotes transpose, or by $\mathbf{y}_{(j)} = (y_1, \cdots, y_j)'$ if $i = 1$. If y_t is stationary with autocovariances $\gamma_y(k) = \text{Cov}(y_t, y_{t+k})$ that are absolutely summable ($\sum_{-\infty}^{\infty} \mid \gamma_y(k) \mid < \infty$) then the spectral density of y_t is

$$f_y(\lambda) = (2\pi)^{-1} \sum_{-\infty}^{\infty} \gamma_y(k) e^{-i\lambda k} = (2\pi)^{-1} \gamma_y(e^{-i\lambda}), \tag{1.2}$$

where $\gamma_y(\zeta) = \sum_{-\infty}^{\infty} \gamma_y(k) \zeta^k$ is the covariance generating function (CGF) of y_t and ζ denotes a dummy complex variable. It will be convenient to let $[\gamma_y]_{(m)}$ represent

Received July 1982; revised August 1983.

[1] This research was supported by the National Science Foundation (grant SES 84-01460 "Research to Improve the Government-Generated Social Science Data Base") and the U.S. Bureau of the Census through the author's participation in the American Statistical Association/Census Bureau fellowship program. The opinions expressed are those of the author, and do not necessarily reflect those of NSF, ASA, or the Census Bureau. The author is grateful to David Findley for many useful discussions and for pointing out the results of his paper (Findley, 1984).

AMS 1970 *subject classifications.* Primary 62M10; secondary 60G35.

Key words and phrases. Nonstationary time series, signal extraction, conditional expectation, conditional variance, starting values.

the variance matrix of any segment, $\mathbf{y}_{(t+m)}^{(t)}$, of the stationary series y_t of length m:

$$[\gamma_y]_{(m)} = \mathrm{Var}(\mathbf{y}_{(t+m)}^{(t)}) = \begin{bmatrix} \gamma_y(0) & \cdots & \gamma_y(m-1) \\ \vdots & & \vdots \\ \gamma_y(m-1) & \cdots & \gamma_y(0) \end{bmatrix}.$$

If we apply an absolutely summable linear filter $\alpha(B) = \sum_{-\infty}^{\infty} \alpha_j B^j$, where B is the backshift operator, to y_t, the resulting time series has autocovariances $\alpha(B)\alpha(F)\gamma_y(k)$, and we write $[\alpha(B)\alpha(F)\gamma_y]_{(m)}$ for the variance matrix of m successive observations on $\alpha(B)y_t$.

ASSUMPTIONS. For convenience, all random variables in this paper will be assumed to have zero mean. We shall initially deal only with univariate time series that, except where stated otherwise, are jointly normal. Extensions of the results to multivariate and nonnormal time series are discussed in Section 6.

With respect to the decomposition (1.1), we assume that while s_t and n_t can be nonstationary,

$$\delta_s(B)s_t = u_t \text{ and } \delta_n(B)n_t = v_t \tag{1.3}$$

are stationary time series independent of each other, where $\delta_s(B) = 1 - \delta_{s1}B - \cdots - \delta_{s,ds}B^{ds}$ is a polynomial of degree ds in the backshift operator B, and $\delta_n(B)$ is a similar polynomial of degree dn in B. We assume all the zeroes of $\delta_s(\zeta)$ and $\delta_n(\zeta)$ lie on or inside the unit circle—if $\delta_s(\zeta)$ has a zero outside the unit circle it corresponds to a factor of $\delta_s(B)$ that can be inverted and incorporated on the right hand side with u_t, and similarly for $\delta_n(\zeta)$. (Anderson (1971, pages 170–171) notes that factors corresponding to zeroes inside the unit circle can be reversed in time to factors with zeroes outside the unit circle, thus becoming part of the stationary part of the model, although operating backwards in time. We shall not allow this here since it corresponds to assumptions about the generation of time series that are different from those we shall use—which are discussed in Sections 2 and 3.) We let

$$\delta(B) = \delta_c(B)\delta_s^*(B)\delta_n^*(B) \tag{1.4}$$

where $\delta_c(B)$ is the product of the common factors in $\delta_s(B)$ and $\delta_n(B)$, $\delta_s^*(B) = \delta_s(B)/\delta_c(B)$, and $\delta_n^*(B) = \delta_n(B)/\delta_c(B)$. We let d denote the degree of $\delta(B)$. Define $w_t = \delta(B)z_t$, which from (1.3) and (1.4) is given by

$$\delta(B)z_t = w_t = \delta_n^*(B)u_t + \delta_s^*(B)v_t, \tag{1.5}$$

so w_t is a stationary time series with CGF

$$\gamma_w(\zeta) = \delta_n^*(\zeta)\delta_n^*(\zeta^{-1})\gamma_u(\zeta) + \delta_s^*(\zeta)\delta_s^*(\zeta^{-1})\gamma_v(\zeta). \tag{1.6}$$

We assume that w_t is purely nondeterministic so it has an infinite moving average (Wold) representation

$$w_t = \Psi(B)a_t = \textstyle\sum_0^{\infty} \Psi_j a_{t-j}. \tag{1.7}$$

We further assume that the $\gamma_w(k)$ are absolutely summable, and that $\gamma_w(\zeta)$ has no zeroes on the unit circle (which means $f_w(\lambda) = (2\pi)^{-1}\gamma_w(e^{-i\lambda})$, the spectral density of w_t, is never zero). Then (Brillinger 1975, pages 78–79) w_t has an infinite autoregressive representation (invertibility)

$$\Pi(B)w_t = (1 - \textstyle\sum_1^\infty \Pi_j B^j)w_t = a_t \tag{1.8}$$

with $\Pi(B) = \Psi(B)^{-1}$. We also make these assumptions about u_t and v_t. These assumptions will hold, in particular, if u_t, v_t, and w_t all follow stationary, invertible, autoregressive–moving average (ARMA) models.

If $\delta(\zeta)$ and $\gamma_w(\zeta)$ have a common zero, then Findley (1982) shows that the model $\delta(B)z_t = \Psi(B)a_t$ can be simplified by cancelling a factor from both sides and adding to the right hand side a deterministic term that is annihilated by the cancelled factor. Thus, we assume there are no common zeroes within the pairs $\{\delta(\zeta), \gamma_w(\zeta)\}$, $\{\delta_s(\zeta), \gamma_u(\zeta)\}$, and $\{\delta_n(\zeta), \gamma_v(\zeta)\}$, and that any deterministic terms have been subtracted out. Findley (1982) also shows that then $\delta(B)$ is the minimal polynomial that renders z_t stationary, in that it divides any other such polynomial.

PREVIOUS WORK. Signal extraction has been used with nonstationary time series in such areas as actuarial graduation (by Whittaker, see Whittaker and Robinson, 1944, pages 303–316), smoothing (Tiao and Hillmer, 1978), and seasonal adjustment (Cleveland, 1972, Burman, 1980, and Hillmer and Tiao, 1982). Typically, the solution for the stationary case has been borrowed and used in the nonstationary case. If s_t and n_t are both stationary, normal, and independent of each other, and the entire realization $\{z_t\}$ is available, the solution of Kolmogorov and Wiener may be written (Fuller 1976, page 170)

$$E(s_t \mid \{z_t\}) = \gamma_s(F)\gamma_z(F)^{-1}z_t \tag{1.9}$$

where $\gamma_z(\zeta)$ is the CGF of z_t. In the nonstationary case $\gamma_s(\zeta)$ and $\gamma_z(\zeta)$ will not exist, but proceeding formally from (1.3) and (1.4), we are led to consider using R_t defined by

$$R_t = \delta_n^*(B)\delta_n^*(F)\gamma_u(F)\gamma_w(F)^{-1}z_t. \tag{1.10}$$

There has been work done on the properties of R_t in the nonstationary case. Hannan (1967) and Sobel (1967) considered the case where n_t is stationary and $\delta_s(\zeta)$ has all its zeroes on the unit circle. Hannan (1967) showed that R_t minimizes the mean squared error in the class of linear estimators that perfectly predict any sequence p_t that is annihilated by $\delta_s(B)$ (i.e., $\delta_s(B)p_t = 0$). Sobel (1967) established that R_t asymptotically approaches (as $t \to \infty$) the best linear estimator. Cleveland and Tiao (1976) obtained an approximation to $E(s_t \mid z_m, \cdots, z_{m+N})$ for large $m > 0$ when s_t and n_t follow ARIMA models, and $\delta(B)$ is allowed to have factors $(1 - B)^{d_1}(1 - B^c)^{d_2}$. They then noted their approximation approaches R_t as the number of observations, N, grows large, for t not near m or $m + N$. Pierce (1979) examined the behavior of the error series $s_t - R_t$, showing it is stationary if $\delta_s(\zeta)$ and $\delta_n(\zeta)$ have no common zeroes and obtaining its spectral density, and hence $\mathrm{Var}(s_t - R_t)$ in this case. Hannan (1967) had also obtained these for the case of stationary n_t.

The above results leave several questions unanswered. Notice that the exact signal extraction solution, $E(s_t \mid \{z_t\})$, has not been obtained. Also, for some cases, notably when s_t or n_t is explosively nonstationary or when $\delta_s(\zeta)$ and $\delta_n(\zeta)$ have a common zero, not even an approximate solution is available. In Section 4 we obtain the exact solutions to the nonstationary signal extraction problem under two sets of alternative assumptions, which we label A and B. It turns out that $E(s_t \mid \{z_t\}) = R_t$ under our Assumption A when $\delta_s(\zeta)$ and $\delta_n(\zeta)$ have no common zeroes; otherwise, other terms must be added to R_t to get $E(s_t \mid \{z_t\})$.

