BUREAU OF THE CENSUS

STATISTICAL RESEARCH DIVISION REPORT SERIES

SRD Research Report Number: CENSUS/SRD/RR-82/01

SIGNAL EXTRACTION FOR NONSTATIONARY TIME SERIES

by

William Bell

ASA/Census Research Fellow
Statistical Research Division

Recommended by: Myron J. Katzoff

Report completed: June 30, 1982

Report issued: July 27, 1982

SUMMARY

The signal extraction problem in time series is to estimate s_t given observations on $z_t = s_t + n_t$ (signal plus noise). The solution when s_t and n_t are stationary is well known, having been first obtained independently by Kolmogorov and Wiener in the 1940's. Their solution can be written (assuming normality and zero means) as $E(s_t|\{z_t\}) = \gamma_s(F)\gamma_z(F)^{-1}z_t$, where $\gamma_s(F)$ and $\gamma_z(F)$ are the autocovariance generating functions of s_t and z_t and F is the forward-shift operator.

In this paper we allow either s_t or n_t or both to be nonstationary. We consider homogeneous or explosive nonstationarity described by models of the form $\delta(B)z_t = w_t$, where $\delta(B)$, a polynomial in the backshift operator $B = F^{-1}$, has zeroes on or inside the unit circle and w_t is stationary. Similarly, we have $\delta_s(B)s_t = u_t$ and $\delta_n(B)n_t = v_t$. The nonstationary analogue of the solution to the stationary signal extraction problem is $R_t = \delta_n^*(B)\delta_n^*(F)\gamma_u(F)\gamma_w(F)^{-1}z_t$, where $\delta_n^*(B)$ is the product of the factors in $\delta_n(B)$ that are not also in $\delta_s(B)$. R_t has been shown to approximate $E(s_t|\{z_t\})$ in certain cases where s_t and n_t are not both nonstationary by Hannan (1967), Sobel (1967), and Cleveland and Tiao (1976). In this paper we obtain exact expressions for $E(s_t|\{z_t\})$, and allow either or both of s_t and n_t to be nonstationary.

Exact solutions to the nonstationary signal extraction problem require that explicit assumptions be made regarding the generation of the time series $\{z_t\}$,$\{s_t\}$, and $\{n_t\}$. The generation of these series is equivalent to the generation of suitable sets of starting values, $\underset{\sim}{z}_*$, $\underset{\sim}{s}_*$, and $\underset{\sim}{n}_*$, and the associated stationary series $\{w_t\}$,$\{u_t\}$, and $\{v_t\}$. We always assume $\{u_t\}$ and $\{v_t\}$ are generated independently and that $\{w_t\}$ is then obtained from its relation with them. Two alternative assumptions are made regarding the generation of the starting values. Under Assumption A we assume $\underset{\sim}{z}_*$ is generated independ-

ently of $\{u_t\}$,$\{v_t\}$, and $\{w_t\}$, and that $\underline{s}_*$ and $\underline{n}_*$ are then obtained through the solution of the linear equation relating them to $\underline{z}_*$. Under Assumption B we assume $\underline{s}_*$ and $\underline{n}_*$ are generated independently of $\{u_t\}$,$\{v_t\}$, and $\{w_t\}$, that $\{s_t\}$ and $\{n_t\}$ are then generated, and that $\{z_t\}$ is then generated from $z_t = s_t + n_t$. The paper discusses the different consequences of Assumptions A and B.

In the main part of the paper exact expressions for $E(s_t|\{z_t\})$ and $Var(s_t|\{z_t\})$ are obtained under Assumptions A and B. It is found that $E(s_t|\{z_t\}) = R_t$, the "classical" solution, under Assumption A with the additional condition that $\delta_s(B)$ and $\delta_n(B)$ have no common zeroes. For the other cases $E(s_t|\{z_t\})$ can be viewed as R_t plus some adjustment terms. Extensions of the results to the non-Gaussian case, to signal extraction with a finite number of observations, and to the multivariate case are discussed.

SIGNAL EXTRACTION FOR NONSTATIONARY TIME SERIES

1. Introduction

Suppose that

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

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 $\underset{\sim}{y}_{(j)}^{(i)} = (y_i, y_{i+1},\ldots,y_j)'$, where prime denotes transpose, or by $\underset{\sim}{y}_{(j)} = (y_1,\ldots,y_j)'$ if $i = 1$. If y_t is stationary with autocovariances $\gamma_y(k) = \mathrm{Cov}(y_t,y_{t+k})$ that are absolutely summable $(\sum_{-\infty}^{\infty}|\gamma_y(k)|<\infty)$ 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 denotes a dummy complex variable. It will be convenient to let $\left[\gamma_y\right]_{(m)}$ represent the variance matrix of any segment, $\underset{\sim}{y}_{(t+m)}^{(t)}$, of the stationary series y_t of length m:

$$\left[\gamma_y\right]_{(m)} = \text{Var}\,(\underset{\sim}{y}^{(t)}_{(t+m)}) = \begin{bmatrix} \gamma_y(0) & \cdots & \gamma_y(m-1) \\ \vdots & & \vdots \\ \gamma_y(m-1) & \cdots & \gamma_y(0) \end{bmatrix} .$$

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 $\left[\alpha(B)\alpha(F)\gamma_y\right]_{(m)}$ 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 8.

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 $\delta_s(B) = 1 - \delta_{s1}B - \ldots - \delta_{s,ds}B^{ds}$ 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 [2] - 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 $\delta_n(\zeta)$. 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

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

We further assume 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, pp. 78-79) w_t has an infinite autoregressive representation(invertibility)

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

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 canceled factor. Thus, we assume there are no common zeroes within the pairs $\{\delta(\zeta),\gamma_w(\zeta)\}$, $\{\delta_s(\zeta),\gamma_u(\zeta)\}$, and $\{\delta_n(\zeta),\gamma_v(\zeta)\}$.

Findley (1982) also shows that there is a minimal polynomial, $\Delta(B)$ say, that renders z_t stationary - minimal in that $\Delta(B)$ divides any other polynomial which makes z_t stationary. A question arises as to whether $\delta(B) = \Delta(B)$. Findley (1982) notes that this is the case, as is easily seen from the following argument. Suppose $\delta(B) = \delta_1(B)\Delta(B)$ where $\delta_1(B)$ is not 1 and $\Delta(B)z_t = r_t$ is stationary. Then $w_t = \delta(B)z_t = \delta_1(B)r_t$ has CGF $\gamma_w(\zeta) = \delta_1(\zeta)\delta_1(\zeta^{-1})\gamma_r(\zeta)$, which has a zero in common with $\delta(\zeta) = \delta_1(\zeta)\ \Delta(\zeta)$, a contradiction. Hence, $\delta_1(B) = 1$ and $\delta(B) = \Delta(B)$.

Previous Work

Signal extraction has been used with nonstationary time series in such areas as actuarial graduation (by Whittaker, see Whittaker and Robinson 1944, pp. 303-316), smoothing (Tiao and Hillmer 1978), and seasonal adjustment (Burman 1980). Typically, the solution for the stationary case has been borrowed and used in the nonstationary case. It 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, p. 170)

$$E(s_t | \{z_t\}) = \gamma_s(F)\gamma(F)^{-1}z_t \qquad (1.9)$$

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

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

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 $\delta_s(B)$ 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 $E(s_t|z_m,\ldots,z_{m+N})$ for large $m > 0$ when s_t and n_t follow ARMA 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.

The main results of this paper are in Section 5 where we give $E(s_t|\{z_t\})$, and in Section 6 where we give $Var(s_t|\{z_t\})$, under two sets of assumptions about starting values. These assumptions are important and are discussed in Section 3. Prior to this we discuss generation of nonstationary time series in Section 2.

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 $\delta(B)$ were outside the unit circle (so z_t would be stationary) then we could write

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

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

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

Unfortunately, when $\delta(B)$ 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 $(z_1, \ldots, z_d)' = \underset{\sim}{z}_*$. Given $\underset{\sim}{z}_*$ and $\{w_t\}$, the remaining z_t's are easily generated recursively from

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

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

(Notice $\delta_d^{-1} \neq 0$ 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 $\{\underset{\sim}{z}_*, \{w_t\}\}$ 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 $i<0$)

$$A_{1,t} = \xi_{t-1} - \xi_{t-2}\delta_1 - \cdots - \xi_{t-d}\delta_{d-1}$$

$$\vdots \tag{2.5}$$

$$A_{d-1,t} = \xi_{t-d+1} - \xi_{t-d}\delta_1$$

$$A_{d,t} = \xi_{t-d} \, .$$

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 $t=1,\ldots,d$, $A_{j,t} = 1$ 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. These results allow us to prove the following theorem.

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 $\underset{\sim}{z}_* = (z_1,\ldots,z_d)'$. Then, for $t>d$

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

where $\underset{\sim}{A}_t' = (A_{1,t},\ldots,A_{d,t})$ and the $A_{j,t}$'s are given by (2.5).

Proof. Since $A_{j,d+1} = \delta_{d+1-j}$ the result holds for $t = d+1$. We proceed by induction. Assuming the result holds through some $t>d$, for $t+1$ we get

$$\begin{aligned} z_{t+1} &= \delta_1 z_t + \cdots + \delta_d z_{t+1-d} + w_{t+1} \\ &= \delta_1\{\underset{\sim}{A}_t' \underset{\sim}{z}_* + w_t + \xi_1 w_{t-1} + \cdots + \xi_{t-1} w_1\} + \cdots \\ &+ \delta_d\{\underset{\sim}{A}_{t+1-d}' \underset{\sim}{z}_* + w_{t+1-d} + \xi_1 w_{t-d} + \cdots + \xi_{t-d} w_1\} + w_{t+1}\,. \end{aligned}$$

The coefficient of w_{t+1-i} for $i=1,\ldots,t+1-d$ is $\delta_1\xi_{i-1} + \delta_2\xi_{i-2} + \cdots + \delta_d\xi_{i-d} = \xi_i$ from (2.2), and the coefficient of w_{t+1} is $\xi_0 = 1$. The row vector coefficient of $\underset{\sim}{z}_*$ is $\delta_1 \underset{\sim}{A}_t' + \cdots + \delta_d \underset{\sim}{A}_{t-d}' = \underset{\sim}{A}_{t+1}'$. Thus, $z_{t+1} = \underset{\sim}{A}_{t+1}' \underset{\sim}{z}_* + \sum_{i=0}^{t+1-d} \xi_i w_{t+1-i}$.

QED.

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. Lemma 1 in Section 4 gives a result relating the A_{jt}'s and ξ_i's.

We could obtain an analogous backward representation for z_t for $t \leq 0$ involving the starting values $\underset{\sim}{z}_*$ and w_j, $j \leq d$. 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 <u>on</u> 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 $z_d, \ldots, z_1$, now going backwards in time, we get the following backward representation for z_t for $t \leq 0$:

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

3. <u>Assumptions About Starting Values in Signal Extraction</u>

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 $\underset{\sim}{z}_* = (z_1, \ldots, z_d)'$, $\underset{\sim}{s}_* = (s_1, \ldots, s_{ds})'$, and $\underset{\sim}{n}_* = (n_1, \ldots, n_{dn})'$, 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

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

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

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

where

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

$$
\underset{dx(d-ds)}{C_1} = \begin{bmatrix} & 0_{dsx(d-ds)} & \\ -\xi_o^s & & \\ \vdots & \ddots & \\ \xi_{d-ds-1}^s & \cdots & \xi_o^s \end{bmatrix} \qquad \underset{dx(d-dn)}{C_2} = \begin{bmatrix} & 0_{dnx(d-dn)} & \\ -\xi_o^n & & \\ \vdots & \ddots & \\ \xi_{d-dn-1}^n & \cdots & \xi_o^n \end{bmatrix},
$$

which relates the starting values for $\underset{\sim}{s}_t$ and $\underset{\sim}{n}_t$ ($\underset{\sim}{s}_*$ and $\underset{\sim}{n}_*$) to those for z_t ($\underset{\sim}{z}_*$).

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

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

<u>Assumption B</u>: $\underset{\sim}{s}_*$ and $\underset{\sim}{n}_*$ 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 $\underset{\sim}{z}_*$ is obtained through $z_t = s_t + n_t$, t=1,...,d.

In section A.I. of the Appendix we show that the dx(ds+dn) matrix $\begin{bmatrix} H_1 & H_2 \end{bmatrix}$ 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,...,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, $\begin{bmatrix} H_1 & H_2 \end{bmatrix}^-$, of $\begin{bmatrix} H_1 & H_2 \end{bmatrix}$, so in making Assumption A with dc>0, one must also make an assumption as to which generalized inverse of $\begin{bmatrix} H_1 & H_2 \end{bmatrix}$ 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 $\text{Cov}(s_t, n_j) = -\text{Cov}(n_1, n_j)$ 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 $\underset{\sim}{z}_*$ 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 $w_t, w_{t-1},\ldots$ 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 $z_t = z_1 + \sum_{i=0}^{t-2} w_i$ with $z_1 = s_1 + n_1$, and

$$a_{t+\ell} = (1-\theta B)^{-1}w_{t+\ell} = (1-\theta B)^{-1}u_{t+\ell} + \left[1 - (1-\theta)\sum_{i=1}^{\infty}\theta^{i-1}B^i\right]n_{t+\ell} ,$$

implying $\text{Cov}(z_t,a_{t+\ell}) = \text{Cov}(n_1, n_{t+\ell} - (1-\theta)\sum_{i=1}^{\infty}\theta^{i-1}n_{t+\ell-i}) = -(1-\theta)\,\theta^{t+\ell-2}\,\gamma_n(0)$.