The assumptions we make have to do with the generation of the time series s_t, n_t, and z_t, and their starting values, something that is often ignored in the literature. That our results depend on these assumptions shows that they should not be ignored. We discuss generation of time series in Section 2 and our Assumptions A and B in Section 3. After the results for $E(s_t \mid \{z_t\})$ in Section 4, we obtain results for $\mathrm{Var}(s_t \mid \{z_t\})$ in Section 5. It is seen that the usual results (e.g., Pierce, 1979) give $\mathrm{Var}(s_t \mid \{z_t\})$ under Assumption A when $\delta_s(\zeta)$ and $\delta_n(\zeta)$ have no common zeroes, and otherwise other expressions as given are needed. Finally, in Section 6 we extend our results to signal extraction for a finite set of observations, for nonGaussian time series, for series with known starting values, and for multivariate time series.

For some of the results given here, details are omitted from the proofs. The interested reader is urged to consult Bell (1982) for detailed proofs.

2. Generation of nonstationary time series. A purely nondeterministic stationary time series w_t may be viewed as arising for all t from the Wold decomposition (1.7), given a white noise series $\{a_t\}$. If the zeroes of $\delta(\zeta)$ were outside the unit circle (so z_t would be stationary) then we could write

$$z_t = (1 + \xi_1 B + \xi_2 B^2 + \cdots) w_t \tag{2.1}$$

where $\xi(B) = (1 + \xi_1 B + \xi_2 B^2 + \cdots) = \delta(B)^{-1}$. The ξ's can be obtained by equating coefficients of $B^0, B^1, B^2, \cdots$ in $(\xi_0 + \xi_1 B + \xi_2 B^2 + \cdots)\,\delta(B) = 1$, so that

$$\xi_0 = 1 \quad \xi_i = \textstyle\sum_{k=1}^{\min(d,i)} \delta_k \xi_{i-k} \quad i \ge 1. \tag{2.2}$$

Unfortunately, when $\delta(\zeta)$ contains zeroes that lie on or inside the unit circle, (2.1) will not converge, and z_t cannot be viewed as being generated this way.

To produce $\{z_t\}$ in the nonstationary case we need, in addition to $\{w_t\}$, a suitable set of starting values for z_t. Since $\delta(B)$ is of order d we need d starting values, which we will assume are $(z_1, \cdots, z_d)' = \mathbf{z}_*$. Given $\mathbf{z}_*$ and $\{w_t\}$, the remaining z_t's are easily generated recursively from

$$z_t = \delta_1 z_{t-1} + \cdots + \delta_d z_{t-d} + w_t \quad t > d \tag{2.3}$$

$$z_t = \delta_d^{-1}(z_{t+d} - \delta_1 z_{t+d-1} - \cdots - \delta_{d-1} z_{t+1} - w_{t+d}), \quad t \le 0. \tag{2.4}$$

(Notice $\delta_d^{-1} \ne 0$ or $\delta(B)$ would not be of degree d.) In the stationary case there is a one-to-one correspondence between the collections of random variables $\{z_t\}$ and $\{w_t\}$ through $\delta(B) z_t = w_t$ and (2.1), while in the nonstationary case there is a one-to-one correspondence between $\{z_t\}$ and $\{\mathbf{z}_*, \{w_t\}\}$ through $\delta(B)z_t = w_t$ and (2.3) and (2.4).

We now obtain a representation of z_t for $t > 0$ in the nonstationary case that is analogous to (2.1). Notice the ξ_i's can still be defined by (2.2). For our representation we need the following quantities $A_{j,t}$, defined for $t \geq 1$ by (using $\xi_i = 0$ for $i < 0$)

$$\begin{aligned} A_{1,t} &= \xi_{t-1} - \xi_{t-2}\delta_1 - \cdots - \xi_{t-d}\delta_{d-1} \\ &\vdots \\ A_{d-1,t} &= \xi_{t-d+1} - \xi_{t-d}\delta_1 \\ A_{d,t} &= \xi_{t-d} \,. \end{aligned} \tag{2.5}$$

From $(\xi_0 + \xi_1 B + \xi_2 B^2 + \cdots)\delta(B) = 1$ we see $\delta(B)\xi_i = 0$ for $i \geq 1$. Using this fact, it can be shown that for $t = 1, \cdots, d$, $A_{j,t} = 1$ when $t = j$ and is 0 otherwise, and $A_{j,d+1} = \delta_{d+1-j}$. Also, $\delta(B)\xi_i = 0$ for $i \geq 1$ immediately shows $A_{j,t} = \delta_1 A_{j,t-1} + \cdots + \delta_d A_{j,t-d}$ for $t > d$, so the $A_{j,t}$'s may be computed directly without computing the ξ_i's. A useful result relating the ξ_i's and $A_{j,t}$'s is as follows.

LEMMA 1. *For $t > d$*

$$\left(\textstyle\sum_{i=0}^{t-d-1} \xi_i B^i\right)\delta(B) = 1 - \textstyle\sum_{i=1}^{d} A_{i,t} B^{t-i}.$$

PROOF. On the left hand side above, the coefficient of B^0 is 1 and that of B^j, for $j = 1, \cdots, t-d-1$, is $\delta(B)\xi_i = 0$ (using $\xi_i = 0$ for $i < 0$). The coefficient of B^{t-i} for $i = 1, \cdots, d$ is

$$-\delta_d \xi_{t-d-i} - \cdots - \delta_{d+1-i}\xi_{t-d-1} = \delta_{d-i}\xi_{t-d} + \cdots + \delta_1 \xi_{t-i-1} - \xi_{t-i} = -A_{it}. \quad \square$$

This result allows us to easily prove the following.

THEOREM 1. *Let $\delta(B)z_t = w_t$ where $\delta(B)$ is of degree d and z_t is generated from $\{w_t\}$ and starting values $\mathbf{z}_* = (z_1, \cdots, z_d)'$. Then, for $t > d$*

$$z_t = \mathbf{A}_t' \mathbf{z}_* + \textstyle\sum_{i=0}^{t-d-1} \xi_i w_{t-i}$$

where $\mathbf{A}_t' = (A_{1,t}, \cdots, A_{d,t})$. The result holds for $t = 1, \cdots, d$ if the last term is interpreted as zero.

PROOF. By Lemma 1, for $t > d$

$$\begin{aligned} z_t - \mathbf{A}_t'\mathbf{z}_* = \left(1 - \textstyle\sum_{i=1}^{d} A_{i,t}B^{t-i}\right)z_t &= \left(\textstyle\sum_{i=0}^{t-d-1} \xi_i B^i\right)\delta(B)z_t \\ &= \textstyle\sum_{i=0}^{t-d-1} \xi_i w_{t-i}. \quad \square \end{aligned}$$

As solutions to $\delta(B)A_{jt} = 0$, the behavior of the A_{jt}'s as t increases will depend on the zeroes of $\delta(\zeta)$. If any lie inside the unit circle the A_{jt}'s will exhibit explosive behavior, while if they all lie on the unit circle the A_{jt}'s will either remain bounded (all zeroes distinct) or grow in polynomial fashion (repeated zeroes). The same comments apply to ξ_i as i increases.

We could obtain an analogous backward representation for z_t for $t \le 0$ involving the starting values $\mathbf{z}_*$ and $w_j, j \le d$. The coefficients of the w_j's would be obtained by formally inverting the operator $1 + (\delta_{d-1}/\delta_d)F + \cdots + (\delta_1/\delta_d)F^{d-1} - (1/\delta_d)F^d$ (see 2.4). An important special case occurs when all the zeroes of $\delta(B)$ lie *on* the unit circle. In this case it can be shown that $\delta(B) = (-1)^r B^d \delta(F)$ where r is the number of times the factor $(1 - B)$ occurs in $\delta(B)$. From this relation we can write $\delta(F)z_t = x_t$ where $x_t = (-1)^r w_{t+d}$. The $A_{j,t}$'s and ξ_i's that we need will thus be the same as above, and using the starting values $z_d, \cdots, z_1$, now going backwards in time, we get the following backward representation for z_t for $t \le 0$:

$$z_t = (A_{d,d+1-t}, \cdots, A_{1,d+1-t})\mathbf{z}_* + \textstyle\sum_{i=0}^{-t} \xi_i x_{t+i}. \tag{2.6}$$

3. Assumptions about starting values in signal extraction. In doing signal extraction we must make assumptions about the generation of the three time series $\{z_t\}$, $\{s_t\}$, and $\{n_t\}$. Generating these series is equivalent to generating their starting values and the series $\{w_t\}$, $\{u_t\}$, and $\{v_t\}$ (see (1.3)). We shall always assume that the series $\{u_t\}$ is generated independently of the series $\{v_t\}$, and that each w_t is then obtained from (1.5). The starting values we need are $\mathbf{z}_* = (z_1, \cdots, z_d)'$, $\mathbf{s}_* = (s_1, \cdots, s_{ds})'$, and $\mathbf{n}_* = (n_1, \cdots, n_{dn})'$, where ds is the order of $\delta_s(B)$ and dn is the order of $\delta_n(B)$. There are thus $d + ds + dn = 2d + dc$ starting values, where dc, the order of $\delta_c(B)$, is the number of common factors in $\delta_s(B)$ and $\delta_n(B)$.