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 here 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)). These results reflect the fact that our arbitrary assumption that the starting values for z_t occur at time points $1,\ldots,d$ is important. This situation is disturbing, but we obviously must make some assumption of this sort.

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. Of course, assumptions other than A or B could be used. In Section 5 we will obtain signal extraction results under both Assumptions A and B.

4. Preliminary Results

Before proceeding to the signal extraction results in Section 5, we give some preliminary results that will be needed there.

Lemma 1. Given $\delta(B)$ let the ξ_i be as defined in (2.2) and the A_{it} as defined in (2.5). Then for $t>d$

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

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

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

Lemma 2. Let Y_t and X_t be jointly stationary Gaussian time series with cross covariances $\gamma_{yx}(k) = E(Y_t X_{t+k})$. Then

$$E(Y_t|\{X_t\}) = \gamma_{yx}(F)\gamma_x(F)^{-1}X_t$$

where $\gamma_{yx}(\zeta) = \sum_{-\infty}^{\infty}\gamma_{yx}(k)\zeta^k$.

This result is given in Brillinger (1975, theorems 8.3.1 and 8.3.2). Notice that in the stationary signal extraction problem (with $\{s_t\}$ independent of $\{n_t\}$) $\gamma_{zs}(k) = \gamma_s(k)$ so that Lemma 2 immediately gives the stationary signal extraction result (1.9).

For the next theorem let I be an arbitrary index set, countable or uncountable, and let Y and $\{X_i\} = \{X_i, i\varepsilon I\}$ be elements of $\mathcal{L}$, the Hilbert space of zero mean, finite variance random variables over some given probability space. We temporarily drop our assumption that they be normally distributed. Let $\mathcal{L}\{X_i\}$ denote the closed linear subspace generated by the X_i's. We call $\tilde{Y}$ a linear predictor of Y (given the X_i's) if $\tilde{Y}\varepsilon\ \mathcal{L}\{X_i\}$. The best linear predictor, $\hat{Y}$, of Y is the almost surely (a.s.) unique element of $\mathcal{L}\{X_i\}$ that minimizes $E[(Y-\hat{Y})^2]$ -- this is the projection of Y on $\mathcal{L}\{X_i\}$.

Theorem 2. Let $\tilde{Y}\varepsilon\ \mathcal{L}\{X_i, i\varepsilon I\}$. The following statements are equivalent.

1. $\tilde{Y} = \hat{Y}$, the (a.s.) unique best linear predictor of Y.
2. $\tilde{\varepsilon} = Y-\tilde{Y}$ is uncorrelated with X_i, for all $i\varepsilon\ I$.
3. $\tilde{Y} = E(Y|\{X_i, i\varepsilon\ I\})$ when Y and $\{X_i, i\varepsilon I\}$ are jointly normally distributed.

This theorem is essentially stated and proved in the book by Gikhman and Skorokhod (1969, see their theorem 2, p.229, and section 3.3, pp.111-118). We use this theorem to prove the next result.

Lemma 3. Assume Y, $\{X_i\}$, $\{W_j\}$ (j in some index set J) are jointly normal and in $\mathcal{L}$. Then

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

Note. This result has been used by Sobel (1967) without proof. It is the infinite dimensional analogue to orthogonalizing the independent variables in finite dimensional regression.

Proof. We prove this by proving the analogous result for best linear predictions and using Theorem 2. Let $\hat{Y}\{X_i\}$ be the projection of Y on $\mathcal{L}\{X_i\}$, $\hat{W}_j\{X_i\}$ the projection of W_j on $\mathcal{L}\{X_i\}$, and $U_j = W_j - \hat{W}_j\{X_i\}$, $j \varepsilon J$. We first notice that

$$\mathcal{L}(\{X_i\},\{W_j\}) = \mathcal{L}(\{X_i\},\{W_j\},\{U_j\}) = \mathcal{L}(\{X_i\}\{U_j\}), \tag{4.1}$$

the first equality being true since $\hat{W}_j\{X_i\} \varepsilon \mathcal{L}\{X_i\}$ implies $U_j \varepsilon \mathcal{L}(\{X_i\},\{W_j\})$ for all j, and the second since $W_j = U_j + \hat{W}_j\{X_i\} \varepsilon \mathcal{L}(\{X_i\},\{U_j\})$ for all j.

Now consider $\varepsilon = Y - \hat{Y}\{X_i\} - \hat{Y}\{U_j\}$. By Theorem 2, $Y-\hat{Y}\{X_i\}$ is uncorrelated with X_i for all i, and U_j is uncorrelated with X_i for all i and j. $\hat{Y}\{U_j\} \varepsilon \mathcal{L}\{U_j\}$ so it is either a linear combination of the U_j, or it is a mean squared limit of such random variables (Robinson 1959, p.49). In either case $\hat{Y}\{U_j\}$ can be shown to be uncorrelated with X_i for all i. Thus, ε is uncorrelated with X_i for all i. It follows in a similar fashion that ε is uncorrelated with U_j for all j. Thus, by Theorem 2 $\hat{Y}\{X_i\} + \hat{Y}\{U_j\}$ is the projection of Y on $\mathcal{L}(\{X_i\},\{U_j\})$, hence by (4.1) the projection of Y on $\mathcal{L}(\{X_i\}, \{W_j\})$. By Theorem 2 again the result holds for conditional expectations under normality. QED

5. Signal Extraction for Nonstationary Time Series

We now return to the signal extraction problem and obtain a general expression for $E(s_t|\{z_t\})$ in the nonstationary case. In subsections 5.1 and 5.2 we obtain specific results under Assumptions A and B respectively. The results are given for $E(s_t|\{z_t\})$ 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 $E(n_t|\{z_t\})$, which can be done by relabelling, and then compute $E(s_t|\{z_t\})$ as $z_t - E(n_t|\{z_t\})$.

From the discussion in Section 2, $E(s_t|\{z_t\}) = E(s_t|\underset{\sim}{z}_*,\{w_t\})$, so by (3.1), the linearity of conditional expectations, and Lemma 3

$$E(s_t|\{z_t\}) = \underset{\sim}{A}_t^{s\prime}[E(\underset{\sim}{s}_*|\{w_t\}) + E(\underset{\sim}{s}_*|\underset{\sim}{z}_*-E(\underset{\sim}{z}_*|\{w_t\}))]$$
$$+ \sum_{i=0}^{t-ds-1} \xi_i^s\left[E(u_{t-i}|\{w_t\}) + E(u_{t-i}|\underset{\sim}{z}_* - E(\underset{\sim}{z}_*|\{w_t\}))\right]. \tag{5.1}$$

By Lemma 2

$$\begin{aligned} E(u_{t-i}|\{w_t\}) &= \gamma_{uw}(F)\gamma_w(F)^{-1}w_{t-i} \\ &= \delta_n{*}(F)\gamma_u(F)\gamma_w(F)^{-1}\delta_s(B)\delta_n{*}(B)z_{t-i} \\ &= \delta_s(B)\, R_{t-i}. \end{aligned} \tag{5.2}$$

The second equality in (5.2) uses (1.4) and the result $\gamma_{uw}(\zeta) = \delta_n{*}(\zeta)\gamma_u(\zeta)$, which follows from (1.5). The last equality uses (1.10). Then, applying Lemma 1

$$\begin{aligned} \sum_{i=0}^{t-ds-1} \xi_i^s E(u_{t-i}|\{w_t\}) &= \left(\sum_{i=0}^{t-ds-1} \xi_i^s B^i\right)\delta_s(B)R_t \\ &= \left[1 - \sum_{i=1}^{ds} A_{i,t}^s B^{t-i}\right]R_t \\ &= R_t - \underset{\sim}{A}_t^{s\prime}\underset{\sim}{R}_{(ds)}. \end{aligned} \tag{5.3}$$

Using (5.3),(5.1) can be written as

$$E(s_t|\{z_t\}) = \underset{\sim}{A}_t^{s\prime}\{E(\underset{\sim}{s}_*|\{z_t\}) - \underset{\sim}{R}_{(ds)}\} + R_t \tag{5.4}$$

$$+ \sum_{i=0}^{t-ds-1} \xi_i^s E(u_{t-i}|\underset{\sim}{z}_* - E(\underset{\sim}{z}_*|\{w_t\})) .$$

Notice that (5.4) includes R_t, the nonstationary analogue of the stationary solution (see (1.10)). R_t has been rather widely used in nonstationary signal extraction and has been shown (Sobel 1967, Cleveland and Tiao 1976) to approximate $E(s_t|\{z_t\})$ in certain cases. However, (5.4) also includes an adjustment for the effect of the deviation of $E(s_j|\{z_t\})$ from R_j for $j=1,\ldots,ds$, plus an adjustment for what $\underset{\sim}{z}_*$ 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 values s_j $(j=1,\ldots,ds)$, and $\underset{\sim}{z}_*$ contains no information on the u_{t-i}'s beyond that in $\{w_t\}$, then R_t will be $E(s_t|\{z_t\})$.

Actually, the steps in (5.2) above need 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 (5.2) (this is not obvious since z_t is nonstationary.) Conditions under which these things hold are given in section A.II of the Appendix. Here we merely note that these conditions will hold, in particular, if all the zeroes of $\delta(\zeta)$ are on the unit circle, $\gamma_w(\zeta)$ 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 (5.2). 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 $E(u_{t-i}|\{w_t\})$ in (5.1) and proceeding from there instead of from (5.4). This approach is used in Section 6 when we obtain $\mathrm{Var}(s_t|\{z_t\})$.

5.1 Signal Extraction Under Assumption A

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

Theorem 3. Make Assumption A so that $\underset{\sim}{z}_*$ is independent of $\{w_t\}$. Also assume $\delta_s(B)$ and $\delta_n(B)$ have no common zeroes so $\delta(B) = \delta_s(B)\delta_n(B)$. Then

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

Proof. The signal extraction error is

$$\begin{aligned}\varepsilon_t &= s_t - R_t\\ &= s_t - \delta_n(B)\delta_n(F)\gamma_u(F)\gamma_w(F)^{-1}(s_t + n_t)\\ &= \delta_s(B)\delta_s(F)\gamma_v(F)\gamma_w(F)^{-1}s_t - \delta_n(B)\delta_n(F)\gamma_u(F)\gamma_w(F)^{-1}n_t\\ &= \delta_s(F)\gamma_v(F)\gamma_w(F)^{-1}u_t - \delta_n(F)\gamma_u(F)\gamma_w(F)^{-1}v_t \qquad (5.5)\end{aligned}$$

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 $f_u(\lambda)\delta_s(e^{-i\lambda})$ (see 1.5), so from (5.5) the cross spectral density of ε_t with w_t is

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

This shows ε_t is uncorrelated with every w_j, and since $\underset{\sim}{z}_*$ is assumed independent of $\{u_t\}$ and $\{v_t\}$, ε_t is also uncorrelated with $\underset{\sim}{z}_*$. Thus, ε_t is uncorrelated with $\{z_t\}$, and by Theorem 2 the proof is complete. QED

We now consider the case where $\delta_s(B)$ and $\delta_n(B)$ 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 (5.4), we notice that the independence of $\underset{\sim}{z}_*$ and $\{w_t\}$ under Assumption A implies that $E(\underset{\sim}{z}_*|\{w_t\}) = \underset{\sim}{0}$, so the last term in (5.4) involves $E(u_{t-i}|\underset{\sim}{z}_*)$'s, which are all zero since $\underset{\sim}{z}_*$ is independent of $\{u_t\}$. Hence, the last term in (5.4) drops out and we only need to evaluate $E(\underset{\sim}{s}_*|\{z_t\}) = E(\underset{\sim}{s}_*|\underset{\sim}{z}_*) + E(\underset{\sim}{s}_*|\{w_t\})$. From (3.2) we see that

$$\begin{bmatrix} H_1 & H_2 \end{bmatrix} \begin{bmatrix} E(\underset{\sim}{s}_*|\{z_t\}) \\ E(\underset{\sim}{n}_*|\{z_t\}) \end{bmatrix} = \underset{\sim}{z}_* - C_1E(\underset{\sim}{u}_{(d)}^{(ds+1)}|\{w_t\}) - C_2E(\underset{\sim}{v}_{(d)}^{(dn+1)}|\{w_t\}) \ . \quad (5.6)$$

To obtain $E(\underset{\sim}{s}_*|\{z_t\})$ and $E(\underset{\sim}{n}_*|\{z_t\})$ we would solve (5.6) in the same way we assume (3.2) was solved in generating $\underset{\sim}{s}_*$ and $\underset{\sim}{n}_*$, i.e., we make the same choice of $\begin{bmatrix} H_1 & H_2 \end{bmatrix}^-$.