Notice that Theorem 1 implies that

$$\begin{aligned} s_t &= \mathbf{A}_t^{s\prime}\, \mathbf{s}_* + \textstyle\sum_{i=0}^{t-ds-1} \xi_i^s u_{t-i}, \quad t > 0 \\ n_t &= \mathbf{A}_t^{n\prime}\, \mathbf{n}_* + \textstyle\sum_{i=0}^{t-dn-1} \xi_i^n v_{t-i}, \quad t > 0 \end{aligned} \tag{3.1}$$

where the ξ_i^s and $\mathbf{A}_t^s$ are obtained from $\delta_s(B)$ in the same way that ξ_i and $\mathbf{A}_t$ were obtained above from $\delta(B)$, and similarly for the ξ_i^n and $\mathbf{A}_t^n$. From (3.1) we get

$$\mathbf{z}_* = [H_1 H_2]\begin{bmatrix}\mathbf{s}_* \\ \mathbf{n}_*\end{bmatrix} + C_1 \begin{bmatrix} u_{ds+1} \\ \vdots \\ u_d \end{bmatrix} + C_2 \begin{bmatrix} v_{dn+1} \\ \vdots \\ v_d \end{bmatrix} \tag{3.2}$$

where

$$\underset{d\times ds}{H_1} = \begin{bmatrix} I_{ds} \\ \mathbf{A}_{ds+1}^{s\prime} \\ \vdots \\ \mathbf{A}_d^{s\prime} \end{bmatrix} \qquad \underset{d\times dn}{H_2} = \begin{bmatrix} I_{dn} \\ \mathbf{A}_{dn+1}^{n\prime} \\ \vdots \\ \mathbf{A}_d^{n\prime} \end{bmatrix}$$

$$\underset{d\times(d-ds)}{C_1} = \begin{bmatrix} & O_{ds\times(d-ds)} & \\ \xi_0^s & & \\ \vdots & \ddots & \\ \xi_{d-ds-1}^s & \cdots & \xi_0^s \end{bmatrix} \qquad \underset{d\times(d-dn)}{C_2} = \begin{bmatrix} & O_{dn\times(d-dn)} & \\ \xi_0^n & & \\ \vdots & \ddots & \\ \xi_{d-dn-1}^n & \cdots & \xi_0^n \end{bmatrix}$$

which relates the starting values for s_t and n_t ($\mathbf{s}_*$ and $\mathbf{n}_*$) to those for z_t ($\mathbf{z}_*$).

We will assume the starting values are generated in one of two ways.

ASSUMPTION A. $\mathbf{z}_*$ is generated independently of $\{u_t\}$, $\{v_t\}$, and hence $\{w_t\}$. Then $\mathbf{s}_*$ and $\mathbf{n}_*$ are obtained by solving (3.2). When $dc > 0$ the solution will not be unique (see comments below).

ASSUMPTION B. $\mathbf{s}_*$ and $\mathbf{n}_*$ are generated independently of each other and of $\{u_t\}$, $\{v_t\}$, and hence $\{w_t\}$. Then $\{s_t\}$ and $\{n_t\}$ are generated in the same way as (2.3) and (2.4), and $\mathbf{z}_*$ is obtained through $z_t = s_t + n_t$, $t = 1, \cdots, d$.

Bell (1982) shows that the $d \times (ds + dn)$ matrix $[H_1H_2]$ in (3.2) has rank d. Under Assumption A, if $\delta_s(B)$ and $\delta_n(B)$ have no common factors ($dc = 0$), then $ds + dn = d$ and the solution to (3.2) is unique. A simple case of this occurs when n_t is stationary so it requires no starting values ($dn = 0$) and $s_t = z_t - n_t$, $t = 1, \cdots, d$. If $\delta_s(B)$ and $\delta_n(B)$ do have common factors ($dc > 0$), then (3.2) has d equations and $ds + dn = d + dc$ unknowns, so multiple solutions exist. Each solution corresponds to a particular choice of generalized inverse, $[H_1H_2]^-$, of $[H_1H_2]$, so in making Assumption A with $dc > 0$, one must also make an assumption as to which generalized inverse of $[H_1H_2]$ is used in solving (3.2).

Assumptions A and B have different implications. Assumption B implies that s_t and n_j are independent for all t and j, an assumption usually made in signal extraction. This is not the case under Assumption A. For example, if n_t is stationary and $(1 - B)s_t = u_t$, then $s_1 = z_1 - n_1$ and $s_t = s_1 + \sum_{i=0}^{t-2} u_{t-i}$ for $t > 1$, and $\mathrm{Cov}(s_t, n_j) = -\mathrm{Cov}(n_1, n_j)$ for all j, which need not be zero for any j. Under Assumption A the stationary filtered series $u_t = \delta_s(B)s_t$ and $v_t = \delta_n(B)n_t$ are still independent, but correlation between s_t and n_t can be generated through their starting values.

However, Assumption A has one advantage over Assumption B. Under Assumption A, $\mathbf{z}_*$ is assumed independent of $\{w_t\}$, so it is independent of any $a_t = \Pi(B)w_t$. Since $a_{t+\ell}$ is independent of $w_t, w_{t-1}, \cdots$ for any $\ell > 0$, it follows from the expression for z_t in Theorem 1 that, for $t > 0$, z_t and $a_{t+\ell}$ are independent for any $\ell > 0$. This is typically assumed in modeling and forecasting the observed series z_t, but it does not generally hold under Assumption B. For example, under Assumption B suppose $(1 - B)s_t = u_t$ and n_t are both white noise, so $w_t = u_t + (1 - B)n_t$ is moving average of order one, i.e., $w_t = (1 - \theta B)a_t$. Then $z_t = z_1 + \sum_{i=0}^{t-2} w_{t-i}$ with $z_1 = s_1 + n_1$, and $a_{t+\ell} = (1 - \theta B)^{-1}w_{t+\ell} = (1 - \theta B)^{-1}u_{t+\ell} + [1 - (1 - \theta) \sum_{i=1}^{\infty} \theta^{i-1}B^i]\, n_{t+\ell}$, implying $\mathrm{Cov}(z_t, a_{t+\ell}) = \mathrm{Cov}(n_1, n_{t+\ell} - (1 - \theta) \sum_{i=1}^{\infty} \theta^{i-1}n_{t+\ell-i}) = -(1 - \theta)\, \theta^{t+\ell-2}\gamma_n(0)$.

It should be noted that it does not seem possible in general to make assumptions so that z_t and $a_{t+\ell}$ ($\ell > 0$) are independent for all t. Under Assumption A we get this only for $t > 0$. There is an analogous result using a backward representation for z_t (such as (2.6)), which states that, under Assumption A, for $t \leq 0$, z_t is independent of the backward innovation at time $t - \ell$ for $\ell > 0$ ($\tilde{a}_{t-\ell} = \Pi(F)x_{t-\ell}$ in (2.6)).

In Section 4 we obtain signal extraction results under both Assumptions A and B. The results under the two sets of assumptions differ, reflecting the fact that assumptions about starting values are important. A choice between Assumptions A and B for any given problem will depend on the problem. If there is reason to believe that z_t was actually generated by two independent components s_t and n_t, then Assumption B may be preferred. On the other hand, if the components s_t and n_t are really just artificial constructs (as in seasonal adjustment), then Assumption A may have more appeal. We do not intend to suggest that A and B are the only reasonable assumptions; other assumptions could also be used. Rather, the purposes of this paper are (i) to show how exact nonstationary signal extraction results can be obtained given assumptions about the starting values, and (ii) to demonstrate that such assumptions are required to obtain results, a fact that has not previously been appreciated.

4. Signal extraction for nonstationary time series. We now return to the signal extraction problem and obtain a general expression for $E(s_t \mid \{z_t\})$ in the nonstationary case. In subsections 4.1 and 4.2 we obtain specific results under Assumptions A and B, respectively. The results are given for $E(s_t \mid \{z_t\})$ for $t \geq 1$, but analogous results for $t \leq 0$ could be obtained using a backward representation for s_t (see Section 2). In some cases it may be easier to apply the results here to get $E(n_t \mid \{z_t\})$, which can be done by relabeling, and then compute $E(s_t \mid \{z_t\})$ as $z_t - E(n_t \mid \{z_t\})$.

To begin we need the following lemma which is proved in Bell (1982).

LEMMA 2. *Let I and J be index sets, countable or uncountable, and assume Y, $\{X_i\} = \{X_i, i \in I\}$, and $\{W_j\} = \{W_j, j \in J\}$ are jointly normal, with zero means and finite variances. Then*

$$E(Y \mid \{X_i\}, \{W_j\}) = E(Y \mid \{X_i\}) + E(Y \mid \{W_j - E(W_j \mid \{X_i\})\}).$$

From the discussion in Section 2, $E(s_t \mid \{z_t\}) = E(s_t \mid \mathbf{z}_*, \{w_t\})$, so by (3.1), the linearity of conditional expectations, and Lemma 2

$$\begin{aligned} E(s_t \mid \{z_t\}) &= \mathbf{A}_t^{s\prime}[E(\mathbf{s}_* \mid \{w_t\}) + E(\mathbf{s}_* \mid \mathbf{z}_* - E(\mathbf{z}_* \mid \{w_t\}))] \\ &\quad + \textstyle\sum_{i=0}^{t-ds-1} \xi_i^s [E(u_{t-i} \mid \{w_t\}) + E(u_{t-i} \mid \mathbf{z}_* - E(\mathbf{z}_* \mid \{w_t\}))]. \end{aligned} \tag{4.1}$$

Since u_t and w_t are stationary, it is well known that (e.g., Brillinger, 1975, Theorems 8.3.1 and 8.3.2) $E(u_{t-i} \mid \{w_t\}) = \gamma_{uw}(F)\gamma_w(F)^{-1}w_{t-i}$. The stationary signal extraction result (1.9) is a special case of this. From (1.5), $\gamma_{uw}(F) = \delta_n^*(F)\gamma_u(F)$ so from this and (1.10)

$$E(u_{t-i} \mid \{w_t\}) = \delta_n^*(F)\gamma_u(F)\gamma_w(F)^{-1}\delta_s(B)\delta_n^*(B)z_{t-i} = \delta_s(B)R_{t-i}. \tag{4.2}$$