It remains for us to simplify the right hand side of (5.6). From (3.2) we see that the j^{th} element of $C_1E(\underset{\sim}{u}_{(d)}^{(ds+1)}|\{w_t\})$ is zero for j=1,...,ds, and by (5.3) is

$$\sum_{i=0}^{j-ds-1} \xi_i^s E(u_{j-i}|\{w_t\}) = R_j - \underset{\sim}{A}_j^{s\prime}\underset{\sim}{R}_{(ds)} \quad (5.7)$$

for j=ds+1,...,d. The j^{th} element of $C_2E(\underset{\sim}{v}_{(d)}^{(dn+1)}|\{w_t\})$ is also zero for j=1, ...,dn. Analogous to (5.2) we have $E(v_{j-i}|\{w_t\}) = \delta_s^*(F)\gamma_v(F)\gamma_w(F)^{-1}\delta_n(B)\delta_s^*(B)z_{j-i}$, and since (1.6) and (1.10) imply that

$$\delta_s^*(B)\delta_s^*(F)\gamma_v(F)\gamma_w(F)^{-1}z_t = z_t - R_t \ , \quad (5.8)$$

we get

$$\sum_{i=0}^{j-dn-1} \xi_i^n E(v_{j-i}|\{w_t\}) = (z_j - R_j) - \underset{\sim}{A}_j^{n\prime}(\underset{\sim}{z}_{(dn)} - \underset{\sim}{R}_{(dn)}) \quad (5.9)$$

for $j=dn+1,\ldots,d$. We now have that

$$C_1E(\underset{\sim}{u}_{(d)}^{(ds+1)}|\{w_t\}) + C_2E(\underset{\sim}{v}_{(d)}^{(dn+1)}|\{w_t\}) =$$

$$\begin{bmatrix} 0 \\ \cdot \\ \cdot \\ 0 \\ R_{ds+1} \\ \cdot \\ \cdot \\ R_d \end{bmatrix} - \begin{bmatrix} \underset{\sim}{0}' \\ \cdot \\ \cdot \\ \underset{\sim}{0}' \\ \underset{\sim}{A}_{ds+1}^{s\prime} \\ \cdot \\ \cdot \\ \underset{\sim}{A}_d^{s\prime} \end{bmatrix} \begin{bmatrix} R_1 \\ \cdot \\ \cdot \\ R_{ds} \end{bmatrix} + \begin{bmatrix} 0 \\ \cdot \\ \cdot \\ 0 \\ z_{dn+1}-R_{dn+1} \\ \cdot \\ \cdot \\ z_d-R_d \end{bmatrix} - \begin{bmatrix} \underset{\sim}{0}' \\ \cdot \\ \cdot \\ \underset{\sim}{0}' \\ \underset{\sim}{A}_{dn+1}^{n\prime} \\ \cdot \\ \cdot \\ \underset{\sim}{A}_d^{n\prime} \end{bmatrix} \begin{bmatrix} z_1-R_1 \\ \cdot \\ \cdot \\ z_{dn}-R_{dn} \end{bmatrix}$$

$$= \{I_d - [H_1 \; 0_{dx(d-ds)}]\} \begin{bmatrix} R_1 \\ \cdot \\ \cdot \\ \cdot \\ R_d \end{bmatrix} + \{I_d - [H_2 \; 0_{dx(d-dn)}]\} \begin{bmatrix} z_1-R_1 \\ \cdot \\ \cdot \\ \cdot \\ z_d-R_d \end{bmatrix}$$

$$= \underset{\sim}{z}_* - H_1\underset{\sim}{R}_{(ds)} - H_2\underset{\sim}{z}_{(dn)} + H_2\underset{\sim}{R}_{(dn)} \; . \tag{5.10}$$

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

$$\underset{dxm}{H_3} = H_1 - \begin{bmatrix} H_2 \; 0_{dx(ds-dn)} \end{bmatrix} \quad ds \geq dn$$
$$= \begin{bmatrix} H_1 \; 0_{dx(dn-ds)} \end{bmatrix} - H_2 \quad ds < dn \; . \tag{5.11}$$

We can now write (5.6) as

$$\begin{bmatrix} H_1 & H_2 \end{bmatrix} \begin{bmatrix} E(\underset{\sim}{s}_*|\{z_t\}) \\ E(\underset{\sim}{n}_*|\{z_t\}) \end{bmatrix} = H_2\underset{\sim}{z}_{(dn)} + H_3\underset{\sim}{R}_{(m)} \; . \tag{5.12}$$

We summarize our results in the following theorem.

<u>Theorem 4</u>. Make Assumption A and assume $\delta_s(B)$ and $\delta_n(B)$ have at least one common zero. Then for $t > ds$

$$E(s_t|\{z_t\}) = \underset{\sim}{A}_t^{s\prime}\{E(\underset{\sim}{s}_*|\{z_t\}) - \underset{\sim}{R}_{(ds)}\} + R_t$$

where $R_t = \delta_n^*(B)\delta_n^*(F)\gamma_u(F)\gamma_w(F)^{-1}z_t$. $E(\underset{\sim}{s}_*|\{z_t\})$ is obtained by solving (5.12) in the same way it is assumed that (3.2) is solved in generating $\underset{\sim}{s}_*$ and $\underset{\sim}{n}_*$, i.e. we make the same choice of $\begin{bmatrix} H_1 & H_2 \end{bmatrix}^-$.

5.2 Signal Extraction Under Assumption B

Under Assumption B $\underset{\sim}{s}_*$ and $\{w_t\}$ are independent, so that $E(\underset{\sim}{s}_*|\{w_t\}) = \underset{\sim}{0}$ and (5.4) becomes

$$E(s_t|\{z_t\}) = \underset{\sim}{A}_t^{s'}\{E[\underset{\sim}{s}_*|\underset{\sim}{z}_*-E(\underset{\sim}{z}_*|\{w_t\})]-\underset{\sim}{R}_{(ds)}\}+R_t+\sum_{i=0}^{t-ds-1}\xi_i^s E[u_{t-i}|\underset{\sim}{z}_*-E(\underset{\sim}{z}_*|\{w_t\})]. \tag{5.13}$$

For the case where n_t is stationary and $\delta_s(B)$ has no zeroes inside the unit circle, Sobel (1967) establishes that R_t converges to $E(s_t|\{z_t\})$ as $t\to\infty$. Cleveland and Tiao (1976) similarly show that R_t approximates $E(s_t|\underset{\sim}{z}^{(m)}_{(m+N)})$ for $m\leq t\leq m+N$ with m and N large and t not near m or m+N. We now show how to evaluate (5.13) exactly by obtaining $\underset{\sim}{z}_* - E(\underset{\sim}{z}_*|\{w_t\})$, its variance matrix, and its covariances with $\underset{\sim}{s}_*$ and $\sum_{i=0}^{t-ds-1}\xi_i^s u_{t-i}$.

To evaluate $\underset{\sim}{z}_* - E(\underset{\sim}{z}_*|\{w_t\})$ under Assumption B we notice $\underset{\sim}{n}_*$ and $\underset{\sim}{s}_*$ are both independent of $\{w_t\}$ so $E([\underset{\sim}{s}_*' \ \underset{\sim}{n}_*']|\{w_t\}) = \underset{\sim}{0}'$. Then, by (3.2), (5.10), and (5.11), we get

$$\underset{\sim}{z}_* - E(\underset{\sim}{z}_*|\{w_t\}) = H_2\underset{\sim}{z}_{(dn)} + H_3\underset{\sim}{R}_{(m)}. \tag{5.14}$$

To compute the variance matrix of $\underset{\sim}{z}_* - E(\underset{\sim}{z}_*|\{w_t\})$ we need a different expression than (5.14). For t>ds, from (3.1), (5.2), (1.5), and (1.6) we get

$$\begin{aligned} s_t - E(s_t|\{w_t\}) &= \underset{\sim}{A}_t^{s'}\underset{\sim}{s}_* + (\sum_{i=0}^{t-ds-1}\xi_i^s B^i)\Big[u_t - \\ &\quad \delta_n^*(F)\gamma_u(F)\gamma_w(F)^{-1}(\delta_n^*(B)u_t+\delta_s^*(B)v_t)\Big] \\ &= \underset{\sim}{A}_t^{s'}\underset{\sim}{s}_* + (\sum_{i=0}^{t-ds-1}\xi_i^s B^i)\delta_s^*(B)\Big[\delta_s^*(F)\gamma_v(F)\gamma_w(F)^{-1}u_t \\ &\quad - \delta_n^*(F)\gamma_u(F)\gamma_w(F)^{-1}v_t\Big] \end{aligned} \tag{5.15}$$

Similarly, for t > dn

$$n_t - E(n_t|\{w_t\}) = \underset{\sim}{A}_t^{n}{}'\underset{\sim}{n}_* + (\sum_{i=0}^{t-dn-1} \xi_i^n B^i)\delta_n^*(B)\Big[\delta_n^*(F)\gamma_u(F)\gamma_w(F)^{-1}v_t$$

$$- \delta_s^*(F)\gamma_v(F)\gamma_w(F)^{-1}u_t\Big] . \qquad (5.16)$$

Both (5.15) and (5.16) involve the stationary time series X_t defined by

$$X_t = \delta_s^*(F)\gamma_v(F)\gamma_w(F)^{-1}u_t - \delta_n^*(F)\gamma_u(F)\gamma_w(F)^{-1}v_t , \qquad (5.17)$$

which has spectral density

$$f_x(\lambda) = |\delta_s^*(e^{i\lambda})|^2 f_v(\lambda)f_w(\lambda)^{-1}f_u(\lambda)f_w(\lambda)^{-1}f_v(\lambda)$$

$$+ |\delta_n^*(e^{i\lambda})|^2 f_u(\lambda)f_w(\lambda)^{-1}f_v(\lambda)f_w(\lambda)^{-1}f_u(\lambda) .$$

This 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

$$\gamma_x(\zeta) = \gamma_u(\zeta)\gamma_w(\zeta)^{-1}\gamma_v(\zeta) . \qquad (5.18)$$

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 $\gamma_x(k)$.

For $t \le ds$ $s_t - E(s_t|\{w_t\}) = s_t$, and for $t \le dn$ $n_t - E(n_t|\{w_t\}) = n_t$. Taking these results along with (5.15), (5.16), and (5.17) gives

$$\underset{\sim}{z}_* - E(\underset{\sim}{z}_*|\{w_t\}) = H_1\underset{\sim}{s}_* + H_2\underset{\sim}{n}_* + G_3 \underset{\sim}{X}_{(d)} \qquad (5.19)$$

where $G_3 = G_1 - G_2$ with the dxd lower triangular matrices G_1 and G_2 defined as follows. The first ds rows of $G_1 = (g_{ij}^{(1)})$ consist of all zeroes. For $t>ds$ $g_{tj}^{(1)}$ is the coefficient of B^{t-j} in $(\sum_{i=0}^{t-ds-1} \xi_i^s B^i)\delta_s^*(B)$, which is a polynomial in B of degree t-dc-l. So

$$g_{tt}^{(1)} = 1 \quad g_{tj}^{(1)} = \begin{cases} 0 & j>t \text{ or } j\le dc \\ \text{coefficient of } B^{t-j} \text{ in } (\sum_{i=0}^{t-ds-1} \xi_i^s B^i)\delta_s^*(B) & dc<j<t. \end{cases} \quad (5.20)$$

Similarly, the first dn rows of $G_2 = (g_{ij}^{(2)})$ consist of all zeroes and for $t > dn$

$$g_{tt}^{(2)} = 1 \quad g_{tj}^{(2)} = \begin{cases} 0 & j>t \text{ or } j\le dc \\ \text{coefficient of } B^{t-j} \text{ in } (\sum_{i=0}^{t-dn-1} \xi_i^n B^i)\delta_n^*(B) & dc<j<t. \end{cases} \quad (5.21)$$

Let Ω denote $\mathrm{Var}(\tilde{z}_*-E(\tilde{z}_*|\{w_t\}))$. Since $\tilde{s}_*$, $\tilde{n}_*$, and $\tilde{X}_{(d)}$ are independent, from (5.19) we get

$$\Omega \equiv \mathrm{Var}(\tilde{z}_*-E(\tilde{z}_*|\{w_t\})) = H_1\mathrm{Var}(\tilde{s}_*)H_1' + H_2\mathrm{Var}(\tilde{n}_*)H_2' + G_3\left[\gamma_x\right]_{(d)}G_3' . \quad (5.22)$$

The independence of $\tilde{s}_*$, $\tilde{n}_*$, and $\{X_t\}$ yields that (from (5.19))

$$\mathrm{Cov}(\tilde{s}_*,\tilde{z}_*-E(\tilde{z}_*|\{w_t\})) = \mathrm{Var}(\tilde{s}_*)H_1' . \quad (5.23)$$

Then, from (5.22), (5.23), (5.14), and the result that $E(\tilde{Y}_1|\tilde{Y}_2) = \mathrm{Cov}(\tilde{Y}_1,\tilde{Y}_2)\mathrm{Var}(\tilde{Y}_2)^{-1}\tilde{Y}_2$ for zero mean normal random vectors (Brillinger 1975, p. 292)

$$E(\tilde{s}_*|\tilde{z}_*-E(\tilde{z}_*|\{w_t\})) = \mathrm{Var}(\tilde{s}_*)H_1'\Omega^{-1}\{H_2\tilde{z}_{(dn)} + H_3\tilde{R}_{(m)}\} . \quad (5.24)$$

If Ω is singular we can use any generalized inverse of it in (5.24).