By Lemma 1, $(\sum_{i=0}^{t-ds-1} \xi_i^s B^i)\delta_s(B) = 1 - \sum_{i=1}^{ds} A_{i,t}^s B^{t-i}$, so from (4.2)

$$\textstyle\sum_{i=0}^{t-ds-1} \xi_i^s E(u_{t-i} \mid \{w_t\}) = R_t - \mathbf{A}_t^{s\prime}\mathbf{R}_{(ds)}. \tag{4.3}$$

Using (4.3), (4.1) can be written as

$$\begin{aligned} E(s_t \mid \{z_t\}) &= \mathbf{A}_t^{s\prime}[E(\mathbf{s}_* \mid \{z_t\}) - \mathbf{R}_{(ds)}] + R_t \\ &\quad + \textstyle\sum_{i=0}^{t-ds-1} \xi_i^s E(u_{t-i} \mid \mathbf{z}_* - E(\mathbf{z}_* \mid \{w_t\})). \end{aligned} \tag{4.4}$$

Notice that (4.4) includes R_t, the nonstationary analogue of the stationary solution. However, (4.4) also includes an adjustment for the effect of the deviation of $E(s_j \mid \{z_t\})$ from R_j for $j = 1, \cdots, ds$, plus an adjustment for what $\mathbf{z}_*$ has to say about $\sum_{i=0}^{t-ds-1} \xi_i^s u_{t-i}$ beyond the information in $\{w_t\}$. If it happens that R_j is correct at the starting vaues $s_j (j = 1, \cdots, ds)$, and $\mathbf{z}_*$ contains no information on the u_{t-i}'s beyond that in $\{w_t\}$, then R_t will be $E(s_t \mid \{z_t\})$.

Actually, (4.2) above needs to be justified. Specifically, we need to know that (i) R_t exists, in that when we compute the filter $\delta_n^*(B)\delta_n^*(F)\gamma_u(F)\gamma_w(F)^{-1}$ and apply it to z_t, we get something that converges in mean square, and (ii) we can interchange operators like $\delta_s(B)$, $\gamma_u(F)$, and $\gamma_w(F)^{-1}$ in (4.2) (this is not obvious since z_t is nonstationary.) Conditions under which these things hold are given in Bell (1982). Here we merely note that these conditions will hold, in particular, if all the zeroes of $\delta(\zeta)$ are on the unit circle, $\gamma_w(\xi)$ is nonzero on the unit circle (as we always assume), and $\gamma_u(k)$ and $\gamma_w(k)$ decrease exponentially to zero as $|k| \to \infty$ (such as when u_t, v_t and w_t follow autoregressive-moving average models). If $\delta(\zeta)$ has zeroes inside the unit circle, we must be more careful. For the rest of this paper we will assume the required conditions are satisfied so that we can manipulate things as in (4.2). If these conditions are not satisfied, all is not lost. We can still do nonstationary signal extraction by substituting $\delta_n^*(F)\gamma_u(F)\gamma_w(F)^{-1}w_{t-i}$ directly for $E(u_{t-i} \mid \{w_t\})$ in (4.1) and proceeding from there instead of from (4.4). This approach is used in Section 5 when we obtain $\operatorname{Var}(s_t \mid \{z_t\})$.

4.1 *Signal extraction under Assumption* A. Under Assumption A, when $\delta_s(\zeta)$ and $\delta_n(\zeta)$ have no common zeroes, R_t actually is the solution to the nonstationary signal extraction problem, as is established in the following theorem.

THEOREM 2. *Make Assumption* A *so that* $\mathbf{z}_*$ *is independent of* $\{w_t\}$. *Also assume* $\delta_s(\zeta)$ *and* $\delta_n(\zeta)$ *have no common zeroes so* $\delta(B) = \delta_s(B)\delta_n(B)$. *Then*

$$E(s_t \mid \{z_t\}) = R_t = \delta_n(B)\delta_n(F)\gamma_u(F)\gamma_w(F)^{-1}z_t.$$

PROOF. The signal extraction error is

$$\begin{aligned}
\varepsilon_t &= s_t - R_t \\
&= s_t - \delta_n(B)\delta_n(F)\gamma_u(F)\gamma_w(F)^{-1}(s_t + n_t) \\
&= \delta_s(B)\delta_s(F)\gamma_v(F)\gamma_w(F)^{-1}s_t - \delta_n(B)\delta_n(F)\gamma_u(F)\gamma_w(F)^{-1}n_t \\
&= \delta_s(F)\gamma_v(F)\gamma_w(F)^{-1}u_t - \delta_n(F)\gamma_u(F)\gamma_w(F)^{-1}v_t
\end{aligned} \tag{4.5}$$

using the fact that (from 1.6) $\gamma_w(F) = \delta_n(B)\delta_n(F)\gamma_u(F) + \delta_s(B)\delta_s(F)\gamma_v(F)$ in the second line. The cross spectral density of u_t with w_t is $f_u(\lambda)\delta_n(e^{-i\lambda})$, and that of v_t with w_t is $f_v(\lambda)\delta_s(e^{-i\lambda})$ (see 1.5), so from (4.5) the cross spectral density of ε_t with w_t is

$$\begin{aligned}
f_{\varepsilon w}(\lambda) &= \delta_s(e^{-i\lambda})f_v(\lambda)f_w(\lambda)^{-1}f_u(\lambda)\delta_n(e^{-i\lambda}) \\
&\quad - \delta_n(e^{-i\lambda})f_u(\lambda)f_w(\lambda)^{-1}f_v(\lambda)\delta_s(e^{-i\lambda}) = 0.
\end{aligned}$$

This shows ε_t is uncorrelated with every w_j, and since $\mathbf{z}_*$ is assumed independent of $\{u_t\}$ and $\{v_t\}$, ε_t is also uncorrelated with $\mathbf{z}_*$. Thus, ε_t is uncorrelated with $\{z_t\}$. Since we are assuming joint normality, by results in Gihman and Skorohod (1980, pages 273–274) the theorem is proved. □

We now consider the case where $\delta_s(\zeta)$ and $\delta_n(\zeta)$ have common zeroes so $\delta(B) = \delta_s^*(B)\delta_c(B)\delta_n^*(B)$, where $\delta_c(B)$ is the product of the factors in $\delta(B)$ that are in both $\delta_s(B)$ and $\delta_n(B)$. Returning to (4.4), we notice that the independence of $\mathbf{z}_*$ and $\{w_t\}$, and z_* and $\{u_t\}$, under Assumption A implies that the last term in (4.4) drops out. Thus, we only need to evaluate $E(\mathbf{s}_* \mid \{z_t\}) = E(\mathbf{s}_* \mid \mathbf{z}_*) + E(\mathbf{s}_* \mid \{w_t\})$. From (3.2) we see that

$$(4.6)\quad [H_1H_2]\begin{bmatrix} E(\mathbf{s}_* \mid \{z_t\}) \\ E(\mathbf{n}_* \mid \{z_t\}) \end{bmatrix} = \mathbf{z}_* - C_1E(\mathbf{u}_{(d)}^{(ds+1)} \mid \{w_t\}) - C_2E(\mathbf{v}_{(d)}^{(dn+1)} \mid \{w_t\}).$$

The jth element of $C_1E(\mathbf{u}_{(d)}^{(ds+1)} \mid \{w_t\})$ is zero for $j = 1, \cdots, ds$ and by (4.3) is

$$\textstyle\sum_{i=0}^{j-ds-1} \xi_i^s E(u_{j-i} \mid \{w_t\}) = R_j - \mathbf{A}_j^{s\prime}\mathbf{R}_{(ds)}$$

for $j = ds + 1, \cdots, d$. Similarly, the jth element of $C_2E(\mathbf{v}_{(d)}^{(dn+1)} \mid \{w_t\})$ is also zero for $j = 1, \cdots, dn$ and is

$$\textstyle\sum_{i=0}^{j-dn-1} \xi_i^n E(v_{j-i} \mid \{w_t\}) = (z_j - R_j) - \mathbf{A}_j^{n\prime}(\mathbf{z}_{(dn)} - \mathbf{R}_{(dn)})$$

for $j = dn + 1, \cdots, d$. After some algebra, we get

$$(4.7)\quad \begin{aligned} &C_1E(\mathbf{u}_{(d)}^{(ds+1)} \mid \{w_t\}) + C_2E(\mathbf{v}_{(d)}^{(dn+1)} \mid \{w_t\}) \\ &\qquad = \mathbf{z}_* - H_1\mathbf{R}_{(ds)} - H_2\mathbf{z}_{(dn)} + H_2\mathbf{R}_{(dn)}. \end{aligned}$$

Let $m = \max(ds, dn)$ and define

$$(4.8)\quad \begin{aligned} \underset{d\times m}{H_3} &= H_1 - [H_2 O_{d\times(ds-dn)}], \quad ds \geq dn \\ &= [H_1 O_{d\times(dn-ds)}] - H_2, \quad ds < dn. \end{aligned}$$

We can now write (4.6) as

$$(4.9)\quad [H_1H_2]\begin{bmatrix} E(\mathbf{s}_* \mid \{z_t\}) \\ E(\mathbf{n}_* \mid \{z_t\}) \end{bmatrix} = H_2\mathbf{z}_{(dn)} + H_3\mathbf{R}_{(m)}.$$

We summarize our results in the following theorem.