Finally, we consider $E(\sum_{i=0}^{t-ds-1} \xi_i^s u_{t-i}|\tilde{z}_* - E(\tilde{z}_*|\{w_t\}))$. Notice $\sum_{i=0}^{t-ds-1} \xi_i^s u_{t-i}$ is independent of $\tilde{s}_*$ and $\tilde{n}_*$, so by (5.19) we only need consider its covariance with $\tilde{X}_{(d)}$. From (5.17) we see

$$\begin{aligned} \mathrm{Cov}(\sum_{i=0}^{t-ds-1} \xi_i^s u_{t-i}, X_j) &= \mathrm{Cov}((\sum_{i=0}^{t-ds-1} \xi_i^s B^i)u_t,\ \delta_s^*(F)\gamma_v(F)\gamma_w(F)^{-1}u_j) \\ &= (\sum_{i=0}^{t-ds-1} \xi_i^s F^i)\delta_s^*(F)\mathrm{Cov}(u_t,\gamma_v(F)\gamma_w(F)^{-1}u_j) \end{aligned} \quad (5.25)$$

where F operates on j in (5.25). The cross spectral density between the time series u_t and $\gamma_v(F)\gamma_w(F)^{-1}u_t$ is $f_u(\lambda)f_w(\lambda)^{-1}f_v(\lambda) = f_x(\lambda)$ so that $Cov(u_t,\gamma_v(F)\gamma_w(F)^{-1}u_j) = \gamma_x(j-t)$. Using this and (5.20) we write (5.25) as (F applies to j)

$$\begin{aligned} Cov(\sum_{i=0}^{t-ds-1} \xi_i^s u_{t-i}, X_j) &= (\sum_{i=dc+1}^{t} g_{ti}^{(1)} F^{t-i})\gamma_x(j-t) \\ &= \sum_{i=dc+1}^{t} g_{ti}^{(1)}\gamma_x(j-i), \end{aligned} \tag{5.26}$$

where for t>d, $g_{ti}^{(1)}$ is still defined by (5.20). Taking (5.26) for j=1,...,d we obtain

$$Cov(\sum_{i=0}^{t-ds-1} \xi_i^s u_{t-i}, \underset{\sim}{X}_{(d)}) = \begin{bmatrix} 0 & \cdots & 0 & g_{t,dc+1}^{(1)} & \cdots & g_{tt}^{(1)} \end{bmatrix} \begin{bmatrix} \gamma_x(0) & \cdots & \gamma_x(d-1) \\ \vdots & & \vdots \\ \gamma_x(1-t) & \cdots & \gamma_x(d-t) \end{bmatrix}$$

We thus have

$$\begin{aligned} E(\sum_{i=0}^{t-ds-1} \xi_i^s u_{t-i} | \underset{\sim}{z}_* - E(\underset{\sim}{z}_* | \{w_t\})) &= Cov(\sum_{i=0}^{t-ds-1} \xi_i^s u_{t-i}, \underset{\sim}{X}_{(d)}) \times \\ & G_3' \, \Omega^{-1}\{H_2 \underset{\sim}{Z}_{(dn)} + H_3 \underset{\sim}{R}_{(m)}\} \end{aligned} \tag{5.28}$$

and use (5.27) in evaluating (5.28).

We summarize our results in a theorem.

<u>Theorem 5</u>. Make Assumption B so that $\underset{\sim}{s}_*$ and $\underset{\sim}{n}_*$ are independent of each other and of $\{w_t\}$. Then, for t > ds

$$E(s_t|\{z_t\}) = \underset{\sim}{A}^{s\,'}_t \{ E[\underset{\sim}{s}_* | \underset{\sim}{z}_* - E(\underset{\sim}{z}_* | \{w_t\})] - \underset{\sim}{R}_{(ds)} \}$$

$$+R_t + E\left[\sum_{i=0}^{t-ds-1} \xi_i^s u_{t-i} \,|\, \underset{\sim}{z}_* - E(\underset{\sim}{z}_* | \{w_t\})\right]$$

where $R_t = \delta_n^*(B)\delta_n^*(F)\gamma_u(F)\gamma_w(F)^{-1}z_t$, $E\left[\underset{\sim}{s}_* | \underset{\sim}{z}_* - E(\underset{\sim}{z}_* | \{w_t\})\right]$

is given by (5.24), and $E\left[\sum_{i=0}^{t-ds-1} \xi_i^s u_{t-i} \,|\, \underset{\sim}{z}_* - E(\underset{\sim}{z}_* | \{w_t\})\right]$ is given by (5.28), using (5.27) and (5.20). The covariance generating function $\gamma_x(\zeta)$ given by (5.18) may be used to compute the $\gamma_x(k)$ needed.

If $\delta_s(B)$ and $\delta_n(B)$ 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 by Lemma 1

$$\left(\sum_{i=0}^{t-ds-1} \xi_i^s B^i\right) \delta_s^*(B) = 1 - \sum_{i=1}^{ds} A_{it}^s B^{t-i}$$

$$\left(\sum_{i=0}^{t-dn-1} \xi_i^n B^i\right) \delta_n^*(B) = 1 - \sum_{i=1}^{dn} A_i^n B^{t-i}$$

It then follows that $(m = \max(ds, dn))$

$$G_3 \underset{\sim}{X}_{(d)} = \left\{ \begin{bmatrix} 0 & 0 \\ -\underset{\sim}{A}^{s\,'}_{ds+1} & 1 \\ \vdots & \ddots \\ -\underset{\sim}{A}^{n\,'}_{dn+1} & \quad 1 \end{bmatrix} - \begin{bmatrix} 0 & 0 \\ -\underset{\sim}{A}^{n\,'}_{dn+1} & 1 \\ \vdots & \ddots \\ -\underset{\sim}{A}^{n\,'}_{d} & \quad 1 \end{bmatrix} \right\} \underset{\sim}{X}_{(d)} = -H_3 \underset{\sim}{X}_{(m)}, \quad (5.29)$$

and hence $\underset{\sim}{z}_* - E(\underset{\sim}{z}_* | w_t\}) = H_1 \underset{\sim}{s}_* + H_2 \underset{\sim}{n}_* - H_3 \underset{\sim}{X}_{(m)}$.

We thus have

$$\Omega \equiv \mathrm{Var}(\underset{\sim}{z}_* - E(\underset{\sim}{z}_* | \{w_t\})) = H_1 \mathrm{Var}(\underset{\sim}{s}_*) H_1' + H_2 \mathrm{Var}(\underset{\sim}{n}_*) H_2' + H_3 [\gamma_x]_{(m)} H_3' \quad (5.30)$$

in place of (5.22), and substitute $-H_3$ for G_3 in (5.28). In addition, (5.25) leads to

$$\mathrm{Cov}\left(\sum_{i=0}^{t-ds-1} \xi_i^s u_{t-i}, X_j\right) = \left(1 - \sum_{i=1}^{ds} A_{it}^s F^{t-i}\right) \gamma_x(j-t)$$

$$= \gamma_x(j-t) - \sum_{i=1}^{ds} A_{it}^s \gamma_x(j-i).$$

We thus replace (5.27) by

$$\mathrm{Cov}\left(\sum_{i=0}^{t-ds-1} \xi_i^s u_{t-i}, \underset{\sim}{X}_{(m)}\right) = \left[\gamma_x(1-t) \ \dots \ \gamma_x(m-t)\right] - \underset{\sim}{A}_t^{s\prime} \begin{bmatrix} \gamma_x(0) & \dots & \gamma_x(m-1) \\ \vdots & & \vdots \\ \gamma_x(1-ds) & \dots & \gamma_x(m-ds) \end{bmatrix} \quad (5.31)$$

and use this in place of $\mathrm{Cov}\left(\sum_{i=0}^{t-ds-1} \xi_i^s u_{t-i}, \underset{\sim}{X}_{(d)}\right)$ in (5.28).

6. Variances of Signal Extraction Errors

In many applications of signal extraction we want to compute not only the estimate $E(s_t | \{z_t\})$, but also the conditional variance, $\mathrm{Var}(s_t|\{z_t\})$. This is the same as the variance of the signal extraction error, $\varepsilon_t = s_t - E(s_t|\{z_t\})$. One setting where this is important is in seasonal adjustment by signal extraction; the results given here could be used to compute the conditional variance of the seasonally adjusted data.

When R_t is used instead of $E(s_t|\{z_t\})$, Hannan (1967) gives the variance of the resulting error, $\mathrm{Var}(s_t - R_t)$, 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 7 and the discussion following it in the next subsection).

To obtain Var $(s_t|\{z_t\})$ notice that by Lemma 3 we may write

$$s_t = \varepsilon_t + E(s_t|\{z_t\}) \tag{6.1}$$

$$= \varepsilon_t + E(s_t|\{w_t\}) + E\left[s_t|z_* - E(z_*|\{w_t\})\right]. \tag{6.2}$$

By Theorem 2, the terms on the right hand side in (6.2) are independent (and similarly for (6.1)). Therefore,

$$\mathrm{Var}(s_t) = \mathrm{Var}(\varepsilon_t) + \mathrm{Var}(E(s_t|\{w_t\})) + \mathrm{Var}(E\left[s_t|\underset{\sim}{z}_* - E(\underset{\sim}{z}_*|\{w_t\})\right]),$$

and we may compute $\mathrm{Var}(s_t|\{z_t\}) = \mathrm{Var}(\varepsilon_t)$ as

$$\mathrm{Var}(s_t|\{z_t\}) = \mathrm{Var}(s_t) - \mathrm{Var}(E(s_t|\{z_t\})) \tag{6.3}$$

$$= \mathrm{Var}(s_t) - \mathrm{Var}(E(s_t|\{w_t\})) - \mathrm{Var}(E\left[s_t|\underset{\sim}{z}_* - E(\underset{\sim}{z}_*|\{w_t\})\right]). \tag{6.4}$$

(6.3) and (6.4) will still hold if we replace s_t by a vector , say $(s_1,\ldots,s_t)' = \underset{\sim}{s}_{(t)}$, so we can use (6.3) and (6.4) to compute variances and covariances of the s_t's conditional on $\{z_t\}$. We now show how to evaluate the required terms in (6.3) and (6.4) under Assumptions A and B.

6.1 Variances Under Assumption A

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

To begin, we notice from (3.1) that

$$\underset{\sim}{s}_{(t)} = H_{1t}\,\underset{\sim}{s}_* + C_{1t}\,\underset{\sim}{u}_{(t)}^{(ds+1)} \tag{6.5}$$

where

$$H_{1t} = \begin{bmatrix} I_{ds} \\ A^{s'}_{ds+1} \\ \vdots \\ A^{s'}_{t} \end{bmatrix} \qquad C_{1t} = \begin{bmatrix} & 0_{dsx(t-ds)} & \\ \xi^s_o & & \\ \vdots & \ddots & \\ \xi^s_{t-ds-1} & \cdots & \xi^s_o \end{bmatrix}$$

Notice $H_{1d} = H_1$ and $C_{1d} = C_1$(see 3.2)). From (3.2)

$$s_* = \begin{bmatrix} I_{ds} & 0_{dsxdn} \end{bmatrix} \begin{bmatrix} s_* \\ n_* \end{bmatrix}$$

$$= \begin{bmatrix} I_{ds} & 0_{dsxdn} \end{bmatrix} \begin{bmatrix} H_1 & H_2 \end{bmatrix}^- \{z_* - C_1\, u^{(ds+1)}_{(d)} - C_2\, v^{(dn+1)}_{(d)}\}$$

so that

$$s_{(t)} = K_1\, z_* + K_2\, u^{(ds+1)}_{(t)} + K_3\, v^{(dn+1)}_{(d)} \qquad (6.6)$$

where

$$K_1 = H_{1t} \begin{bmatrix} I_{ds} & 0_{dsxdn} \end{bmatrix} \begin{bmatrix} H_1 & H_2 \end{bmatrix}^-$$

$$K_2 = C_{1t} - \begin{bmatrix} K_1 C_1 & 0_{tx(t-d+ds)} \end{bmatrix}$$

$$K_3 = -K_1\, C_2.$$

The terms on the right hand side of (6.6) are independent under Assumption A

so

$$\mathrm{Var}(s_{(t)}) = K_1\, \mathrm{Var}(z_*)\, K_1' + K_2 \begin{bmatrix} \Upsilon_u \end{bmatrix}_{(t-ds)} K_2' + K_3 \begin{bmatrix} \Upsilon_v \end{bmatrix}_{(d-dn)} K_3' \qquad (6.7)$$

To compute $\mathrm{Var}(E(s_{(t)}|\{z_t\}))$ we notice from (6.6) that

$$E(s_{(t)}|\{z_t\}) = K_1\, z_* + K_2\, P^{(ds+1)}_{(t)} + K_3\, Q^{(dn+1)}_{(d)} \qquad (6.8)$$

where (see (5.2))

$$P_t = E(u_t|\{w_t\}) = \delta_n^*(F)\ \gamma_u(F)\ \gamma_w(F)^{-1}w_t \qquad (6.9)$$

$$Q_t = E(v_t|\{w_t\}) = \delta_s^*(F)\ \gamma_v(F)\ \gamma_w(F)^{-1}w_t.$$

(We could use (6.8) and (6.9) in place of the result of Theorem 4 to do the signal extraction, but the latter requires less computation.)