THEOREM 3. *Make Assumption* A *and assume $\delta_s(\zeta)$ and $\delta_n(\zeta)$ have at least one common zero. Then for $t > ds$*

$$E(s_t \mid \{z_t\}) = \mathbf{A}_t^{s\prime}\{E(\mathbf{s}_* \mid \{z_t\}) - \mathbf{R}_{(ds)}\} + R_t.$$

$E(\mathbf{s}_* \mid \{z_t\})$ is obtained by solving (4.9) in the same way it is assumed that (3.2) is solved in generating $\mathbf{s}_*$ and $\mathbf{n}_*$, i.e., we make the same choice of $[H_1H_2]^-$.

4.2 *Signal extraction under Assumption* B. Under Assumption B, $\mathbf{s}_*$ and $\{w_t\}$

are independent, so that $E(\mathbf{s}_* \mid \{w_t\}) = \mathbf{0}$ and (4.4) becomes

$$E(s_t \mid \{z_t\}) = \mathbf{A}_t^{s\prime}\{E[\mathbf{s}_* \mid \mathbf{z}_* - E(\mathbf{z}_* \mid \{W_t\})] - \mathbf{R}_{(ds)}\} + R_t + \textstyle\sum_{i=0}^{t-ds-1} \xi_i^s E[u_{t-i} \mid \mathbf{z}_* - E(\mathbf{z}_* \mid \{w_t\})]. \tag{4.10}$$

For the case where n_t is stationary and $\delta_s(\zeta)$ has no zeroes inside the unit circle, Sobel (1967) establishes that R_t converges to $E(s_t \mid \{z_t\})$ as $t \to \infty$. Cleveland and Tiao (1976) similarly show that R_t approximates $E(s_t \mid \mathbf{z}_{(m+N)}^{(m)})$ for m and N large and t between and not near m and $m + N$. We now show how to evaluate (4.10) exactly by obtaining $\mathbf{z}_* - E(\mathbf{z}_* \mid \{w_t\})$, its variance matrix, and its covariances with $\mathbf{s}_*$ and $\sum_{i=0}^{t-ds-1} \xi_i^s u_{t-i}$.

To evaluate $\mathbf{z}_* - E(\mathbf{z}_* \mid \{w_t\})$ under Assumption B we notice $\mathbf{n}_*$ and $\mathbf{s}_*$ are both independent of $\{w_t\}$ so $E([\mathbf{s}_*' \mathbf{n}_*'] \mid \{w_t\}) = \mathbf{0}'$. Then, by (3.2), (4.7), and (4.8), we get

$$\mathbf{z}_* - E(\mathbf{z}_* \mid \{w_t\}) = H_2\mathbf{z}_{(dn)} + H_3\mathbf{R}_{(m)}. \tag{4.11}$$

To compute the variance matrix of $\mathbf{z}_* - E(\mathbf{z}_* \mid \{w_t\})$ we need a different expression than (4.11). For $t > ds$, from (3.1), (4.2), (1.5), and (1.6) we get

$$\begin{aligned} s_t - E(s_t \mid \{w_t\}) &= \mathbf{A}^{s\prime}\mathbf{s}_* + (\textstyle\sum_{i=0}^{t-ds-1} \xi_i^s B^i) \\ &\quad \cdot [u_t - \delta_n^*(F)\gamma_u(F)\gamma_w(F)^{-1}(\delta_n^*(B)u_t + \delta_s^*(B)v_t)] \\ &= \mathbf{A}_t^{s\prime}\mathbf{s}_* + (\textstyle\sum_{i=0}^{t-ds-1} \xi_i^s B^i)\delta_s^*(B)X_t \end{aligned} \tag{4.12}$$

where

$$X_t = \delta_s^*(F)\gamma_v(F)\gamma_w(F)^{-1}u_t - \delta_n^*(F)\gamma_u(F)\gamma_w(F)^{-1}v_t.$$

Similarly, for $t > dn$ we can show that

$$n_t - E(n_t \mid \{w_t\}) = \mathbf{A}_t^{n\prime}\mathbf{n}_* - (\textstyle\sum_{i=0}^{t-dn-1} \xi_i^n B^i)\delta_n^*(B)X_t. \tag{4.13}$$

For $t = 1, \cdots, ds$, $E(s_t \mid \{w_t\}) = 0$, so for $t \geq 1$ we may write

$$s_t - E(s_t \mid \{w_t\}) = \mathbf{A}_t^{s\prime}\mathbf{s}_* + \textstyle\sum_{j=1}^{t} g_{tj}^{(1)} X_j$$

where for all t, j we define

$$g_{tj}^{(1)} = \text{coefficient of } B^{t-j} \text{ in } (\textstyle\sum_{i=0}^{t-ds-1} \xi_i^s B^i)\delta_s^*(B). \tag{4.14}$$

Notice $g_{tj}^{(1)}$ is zero for $t < j$, $t \leq ds$, and $j \leq dc$. We thus have $\mathbf{s}_{(d)} - E(\mathbf{s}_{(d)} \mid \{w_t\}) = H_1\mathbf{s}_* + G_1\mathbf{X}_{(d)}$ where $G_1 = (g_{tj}^{(1)})$ is a $d \times d$ lower triangular matrix. Similarly $\mathbf{n}_{(d)} - E(\mathbf{n}_{(d)} \mid \{w_t\}) = H_2\mathbf{n}_* - G_2\mathbf{X}_{(d)}$ where $G_2 = (g_{tj}^{(2)})$ and

$$g_{tj}^{(2)} = \text{coefficient of } B^{t-j} \text{ in } (\textstyle\sum_{i=0}^{t-dn-1} \xi_i^n B^i)\delta_n^*(B). \tag{4.15}$$

Therefore, letting $G_3 = G_1 - G_2$

$$\mathbf{z}_* - E(\mathbf{z}_* \mid \{w_t\}) = H_1\mathbf{s}_* + H_2\mathbf{n}_* + G_3\mathbf{X}_{(d)}. \tag{4.16}$$

Since $\mathbf{s}_*$, $\mathbf{n}_*$, and X_t are independent, letting Ω denote $\mathrm{Var}(\mathbf{z}_* - E(\mathbf{z}_* \mid \{w_t\}))$, we see

$$\Omega = H_1 \mathrm{Var}(\mathbf{s}_*) H_1' + H_2 \mathrm{Var}(\mathbf{n}_*) H_2' + G_3 [\gamma_x]_{(d)} G_3' \tag{4.17}$$

$$\mathrm{Cov}(\mathbf{s}_*, \mathbf{z}_* - E(\mathbf{z}_* \mid \{w_t\})) = \mathrm{Var}(\mathbf{s}_*) H_1'. \tag{4.18}$$

From (4.11), (4.17), (4.18), and standard results on conditional expectations for zero mean normal random vectors

$$E(\mathbf{s}_* \mid \mathbf{z}_* - E(\mathbf{z}_* \mid \{w_t\})) = \mathrm{Var}(\mathbf{s}_*) H_1' \Omega^{-1} \{H_2 \mathbf{z}_{(dn)} + H_3 \mathbf{R}_{(m)}\}. \tag{4.19}$$

To compute this we need the autocovariances for X_t. The spectral density for X_t can be shown to be equal to $f_x(\lambda) = f_u(\lambda) f_w(\lambda)^{-1} f_v(\lambda)$. The autocovariances, $\gamma_x(k)$, can be computed by Fourier transforming $f_x(\lambda)$ or by expanding the CGF

$$\gamma_x(\zeta) = \gamma_u(\zeta)\gamma_w(\zeta)^{-1}\gamma_v(\zeta). \tag{4.20}$$

If u_t, v_t, and hence X_t follow autoregressive-moving average models the techniques discussed in McLeod (1975, 1977) for the univariate case and in Nicholls and Hall (1979) for the multivariate case can be used to compute the $\gamma_x(k)$.

Finally, we consider $E(\sum_{i=0}^{t-ds-1} \xi_i^s u_{t-i} \mid \mathbf{z}_* - E(\mathbf{z}_* \mid \{w_t\}))$. Notice $\sum_{i=0}^{t-ds-1} \xi_i^s u_{t-i}$ is independent of $\mathbf{s}_*$ and $\mathbf{n}_*$, so by (4.16) we only need consider its covariance with $\mathbf{X}_{(d)}$. The cross spectral density between the time series u_t and X_t is $f_u(\lambda) f_w(\lambda)^{-1} f_v(\lambda) \delta_s^*(e^{i\lambda}) = f_x(\lambda)\delta_s^*(e^{i\lambda})$, so that $\mathrm{Cov}(u_t, X_j) = \delta_s^*(F)\gamma_x(j - t)$. Using this and (4.14) we write (F applies to j)

$$\begin{aligned} \mathrm{Cov}\left(\textstyle\sum_{i=0}^{t-ds-1} \xi_i^s u_{t-i}, X_j\right) &= \left(\textstyle\sum_{i=0}^{t-ds-1} \xi_i^s F^i\right)\delta_s^*(F)\gamma_x(j - t) \\ &= \textstyle\sum_{i=dc+1}^{t} g_{ti}^{(1)} \gamma_x(j - i). \end{aligned} \tag{4.21}$$

Taking (4.21) for $j = 1, \cdots, d$ we obtain

$$\begin{aligned} &\mathrm{Cov}\left(\textstyle\sum_{i=0}^{t-ds-1} \xi_i^s u_{t-i}, \mathbf{X}_{(d)}\right) \\ &\quad = [0 \cdots 0 \; g_{t,dc+1}^{(1)} \cdots g_{tt}^{(1)}] \begin{bmatrix} \gamma_x(0) & \cdots & \gamma_x(d-1) \\ \vdots & & \vdots \\ \gamma_x(1-t) & \cdots & \gamma_x(d-t) \end{bmatrix}. \end{aligned} \tag{4.22}$$

We thus have

$$\begin{aligned} &E\left(\textstyle\sum_{i=0}^{t-ds-1} \xi_i^s u_{t-i} \mid \mathbf{z}_* - E(\mathbf{z}_* \mid \{w_t\})\right) \\ &\quad = \mathrm{Cov}\left(\textstyle\sum_{i=0}^{t-ds-1} \xi_i^s u_{t-i}, \mathbf{X}_{(d)}\right) \times G_3' \Omega^{-1}\{H_2 \mathbf{z}_{(dn)} + H_3 \mathbf{R}_{(m)}\} \end{aligned} \tag{4.23}$$

and use (4.22) in evaluating (4.23).

We summarize our results in a theorem.

THEOREM 4. *Make Assumption* B *so that* $\mathbf{s}_*$ *and* $\mathbf{n}_*$ *are independent of each* *other and of* $\{w_t\}$. *Then, for* $t > ds$

$$E(s_t \mid \{z_t\}) = \mathbf{A}_t^{s\prime}\{E[\mathbf{s}_* \mid \mathbf{z}_* - E(\mathbf{z}_* \mid \{w_t\})] - \mathbf{R}_{(ds)}\} + R_t + E(\textstyle\sum_{i=0}^{t-ds-1} \xi_i^s u_{t-i} \mid \mathbf{z}_* - E(\mathbf{z}_* \mid \{w_t\}))$$

where $R_t = \delta_n^*(B)\delta_n^*(F)\gamma_u(F)\gamma_w(F)^{-1}z_t$, $E[\mathbf{s}_* \mid \mathbf{z}_* - E(\mathbf{z}_* \mid \{w_t\})]$ *is given by* (4.19), *and* $E(\sum_{i=0}^{t-ds-1} \xi_i^s u_{t-i} \mid \mathbf{z}_* - E(\mathbf{z}_* \mid \{w_t\}))$ *is given by* (4.23), *using* (4.22) *and* (4.14). *The covariance generating function* $\gamma_x(\xi)$ *given by* (4.20) *may be used to compute the* $\gamma_x(k)$ *needed.*

If $\delta_s(\zeta)$ and $\delta_n(\zeta)$ have no common zeroes, the above results simplify somewhat. We first notice that $\delta_s^*(B) = \delta_s(B)$ and $\delta_n^*(B) = \delta_n(B)$, so that (4.14) shows $g_{tj}^{(1)}$ to be the coefficient of B^{t-j} in $1 - \sum_{i=1}^{ds} A_{it}^s B^{t-i}$ and similarly for $g_{tj}^{(2)}$. Bell (1982) then shows that $G_3\mathbf{X}_{(d)} = -H_3\mathbf{X}_{(m)}$ $(m = \max(ds, dn))$, so we may substitute $-H_3$ for G_3 and m for d in (4.17) and (4.23). In addition, we may replace (4.22) by

$$\operatorname{Cov}(\textstyle\sum_{i=0}^{t-ds-1} \xi_i^s u_{t-i}, \mathbf{X}_{(m)}) = [\gamma_x(1-t) \ \cdots \ \gamma_x(m-t)] - \mathbf{A}_t^{s\prime} \begin{bmatrix} \gamma_x(0) & \cdots & \gamma_x(m-1) \\ \vdots & & \vdots \\ \gamma_x(1-ds) & \cdots & \gamma_x(m-ds) \end{bmatrix}$$

and use this in (4.23).

5. Variances of signal extraction errors. In many applications of signal extraction we want to compute not only the estimate $E(s_t \mid \{z_t\})$, but also the conditional variance, $\operatorname{Var}(s_t \mid \{z_t\})$. This is the same as the variance of the signal extraction error, $\varepsilon_t = s_t - E(s_t \mid \{z_t\})$. When R_t is used instead of $E(s_t \mid \{z_t\})$, Hannan (1967) gives the variance of the resulting error, $\operatorname{Var}(s_t - R_t)$, for the case where n_t is stationary. For this and other cases the properties of $s_t - R_t$ have been more extensively investigated by Pierce (1979) (see Theorem 6 and the discussion following it in the next subsection).

To obtain $\operatorname{Var}(s_t \mid \{z_t\})$, notice that by Lemma 2 we may write

$$s_t = \varepsilon_t + E(s_t \mid \{z_t\}) \tag{5.1}$$

$$= \varepsilon_t + E(s_t \mid \{w_t\}) + E(s_t \mid \mathbf{z}_* - E(\mathbf{z}_* \mid \{w_t\})). \tag{5.2}$$

By results in Gihman and Skorohod (1980, pages 273–274), we can see that the terms on the right hand side in (5.1) and (5.2) are independent, so we may compute $\operatorname{Var}(s_t \mid \{z_t\}) = \operatorname{Var}(\varepsilon_t)$ as

$$\operatorname{Var}(s_t \mid \{z_t\}) = \operatorname{Var}(s_t) - \operatorname{Var}(E(s_t \mid \{z_t\})) \tag{5.3}$$

$$= \operatorname{Var}(s_t) - \operatorname{Var}(E(s_t \mid \{w_t\})) - \operatorname{Var}(E(s_t \mid \{\mathbf{z}_* - E(\mathbf{z}_* \mid \{w_t\})\})). \tag{5.4}$$

(5.3) and (5.4) will still hold if we replace s_t by a vector, say $(s_1, \cdots, s_t)' = \mathbf{s}_{(t)}$, so we can use (5.3) and (5.4) to compute variances and covariances of the s_t's conditional on $\{z_t\}$. We now show how to evaluate the required terms in (5.3) and (5.4) under Assumptions A and B.

5.1 *Variances under Assumption* A. We start with the general case under Assumption A where $\delta_s(\zeta)$ and $\delta_n(\zeta)$ may have common zeroes. The case where $\delta_s(\zeta)$ and $\delta_n(\zeta)$ do not have common zeroes is much simpler and will be discussed later.

To begin, we notice from (3.1) that

$$\mathbf{s}_{(t)} = H_{1t}\mathbf{s}_* + C_{1t}\mathbf{u}_{(t)}^{(ds+1)} \tag{5.5}$$

where

$$H_{1t} = \begin{bmatrix} I_{ds} \\ \mathbf{A}_{ds+1}^{s\prime} \\ \vdots \\ \mathbf{A}_t^{s\prime} \end{bmatrix} \qquad C_{1t} = \begin{bmatrix} & O_{ds\times(t-ds)} & \\ \xi_0^s & & \\ \vdots & \ddots & \\ \xi_{t-ds-1}^s & \cdots & \xi_0^s \end{bmatrix}.$$

Notice $H_{1d} = H_1$ and $C_{1d} = C_1$. Solving for $\mathbf{s}_*$ from (3.2) and substituting in (5.5), we can show that

$$\mathbf{s}_{(t)} = K_1\mathbf{z}_* + K_2\mathbf{u}_{(t)}^{(ds+1)} + K_3\mathbf{v}_{(d)}^{(dn+1)} \tag{5.6}$$

where

$$K_1 = H_{1t}[I_{ds}\ O_{ds\times dn}][H_1 H_2]^-$$
$$K_2 = C_{1t} - [K_1C_1\ O_{t\times(t-d+ds)}]$$
$$K_3 = -K_1C_2.$$

The terms on the right hand side of (5.6) are independent under Assumption A so

$$\mathrm{Var}(\mathbf{s}_{(t)}) = K_1\mathrm{Var}(\mathbf{z}_*)K_1' + K_2[\gamma_u]_{(t-ds)}K_2' + K_3[\gamma_v]_{(d-dn)}K_3'. \tag{5.7}$$

To compute $\mathrm{Var}(E(\mathbf{s}_{(t)}\,|\,\{z_t\}))$ we notice from (5.6) that

$$E(\mathbf{s}_{(t)}\,|\,\{z_t\}) = K_1\mathbf{z}_* + K_2\mathbf{P}_{(t)}^{(ds+1)} + K_3\mathbf{Q}_{(d)}^{(dn+1)} \tag{5.8}$$

where (see (4.2))

$$\begin{aligned} P_t &= E(u_t\,|\,\{w_t\}) = \delta_n^*(F)\gamma_u(F)\gamma_w(F)^{-1}w_t \\ Q_t &= E(v_t\,|\,\{w_t\}) = \delta_s^*(F)\gamma_v(F)\gamma_w(F)^{-1}w_t. \end{aligned} \tag{5.9}$$

(We could use (5.8) and (5.9) in place of the results of Theorems 2 and 3 to do the signal extraction under Assumption A. This would be necessary if, as was mentioned in Section 4, (4.2) should not hold.) It can be shown that

$$\begin{aligned} \gamma_P(k) &= \gamma_u(k) - \delta_s^*(B)\delta_s^*(F)\gamma_x(k), \quad \gamma_{PQ}(k) = \delta_n^*(B)\delta_s^*(F)\gamma_x(k) \\ \gamma_Q(k) &= \gamma_v(k) - \delta_n^*(B)\delta_n^*(F)\gamma_x(k). \end{aligned} \tag{5.10}$$

Under Assumption A $\mathbf{z}_*$ is independent of $\{P_t\}$ and $\{Q_t\}$, so from (5.8) we get

$$\begin{aligned}\operatorname{Var}(E(\mathbf{s}_{(t)} \mid \{z_t\})) = {} & K_1 \operatorname{Var}(\mathbf{z}_*) K_1' + K_2[\gamma_P]_{(t-ds)} K_2' \\ & + K_3[\gamma_Q]_{(d-dn)} K_3' + K_2 \operatorname{Cov}(\mathbf{P}_{(t)}^{(ds+1)}, \mathbf{Q}_{(d)}^{(dn+1)}) K_3' \\ & + K_3 \operatorname{Cov}(\mathbf{Q}_{(d)}^{(dn+1)}, \mathbf{P}_{(t)}^{(ds+1)}) K_2'.\end{aligned} \tag{5.11}$$

Using (5.3), we obtain the following theorem.