It can be shown that

$$f_P(\lambda) = f_u(\lambda) - |\delta_s^*(e^{i\lambda})|^2 f_x(\lambda) \qquad f_{pq}(\lambda) = \delta_n^*(e^{-i\lambda}) f_x(\lambda)\delta_s^*(e^{i\lambda})$$

$$f_Q(\lambda) = f_v(\lambda) - |\delta_n^*(e^{i\lambda})|^2 f_x(\lambda),$$

so that auto- and cross-covariances may be computed as

$$\gamma_P(k) = \gamma_u(k) - \delta_s^*(B)\ \delta_s^*(F)\ \gamma_x(k) \qquad \gamma_{PQ}(k) = \delta_n^*(B)\ \delta_s^*(F)\gamma_x(k) \qquad (6.10)$$

$$\gamma_Q(k) = \gamma_v(k) - \delta_n^*(B)\ \delta_n^*(F)\ \delta_x(k).$$

Under Assumption A $\underset{\sim}{z}_*$ is independent of $\{P_t\}$ and $\{Q_t\}$, so from (6.8) we get

$$\begin{aligned} \mathrm{Var}(E(\underset{\sim}{s}_{(t)}|\{z_t\})) &= K_1 \mathrm{Var}(\underset{\sim}{z}_*)\ K_1' + K_2 \left[\gamma_P\right]_{(t-ds)} K_2' \\ &+ K_3 \left[\gamma_Q\right]_{(d-dn)} K_3' \qquad (6.11) \\ &+ K_2 \begin{bmatrix} \gamma_{PQ}(dn-ds) & \cdots & \gamma_{PQ}(d-ds-1) \\ \vdots & & \vdots \\ \gamma_{PQ}(dn+1-t) & \cdots & \gamma_{PQ}(d-t) \end{bmatrix} K_3' \\ &+ K_3 \begin{bmatrix} \gamma_{PQ}(dn-ds) & \cdots & \gamma_{PQ}(dn+1-t) \\ \vdots & & \vdots \\ \gamma_{PQ}(d-ds-1) & \cdots & \gamma_{PQ}(d-t) \end{bmatrix} K_3'. \end{aligned}$$

Using (6.3), we obtain the following theorem.

Theorem 6. Under Assumption A when $\delta_s(B)$ and $\delta_n(B)$ have at least one common zero

$$\mathrm{Var}(\underset{\sim}{s}_{(t)}|\{z_t\}) = K_2\left[\delta_s^*(B)\ \delta_s^*(F)\ \gamma_x\right]_{(t-ds)} K_2' + K_3\left[\delta_n^*(B)\ \delta_n^*(F)\ \gamma_x\right]_{(d-dn)} K_3' - K_2\begin{bmatrix}\gamma_{PQ}(dn-ds) & \cdots & \gamma_{PQ}(d-ds-1)\\ \vdots & & \vdots\\ \gamma_{PQ}(dn+1-t) & \cdots & \gamma_{PQ}(d-t)\end{bmatrix} K_3' - K_3\begin{bmatrix}\gamma_{PQ}(dn-ds) & \cdots & \gamma_{PQ}(dn+1-t)\\ \vdots & & \vdots\\ \gamma_{PQ}(d-ds-1) & \cdots & \gamma_{PQ}(d-t)\end{bmatrix} K_2' .$$

The $\gamma_x(k)$ may be computed from (5.18) and the $\gamma_{PQ}(k)$ from (6.10).

Proof. When we subtract (6.11) from (6.7) to apply (6.3) the $K_1\mathrm{Var}(\underset{\sim}{z}_*)K_1'$ term cancels. Also, by (6.10) the terms involving $\gamma_u(k)$'s and $\gamma_v(k)$'s cancel, giving the result. QED

When $\delta_s(B)$ and $\delta_n(B)$ have no common zeroes there is a far simpler approach to computing $\mathrm{Var}(\underset{\sim}{s}_{(t)}|\{z_t\})$ than the above. From Theorem 3, (5.5), and (5.17) the signal extraction error in this case is

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

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

<u>Theorem 7.</u> Under Assumption A, if $\delta_s(B)$ and $\delta_n(B)$ have no common zeroes, then $\mathrm{Var}(\underset{\sim}{s}_{(t)}|\{z_t\}) = \mathrm{Var}(\underset{\sim}{X}_{(t)}) = \left[\gamma_x\right]_{(t)}$, the elements of which may be computed using (5.18).

This result has been given by Pierce (1979), who examines the behavior of $s_t - R_t$ when $\delta_s(B)$ and $\delta_n(B)$ do and do not have common zeroes. However, Pierce's statement that when $\delta_s(B)$ and $\delta_n(B)$ 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 $s_t - E(s_t|\{z_t\})$ will be nonstationary when $\delta_s(B)$ and $\delta_n(B)$ have common zeroes (the latter will always be non-stationary under Assumption B) they will both have finite mean square, as is easy to see from the results in this paper (including Theorems 6. and 8.).

It is interesting to note that under Assumption A $\mathrm{Var}(\underset{\sim}{s}_{(t)}|\{z_t\})$ does not involve $\mathrm{Var}(\underset{\sim}{z}_*)$. For that matter, neither does $E(\underset{\sim}{s}_t|\{z_t\})$ - see Theorems 3 and 4. This is true whether or not $\delta_s(B)$ and $\delta_n(B)$ have common zeroes. Thus, when making Assumption A we need not concern ourselves with $\mathrm{Var}(\underset{\sim}{z}_*)$. The situation under Assumption B is different: there we must know $\mathrm{Var}(\underset{\sim}{s}_*)$ and $\mathrm{Var}(\underset{\sim}{n}_*)$ (see Theorems 5. and 8.).

<u>6.2 Variances Under Assumption B.</u>

Under Assumption B. $\underset{\sim}{s}_*$ and $\{u_t\}$ are independent, so from (6.5)

$$\mathrm{Var}(\underset{\sim}{s}_{(t)}) = H_{1t}\,\mathrm{Var}(\underset{\sim}{s}_*)\,H_{1t}' + C_{1t}\left[\gamma_u\right]_{(t-ds)}C_{1t}'. \tag{6.12}$$

$\underset{\sim}{s}_*$ is also independent of $\{w_t\}$, so from (6.5) and (6.9)

$$E(\underset{\sim}{s}_{(t)}|\{w_t\}) = C_{1t}\,\underset{\sim}{\rho}^{(ds-1)}_{(t)}$$

so that

$$\mathrm{Var}(E(\underset{\sim}{s}_{(t)} \mid \{w_t\})) = C_{1t} \left[\gamma_p\right]_{(t-ds)} C_{1t}'. \tag{6.13}$$

The remaining term we need is

$\mathrm{Var}(E[\, \underset{\sim}{s}_{(t)} \mid \underset{\sim}{z}_* - E(\underset{\sim}{z}_* \mid \{w_t\}) \,])$, which is

$$\mathrm{Cov}(\underset{\sim}{s}_{(t)}, \underset{\sim}{z}_* - E(\underset{\sim}{z}_* \mid \{w_t\}))\, \Omega^{-1} \mathrm{Cov}(\underset{\sim}{s}_{(t)}, \underset{\sim}{z}_* - E(\underset{\sim}{z}_* \mid \{w_t\}))' \tag{6.14}$$

where $\Omega = \mathrm{Var}(\underset{\sim}{z}_* - E(\underset{\sim}{z}_* \mid \{w_t\}))$ is given by (5.22), or by (5.30) if $\delta_s(B)$ and $\delta_n(B)$ have no common zeroes . From (5.19) and (6.5) we find that

$$\mathrm{Cov}(\underset{\sim}{s}_{(t)}, \underset{\sim}{z}_* - E(\underset{\sim}{z}_* \mid \{w_t\})) = H_{1t}\mathrm{Var}(\underset{\sim}{s}_*)\, H_1' + \mathrm{Cov}(C_{1t}\underset{\sim}{u}_{(t)}^{(ds+1)}, \underset{\sim}{X}_{(d)})G_3' \tag{6.15}$$

Now let G_{1t} be the txt lower triangular matrix

$$G_{1t} = \left[\begin{array}{c|cccc} \multicolumn{5}{c}{0_{ds \times t}} \\ \hline & g_{ds+1,dc+1}^{(1)} & \cdots & g_{ds+1,ds+1}^{(1)} & \\ 0_{(t-ds)\times dc} & \vdots & & \ddots & \\ & g_{t,dc+1}^{(1)} & \cdots & \cdots & g_{tt}^{(1)} \end{array}\right]$$

where the $g_{tj}^{(I)}$'s are defined in (5.20). Notice $G_{1d} = G_I$ (see discussion preceeding (5.20)). It may be seen from (6.5) and (5.27) that

$$\mathrm{Cov}\,(C_{1t}\, \underset{\sim}{u}_{(t)}^{(ds+1)}, \underset{\sim}{X}_{(d)}) = G_{1t} \begin{bmatrix} \gamma_x(0) & \cdots & \gamma_x(d-1) \\ \vdots & & \vdots \\ \gamma_x(1-t) & \cdots & \gamma_x(d-t) \end{bmatrix}$$

so that

$$\text{Cov}(\underset{\sim}{s}_{(t)}, \underset{\sim}{z}_* - E(\underset{\sim}{z}_* | \{w_t\})) = H_{1t}\text{Var}(\underset{\sim}{s}_*)\, H_1{}' + C_{1t} \begin{bmatrix} \gamma_x(0) & \cdots & \gamma_x(d-1) \\ \vdots & & \vdots \\ \gamma_x(1-t) & \cdots & \gamma_x(d-t) \end{bmatrix} G_3{}'. \tag{6.16}$$

We use (6.16) to evaluate (6.14).

Following (6.4) we subtract (6.13) and (6.14) from (6.12) to obtain our result.

<u>Theorem 8</u>. Under Assumption B

$$\text{Var}(\underset{\sim}{s}_{(t)} | \{z_t\}) = H_{1t}\text{Var}(\underset{\sim}{s}_*)\, H_{1t}' + C_{1t}\left[\delta_s^*(B)\delta_s^*(F)\gamma_x\right]_{(t-ds)}C_{1t}'$$

$$- \text{Cov}(\underset{\sim}{s}_{(t)}, \underset{\sim}{z}_* - E(\underset{\sim}{z}_* | \{w_t\}))\, \Omega^{-1}\text{Cov}(\underset{\sim}{s}_{(t)}, \underset{\sim}{z}_* - E(\underset{\sim}{z}_* | \{w_t\}))$$

where Ω is given by (5.22) and $\text{Cov}(\underset{\sim}{s}_{(t)}, \underset{\sim}{z}_* - E(\underset{\sim}{z}_* | \{w_t\}))$ by (6.16).

<u>Proof</u>. In subtracting (6.13) from (6.12) the terms involving γ_u's cancel by (6.10). QED

If $\delta_s(B)$ and $\delta_n(B)$ have no common zeroes then Ω is given by (5.30) and we can simplify (6.16). By (5.19) and (5.29) $\underset{\sim}{z}_* - E(\underset{\sim}{z}_* | \{w_t\}) = H_1\underset{\sim}{s}_* + H_2\, \underset{\sim}{n}_* - H_3\, \underset{\sim}{X}_{(m)}$ so by (6.5)

$$\text{Cov}(\underset{\sim}{s}_{(t)}, \underset{\sim}{z}_* - E(\underset{\sim}{z}_* | \{w_t\})) = H_{1t}\, \text{Var}(\underset{\sim}{s}_*)\, H_1{}' - \text{Cov}(C_{1t}\underset{\sim}{u}_{(t)}^{(ds+1)}, \underset{\sim}{X}_{(m)})\, H_3{}'.$$

Now by (5.31)

$$\mathrm{Cov}(C_{1t}\underset{\sim}{u}_{(t)}^{(ds+1)},\underset{\sim}{X}_{(m)}) = \begin{bmatrix} 0_{ds\times m} \\ \hline \gamma_x(-ds) & \cdots & \gamma_x(m-ds-1) \\ \vdots & & \vdots \\ \gamma_x(1-t) & \cdots & \gamma_x(m-t) \end{bmatrix} - \begin{bmatrix} 0_{ds\times ds} \\ \hline \underset{\sim}{A}_{ds+1}^{s'} \\ \vdots \\ \underset{\sim}{A}_{t}^{s'} \end{bmatrix} \begin{bmatrix} \gamma_x(0) & \cdots & \gamma_x(m-1) \\ \vdots & & \vdots \\ \gamma_x(1-ds) & \cdots & \gamma_x(m-ds) \end{bmatrix}$$

$$= \left\{ \begin{bmatrix} 0_{ds\times t} \\ \hline 0_{(t-ds)\times ds} \;\big|\; I_{t-ds} \end{bmatrix} - \begin{bmatrix} 0_{ds\times t} \\ \hline \begin{matrix} \underset{\sim}{A}_{ds+1}^{s'} \\ \vdots \\ \underset{\sim}{A}_{t}^{s'} \end{matrix} \;\Bigg|\; 0_{(t-ds)\times(t-ds)} \end{bmatrix} \right\} \begin{bmatrix} \gamma_x(0) & \cdots & \gamma_x(m-1) \\ \vdots & & \vdots \\ \gamma_x(1-ds) & \cdots & \gamma_x(m-ds) \\ \hline \gamma_x(-ds) & \cdots & \gamma_x(m-ds-1) \\ \vdots & & \vdots \\ \gamma_x(1-t) & \cdots & \gamma_x(m-t) \end{bmatrix}$$

Combining these results and recalling the definition of H_{1t} we now see that

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

when $\delta_s(B)$ and $\delta_n(B)$ have no common zeroes.

7. Signal Extraction With a Finite Number of Observations

Suppose that a finite stretch of the time series, say $\underset{\sim}{z}_{(N)} = (z_1,\ldots,z_N)'$, is available instead of the complete realization $\{z_t\}$. Cleveland(1972) showed that $E(s_t|\underset{\sim}{z}_{(N)})$ can be obtained by replacing unknown z_j's in $E(s_t|\{z_t\})$ by $E(z_j|\underset{\sim}{z}_{(N)})$, which are forecasted ($j>N$) or backcasted ($j\le 0$) values.