THEOREM 5. *Under Assumption* A *when* $\delta_s(\zeta)$ *and* $\delta_n(\zeta)$ *have at least one common zero*

$$\begin{aligned}\operatorname{Var}(\mathbf{s}_{(t)} \mid \{z_t\}) = {} & K_2[\delta_s^*(B)\delta_s^*(F)\gamma_x]_{(t-ds)} K_2' \\ & + K_3[\delta_n^*(B)\delta_n^*(F)\gamma_x]_{(d-dn)} K_3' \\ & - K_2 \begin{bmatrix} \gamma_{PQ}(dn-ds) & \cdots & \gamma_{PQ}(d-ds-1) \\ \vdots & & \vdots \\ \gamma_{PQ}(dn+1-t) & \cdots & \gamma_{PQ}(d-t) \end{bmatrix} K_3' \\ & - K_3 \begin{bmatrix} \gamma_{PQ}(dn-ds) & \cdots & \gamma_{PQ}(dn+1-t) \\ \vdots & & \vdots \\ \gamma_{PQ}(d-ds-1) & \cdots & \gamma_{PQ}(d-t) \end{bmatrix} K_2'.\end{aligned}$$

The $\gamma_x(k)$ *may be computed from* (4.20) *and the* $\gamma_{PQ}(k)$ *from* (5.10).

PROOF. Subtract (5.11) from (5.7) and use (5.10). □

When $\delta_s(\zeta)$ and $\delta_n(\zeta)$ have no common zeroes there is a far simpler approach to computing $\operatorname{Var}(\mathbf{s}_{(t)} \mid \{z_t\})$ than the above. From Theorem 3, (4.5), and (4.12) the signal extraction error in this case is

$$s_t - E(s_t \mid \{z_t\}) = s_t - R_t = X_t.$$

(Note X_t does not equal $s_t - R_t$ if $\delta_s(\zeta)$ and $\delta_n(\zeta)$ have common zeroes.) Thus, we have

THEOREM 6. *Under Assumption* A, *if* $\delta_s(\zeta)$ *and* $\delta_n(\zeta)$ *have no common zeroes, then* $\operatorname{Var}(\mathbf{s}_{(t)} \mid \{z_t\}) = \operatorname{Var}(\mathbf{X}_{(t)}) = [\gamma_x]_{(t)}$, *the elements of which may be computed using* (4.20).

This result has been given by Pierce (1979), who examines the behavior of $s_t - R_t$ when $\delta_s(\zeta)$ and $\delta_n(\zeta)$ do and do not have common zeroes. However, Pierce's statement that when $\delta_s(\zeta)$ and $\delta_n(\zeta)$ have common zeroes the mean squared signal extraction error does not exist (i.e., that it is infinite) is incorrect. Although both

$s_t - R_t$ and $s_t - E(s_t \mid \{z_t\})$ will be nonstationary when $\delta_s(\zeta)$ and $\delta_n(\zeta)$ have common zeroes (the latter will always be nonstationary under Assumption B) they will both have finite mean square, as is easy to see from the results in this paper (including Theorems 5 and 7).

It is interesting to note that under Assumption A, $\mathrm{Var}(\mathbf{s}_{(t)} \mid \{z_t\})$ does not involve $\mathrm{Var}(\mathbf{z}_*)$. For that matter, neither does $E(\mathbf{s}_t \mid \{z_t\})$—see Theorems 2 and 3. This is true whether or not $\delta_s(\zeta)$ and $\delta_n(\zeta)$ have common zeroes. Thus, when making Assumption A we need not concern ourselves with $\mathrm{Var}(\mathbf{z}_*)$. The situation under Assumption B is different: there we must know $\mathrm{Var}(\mathbf{s}_*)$ and $\mathrm{Var}(\mathbf{n}_*)$.

5.2 *Variances under Assumption* B. Under Assumption B, $\mathbf{s}_*$ and $\{u_t\}$ are independent, so from (5.5)

$$\mathrm{Var}(\mathbf{s}_{(t)}) = H_{1t}\mathrm{Var}(\mathbf{s}_*)H_{1t}' + C_{1t}[\gamma_u]_{(t-ds)}C_{1t}'. \tag{5.12}$$

$\mathbf{s}_*$ is also independent of $\{w_t\}$, so from (5.5) and (5.9)

$$E(\mathbf{s}_{(t)} \mid \{w_t\}) = C_{1t}\mathbf{P}_{(t)}^{(ds+1)}$$

so that

$$\mathrm{Var}(E(\mathbf{s}_{(t)} \mid \{w_t\})) = C_{1t}[\gamma_P]_{(t-ds)}C_{1t}'. \tag{5.13}$$

The remaining term we need is $\mathrm{Var}(E[\mathbf{s}_{(t)} \mid \mathbf{z}_* - E(\mathbf{z}_* \mid \{w_t\})])$, which is

$$\mathrm{Cov}(\mathbf{s}_{(t)}, \mathbf{z}_* - E(\mathbf{z}_* \mid \{w_t\}))\ \Omega^{-1}\mathrm{Cov}(\mathbf{s}_{(t)}, \mathbf{z}_* - E(\mathbf{z}_* \mid \{w_t\}))' \tag{5.14}$$

where $\Omega = \mathrm{Var}(\mathbf{z}_* - E(\mathbf{z}_* \mid \{w_t\}))$ is given by (4.17). It can be shown that

$$\begin{aligned}&\mathrm{Cov}(\mathbf{s}_{(t)}, \mathbf{z}_* - E(\mathbf{z}_* \mid \{w_t\}))\\ &\quad = H_{1t}\mathrm{Var}(\mathbf{s}_*)H_1' + G_{1t}\begin{bmatrix}\gamma_x(0) & \cdots & \gamma_x(d-1)\\ \vdots & & \vdots\\ \gamma_x(1-t) & \cdots & \gamma_x(d-t)\end{bmatrix}G_3'\end{aligned} \tag{5.15}$$

where $G_{1t} = (g_{tj}^{(1)})$ is $t \times t$ lower triangular (note $G_{1d} = G_1$).

Following (5.4) we subtract (5.13) and (5.14) from (5.12) to obtain our result.

THEOREM 7. *Under Assumption* B

$$\begin{aligned}\mathrm{Var}(\mathbf{s}_{(t)} \mid \{z_t\}) &= H_{1t}\mathrm{Var}(\mathbf{s}_*)H_{1t}' + C_{1t}[\delta_s^*(B)\delta_s^*(F)\gamma_x]_{(t-ds)}C_{1t}'\\ &\quad - \mathrm{Cov}(\mathbf{s}_{(t)}, \mathbf{z}_* - E(\mathbf{z}_* \mid \{w_t\}))\\ &\quad \cdot \Omega^{-1}\mathrm{Cov}(\mathbf{s}_{(t)}, \mathbf{z}_* - E(\mathbf{z}_* \mid \{w_t\}))\end{aligned}$$

where Ω *is given by* (4.17) *and* $\mathrm{Cov}(\mathbf{s}_{(t)}, \mathbf{z}_* - E(\mathbf{z}_* \mid \{w_t\}))$ *by* (5.15).

If $\delta_s(\zeta)$ and $\delta_n(\zeta)$ have no common zeroes we substitute $-H_3$ for G_3 and m for

d in (4.17). Also, Bell (1982) shows that (5.15) then simplifies to

$$\mathrm{Cov}(\mathbf{s}_{(t)}, \mathbf{z}_* - E(\mathbf{z}_* \mid \{w_t\})) = H_{1t}\mathrm{Var}(\mathbf{s}_*)H_1'$$
$$- \{I_t - [H_{1t} \mid O_{t\times(t-ds)}]\} \begin{bmatrix} \gamma_x(0) & \cdots & \gamma_x(m-1) \\ \vdots & & \vdots \\ \gamma_x(1-t) & \cdots & \gamma_x(m-t) \end{bmatrix} H_3'.$$

6. Extensions of results.

6.1 *Signal extraction with a finite number of observations.* In practice a finite stretch of the time series, say $\mathbf{z}_{(N)} = (z_1, \cdots, z_N)'$, will be available rather than the complete realization $\{z_t\}$. Cleveland (1972) observed that $E(s_t \mid \mathbf{z}_{(N)}) = E(E(s_t \mid \{z_t\}) \mid \mathbf{z}_{(N)})$, so that $E(s_t \mid \mathbf{z}_{(N)})$ can be obtained by replacing unknown z_j's in $E(s_t \mid \{z_t\})$ by $E(z_j \mid \mathbf{z}_{(N)})$, which are forecasted ($j > N$) or backcasted ($j \le 0$) values. Bell (1980) established that as long as the expression for $E(s_t \mid \{z_t\})$, which is linear in the z_t's, converges in mean square, then this procedure converges pointwise to $\mathrm{Cov}(s_t, \mathbf{z}_{(N)})\mathrm{Var}(\mathbf{z}_{(N)})^{-1}\mathbf{z}_{(N)} = E(s_t \mid \mathbf{z}_{(N)})$. To apply this procedure, we need to be able to compute the forecasts and backcasts. Bell (1982) observes that the usual forecasting and backcasting procedures (discussed, for example, in Box and Jenkins, 1976) are correct under Assumption A, but incorrect under Assumption B. For ARIMA models, Burman (1980) shows how to avoid the need to compute a large number of forecasts and backcasts by using an algorithm due to G. Tunnicliffe Wilson (this would apply under Assumption A when $\delta_s(\zeta)$ and $\delta_n(\zeta)$ have no common zeroes).

A convenient means of obtaining $E(s_t \mid \mathbf{z}_{(N)})$ under Assumption B when s_t and n_t follow autoregressive-moving average models, is to put the signal plus noise model in state-space form and use the Kalman filter/smoother (see Meditch, 1969). Kitagawa (1981) illustrates how to do this for some particular models. The Kalman filter/smoother can also be used under Assumption A, although one must be careful then to choose a state representation such that the state vector at time t is independent of the process noise in the state equation at time $t + 1$.

Along with $E(s_t \mid \mathbf{z}_{(N)})$, the Kalman filter/smoother directly produces conditional variances, $\mathrm{Var}(s_t \mid \mathbf{z}_{(N)})$; conditional covariances, $\mathrm{Cov}(s_t, s_j \mid \mathbf{z}_{(N)})$, can also be obtained. This is important since $\mathrm{Var}(s_t \mid \{z_t\})$ as given in Section 5 will differ from $\mathrm{Var}(s_t \mid \mathbf{z}_{(N)})$ for any t for which $E(s_t \mid \{z_t\})$ is appreciably affected by z_j's outside of $z_1, \cdots, z_N$-typically for t near 1 or N. An alternative to the Kalman filter/smoother under Assumption A when $\delta_s(\zeta)$ and $\delta_n(\zeta)$ have no common zeroes is to use the results of Pierce (1979) to produce $\mathrm{Var}(s_t \mid z_N, z_{N-1}, \cdots, z_0, \cdots)$. Hillmer (1982) gives an approach that can be used to get $\mathrm{Var}(s_t \mid \mathbf{z}_{(N)})$.

6.2 *Linear projection results for the nonGaussian case.* By results of Gihman and Skorohod (1980, pages 273–274) the results of Section 4 produce $\hat{s}_t$, the linear function of the observed z_t's which minimizes $E[(s_t - \hat{s}_t)^2]$, whether or not the series involved are normal. The results of Section 5 produce variances and covariances for the time series $s_t - \hat{s}_t$, which are not the conditional variances and covariances without the normality assumption.

6.3 *The case of known starting values.* In some cases the starting values $\mathbf{s}_*$ and $\mathbf{n}_*$ might be known, fixed quantities. Then they are independent of $\{u_t\}$, $\{v_t\}$, and $\{w_t\}$, so we have Assumption B with $\text{Var}(\mathbf{s}_*) = 0$ and $\text{Var}(\mathbf{n}_*) = 0$. We remove the effects of the known starting values by considering the decomposition $\mathring{z}_t = \mathring{s}_t + \mathring{n}_t$, where for $t > 0$

$$\dot{s}_t = \begin{cases} 0 & t = 1, \cdots, ds \\ s_t - \mathbf{A}_t^{s\prime}\mathbf{s}_* & t > ds \end{cases} \qquad \dot{n}_t = \begin{cases} 0 & t = 1, \cdots dn \\ n_t - \mathbf{A}_t^{n\prime}\mathbf{n}_* & t > dn \end{cases}$$

$$\dot{z}_t = z_t - \mathbf{A}_t^{s\prime}\mathbf{s}_* - \mathbf{A}_t^{n\prime}\mathbf{n}_*$$

with analogous definitions for $t \leq 0$. We then apply the results of Section 4.2 to $\dot{z}_t$ to get $E(\dot{s}_t \mid \{z_t\})$, to which we add $\mathbf{A}_t^{s\prime}\mathbf{s}_*$ (for $t > 0$) to produce $E(s_t \mid \{z_t\})$. In this case $\text{Var}(\dot{\mathbf{z}}_* - E(\dot{\mathbf{z}}_* \mid \{w_t\}))$ may very well be singular, so that we must use a generalized inverse of it.

6.4 *Extensions to the multivariate case.* In the multivariate case we have

$$\mathbf{z}_t = \mathbf{s}_t + \mathbf{n}_t, \qquad t = 0, \pm 1, \pm 2, \cdots$$

where $\mathbf{z}_t$, $\mathbf{s}_t$, and $\mathbf{n}_t$ are $k \times 1$ random vectors with

$$\delta(B)\mathbf{z}_t = \mathbf{w}_t, \quad \delta_s(B)\mathbf{s}_t = \mathbf{u}_t, \quad \delta_n(B)\mathbf{n}_t = \mathbf{v}_t$$

jointly stationary $k \times 1$ vector time series. An important special case occurs when $\delta(B)$, $\delta_s(B)$, and $\delta_n(B)$ remain *scalar* operators, so that (1.4) and (1.5) still hold. The results and proofs in this paper have been presented in a way that allows them to be used in this particular multivariate case with little or no modification. For example, Theorem 2 is still correct with $\gamma_u(F)$ and $\gamma_w(F)$ the $k \times k$ matrix covariance generating functions of $\mathbf{u}_t$ and $\mathbf{w}_t$. The starting values require somewhat special consideration, for which see Bell (1982).

The general case where $\delta(B)$, $\delta_s(B)$, and $\delta_n(B)$ are $k \times k$ matrix operators is more difficult and we have chosen not to treat it here. In this case the relationship between $\delta(B)$, $\delta_s(B)$, and $\delta_n(B)$ is not clear—(1.4) need not hold. One may be able to obtain results in a manner analogous to that used here for certain special cases, such as when $\mathbf{n}_t$ is stationary.

REFERENCES

ANDERSON, T. W. (1971). *The Statistical Analysis of Time Series.* Wiley, New York.

BELL, W. R. (1980). Multivariate time series: smoothing and backward models. Ph.D. thesis, University of Wisconsin–Madison.

BELL, W. R. (1980). Signal extraction for nonstationary time series. SRD Research Report Number: CENSUS/SRD/RR-82/01, U. S. Bureau of the Census.

BOX, G. E. P. and JENKINS, G. M. (1976). *Time Series Analysis: Forecasting and Control.* Holden-Day, San Francisco.

BRILLINGER, D. R. (1975). *Time Series: Data Analysis and Theory.* Holt, New York.

BURMAN, J. P. (1980). Seasonal adjustment by signal extraction. *J. Roy. Statist. Soc. Ser. A* **143** 321–337.

CLEVELAND, W. P. (1972). Analysis and forecasting of seasonal time series. Ph.D. thesis, University of Wisconsin–Madison.

CLEVELAND, W. P., and TIAO, G. C. (1976). Decomposition of seasonal time series: a model for the Census X-11 program. *J. Amer. Statist. Assoc.* **71** 581–587.

FINDLEY, D. F. (1984). On backshift-operator polynomial transformations to stationarity for nonstationary time series and their aggregates. *Commun. Statist. Theor. Meth.* **A13**, to appear.

FULLER, W. A. (1976). *Introduction to Statistical Time Series.* Wiley, New York.

GIHMAN, I. I., and SKOROHOD, A. V. (1980). *The Theory of Stochastic Processes* I. Springer, Berlin.

HANNAN, E. J. (1967). Measurement of a wandering signal amid noise. *J. Appl. Probab.* **4** 90–102.

HILLMER, S. C. (1982). Confidence intervals for seasonally adjusted data. Proceedings of the American Statistical Association, Bus. and Econ. Statis. Section, 409–414.

HILLMER, S. C. AND TIAO, G. C. (1982). An ARIMA-model-based approach to seasonal adjustment. *J. Amer. Statist. Assoc.* **77** 63–70.

KITAGAWA, G. (1981). A nonstationary time series model and its fitting by a recursive filter. *J. Time Ser. Anal.* **2** 103–116.

KOLMOGOROV, A. N. (1939). Sur l'interpolation et extrapolation des suites stationnaires. *C. R. Acad. Sci. Paris* **208** 2043–2045.

KOLMOGOROV, A. N. (1941). Interpolation und extrapolation von stationären zufälligen Folgen. *Bull. Acad. Sci. U.R.S.S. Ser. Math.* **5** 3–14.

MCLEOD, I. (1975). Derivation of the theoretical autocovariance function of autoregressive-moving average time series. *Appl. Statist.* **24** 255–256.

MCLEOD, I. (1977). Correction to derivation of the theoretical autocovariance function of autoregressive-moving average time series. *Appl. Statist.* **26** 194.

MEDITCH, J. S. (1969). *Stochastic Optimal Linear Estimation and Control.* McGraw, New York.

NICHOLLS, D. F. and HALL, A. D. (1979). The exact likelihood function of multivariate autoregressive-moving average models. *Biometrika* **66** 259–264.

PIERCE, D. A. (1979). Signal extraction error in nonstationary time series. *Ann. Statist.* **7** 1303–1320.

SOBEL, E. L. (1967). Prediction of a noise-distorted, multivariate, non-stationary signal. *J. Appl. Probab.* **4** 330–342.

TIAO, G. C., and HILLMER, S. C. (1978). Some consideration of decomposition of a time series. *Biometrika* **65** 497–502.

WHITTAKER, E. T. and ROBINSON, G. (1944). *The Calculus of Observations* (4th ed.). Blackie, London.

WIENER, N. (1949). *The Extrapolation, Interpolation and Smoothing of Stationary Time Series With Engineering Applications.* Wiley, New York.

STATISTICAL RESEARCH DIVISION
U.S. DEPARTMENT OF COMMERCE
BUREAU OF THE CENSUS
WASHINGTON, D.C. 20233