However, he did not consider whether this procedure converges. Bell (1980) established that as long as the expression for $E(s_t|\{z_t\})$, which is linear in the z_t's, converges in mean square, then this procedure converges pointwise to $\mathrm{Cov}(s_t, \underset{\sim}{z}_{(N)})\ \mathrm{Var}(\underset{\sim}{z}_{(N)})^{-1}\underset{\sim}{z}_{(N)} = E(s_t|\underset{\sim}{z}_{(N)})$.

To apply the above result we need to be able to compute $E(z_j|\underset{\sim}{z}_{(N)})$ for as many z_j's as exert an appreciable effect on $E(s_t|\{z_t\})$. Notice from $\delta(B)z_t = w_t$ and Lemma 3 that

$$E(z_{n+\ell}|\underset{\sim}{z}_{(N)}) = \delta_1 E(z_{n+\ell-1}|\underset{\sim}{z}_{(N)}) + \ldots + \delta_d E(z_{n+\ell-d}|\underset{\sim}{z}_{(N)}) \tag{7.1}$$

$$+ E(w_{n+\ell}|\underset{\sim}{w}^{(d+1)}_{(N)}) + E(w_{n+\ell}|\underset{\sim}{z}_* - E(\underset{\sim}{z}_*|\underset{\sim}{w}^{(d+1)}_{(N)})),$$

there being a one - to - one correspondence between $\underset{\sim}{z}_{(N)}$ and $(\underset{\sim}{z}_*, \underset{\sim}{w}^{(d+1)}_{(N)})$.

If the last term in (7.1) is zero, then computing (7.1) recursively for $\ell = 1,2,\ldots$ leads to the usual forecasting procedure discussed, for example, in Box and Jenkins (1976). Under Assumption A $\underset{\sim}{z}_*$ and $\{w_t\}$ are independent so the last term in (7.1) is zero, but this is not the case under Assumption B. Thus, the usual forecasting procedure is correct under Assumption A but is incorrect under Assumption B. This reflects the problem noted earlier under Assumption B that $a_{t+\ell}$ $(\ell > 0)$ can be correlated with z_t through the starting values $\underset{\sim}{z}_*$. Analogously, the usual back-casting procedure is correct under Assumption A and incorrect under Assumption B.

A convenient means of obtaining $E(s_t|\underset{\sim}{z}_{(N)})$ 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, and also briefly discusses the filtering and smoothing procedure. To use the Kalman filter/smoother with general ARMA models one must choose a state space representation corresponding to the model and compute the covariance matrix of the initial state vector. Akaike (1974) gives a state space representation which can be used (it can be extended to nonstationary models), although other choices can be made. Some of our results given earlier should prove useful in computing the covariance matrix of the initial state vector.

Along with $E(s_t|\underset{\sim}{z}_{(N)})$, the Kalman filter/smoother directly produces conditional variances, $Var(s_t|\underset{\sim}{z}_{(N)})$; conditional covariances, $Cov(s_t,s_j|\underset{\sim}{z}_{(N)})$, can also be obtained. This is important since $Var(s_t|\{z_t\})$ as given in Section 6. will differ from $Var(s_t|\underset{\sim}{z}_{(N)})$ for any t for which $E(s_t|\{z_t\})$ is appreciably affected by z_j's outside of $z_1,\ldots,z_N$ - typically for t near 1 or N.

The Kalman filter/smoother can also be used under Assumption A. In doing so the properties of $\underset{\sim}{s}_*$ and $\underset{\sim}{n}_*$ as solutions to (3.2) with $\underset{\sim}{z}_*,\{u_t\}$, and $\{v_t\}$ independent must be taken into account. In particular, this will affect the covariance matrix of the intital state vector. A word of caution is in order here: to use the Kalman filter one must be sure that the state vector at time t is independent of the process noise series in the state equation at time t + 1 (which is not the same as n_{t+1}).

An alternative to the Kalman filter/smoother for producing $Var(s_t|\underset{\sim}{z}_{(N)})$ under Assumption A when $\delta_s(B)$ and $\delta_n(B)$ have no common zeroes is to use the results of Pierce(1979), who gives the spectral density of $s_t - E(s_t|z_N,z_{N-1},\ldots,z_0,\ldots)$ in this case. If the weights for R_t in (1.10) die out rapidly enough so that $R_t = E(s_t | \{z_t\})$ is not influenced much by z_0, z_{-1}, ... say for $t > N/2$, then Pierce's results can be used directly for $t \geq N/2$, and can be turned around to be used for $t < N/2$.

8. Extensions of Results

8.1 Linear Projection Results for the Non-Gaussian Case

By Theorem 2. the results of Section 5. produce $\hat{s}_t$, the linear function of the observed z_t's which minimizes $E\left[(s_t - \hat{s}_t)^2\right]$, whether or not the series involved are normal. The results of Section 6. produce variances and covariances for the time series $s_t - \hat{s}_t$; note these are not the conditional variances and covariances without the normality assumption. Similar reinterpretations apply to the other results in this paper.

8.2 The Case of Known Starting Values

In some cases the starting values $\underset{\sim}{s}_*$ and $\underset{\sim}{n}_*$ might be known quantities. Since the results in this paper assume all random variables have zero mean, we must subtract the effect of the starting values from each z_t to produce a new series, say $\dot{z}_t$. For $t = 1,\ldots,d$ we get $\dot{\underset{\sim}{z}}_* = (\dot{z}_1,\ldots,\dot{z}_d)' = \underset{\sim}{z}_* - H_1\underset{\sim}{s}_* - H_2\underset{\sim}{n}_*$ by (3.2), and for $t > d$ we get $\dot{z}_t = z_t - \underset{\sim}{A}_t^{s'}\underset{\sim}{s}_* - \underset{\sim}{A}_t^{n'}\underset{\sim}{n}_*$ by (3.1). The effects of $\underset{\sim}{s}_*$ and $\underset{\sim}{n}_*$ on z_t for $t \leq 0$ can be similarly removed by considering the generation of s_t and n_t for $t \leq 0$. We now have the decomposition $\dot{z}_t = \dot{s}_t + \dot{n}_t$ where

$$\dot{s}_t = \begin{cases} 0 & t=1,\ldots,ds \\ s_t - \underset{\sim}{A}_t^{s'}\underset{\sim}{s}_* & t > ds \end{cases} \qquad \dot{n}_t = \begin{cases} 0 & t=1,\ldots dn \\ n_t - \underset{\sim}{A}_t^{n'}\underset{\sim}{n}_* & t > dn \end{cases}$$

with analogous definitions for $t \leq 0$. Notice that $\dot{s}_t$ and $\dot{n}_t$ have zero mean and known starting values $\dot{\underset{\sim}{s}}_* = \underset{\sim}{0}$ and $\dot{\underset{\sim}{n}}_* = \underset{\sim}{0}$. Since $\dot{\underset{\sim}{s}}_*$ and $\dot{\underset{\sim}{n}}_*$ are degenerate random vectors they are independent of $\{u_t\}$, $\{v_t\}$, and $\{w_t\}$, so the decomposition $\dot{z}_t = \dot{s}_t + \dot{n}_t$ falls under Assumption B with $\mathrm{Var}\ (\dot{\underset{\sim}{s}}_*) = \underset{\sim}{0}$ and $\mathrm{Var}(\dot{\underset{\sim}{n}}_*) = \underset{\sim}{0}$. We apply the results of Section 5.2 to $\dot{\underset{\sim}{z}}_t$ to get $E(\dot{s}_t|\{z_t\})$, to which we add $\underset{\sim}{A}_t^{s'} \underset{\sim}{s}_*$ (for $t > 0$) to produce $E(s_t|\{z_t\})$. In this case $\mathrm{Var}(\dot{\underset{\sim}{z}}_* - E(\dot{\underset{\sim}{z}}_*|\{w_t\}))$ may very well be singular, so that we must use a generalized inverse of it.

We can obviously use a similar approach when $\underset{\sim}{s}_*$ and $\underset{\sim}{n}_*$ have (known) nonzero means, but are not themselves known exactly ($\mathrm{Var}(\underset{\sim}{s}_*)$ and $\mathrm{Var}(\underset{\sim}{n}_*)$ nonzero).

8.3 Extensions to the Multivariate Case

In the multivariate case we have

$$\underset{\sim}{z}_t = \underset{\sim}{s}_t + \underset{\sim}{n}_t \qquad t = 0, \pm 1, \pm 2, \ . \ . \ .$$

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

$$\delta(B)\dot{\underset{\sim}{z}}_t = \underset{\sim}{w}_t \qquad \delta_s(B)\underset{\sim}{s}_t = \underset{\sim}{u}_t \qquad \delta_n(B)\underset{\sim}{n}_t = \underset{\sim}{v}_t$$

jointly stationary $k \times 1$ vector time series. An important special case occurs when $\delta(B)$, $\delta_s(B)$, and $\delta_n(B)$ remain <u>scalar</u> 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 3. is still correct with $\gamma_u(F)$ and $\gamma_w(F)$ the $k \times k$ matrix covariance generating functions of $\underset{\sim}{u}_t$ and $\underset{\sim}{w}_t$.

One aspect of the multivariate case that requires special consideration is the treatment of starting values. Define the vectors of starting values

$$Z_* = \text{vec}\left(z_1, \ldots, z_d\right) = \begin{bmatrix} z_1 \\ \vdots \\ z_d \end{bmatrix}$$

$$S_* = \text{vec}\left(s_1, \ldots, s_{ds}\right) \quad N_* = \text{vec}\left(n_1, \ldots, n_{dn}\right).$$

The first equation in (3.1) becomes

$$s_t = \sum_{j=1}^{ds} A^s_{t,j}\, s_j + \sum_{i=0}^{t-ds-1} \xi^s_i\, u_{t-i} \qquad t > ds$$

$$= (A^{s'}_t \otimes I_k)\, S_* + \sum_{i=0}^{t-ds-1} \xi^s_i\, u_{t-i}$$

where $\otimes$ denotes the Kronecker product. (3.2) becomes

$$Z_* = \left(\begin{bmatrix} H_1 & H_2 \end{bmatrix} \otimes I_k\right) \begin{bmatrix} S_* \\ N_* \end{bmatrix} + (C_1 \otimes I_k) \begin{bmatrix} u_{ds+1} \\ \vdots \\ u_d \end{bmatrix} + (C_2 \otimes I_k) \begin{bmatrix} v_{dn+1} \\ \vdots \\ v_d \end{bmatrix} \qquad (8.1)$$

Analogous to (5.12), under Assumption A we get

$$\left(\begin{bmatrix} H_1 & H_2 \end{bmatrix} \otimes I_k\right) \begin{bmatrix} E(S_* \mid \{z_t\}) \\ E(N_* \mid \{z_t\}) \end{bmatrix} = (H_1 \otimes I) \begin{bmatrix} z_1 \\ \vdots \\ z_{ds} \end{bmatrix} - H_3 \begin{bmatrix} R_1 \\ \vdots \\ R_m \end{bmatrix} \qquad (8.2)$$

Since $\begin{bmatrix} H_1 & H_2 \end{bmatrix}$ has rank d (see Appendix), the $d \cdot k \times (d+dc) \cdot k$ matrix $\begin{bmatrix} H_1 & H_2 \end{bmatrix} \otimes I$ has rank $d \cdot k$ and (8.2) can be solved. As before, we would solve it in the same way as we assume (8.1) is solved for S_* and N_*, that is, we make the same choice of $\left(\begin{bmatrix} H_1 & H_2 \end{bmatrix} \otimes I\right)^-$. Other expressions in the paper involving starting values are easily modified in a similar fashion

The general case where $\delta(B)$, $\delta_s(B)$, and $\delta_n(B)$ are kxk matrix operators is more difficult and we have chosen not to treat it in this paper. 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 analoguous to that used here for certain special cases, such as when $\underset{\sim}{n}_t$ is stationary.

Appendix

A.I. Rank of the Matrix $\begin{bmatrix} H_1 & H_2 \end{bmatrix}$

We show the matrix $\begin{bmatrix} H_1 & H_2 \end{bmatrix}$ has rank d.

Proof. Let r = ds + dn = d +dc and consider the augmented r x r matrix.

$$\begin{bmatrix} H_{1r} & H_{2r} \end{bmatrix} = \left[\begin{array}{ccc|ccc} 1 & & & 1 & & \\ & \ddots & & & \ddots & \\ & & 1 & & & 1 \\ A^s_{ds+1,1} & \cdots & A^s_{ds+1,ds} & A^n_{dn+1,1} & \cdots & A^n_{dn+1,dn} \\ & \vdots & & & \vdots & \\ A^s_{r,1} & \cdots & A^s_{r,ds} & A^n_{r,1} & \cdots & A^n_{r,dn} \end{array}\right]$$

We show this matrix has rank d. First consider the r x dn matrix

$$\Delta_s = \begin{bmatrix} 0 & 0 & & -\delta_{s,ds} \\ \vdots & \vdots & & \vdots \\ \vdots & 0 & & \vdots \\ 0 & -\delta_{s,ds} & \cdots & -\delta_{s,1} \\ -\delta_{s,ds} & \vdots & & 1 \\ \vdots & -\delta_{s,1} & & 0 \\ -\delta_{s,1} & 1 & & \vdots \\ 1 & 0 & & 0 \end{bmatrix}$$

The dn columns of Δ_s are linearly independent and, since

$\delta_s(B)\, A^s_{ti} = 0 \quad i = 1,\ldots,ds$ for $t > ds$ (see the discussion following (2.5)), the columns of Δ_s are orthogonal to the ds columns of H_{1r}, which are also linearly independent. Thus, $\mathcal{S}(\Delta_s) = \mathcal{S}(H_{1r})^{\perp}$ where $\mathcal{S}(\cdot)$ denotes the space spanned by the columns of a matrix and $\mathcal{S}(\cdot)^{\perp}$ denotes the orthogonal complement of $\mathcal{S}(\cdot)$ in r- dimensional Euclidean space.

Consider the homogeneous difference equation

$$\delta_s(B)\alpha_t = \alpha_t - \delta_{s,1}\alpha_{t-1} - \cdots - \delta_{s,ds}\alpha_{t-ds} = 0. \tag{A.1}$$

Let the distinct zeroes of $\delta_s(\omega)$ be $\omega_1,\ldots,\omega_k$ with multiplicities $m_1,\ldots,m_k$ where $m_1+\ldots+m_k = ds$. It is well known (see Henrici 1974, pp. 584-587) that the space of solutions to (A.1) has dimension ds, and that the sequences $(t \geq 0)$

$$t^j\omega_\ell^{-t} \quad j = 0,\ldots,m_\ell-1 \ ; \ \ell = 1,\ldots,k$$

provide a (linearly independent) basis for this space. Take the first r elements of each of these ds sequences to form the ds columns of an r x ds matrix; call this matrix $\Omega_s = (\omega^s_{it})$. If $\sum_{i=1}^{ds} c_i\omega^s_{it} = 0$ for $t = 1,\ldots,r$, then since $\omega^s_{it} = \delta_{s,1}\omega^s_{i,t-1} +\ldots+ \delta_{s,ds}\,\omega^s_{i,t-ds}$ for $t > r$, we easily see $\sum_{i=i}^{ds} c_i\,\omega^s_{it}=0$ for all t, which implies $c_1 =\ldots= c_{ds}=0$. Hence, the ds columns of Ω_s are linearly independent. Also, the ds columns of Ω_s are orthogonal to the dn columns of Δ_s, so $\mathcal{S}(\Omega_s) = \mathcal{S}(\Delta_s)^{\perp} = (\mathcal{S}(H_{1r})^{\perp})^{\perp} = (H_{1r})$.

We define $\Omega_n=(\omega^n_{it})$ to be the r×dn matrix analogous to Ω_s whose columns come from solutions to

$$\delta_n(B)\ \eta_t = \eta_t - \delta_{n,1}\eta_{t-1}-\cdots-\delta_{n,dn}\ \eta_{t-dn} =0. \tag{A.2}$$

Then $\mathcal{S}(\Omega_n) = \mathcal{S}(H_{2r})$ and $\mathcal{S}\begin{bmatrix}\Omega_s & \Omega_n\end{bmatrix} = \mathcal{S}\begin{bmatrix}H_{1r} & H_{2r}\end{bmatrix}$. Since

$$\delta(B) = \delta_c(B)\ \delta_s^*(B)\ \delta_n^*(B)$$

$$\delta_s(B) = \delta_c(B)\delta_s^*(B) \qquad \delta_n(B) = \delta_c(B)\delta_n^*(B)$$

(where $\delta_s^*(B)$ and $\delta_n^*(B)$ have no common factors) any solution to (A.1) or (A.2) is obviously a solution to $\delta(B)\ \beta_t = 0$. It is then easy to see that the columns of $\begin{bmatrix}\Omega_s & \Omega_n\end{bmatrix}$ come from the d linearly independent solutions to $\delta(B)\beta_t=0$ of the form $t^j\zeta_\ell^{-t}$ $j = 0,..,\mu_\ell-1$, $\ell = 1,...,h$, where $\zeta_1,...,\zeta_h$ are the distinct zeroes of $\delta(\zeta)$ with multiplicities $\mu_1,..., \mu_h$ ($\sum_1^h \mu_i = d$). Of these $r = d+dc$ columns d will be distinct and linearly independent, and the dc columns corresponding to $\delta_c(B)$ will occur twice. Thus, $\begin{bmatrix}\Omega_s & \Omega_n\end{bmatrix}$ has rank d, and so does $\begin{bmatrix}H_{1r} & H_{2r}\end{bmatrix}$. Finally, $\delta_s(B)A_{tj}^s = 0$ for $t > ds$ implies $\delta(B)A_{tj}^s = 0$ for $t > d$, and similarly $\delta(B)A_{tj}^n = 0$ for $t > d$, so that

$$\left[\begin{array}{c|c} I_d & 0_{d\times dc} \\ \hline \begin{matrix}-\delta_d \cdot\ \cdot\ \cdot\ \cdot -\delta_1 \\ -\delta_d\ \cdot\ \cdot -\delta_2 \\ \\ -\delta_d \cdot\ \cdot \end{matrix} & \begin{matrix}1 & & & \\ -\delta_1 & 1 & & \\ & \ddots & \ddots & \\ \cdot\ \cdot\ \cdot & & -\delta_1 & 1\end{matrix}\end{array}\right]\begin{bmatrix}H_{1r} & H_{2r}\end{bmatrix} = \begin{bmatrix}H_1 & H_2 \\ 0 & 0\end{bmatrix}$$

This shows that rank $\begin{bmatrix}H_1 & H_2\end{bmatrix} = d$. QED

A. II Filtering Nonstationary Time Series

Although there has been considerable work on the linear filtering of stationary time series (see Hannan 1970), little seems to have been done on filtering nonstationary time series. We present two basic results on this and apply them to the nonstationary signal extraction problem.

Theorem 9. Suppose $\delta(B)z_t = w_t$ is stationary, and the distinct zeroes of $\delta(\zeta) = 1 - \delta_1\zeta - \ldots - \delta_d\zeta^d$ (which are all on or inside the unit circle) are $\zeta_1,\ldots,\zeta_h$ of multiplicities $\mu_1,\ldots,\mu_h$. Let $|\zeta_{min}| = \min(|\zeta_1|,\ldots,|\zeta_h|)$ and $m = \max(\mu_1,\ldots,\mu_h)$. Then $\alpha(F)z_t = \sum_{-\infty}^{\infty} \alpha_j \, z_{t+j}$ converges in mean square under either of the following conditions.

a. If all the zeroes of $\delta(\zeta)$ lie on the unit cirle ($|\zeta_{min}| = 1$), $\sum_{-\infty}^{\infty} |j|^m |\alpha_j| < \infty$ ($\{\alpha_j\}$ is m - summable).

b. If $\delta(\zeta)$ has a zero inside the unit circle ($|\zeta_{min}| < 1$), $\alpha(\zeta) = \sum_{-\infty}^{\infty} \alpha_j\zeta^j$ is analytic (i.e. converges absolutely) on some annulus $\rho < |\zeta| < \rho^{-1}$ with $\rho < |\zeta_{min}| < 1$.

Proof. We prove the theorem using two lemmas.

Lemma 4. Let $\{\beta_k\}_1^\infty$ be a sequence of real numbers and $\{Y_k\}_1^\infty$ a sequence of zero mean random variables.

a. If $\sum_1^\infty |\beta_k| \, k^m < \infty$ and $Var(Y_k) \leq M_1 \, k^{2m}$ for some M_1, then $\sum_1^\infty \beta_k Y_k$ converges in mean square.

b. If $\sum_{1}^{\infty} \beta_k \zeta^k$ converges absolutely (i.e. is analytic) for

$|\zeta| < \tau_2$ and $\mathrm{Var}(Y_k) \leq M_2 \tau_1^{2k}$ for some M_2 where $0 < \tau_1 < \tau_2$,

$\sum_{1}^{\infty} \beta_k Y_k$ converges in mean square.

Proof of Lemma 4. Consider $\sum_{\ell}^{n} \beta_k Y_k$ $(0<\ell<n)$. By repeated application of Minkowski's inequality (Rudin 1974, p. 65)

$$\mathrm{Var}\Big(\sum_{\ell}^{n} \beta_k Y_k\Big) \leq \left[\sum_{\ell}^{n} |\beta_k| \; \mathrm{Var}(Y_k)^{\frac{1}{2}}\right]^2 . \tag{A.3}$$

For a. we see $\mathrm{Var}\Big(\sum_{\ell}^{n} \beta_k Y_k\Big) \leq M_1 \left[\sum_{\ell}^{n} |\beta_k| k^m\right]^2 \to 0$ as $\ell, n \to \infty$. For b. we see $\mathrm{Var}\Big(\sum_{\ell}^{n} \beta_k Y_k\Big) \leq M_2 \left[\sum_{\ell}^{n} \beta_k \tau_1^k\right]^2 \to 0$ as $\ell, n \to \infty$. QED

Lemma 5. Consider $\mathrm{Var}(z_t)$ for $t > 0$.

a. If all the zeroes of $\delta(\zeta)$ lie on the unit circle, then

$\mathrm{Var}(z_t) \leq M_1 t^{2m}$ for some M_1, for all $t > 0$.

b. For any $\tau > |\zeta_{min}|^{-1}$, $\mathrm{Var}(z_t) \leq M_2 \tau^{2t}$ for some M_2, for all $t > 0$.

Proof of Lemma 5. By Theorem 1, for $t > d$

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

We know (see discussion following (2.5))

$$\delta(B)\, A_{jt} = 0 \quad t > d \qquad \delta(B)\xi_i = 0 \quad i \geq 1.$$

By results on solutions to difference equations (Henrici 1974, pp. 584-587)

$$A_{jt} = \sum_{k=1}^{h} p_{jk}(t)\zeta_k^{-t} \quad j=1,\ldots,d \qquad \xi_i = \sum_{k=1}^{h} p_k(i)\zeta_k^{-i}$$

for $t > 0$ and $i \geq 0$, where $p_k(\cdot)$, $p_{jk}(\cdot)$ $j = 1,\ldots,d$, $k = 1,\ldots,h$ are polynomials of degree μ_k-1.

Consider first the case where all zeroes of $\delta(\zeta)$ lie on the unit circle. Then $|\zeta_k| = 1$, $k = 1,\ldots,h$, and since

$$|p_{jk}(t)| = \left|p_{jk0} + p_{jk1}t+\ldots+ p_{jk,\mu_k-1}t^{\mu_k-1}\right| ,$$

$$\leq (|p_{jk0}| + |p_{jk1}|+\ldots+ |p_{jk,\mu_k-1}|)\, t^{\mu_k-1}$$

we see that for $t > 0$

$$|A_{jt}| \leq \sum_{k=1}^{h} |p_{jk}(t)|\,|\zeta_k^{-t}|$$

$$\leq \sum_{k=1}^{h} M_{jk}t^{\mu_k-1}\cdot 1 \qquad M_{jk} = |p_{jk0}|+\ldots+|p_{jk,\mu_k-1}|$$

$$\leq M_j^* t^{m-1} . \qquad M_j^* = M_{j1}+\ldots+M_{jh}$$

Therefore, using Minkowski's inequality

$$\mathrm{Var}(\underset{\sim}{A}_t{}'\underset{\sim}{z}_*) \leq \left[\sum_{j=1}^{d} |A_{jt}|\ \mathrm{Var}(z_j)^{\frac{1}{2}}\right]^2 \leq M_A t^{2(m-1)}$$

where $M_A = \left[\sum_{j=1}^{d} M_j^*\ \mathrm{Var}(z_j)^{\frac{1}{2}}\right]^2$. Similarly $|\xi_i| \leq M^* i^{m-1}$ for $i > 0$ so

$$\mathrm{Var}\Big(\sum_{i=0}^{t-d-1} \xi_i w_{t-i}\Big) \leq \left[\sum_{i=0}^{t-d-1} |\xi_i|\gamma_w(0)^{\frac{1}{2}}\right]^2$$

$$\leq (M^*)^2 \gamma_w(0) \left[\sum_{i=0}^{t-d-1} i^{m-1}\right]^2$$

$$\leq M_\xi t^{2m}. \qquad M_\xi = (M^*)^2 \gamma_w(0)$$

Finally, using Minkowski's inequality again

$$\mathrm{Var}(z_t) \leq \left[\mathrm{Var}(\underset{\sim}{A}_t' \underset{\sim}{z}_*)^{\frac{1}{2}} + \mathrm{Var}(\sum_{i=0}^{t-d-1} \xi_i w_{t-i})^{\frac{1}{2}}\right]^2$$

$$\leq \left[M_A^{\frac{1}{2}} t^{m-1} + M_\xi^{\frac{1}{2}} t^m\right]^2$$

$$< M_1 t^{2m} \qquad M_1 = \left[M_A^{\frac{1}{2}} + M_\xi^{\frac{1}{2}}\right]^2.$$

Now consider the case where $\delta(\zeta)$ has a root inside the unit circle. Take any $\tau_1 > |\xi_{min}|^{-1} > 1$. Then, for $j = 1,\ldots,d$

$$|A_{jt}| \leq \sum_{k=1}^{h} |p_{jk}(t)||\zeta_k^{-t}| \leq \sum_{k=1}^{h} |p_{jk}(t)|\tau_1^t \leq M_j^* t^{m-1}\tau_1^t$$

where M_j^* is as before. By a lemma given by Fuller (1976,p.91, problem 24) there exist constants $\tilde{M}_j$ $j=1,\ldots,d$ such that $|A_{jt}| \leq M_j^* t^{m-1}\tau_1^t < \tilde{M}_j \tau_2^t$ for any $\tau_2 > \tau_1$. Similarly $|\xi_i| < \tilde{M} \tau_2^i$ for some $\tilde{M}$. Now

$$\mathrm{Var}(\underset{\sim}{A}_t' \underset{\sim}{z}_*) \leq \left[\sum_{j=1}^{d} |A_{jt}| \mathrm{Var}(z_j)^{\frac{1}{2}}\right]^2 \leq \tilde{M}_A \tau_2^{2t}$$

where $\tilde{M}_A = \left[\sum_{j=1}^{d} \tilde{M}_j \mathrm{Var}(z_j)^{\frac{1}{2}}\right]^2$, and

$$\text{Var}\left(\sum_{i=0}^{t-d-1} \xi_i w_{t-i}\right) \leq \left[\sum_{i=0}^{t-d-1} |\xi_i| \, \gamma_w(0)^{\frac{1}{2}}\right]^2$$

$$\leq \tilde{M}^2 \gamma_w(0) \left[\sum_{i=0}^{t-d-1} \tau_2^i\right]^2$$

$$\leq \tilde{M}^2 \gamma_w(0)\ (t-d)\ \tau_2^{(t-d-1)2}$$

$$\leq \tilde{M}_\xi \tau_2^{2t}$$

for some $\tilde{M}_\xi$, using Fuller's lemma again. Therefore, for $t>0$

$$\text{Var}(z_t) \leq \left[\tilde{M}_A^{\frac{1}{2}} \tau_2^t + \tilde{M}_\xi^{\frac{1}{2}} \tau_2^t\right]^2 = M_2 \tau_2^{2t}$$

where $M_2 = \left[\tilde{M}_A^{\frac{1}{2}} + \tilde{M}_\xi^{\frac{1}{2}}\right]^2$. Our only requirements on τ_1 and τ_2 are $1 < |\zeta_{min}|^{-1} < \tau_1 < \tau_2$. Then, for any $\tau > |\zeta_{min}|^{-1}$ we can take τ_1 and τ_2 such that $|\zeta_{min}|^{-1} < \tau_1 < \tau_2 < \tau$, and obtain $\text{Var}(z_t) \leq M_2 \tau^{2t}$. QED

We now establish that $\alpha(F)z_t$ converges in mean square when either of the conditions given in Theorem 9 is satisfied. Notice $\sum_{j=-t+1}^{\infty} \alpha_j z_{t+j} = \sum_{k=1}^{\infty} \beta_k z_k$ where $\beta_k = \alpha_{k-t}$. First suppose all the zeroes of $\delta(\zeta)$ lie on the unit circle and $\{\alpha_j\}$ is m-summable. Then by Lemma 5.a $\text{Var}(z_k) \leq M_1 k^{2m}$ for $k > 0$. It is easy to show that $\{\beta_k\}$ is m-summable, so by Lemma 4.a $\sum_1^{\infty} \beta_k z_k$ converges in mean square. Now suppose $\delta(\zeta)$ has a zero inside the unit circle and $\alpha(\zeta)$ is analytic on $\rho < |\zeta| < \rho^{-1}$ with $\rho < |\zeta_{min}| < 1$. Then $\sum_1^{\infty} \beta_k \zeta^k$ is analytic on $|\zeta| < \rho^{-1}$. By Lemma 5.b $\text{Var}(z_k) \leq M_2 \tau^{2k}$ for any τ such that $\rho^{-1} > \tau > |\zeta_{min}|^{-1}$, so by Lemma 4.b $\sum_1^{\infty} \beta_k z_k$ converges in mean square.

To establish the mean square convergence of $\sum_{-\infty}^{-t} \alpha_j z_{t+j} = \sum_{-\infty}^{0} \beta_k z_k$ we must consider the representation of z_t for $t \le 0$, which uses the relation (see (2.4))

$$1 + (\delta_{d-1}/\delta_d)F + \dots + (\delta_1/\delta_d)F^{d-1} - (1/\delta_d F^d)z_t = w_{t+d}.$$

Using this we can obtain bounds on $\mathrm{Var}(z_t)$, which will depend on the zeroes of

$$1 + (\delta_{d-1}/\delta_d)\zeta + \dots + (\delta_1/\delta_d)\zeta^{d-1} - (1/\delta_d)\zeta^d =$$

$$(-1/\delta_d)\zeta^d \left[1-\delta_1\zeta^{-1} - \dots - \delta_{d-1}\zeta^{1-d} - \delta_d\zeta^{-d}\right]$$

$(\zeta \ne 0)$ which are $\zeta_1^{-1}, \dots, \zeta_h^{-1}$ of multiplicities $\mu_1 \dots, \mu_h$.

Now $\zeta_1^{-1}, \dots, \zeta_h^{-1}$ all lie on or outside the unit circle ; consequently, the bounds on $\mathrm{Var}(z_t)$ for $t \le 0$ as a function of $|t|$ are smaller, or at least no larger, than the analogous bounds on $\mathrm{Var}(z_t)$ for $t > 0$ given in Lemma 5. Also $\{\alpha_k\}$ m-summable implies $\sum_{-\infty}^{0} |\beta_k||k|^m < \infty$, and $\alpha(\zeta)$ analytic on $\rho < |\zeta| < \rho^{-1}$ implies $\sum_{-\infty}^{0} \beta_k \zeta^k = \sum_{0}^{\infty} \beta_{-k}\zeta^{-k}$ is analytic on $|\zeta^{-1}| < \rho^{-1}$.

Thus, using the same type of argument as the above, $\sum_{-\infty}^{-t} \alpha_j z_{t+j} = \sum_{-\infty}^{0} \beta_k z_k$ is readily seen to converge in mean square. QED

<u>Theorem 10</u>. Let Y_t and X_t be time series (stationary or nonstationary) related by $\phi(B)Y_t = X_t$ where $\phi(B) = 1-\phi_1 B - \dots - \phi_p B^p$. Suppose $\alpha(F)Y_t$ and $\alpha(F)X_t$ converge in mean square ($\alpha(F) = \sum_{-\infty}^{\infty} \alpha_j F^j$). Then

$$\phi(B)\left[\alpha(F)\ Y_t\right] = \left[\phi(B)\alpha(F)\right]\ Y_t = \alpha(F)\ X_t$$

and these all converge in mean square.

Proof. Let $\beta(F) = \phi(B)\ \alpha(F) = \sum_{-\infty}^{\infty} \beta_k F^k$ so $\beta_k = \alpha_k - \phi_1\alpha_{k+1} - \cdots - \phi_p\alpha_{k+p}$.

Then

$$\sum_{-n}^{N} \beta_k\ Y_{t+k} = \sum_{-n}^{N} (\alpha_k - \phi_1\alpha_{k+1} - \cdots - \phi_p\alpha_{k+p})\ Y_{t+k}$$

$$= \sum_{-n+p}^{N} \alpha_k (Y_{t+k} - \phi_1 Y_{t+k-1} - \cdots - \phi_p Y_{t+k-p})$$

$$+ (\alpha_{-n} - \phi_1\alpha_{-n+1} - \cdots - \phi_{p-1}\alpha_{-n+p-1})\ Y_{t-n} +$$

$$\vdots$$

$$+ \alpha_{-n+p-1}\ Y_{t-n+p-1}$$

$$- (\phi_1\alpha_{N+1} + \cdots + \phi_p\ \alpha_{N+p})\ Y_{t+N} -$$

$$\vdots$$

$$- \phi_p\alpha_{N+1}\ Y_{t+N+1-p}$$

$$= \sum_{-n+p}^{N} \alpha_k X_{t+k} + \text{other terms.}$$

$\alpha(F)\ Y_t$ converges in mean square implies $\alpha_k\ Y_{t+k} \to 0$ in mean square as $k \to \pm\infty$ for any t. Therefore, as $n, N \to \infty$, $\sum_{-n}^{N} \beta_k\ Y_{t+k} \to \alpha(F)\ X_t$ in mean square (the "other terms" all go to zero). This establishes the second equality in the theorem. To get the first write

$$\phi(B) \sum_{-n}^{N} \alpha_k\ Y_{t+k} = \sum_{-n}^{N} \alpha_k\ Y_{t+k} - \phi_1 \sum_{-n}^{N} \alpha_k\ Y_{t+k-1} - \cdots - \phi_p \sum_{-n}^{N} \alpha_k\ Y_{t+k-p}$$

$$\to \alpha(F)Y_t - \phi_1\alpha(F)Y_{t-1} - \cdots - \phi_p\alpha(F)Y_{t-p}$$

$$= \phi(B)\ \left[\alpha(F)Y_t\right]$$

in mean square, as $n, N \to \infty$. QED

An important application of Theorem 10 is to nonstationary Y_t when X_t is stationary and $\phi(\zeta)$ has zeroes on or inside the unit circle. In this case the convergence of $\alpha(F)Y_t$ can be investigated using Theorem 9. Theorem 10 is a weak result here, in the sense that we would prefer not to need the convergence of $\alpha(F)\,Y_t$ to establish that $\left[\phi(B)\alpha(F)\right]Y_t$ converges to $\alpha(F)\left[\phi(B)Y_t\right] = \alpha(F)X_t$. However, examples can be given where $\alpha(F)X_t$ converges but $\left[\phi(B)\alpha(F)\right]Y_t$ does not, because the "other terms" in the proof do not go to zero. Looking for weaker conditions under which $\left[\phi(B)\alpha(F)\right]Y_t$ converges to $\alpha(F)X_t$ is a topic for future research.

We can apply Theorems 9 and 10 to the signal extraction problem. The solutions to the signal extraction problem given in Section 5 all involve

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

If all the zeroes of $\delta(\zeta)$ lie on the unit circle we see from Theorem 9 that R_t will exist if $\{\gamma_u(k)\}$ and $\{\gamma_w(k)\}$ are m-summable, since this implies (Brillinger 1975, p. 78) the coefficients in $\delta_n^*(B)\delta_n^*(F)\gamma_u(F)\gamma_w(F)^{-1}$ are m-summable. This will happen, for example, if u_t follows a stationary, and w_t follows a stationary and invertible, autoregressive-moving average model.

If $\delta(\zeta)$ has a zero inside the unit circle, but $\gamma_u(\zeta)$ is analytic and $\gamma_w(\zeta)$ is analytic and nonzero in $\rho < |\zeta| < \rho^{-1}$ where $\rho < |\zeta_{min}|$, so that $\delta_n^*(\zeta^{-1})\ \delta_n^*(\zeta)\gamma_u(\zeta)\gamma_w(\zeta)^{-1}$ is analytic in $\rho < |\zeta| < \rho^{-1}$, then by Theorem 9 R_t exists. In this case, if u_t and w_t follow autoregressive-moving average models we need the autoregressive polynomial for u_t and the moving average polynomial for w_t to have no zeroes in $\rho < |\zeta| < \rho^{-1}$ to apply Theorem 9 . Assuming one of the conditions in Theorem 9 holds so that R_t exists, we can use Theorem 10 to justify the interchanging of operators that is done in (5.2) through (5.4).

FOOTNOTES

[1] This research was supported by grants from the National Science Foundation and the U.S. Bureau of the Census through the authors participation in the American Statistical Association/Census Bureau fellowship program. The author is grateful to David Findley for many useful discussions and for pointing out the results of his paper (Findley 1982).

AMS 1970 subject classifications, Primary 62M10; secondary 60G35.
Key Words and phrases. Nonstationary time series, signal extraction, conditional expectation, conditional variance, starting values, filtering nonstationary time series.

[2] Anderson(1971,pp.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).

References

Akaike, H. (1974) "Markovian Representation of Stochastic Processes and Its Application to the Analysis of Autoregressive Moving Average Processes," Ann. Inst. Statist. Math., 26, 363-387.

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

Bell, W.R. (1980), "Multivariate Time Series: Smoothing and Backward Models," Ph.D. thesis, University of Wisconsin - Madison.

Box, G.E.P., and Jenkins, G.M., (1976), Time Series Analysis: Forecasting and Control, San Francisco: Holden Day.

Brillinger, D.R. (1975), Time Series: Data Analysis and Theory, New York: Holt, Rinehart, and Winston.

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. (1982) "On Backshift-Operator Polynomial Transformations to Stationarity," submitted for publication.

Fuller, W.A. (1976), Introduction to Statistical Time Series, New York: Wiley.

Gikhman, I.I., and Skorokhod, A.V. (1969), Introduction to the Theory of Random Processes, Philadelphia: W.B. Saunder.

Hannan, E.J. (1967) "Measurement of a Wandering Signal Amid Noise," J. Appl. Probab., 4, 90-102.

__________ (1970), Multiple Time Series, New York: Wiley.

Henrici, P. (1974) Applied and Computational Complex Analysis, Vol. 1, New York: Wiley.

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.

__________ (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.

__________ (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, New York: McGraw-Hill.

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.

Robinson, E.A. (1959), An Introduction to Infinitely Many Variates, New York; Hafner.

Rudin, W. (1974), Real and Complex Analysis, New York: McGraw-Hill.

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.), London: Blackie and Son.

Wiener, N. (1949), The Extrapolation, Interpolation and Smoothing of Stationary Time Series with Engineering Applications, New York: Wiley.